

En Introduktion til Sandsynlighedsregning

9. Udgave

Michael Sørensen

11. juli 2008

0

Forord

Til 2. udgave

Disse forelæsningsnoter trækker i betydelig grad på noter udarbejdet af en række kolleger. Det gælder ikke mindst Tue Tjurs forelæsningsnoter “Sandsynlighedsregning”, som i mange år blev benyttet til Statistik 0 undervisningen, og som har været udgangspunkt for nærværende noter. Visse steder er hele passager taget næsten ordret fra Tue Tjurs noter, ligesom hovedparten af opgaverne stammer derfra. Jeg har imidlertid også hentet mange ideer, formuleringer, eksempler og opgaver fra forelæsningsnoter af Jørgen Hoffmann-Jørgensen, Preben Blæsild, Svend Erik Graversen og Ernst Hansen.

I forhold til den første udgave er de væsentlige ændringer en række nye eksempler, opgaver og figurer. Endvidere er der tilføjet et appendiks om mængdelære, og en del trykfejl og uheldige formuleringer er blevet rettet. Endelig er de fleste kapitler blevet forsynet med en kort sammenfatning.

København, juli 2001
Michael Sørensen

Til 3., 4. og 5. udgave

I forhold til 2. udgave er der kun foretaget mindre ændringer. Et antal trykfejl er blevet rettet, og enkelte steder er forklaringer blevet uddybet eller omformuleret.

København, juni 2004
Michael Sørensen

Til 6., 7., 8. og 9. udgave

I forhold til 5. udgave er der tilføjet et kapitel om Markovkæder med endeligt tilstandsrum. Ellers er et antal trykfejl blevet rettet, og enkelte forklaringer er blevet uddybet eller omformuleret. I 7., 8. og 9. udgave er der foretaget endnu flere mindre rettelser og forbedringer.

København, juli 2008

Michael Sørensen

Indhold

Forord	1
1 Sandsynlighedsregningens grundlag	7
1.1 Indledning	7
1.2 Det endelige sandsynlighedsfelt	12
1.3 Det generelle sandsynlighedsfelt	16
1.4 Betingede sandsynligheder	23
1.5 Stokastisk uafhængighed	30
1.6 Sammenfatning	39
1.7 Opgaver	39
2 Stokastiske variable	45
2.1 Stokastiske variable og deres fordeling	45
2.2 Fordelingsfunktionen	49
2.3 Fler-dimensionale stokastiske variable	56
2.4 Uafhængige stokastiske variable	58
2.5 Sammenfatning	61
2.6 Opgaver	62
3 Fordelinger på endelige mængder	65
3.1 Sandsynlighedsfunktioner	66
3.2 Binomialfordelingen	68
3.3 Den hypergeometriske fordeling	72
3.4 Transformation af fordelinger	75
3.5 Polynomialfordelingen	78
3.6 Uafhængige stokastiske variable	82
3.7 Middelværdi og varians	87

3.8	Kovarians og korrelation	95
3.9	Sammenfatning	104
3.10	Opgaver	105
4	Generelle diskrete fordelinger	114
4.1	Diskrete fordelinger	114
4.2	Transformation af diskrete fordelinger	124
4.3	Uafhængige diskrete stokastiske variable	126
4.4	Middelværdi og varians	130
4.5	Kovarians og korrelation	138
4.6	Betingede fordelinger	139
4.7	Sammenfatning	140
4.8	Opgaver	141
5	En-dimensionale kontinuerte fordelinger	145
5.1	Kontinuerte fordelinger og tætheder	145
5.2	Middelværdi og varians	156
5.3	Normalfordelingen	162
5.4	Transformation af kontinuerte fordelinger på \mathbb{R}	165
5.5	Sammenfatning	172
5.6	Opgaver	173
6	Fler-dimensionale kontinuerte fordelinger	179
6.1	Tætheder og fordelinger på \mathbb{R}^n	179
6.2	Uafhængighed	183
6.3	Transformation af kontinuerte fordelinger på \mathbb{R}^n	187
6.3.1	Reelle transformationer	187
6.3.2	Lineær transformation af en kontinuert fordeling	191
6.4	Middelværdi, varians og kovarians	197
6.5	Kontinuerte betingede fordelinger	200
6.6	Sammenfatning	202
6.7	Opgaver	203
7	Grænsresultater for stokastiske variable	209
7.1	De store tals lov	209
7.2	Den centrale grænseværdisætning	212
7.3	Opgaver	219

8	Normalfordelingens teori	221
8.1	χ^2 -fordelingen og Γ -fordelingen	221
8.2	F-fordelingen og t-fordelingen	225
8.3	Den fler-dimensionale normalfordeling	229
8.3.1	Den fler-dimensionale standard normalfordeling	229
8.3.2	Den generelle fler-dimensionale normalfordeling	233
8.3.3	Den to-dimensionale normalfordeling	235
8.4	Sammenfatning	239
8.5	Opgaver	239
9	Markovkæder med endeligt tilstandsrum	241
9.1	Definition og eksempler	241
9.2	Nogle regneregler	246
9.3	Ergodicitet og stationær fordeling	251
9.4	Sammenfatning	264
9.5	Opgaver	265
10	Nogle stokastiske modeller	269
10.1	Aktier	269
10.2	Poisson processen	271
10.3	Brownsk bevægelse	274
11	Efterskrift	277
A	Om mængder	279
A.1	Mængder og mængdeoperationer	279
A.2	Opgaver	284
B	Om kombinatorik	286
B.1	Fakultetsfunktionen	286
B.2	Nedadstigende faktoriel	287
B.3	Binomialkoefficienter	287
B.4	Polynomialkoefficienter	289
B.5	Opgaver	290
C	Om uendelige rækker	291

D Om integraler	294
D.1 Funktioner af én variabel	294
D.2 Funktioner af flere variable	298
E Tabeller	303
E.1 Standard normalfordelingen	304
E.2 χ^2 -fordelingen	305
E.3 t-fordelingen	306
E.4 F-fordelingen	307
E.5 Det græske alfabet	310
Indeks	311

Kapitel 1

Sandsynlighedsregningens grundlag

1.1 Indledning

I den menneskelige tilværelse forekommer der mange tilfældige og uforudsigelige begivenheder. Skønt dette utvivlsomt gør livet interessant, har man dog i mange situationer brug for metoder til, så vidt det nu er muligt, at håndtere sådanne usikkerhedsmomenter. Det er baggrunden for, at sandsynlighedsregningen og statistikken er blevet udviklet. Disse fags metoder har i vore dage fundet meget bred anvendelse, for eksempel inden for forsikring, finansiering, planlægning og analyse af alle former for data, som er behæftet med usikkerhed.

Sandsynlighedsregningen har sin oprindelse i menneskets spilleglæde. I 40.000 år gamle køkkenmøddinger har man fundet betydelige mængder astragalusknogler fra hov- og klovdyr, der kan benyttes til terningspil, ligesom et ægyptisk gravmaleri fra det første ægyptiske dynasti (ca. 3000 f. Kr.) viser en adelsmand og hans kone, der benytter astragalusknogler i forbindelse med et brætspil. Astragalusknoglen har fire sider. Når den kastes forekommer de to af siderne med sandsynlighed $1/10$, mens de to andre forekommer med sandsynlighed $2/5$. Spil med astragalus-terninger var meget udbredt i antikken, hvor de efterhånden ofte blev lavet af brændt ler, marmor eller bronze. Den første sekssidede terning, man kender, er fundet i det nordlige Irak. Den er fra omkring 3000 f. Kr.

og er lavet af brændt ler.

Mennesker har nok spillet siden tidernes morgen, og det må formodes, at der blandt dygtige spillere har eksisteret en intuitiv fornemmelse for det, vi i dag kalder sandsynligheder. Det er dog ret sent, at en egentlig teori for udfaldet af spil udvikles. Det tidligste skriftlige vidnesbyrd om elementære sandsynlighedsberegninger findes i et i middelalderen meget udbredt opbyggeligt latinsk digt *De Vetula* (Om den Gamle Kone), som tilskrives **Richard de Fournival** (1201 – 1260), en gejstlig ved katedralen i Amiens. I forbindelse med en diskussion af terningspil giver digteren en lang (og korrekt) diskussion af antallet af måder, hvorpå man ved kast med tre terninger kan opnå de forskellige mulige summer af terningernes øjne. I digtet anvendes begrebet sandsynlighed ikke, men der tales om hyppighed af forskellige summer. Det vides ikke nøjagtigt, hvornår man nåede frem til mere klare forestillinger om sandsynligheder og viden om, hvordan de kan beregnes. Muligvis blev intet nedskrevet, fordi viden om disse ting havde betydelig pekuniær betydning.

Den første bog *Liber de Ludo Aleae* om beregning af odds og sandsynligheder for en lang række spilsituationer (terningspil, kortspil og astragalus) blev skrevet af **Geronimo Cardano** (1501 – 1576). Bogen er formentlig blevet til omkring 1550, men Cardano, som var en lidenskabelig spiller, holdt den hemmelig, og den blev først fundet og udgivet i 1663. Cardano gav ikke en abstrakt matematisk definition af sine begreber, og bogen er præget af store tabeller over samtlige mulige udfald i forskellige situationer. Det virkelige gennembrud for sandsynlighedsregningen som matematisk disciplin skete i 1654 i en berømt brevveksling mellem **Pierre de Fermat** (1601 – 1665) og **Blaise Pascal** (1623 – 1662), i hvilken de diskuterede og løste en række komplicerede sandsynligheds-teoretiske problemer i relation til spil. I den forbindelse lagde de det teoretiske fundament for det, vi i dag kalder den elementære sandsynlighedsregning. Fermat var stadsadvokat i Toulouse, og var formentlig alle tiders største amatørmatematiker, mens Pascal, der levede af en mindre formue, udover sin indsats som matematiker, fysiker og astronom også var en betydelig filosof. Hverken Fermat eller Pascal samlede deres resultater i bogform. Fermat havde ikke for vane at publicere sine resultater, og Pascal trak sig kort efter brevvekslingen med Fermat tilbage til et kloster i Port Royal, hvor han resten af livet koncentrerede sig om reli-

gionsfilosofi. Det blev derfor hollænderen **Christiaan Huygens** (1629 – 1695), der efter i 1655 at være kommet til Paris samlede og genbeviste deres resultater, som var i omløb blandt byens matematikere. I 1657 udgav han resultaterne i bogen *De Ratiociniis in Alea Ludo*, der blev lærebog for mange sandsynlighedsteoretikere i den anden halvdel af det 17. århundrede. Nogle af Pascals resultater blev iøvrigt udgivet i 1665 i *Traité du Triangle Arithmétique*, som behandler det vi idag kalder Pascals trekant.

Det næste store skridt fremad blev taget af den schweiziske matematiker **Jakob (Jacques) Bernoulli** (1654 – 1705). Hans bidrag til sandsynlighedsregningen blev publiceret i bogen *Ars Conjectandi*, der blev skrevet i perioden 1690 – 1705, men først blev udgivet i 1713. I denne bog indføres sandsynlighed som et matematisk begreb, og de store tals lov vises. Dette resultat er basis for den frekvensfortolkning af sandsynlighedsregningen, som vi skal komme tilbage til. Løst sagt er resultatet, at den frekvens, hvormed en given hændelse indtræffer, når et forsøg udføres mange gange, er tæt på sandsynligheden for hændelsen.

Den definition af en sandsynlighed, som går tilbage til Cardano, Fermat og Pascal, byggede på formlen

$$\frac{\text{antal gunstige udfald}}{\text{antal mulige udfald}}.$$

Den klassiske sandsynlighedsregning, som er baseret på denne formel, kulminerede med udgivelsen i 1812 af **Pierre Simon Laplace's** (1749 – 1827) bog *Théorie Analytique des Probabilités*, hvori han udtømte teoriens matematiske muligheder. Herefter optræder der en periode med stilstand i den matematiske udvikling.

Derimod skete der en rivende udvikling af sandsynlighedsregningens anvendte side langt uden for spillenes område. Sandsynlighedsregningen blev således anvendt inden for statistik, fysik, forsikringsvidenskab, demografi og andre samfundsvidenskaber (for eksempel kriminologi). Som det ofte sker, var anvendelserne her langt foran en stringent matematisk teori. Den klassiske sandsynlighedsdefinition slår ikke til i disse anvendelser. Dette blev helt klart med den komplicerede sandsynlighedsteoretiske model for aktiekursernes udvikling på børsen i Paris, som i 1900 blev udviklet af **Louis Bachelier** (1870 – 1946) i dennes afhandling *Théorie*

de la Spéculation. Præcis den samme model opstilledes i 1905 uafhængigt af Bacheliers arbejde af **Albert Einstein** (1879 – 1955) til beskrivelse af den såkaldte Brownske bevægelse, som mikroskopiske partikler som støv, pollen og bakterier udfører i væsker og luftarter.

Fra den anden halvdel af det 19. århundrede startede en søgen efter et nyt grundlag for sandsynlighedsregningen, især i den russiske skole. **Pafnuti Lvovich Chebyshev** (1821 – 1894) og dennes to studenter **Andrej Andrejevich Markov** (1856 – 1922) og **Alexander Mikhailovich Lyapunov** (1857 – 1918) indførte de stokastiske variable og studerede deres egenskaber. Derigennem var de i stand til at generalisere nogle af sandsynlighedsregningens vigtigste resultater. I bogen *Grundbegriffe der Wahrscheinlichkeitsrechnung* fra 1933 lagde **Andrej Nikolajevich Kolmogorov** (1903 – 1987) den såkaldte målteori til grund for sandsynlighedsregningen, som derved igen fik et stringent matematisk grundlag. Dette satte en eksplosiv udvikling igang, som endnu ikke er afsluttet, og som har fundet meget bred praktisk anvendelse. Således har det vist sig, at den især i 1970'erne udviklede stokastiske analyse, der giver værktøj til studiet af en bred klasse af stokastiske processer, er som skabt til at fungere som grundlag for den moderne matematiske finansieringsteori.

Imidlertid er vi nu langt udenfor, hvad vi kan beskæftige os med i denne bog. Målteori hører ikke hjemme i et førsteårskursus, så selv om vi følger Kolmogorov noget af vejen, kan vi ikke her give en fremstilling af sandsynlighedsregningen, som er matematisk helt tilfredsstillende. Vi vil imidlertid indføre en række centrale begreber og studere disse. Formålet med bogen er ikke mindst at gennemgå sandsynlighedsregningens helt basale regneregler og give en vis fortrolighed med disse. Dels er dette, hvad man har brug for i elementære anvendelser af sandsynlighedsregning (specielt er det nok til at kunne forstå et introducerende statistikkursus), dels er det den bedste ballast for senere at kunne tilegne sig den mere abstrakte teori baseret på målteori.

Før vi går i gang med teorien, er det på sin plads at filosofere lidt over, hvad en sandsynlighed er. Det er der i tidens løb skrevet meget om. I disse noter vil vi betragte det som sandsynlighedsteoriens opgave at beskrive systemer, hvori der er et element af tilfældighed. Her skal systemer forstås i en meget bred forstand. Der kan for eksempel være tale om fysiske, biologiske eller sociale systemer. Det har erfaringsmæssigt vist sig,

at skønt udfaldet af en enkelt tilfældig begivenhed ikke kan forudsiges, så er der for enhver given hændelse en vis tendens i det betragtede system til, at netop denne hændelse indtræffer. Denne tendens kvantificeres ved et tal mellem nul og en, som kaldes sandsynligheden for hændelsen. Sandsynligheder anses således for at være egenskaber ved det studerede system. De kan ikke observeres direkte, men hvis et forsøg udføres mange gange, viser sandsynlighederne sig som de frekvenser, hvormed forskellige hændelser optræder. Hvis vi for eksempel betragter kast med en ærlig terning, har alle terningens sider samme tendens til at optræde, hvorfor sandsynligheden for hændelsen, at der slås en femmer, er $1/6$. Hvis man slår mange gange med terningen, viser denne sandsynlighed sig ved, at cirka $1/6$ af kastene resulterer i en femmer.

Dette er *sandsynlighedsregningens frekvensfortolkning*. Lad os formulere den lidt mere præcist. Antag at vi udfører et forsøg N gange uafhængigt af hinanden, og at det efter hvert forsøg kan konstateres, om en hændelse, som vi betegner med A , er indtruffet. Vi skal senere definere, hvad vi præcist vil forstå ved uafhængighed af forsøg; på dette sted er den intuitive fornemmelse af, hvad uafhængighed er, tilstrækkelig. Under disse antagelser er sandsynligheden for at A indtræffer lig

$$\lim_{N \rightarrow \infty} \frac{\text{antal forsøg hvor } A \text{ indtræffer}}{N}.$$

I 1931, altså ganske kort før Kolmogorov's bog blev publiceret, opbyggede **Richard von Mises** (1883 – 1953) en stringent sandsynlighedsteori, hvor sådanne grænseværdier var den basale definition af sandsynligheder. Disse ideer er aldrig slået an, og de blev fuldstændig overskygget af Kolmogorovs arbejde. Vi vil derfor som alle andre følge Kolmogorov og i Afsnit 1.3 opstille nogle få basale egenskaber, som sandsynligheder rimeligvis må have. Ud fra disse vil vi argumentere os frem til en række resultater og regneregler og bl.a. bevise, at sandsynlighedsregningens frekvensfortolkning (store tals lov) følger af disse basale antagelser. At man kan bevise dette, er en bekræftelse af, at de oprindelige basale antagelser er en fornuftig definition af sandsynlighed.

At sandsynligheden for at få en femmer ved kast med en ærlig terning er lig $1/6$ følger ved en simpel symmetribetragtning (der er ingen forskel på de seks sider). Som regel er det sværere at beregne sandsynligheden

for en hændelse. Man har typisk behov for både sandsynlighedens regneregler og for at gøre nogle antagelser om det studerede system. Ofte må man også inddrage observationer af systemet, hvorved metoder fra statistikken kommer ind i billedet. For eksempel vil vi, for at kunne beregne sandsynligheden for, at der i en klasse med 25 børn er mindst to børn med samme fødselsdag, gøre nogle antagelser. Vi vil blandt andet antage, at børnenes fødselsdage er uafhængige af hinanden, og at antal fødsler per dag ikke varierer i løbet af året. Sådanne antagelser er ikke nødvendigvis helt korrekte. For eksempel varierer antallet af fødsler per dag lidt med årstiden. Antagelserne skal blot være tilstrækkelig tæt på tingenes faktiske tilstand til, at de beregnede sandsynligheder kan anvendes i praksis. Når man præsenteres for et sandsynlighedsudsagn, bør man derfor altid spørge hvilke antagelser, der ligger bag. Så kan man, hvis man har forstand på sandsynlighedsregning, selv vurdere, om det er rimelige antagelser. Et andet eksempel, som vi skal se på, er beregning af sandsynligheden for, at en mand, som overlever sine første 20 år, vil opnå at blive 60 år. For at beregne denne sandsynlighed, som er af stor interesse i forbindelse med livsforsikring, må man gøre antagelser, både for at kunne anvende sandsynlighedsregningens regler og for at kunne inddrage data fra Statistisk Årbog.

1.2 Det endelige sandsynlighedsfelt

I det tilfælde, hvor der kun er endeligt mange mulige udfald, er det problemløst at opstille en sandsynlighedsteoretisk model. Alt hvad vi behøver er en liste over samtlige mulige udfald og en angivelse af sandsynligheden for ethvert af disse. Dette kan formaliseres på følgende måde.

Ved et *endeligt sandsynlighedsfelt* forstås en endelig mængde $E = \{e_1, \dots, e_k\}$, som kaldes *udfaldsrummet*, og en funktion p fra E ind i intervallet $[0, 1]$, som opfylder, at

$$\sum_{j=1}^k p(e_j) = 1. \quad (1.2.1)$$

Elementerne i E kaldes *udfald*, mens en delmængde af E kaldes en *hændelse*. Funktionen p kaldes *sandsynlighedsfunktionen*. Dens værdi i e_j ,

altså $p(e_j)$, er naturligvis sandsynligheden for udfaldet e_j . Man omtaler denne som *punktsandsynligheden* i e_j .

Eksempel 1.2.1 Betragt situationen, hvor man kaster to terninger. Da består E af de 36 mulige udfald

$$E = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\} = \{1, \dots, 6\} \times \{1, \dots, 6\}.$$

Sandsynlighedsfunktionen er givet ved

$$p(x) = \frac{1}{36} \text{ for alle } x \in E,$$

da alle udfald er lige sandsynlige. □

Eksempel 1.2.2 Betragt nu en klasse med 25 børn. Vi vil interessere os for udsagn vedrørende børnenes fødselsdage, så et udfald består af en vektor af 25 fødselsdage $x = (x_1, \dots, x_{25})$. For at forenkle vore overvejelser ser vi bort fra skudår og antager, at alle år har 365 dage. En fødselsdag kan så angives ved dens nummer i året, d.v.s. som et helt tal mellem 1 og 365. Vi kan derfor benytte udfaldsrummet

$$E = \{1, \dots, 365\}^{25} = \underbrace{\{1, \dots, 365\} \times \dots \times \{1, \dots, 365\}}_{25 \text{ gange}}.$$

Vi specificerer sandsynlighedsfunktionen ved

$$p(x) = 365^{-25} \text{ for alle } x = (x_1, \dots, x_{25}) \in E,$$

d.v.s. vi antager at alle kombinationer af fødselsdage er lige sandsynlige. Her er situationen ikke helt så klar, som den var for et kast med to terninger. At alle kombinationer er lige sandsynlige forudsætter bl.a. at børnene ikke fordeles på klasser efter alderskriterier, at alderskriteriet for optagelse på skolen ikke er blevet ændret det år klassen startede i skolen og at antal fødte per dag ikke er årstidsafhængigt (hvilket ikke er helt rigtigt). Videre antages det, at børnenes fødselsdage er uafhængige af hinanden. Vi skal senere give en definition af begrebet stokastisk uafhængighed, som kan bruges til at præcisere dette udsagn. Her er det nok at bemærke, at vi f.eks. udelukker, at der er et tvillingepar i klassen. □

Hvis A er en hændelse, d.v.s. $A \subseteq E$, så defineres sandsynligheden for A , som vi betegner med $P(A)$, ved

$$P(A) = \sum_{x \in A} p(x). \quad (1.2.2)$$

Sandsynligheden for A er altså summen af alle punktsandsynlighederne hørende til de udfald, som ligger i A . Specielt er $P(\{e_j\}) = p(e_j)$, og da vi definerer summen over ingen elementer til at være nul, er $P(\emptyset) = 0$. Det er ikke vanskeligt at indse (udfør argumentet i detaljer), at

$$0 \leq P(A) \leq 1 \quad (1.2.3)$$

$$P(E) = 1 \quad (1.2.4)$$

$$P(A \cup B) = P(A) + P(B), \quad \text{hvis } A \cap B = \emptyset. \quad (1.2.5)$$

Her er A og B naturligvis hændelser, d.v.s. $A, B \subseteq E$. To mængder, hvis fællesmængde er den tomme mængde, siges at være *disjunkte*. Vi kalder en funktion P fra klassen af delmængder af E ind i de reelle tal, som opfylder (1.2.3) – (1.2.5), et *sandsynlighedsmål*.

I de to eksempler ovenfor var sandsynlighedsfunktionen konstant. Dette er naturligvis ikke altid tilfældet, men det er en særligt pæn situation, som er forudsætningen for den kendte formel for sandsynligheden for en hændelse:

$$\frac{\text{antal gunstige udfald}}{\text{antal mulige udfald}}.$$

Hvis nemlig sandsynlighedsfunktionen p er konstant, følger det af formel (1.2.1), at $p(x) = 1/k$, hvor k er antallet af elementer i E . For en hændelse $A \subseteq E$ følger det derfor af (1.2.2) at

$$P(A) = \sum_{x \in A} \frac{1}{k} = \frac{\#A}{\#E}, \quad (1.2.6)$$

hvor $\#A$ betegner antallet af elementer i A , d.v.s. antallet af gunstige udfald, mens $\#E$ betegner antallet af elementer i E , d.v.s. antallet af mulige udfald. Generelt vil vi fremover for enhver endelig mængde M betegne antallet af elementer i M med $\#M$. Når sandsynlighedsfunktionen p er konstant, er sandsynlighedsmassen fordelt ligeligt over den endelige mængde E . Man taler derfor om *ligefordelingen* på E .

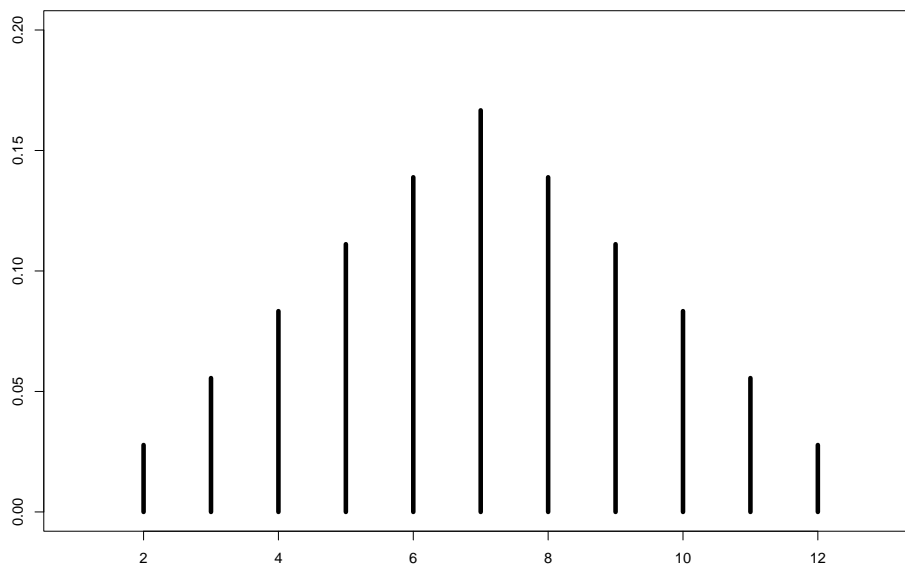
Lad os nu betragte et eksempel, hvor sandsynlighedsfunktionen ikke er konstant.

Eksempel 1.2.3 Betragt igen kast med to terninger, men antag, at vi kun er interesserede i summen af øjnene på de to terninger. Da kan vi benytte udfaldsrummet

$$E = \{2, 3, \dots, 12\},$$

og specificere sandsynlighedsfunktionen

$$\begin{aligned} p(2) = p(12) &= 1/36 \\ p(3) = p(11) &= 1/18 \\ p(4) = p(10) &= 1/12 \\ p(5) = p(9) &= 1/9 \\ p(6) = p(8) &= 5/36 \\ p(7) &= 1/6. \end{aligned}$$



Figur 1.2.1: Sandsynlighedsfunktionen i Eksempel 1.2.3.

Denne specifikation kræver naturligvis et argument. Da modellen i eksempel 1.2.1 er en ligefordeling, kan vi beregne sandsynligheden for at

summen af de to øjne har en bestemt værdi ved hjælp af formel (1.2.6). Således er der kun et enkelt udfald som giver sum to, nemlig (1,1). Derfor er sandsynligheden for at summen er to lig $1/36$. Der er to udfald med sum tre, nemlig (1,2) og (2,1), så sandsynligheden for at summen er tre er lig $2/36 = 1/18$. De øvrige sandsynligheder findes ved tilsvarende overvejelser (udfør nogle af disse). □

1.3 Det generelle sandsynlighedsfelt

Man har ofte brug for sandsynlighedsfelter med uendeligt mange mulige udfald, og i den situation er det mere kompliceret at definere en sandsynlighedsteoretisk model. Hvis antallet af mulige udfald er uendeligt, er det ikke altid tilstrækkeligt blot at angive alle udfald og deres sandsynligheder. Lad os illustrere problemet med et eksempel.

Eksempel 1.3.1 Lad os tænke os, at vi trækker et tilfældigt tal mellem nul og en. Vi vælger altså som udfaldsrum intervallet $E = [0, 1]$, som i hvert fald ikke er en endelig mængde. Det er måske på nuværende tidspunkt ikke helt klart, hvad vi mener med at trække et tilfældigt tal i E , men i hvert fald må alle tal e i E have samme sandsynlighed for at blive udtrukket. Lad os betegne denne sandsynlighed med p . Hvad er sandsynligheden for at udtrække et tal i mængden $\{1/n, 2/n, \dots, (n-1)/n, 1\}$, hvor n er et vilkårligt positivt helt tal? Hvis vi forlanger, at formel (1.2.5) også skal gælde for uendelige sandsynlighedsfelter, ses det ved at benytte denne formel $n-1$ gange, at

$$P(\{1/n, 2/n, \dots, (n-1)/n, 1\}) = P(\{1/n\}) + \dots + P(\{(n-1)/n\}) + P(\{1\}) = np.$$

Da $P(\{1/n, 2/n, \dots, (n-1)/n, 1\})$ skal være en sandsynlighed, d.v.s.

$$0 \leq P(\{1/n, 2/n, \dots, (n-1)/n, 1\}) \leq 1,$$

følger det, at der for alle positive hele tal n gælder, at

$$0 \leq p \leq \frac{1}{n}.$$

Ved at lade n gå mod uendelig ses det, at $p = 0$. I denne situation er det altså ikke særlig nyttigt at angive sandsynligheden for alle mulige udfald. Således er udsagnet, at $P(\{e\}) = 0$ for alle $e \in [0, 1]$, helt uanvendeligt til at sige noget om for eksempel sandsynligheden for at trække et tal mellem $1/4$ og $3/4$. Denne sandsynlighed bør naturligvis være lig en halv.

Det er klart, at vi må gå en anden vej for at definere et sandsynlighedsfelt på en brugbar måde. Vi vil i stedet definere sandsynligheden for at udtrække et tal i et delinterval I af E . En naturlig definition af tilfældig udtrækning af et tal mellem nul og en er at kræve, at sandsynligheden for at trække et tal i intervallet I er proportional med længden af I . Vi antager derfor, at der findes et reelt tal $c \geq 0$, så

$$P(I) = c|I|,$$

hvor $|I|$ betegner længden af I . Vi vil også fremover benytte denne notation for længden af intervaller og af andre delmængder af den reelle akse, for hvilke det giver mening at tale om længden. Det er let at bestemme størrelsen af c , da

$$1 = P([0, 1]) = c|[0, 1]| = c.$$

Det ser altså ud til, at vi kan definere tilfældig udtrækning af et tal mellem nul og en på en fornuftig måde ved at specificere, at

$$P(A) = |A|, \tag{1.3.1}$$

for alle delmængder A af $[0, 1]$, for hvilke det giver mening at tale om længden. Udover intervaller, kan det f. eks. være foreningsmængden af to eller flere intervaller. Da et-punktsmængden $\{e\}$ kan betragtes som et udartet interval $[e, e]$ med længde nul, er (1.3.1) i overensstemmelse med resultatet $P(\{e\}) = 0$, som vi fandt ovenfor.

□

Da det, som vi netop har set, ikke altid er nok at angive samtlige udfald og deres sandsynligheder, må man i stedet angive sandsynligheden for at få et udfald i enhver delmængde af udfaldsrummet (eller i det mindste nogle af delmængderne). Man kan ikke bare angive disse sandsynligheder helt som man har lyst: Sandsynlighederne for forskellige hændelser skal passe sammen. Vi definerer derfor et generelt sandsynlighedsfelt på følgende måde.

Definition 1.3.2 Ved et sandsynlighedsfelt forstås en mængde E , som kaldes udfaldsrummet, en vis klasse \mathcal{E} af delmængder af E , samt en funktion P fra \mathcal{E} ind i intervallet $[0, 1]$. Elementerne i \mathcal{E} skal mindst omfatte både E selv og den tomme mængde \emptyset , mens funktionen P , som kaldes et sandsynlighedsmål, skal opfylde, at

$$P(E) = 1 \quad (1.3.2)$$

$$P(A \cup B) = P(A) + P(B), \quad \text{hvis } A \cap B = \emptyset. \quad (1.3.3)$$

Elementerne i \mathcal{E} kaldes hændelser. Disse er altså delmængder af E . Fortolkningen af sandsynlighedsmålet P er naturligvis, at dets værdi i en hændelse A , d.v.s. $P(A)$, er sandsynligheden for at få et udfald i A . De to betingelser (1.3.2) og (1.3.3) er helt rimelige: Dels skal vi være sikre på at få et udfald i E , dels må sandsynligheden for at få et udfald i enten A eller B være summen af sandsynligheden for et udfald i A og sandsynligheden for et udfald i B , når de to mængder ikke har noget element til fælles. Hvis vi ikke forlangte, at et sandsynlighedsmål opfylder (1.3.3), ville vi ikke kunne bruge det til at komme med konsistente udsagn om forskellige hændelsers sandsynlighed, men ville tværtimod risikere at komme med oplagt forvrøvede udsagn.

Vi har ikke sagt, hvad \mathcal{E} egentlig er for noget. Det skyldes, at vi i disse noter altid vil antage, at \mathcal{E} består af alle delmængder af E . Når E er en endelig mængde, er der ikke noget problem i dette. For visse uendelige mængder, som for eksempel den reelle akse \mathbb{R} og delintervaller af denne, er klassen af alle delmængder imidlertid umådeligt stor. Den er faktisk så stor, at den indeholder visse sære mængder, som man ikke kan tillægge nogen sandsynlighed. Derfor må man, hvis man skal være helt matematisk stringent, "nøjes" med at lade \mathcal{E} være en noget mindre klasse af forholdsvis "pæne" delmængder. De omtalte sære mængder er heldigvis så besynderlige, at man roligt kan sige, at enhver delmængde af \mathbb{R} , som en typisk førsteårs studerende kan forestille sig, kan tillægges en sandsynlighed. Der sker derfor ikke noget ved, at vi her ser bort fra disse problemer, som i praksis er uden betydning for den elementære sandsynlighedsregning. Den antydede problemstilling vil naturligvis blive taget op i et senere kursus, da den har betydning i mere avanceret sandsynlighedsregning.

Fra nu af er en hændelse altså en vilkårlig delmængde af E , hvorfor \mathcal{E} , som vi dermed i dette kursus ikke behøver i definitionen af et sandsynlighedsfelt, ikke vil optræde igen.

Før vi ser nærmere på regnereglerne for et sandsynlighedsmål, skal vi lige have endnu et eksempel på et sandsynlighedsfelt, hvor E er intervallet $[0, 1]$.

Eksempel 1.3.3 Lad $E = [0, 1]$. Vi definerer et sandsynlighedsmål P på E ved for en delmængde A af $[0, 1]$ at sætte

$$P(A) = \int_0^1 1_A(x) 3x^2 dx, \quad (1.3.4)$$

hvor 1_A er den såkaldte *indikatorfunktion* for A , som er defineret ved

$$1_A(x) = \begin{cases} 1 & \text{hvis } x \in A \\ 0 & \text{hvis } x \notin A. \end{cases} \quad (1.3.5)$$

Læg mærke til definitionen af indikatorfunktionen, som vil spille en vigtig rolle fremover.

Sandsynligheden for et delinterval $[a, b]$, hvor $0 \leq a \leq b \leq 1$, er altså

$$P([a, b]) = \int_a^b 3x^2 dx = b^3 - a^3.$$

Specielt er det klart, at $P(E) = 1$, og at en et-punkts mængde $\{a\}$, der jo er lig det udartede interval $[a, a]$, har sandsynlighed nul ligesom i Eksempel 1.3.1. Vi ser endvidere af (1.3.4), at $P(\emptyset) = 0$, idet $1_\emptyset(x)$ er lig med nul for alle $x \in E$.

For alle de hændelser, som vi vil komme ud for, er det klart, at integralet i (1.3.4) er defineret. Hvis for eksempel $A = [a, b] \cup [c, d]$, hvor $0 \leq a \leq b < c \leq d \leq 1$, så er

$$\begin{aligned} P(A) &= \int_a^b 3x^2 dx + \int_c^d 3x^2 dx \\ &= b^3 - a^3 + d^3 - c^3 = P([a, b]) + P([c, d]). \end{aligned}$$

Vi ser, at (1.3.3) er opfyldt. Dette gælder faktisk generelt. Hvis nemlig A og B er disjunkte mængder af $[0, 1]$, er $1_{A \cup B}(x) = 1_A(x) + 1_B(x)$ (overvej

hvorfor dette er rigtigt), og dermed er

$$\begin{aligned} P(A \cup B) &= \int_0^1 1_{A \cup B}(x) 3x^2 dx & (1.3.6) \\ &= \int_0^1 1_A(x) 3x^2 dx + \int_0^1 1_B(x) 3x^2 dx = P(A) + P(B). \end{aligned}$$

Til slut bemærkes det, at funktionen $3x^2$ på $[0, 1]$, som vi her har benyttet til at definere sandsynlighederne, er et eksempel på en såkaldt *sandsynlighedstæthed*. Vi skal i Kapitel 5 vende tilbage til dette begreb. \square

Sætning 1.3.4 *For et sandsynlighedsmål P gælder følgende regneregler, hvor A og B betegner vilkårlige hændelser.*

1) Hvis $B \subseteq A$, er

$$P(A \setminus B) = P(A) - P(B) \quad (1.3.7)$$

og

$$P(B) \leq P(A). \quad (1.3.8)$$

2) Sandsynligheden for den komplementære hændelse til B , d.v.s. $E \setminus B$, er givet ved

$$P(E \setminus B) = 1 - P(B). \quad (1.3.9)$$

3)

$$P(\emptyset) = 0. \quad (1.3.10)$$

4)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.3.11)$$

og

$$P(A \cup B) \leq P(A) + P(B). \quad (1.3.12)$$

Bevis: 1) Hvis $B \subseteq A$, er $A = (A \setminus B) \cup B$, og da $(A \setminus B) \cap B = \emptyset$, følger det af (1.3.3) at $P(A) = P(A \setminus B) + P(B)$, hvilket er ækvivalent med (1.3.7). Uligheden (1.3.8) følger af (1.3.7), da $P(A \setminus B) \geq 0$. 2) Ligningen (1.3.9) er et specialtilfælde af (1.3.7) med $A = E$. 3) Da $E \setminus E = \emptyset$, følger (1.3.10) af (1.3.9) ved at sætte $B = E$. 4) Endelig indses (1.3.11) ved at

bemærke, at $A \cup B = (A \setminus (A \cap B)) \cup B$, hvor selvfølgelig $A \setminus (A \cap B)$ og B er disjunkte. Det følger derfor af (1.3.3) og (1.3.7), at

$$\begin{aligned} P(A \cup B) &= P(A \setminus (A \cap B)) + P(B) \\ &= P(A) - P(A \cap B) + P(B). \end{aligned}$$

Den sidste ulighed er en konsekvens af (1.3.11). □

Bemærk, at (1.3.11) viser, at en formel som den i (1.3.3) ikke gælder uden en betingelse på $A \cap B$. Den forudsætter, at $P(A \cap B) = 0$, hvilket bl.a. er tilfældet, når $A \cap B = \emptyset$.

Trods sin enkelhed, er resultatet (1.3.9) umådeligt nyttigt, når man skal beregne sandsynligheden for en kompliceret hændelse, hvis komplementærhændelse er enkel. Her er et eksempel.

Eksempel 1.2.2 (fortsat). Vi betragter igen en klasse med 25 børn. Vi har i forrige afsnit opstillet og diskuteret en sandsynlighedsteoretisk model, som er egnet til at beregne sandsynligheder for hændelser vedrørende børnenes fødselsdage. Et klassisk spørgsmål er, hvad sandsynligheden er for, at to eller flere børn har samme fødselsdag. Denne er overraskende stor. Lad altså A betegne hændelsen, at to eller flere børn har samme fødselsdag. Dette er jo en temmelig kompliceret hændelse, men den komplimentære hændelse $E \setminus A$ er den meget enkle hændelse, at alle børn har fødselsdag på forskellige dage.

Da sandsynlighedsfunktionen er konstant, kan vi beregne sandsynligheden for $E \setminus A$ ved formlen (1.2.6). Vi skal derfor tælle, hvor mange elementer der er i $E \setminus A$. Dette kan gøres på følgende måde. Der er 365 måder at vælge det første barns fødselsdag på. Når vi har lagt os fast på det første barns fødselsdag, er der, uanset hvilken dag dette er, 364 muligheder tilbage for det næste barns fødselsdag. Ellers ville de to jo ikke have forskellige fødselsdage. Videre er der, når både første og andet barns fødselsdage er valgt, 363 muligheder tilbage til det tredje barn. Således fortsættes til man når det 25. barn, som altid vil have 341 ($= 365 - 24$) muligheder. Vi kan konkludere, at

$$\#(E \setminus A) = 365 \cdot 364 \cdot \dots \cdot 341,$$

således at (1.2.6) giver, at

$$P(E \setminus A) = \frac{\#(E \setminus A)}{\#E} = \frac{365 \cdot 364 \cdot \dots \cdot 341}{365^{25}} = 0.4313.$$

Nu følger det af (1.3.9), at

$$P(A) = 1 - 0.4313 = 0.5687.$$

Man må altså forvente, at der i over halvdelen af alle klasser (med 25 børn) er mindst to børn, som har fødselsdag samme dag. \square

Vi slutter dette afsnit med generaliseringer af (1.3.12) og (1.3.3).

Sætning 1.3.5 *Lad A_1, \dots, A_n være vilkårlige hændelser. Da er*

$$P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n). \quad (1.3.13)$$

Hvis det antages, at A_1, \dots, A_n er indbyrdes disjunkte (d.v.s. $A_i \cap A_j = \emptyset$ for alle $i \neq j$), er

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n). \quad (1.3.14)$$

Bevis: Uligheden (1.3.13) vises ved et induktionsbevis. At udsagnet er korrekt for $n = 2$, er præcis det tidligere viste resultat (1.3.12). Antag nu, at uligheden (1.3.13) gælder for $n - 1$ hændelser A_1, \dots, A_{n-1} , og definer hændelsen $B = A_1 \cup \dots \cup A_{n-1}$. Da er

$$\begin{aligned} P(A_1 \cup \dots \cup A_{n-1} \cup A_n) &= P(B \cup A_n) \\ &\leq P(B) + P(A_n) \\ &= P(A_1 \cup \dots \cup A_{n-1}) + P(A_n) \\ &\leq P(A_1) + \dots + P(A_{n-1}) + P(A_n). \end{aligned}$$

Formel (1.3.14) vises ved induktion på tilsvarende måde (overvej detaljerne). \square

1.4 Betingede sandsynligheder

Eksempel 1.2.1 (fortsat). Vi betragter igen kast med to terninger. Antag at terningerne er blevet kastet uset af os, og at vi kun får oplyst, at de to terninger viser det samme antal øjne. Hvad kan vi på den baggrund sige om sandsynligheden for at summen af øjnene er mindst 11? Det er klart, at den oplysning, vi har fået, må påvirke vores udsagn om sandsynligheden, så svaret er nok ikke $1/12$.

Udfaldsrummet

$$E = \{1, \dots, 6\} \times \{1, \dots, 6\}$$

består af 36 elementer, som alle har sandsynlighed $1/36$, men nu har vi jo fået af vide, at de to terninger viser det samme, d.v.s. at udfaldet tilhører hændelsen

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}.$$

Det er derfor fristende, at betragte A som et nyt udfaldsrum. Da elementerne i det oprindelige udfaldsrum E havde samme sandsynlighed, må det samme gælde for A , så hvert af de 6 elementer i A må have sandsynlighed $1/6$. Da $(6, 6)$ er det eneste udfald i A med sum større end eller lig 11, er et godt bud på den søgte sandsynlighed øjensynligt $1/6$.

Lad os forsøge os med en frekvensfortolkning. Vi tænker os, at vi n gange udfører det forsøg, som består i at slå med to terninger. Med N_A betegner vi antallet af forsøg, i hvilke udfaldet tilhørte A , og med $N_{A \cap B}$ betegner vi antallet af forsøg, i hvilke udfaldet tilhørte både hændelsen A og hændelsen, at summen er større end eller lig 11. Den sidste hændelse betegner vi fra nu af med B . Antallet $N_{A \cap B}$ er altså bare antal forsøg med udfaldet $(6, 6)$. Den relative hyppighed af hændelsen B blandt de forsøg, hvor de to terninger viste det samme, er da

$$\frac{N_{A \cap B}}{N_A} = \frac{f_{A \cap B}}{f_A},$$

hvor $f_{A \cap B} = N_{A \cap B}/n$ og $f_A = N_A/n$ betegner den relative hyppighed af de to hændelser $A \cap B$ og A blandt alle n forsøg. Størrelserne $f_{A \cap B}$ og f_A svarer til vores intuitive opfattelse af sandsynlighederne $P(A \cap B)$

og $P(A)$, og sandsynlighedsregningens frekvensfortolkning siger, at $f_{A \cap B}$ og f_A konvergerer mod henholdsvis $P(A \cap B)$ og $P(A)$, når n går mod uendelig. Brøken ovenfor konvergerer altså mod

$$\frac{P(A \cap B)}{P(A)} = \frac{1/36}{1/6} = \frac{1}{6},$$

som må være et godt bud på den søgte betingede sandsynlighed af hændelsen B givet at vi ved, at udfaldet ligger i A . Dette sidste bud er heldigvis identisk med det tidligere bud. □

Inspireret af dette eksempel giver vi følgende definition. Lad P være et sandsynlighedsmål på udfaldsrummet E , lad A og B være to hændelser, og antag at $P(A) > 0$.

Definition 1.4.1 Den betingede sandsynlighed af B givet A , som vi skriver $P(B|A)$, er defineret ved

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (1.4.1)$$

Fortolkningen af den betingede sandsynlighed er, at $P(B|A)$ er sandsynligheden for at få et udfald i B , når vi ved, at udfaldet ligger i A . Vi siger, at vi betinger med A . Den betingede sandsynlighed er altså forholdet mellem sandsynligheden for den del af B , som ligger i A , og sandsynligheden for hele A .

Eksempel 1.2.1 (fortsat). Igen kastes to terninger, uden at vi ser udfaldet. Vi får at vide, at summen af øjnene er større end eller lig 9, og vil finde den betingede sandsynlighed for, at terningerne viser det samme antal øjne. Hændelsen, at summen af øjnene er større end eller lig 9, er

$$A = \{(3, 6), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\},$$

mens hændelsen, at terningerne viser det samme antal øjne, er

$$B = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}.$$

Da A består af 10 elementer, mens

$$A \cap B = \{(5, 5), (6, 6)\}$$

består af 2 elementer, er

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/36}{10/36} = \frac{1}{5}.$$

□

Vi vil i resten af dette afsnit give nogle nyttige regneregler for betingede sandsynligheder. I det følgende tænker vi os hele tiden, at et sandsynlighedsfelt med udfaldsrum E og sandsynlighedsmål P er givet.

Sætning 1.4.2 *Lad A_1, A_2, \dots, A_n være n hændelser, som opfylder, at $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Da er*

$$P(A_1 \cap \dots \cap A_n) = \tag{1.4.2} P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Bevis: Bemærk først, at betingelsen $P(A_1 \cap \dots \cap A_{n-1}) > 0$ medfører, at alle de betingede sandsynligheder er veldefinerede (overvej dette). Indsæt definitionen af betinget sandsynlighed overalt på højre side. Derved opnås et produkt af brøker, hvor hver tæller (bortset fra den sidste) forkorter ud mod næste brøks nævner. Tilbage står det ønskede resultat.

□

Eksempel 1.4.3 I dette eksempel betragtes livslængden for en dansker af hankøn født i 1989. Som udfaldsrum er det naturligt at vælge $E = \mathbb{N}_0 = \{0, 1, 2, \dots\}$, altså de ikke-negative hele tal. Da vi endnu ikke rigtigt har diskuteret, hvordan man definerer et sandsynlighedsmål på en uendelig mængde som \mathbb{N}_0 , vil vi i stedet vælge $E = \{0, 1, 2, \dots, 500\}$ som udfaldsrum. Det burde ikke give problemer.

Lad os diskutere, hvordan man beregner sandsynligheden for, at en mand, som overlever sine første 20 år, også vil være i live 40 år senere, og altså vil opleve sin 60 års fødselsdag. Spørgsmål af denne art interesserer pensionskasser meget. Lad A_k betegne hændelsen, at manden bliver

mindst k år gammel, d.v.s. $A_k = \{k, k + 1, \dots, 500\}$. Den omtalte sandsynlighed er da den betingede sandsynlighed

$$P(A_{60}|A_{20}) = \frac{P(A_{60} \cap A_{20})}{P(A_{20})} = \frac{P(A_{60})}{P(A_{20})},$$

hvor vi har brugt, at $A_{60} \subseteq A_{20}$. Faktisk gælder der jo, at

$$A_{60} \subseteq A_{59} \subseteq \dots \subseteq A_2 \subseteq A_1 \subseteq A_0 = E,$$

så $A_0 \cap \dots \cap A_k = A_k$, hvorfor formel (1.4.2) giver, at

$$P(A_{60}|A_{20}) = P(A_{21}|A_{20})P(A_{22}|A_{21}) \cdots P(A_{60}|A_{59}).$$

Vi kan altså beregne $P(A_{60}|A_{20})$, hvis vi for de mellemliggende år kender de betingede sandsynligheder af typen “sandsynligheden for at blive $k + 1$ år, givet at man er blevet k år”. En mulig fremgangsmåde er at antage, at de betingede sandsynligheder, $P(A_{k+1}|A_k)$, ikke afhænger af, hvilket år personen er blevet født, og ganske enkelt for alle k bruge de tilsvarende relative hyppigheder observeret f.eks. i 2005-2006. Dødelighedstavlerne i Statistisk Årbog er beregnet ved hjælp af den netop udledte succesive multiplikationsregel.

□

Sætning 1.4.2 er især anvendelig, når man som i eksemplet betragter forløb over tiden. Den næste sætning er nyttig i en anden type situationer, nemlig når man opnår en forenkling af en hændelse ved at opdele den i en række tilfælde. Hermed menes, at det for hvert af disse tilfælde ikke er vanskeligt at finde den betingede sandsynlighed for hændelsen. Eksemplet efter sætningen vil forhåbentlig gøre dette udsagn lidt klarere.

Sætning 1.4.4 *Lad A_1, A_2, \dots, A_n være indbyrdes disjunkte hændelser, som opfylder, at $E = A_1 \cup A_2 \cup \dots \cup A_n$, og at $P(A_i) > 0$ for $i = 1, \dots, n$. For en vilkårlig hændelse B gælder da, at*

$$P(B) = \sum_{j=1}^n P(B|A_j)P(A_j). \quad (1.4.3)$$

Bevis: Da $E = A_1 \cup A_2 \cup \dots \cup A_n$, følger det, at

$$B = E \cap B = (A_1 \cap B) \cup \dots \cup (A_n \cap B),$$

og da $A_1 \cap B, \dots, A_n \cap B$ er indbyrdes disjunkte, giver (1.3.14) at

$$P(B) = \sum_{j=1}^n P(A_j \cap B).$$

Da endvidere

$$P(A_j \cap B) = \frac{P(A_j \cap B)}{P(A_j)} P(A_j) = P(B|A_j)P(A_j),$$

er sætningen vist. □

Eksempel 1.4.5 På et bord står to kasser. Den ene indeholder 50 røde og 50 hvide kugler. Den anden indeholder 20 røde, 80 hvide og 50 sorte kugler. Ved tilfældig lodtrækning (f. eks. med en ærlig mønt) vælges en kasse, og en kugle fra denne udtages tilfældigt. Hvad er sandsynligheden for at trække en rød kugle? Som udfaldsrum kan vi tage $E = \{1, 2\} \times \{\text{rød, hvid, sort}\}$, hvor 1 og 2 nummererer kasserne. Lad K_i være hændelsen, at vi vælger kasse nummer i , d.v.s. $K_i = \{i\} \times \{\text{rød, hvid, sort}\}$ for $i = 1, 2$. Lad tilsvarende R, H og S betegne hændelserne, at vi trækker henholdsvis en rød, en hvid og en sort kugle (f.eks. er $R = \{1, 2\} \times \{\text{rød}\}$). Vi kan umiddelbart sige, at $P(K_1) = P(K_2) = \frac{1}{2}$. Det er også klart, at det er smart at dele op i de to tilfælde, hvor vi har valgt henholdsvis kasse nummer et og kasse nummer to. Det er nemlig let at angive sandsynligheden for at trække en kugle af en bestemt farve, hvis vi ved hvilken kasse, der trækkes fra. Men det betyder netop, at vi let kan angive følgende betingede sandsynligheder:

$$\begin{aligned} P(R|K_1) &= \frac{50}{100} = \frac{1}{2} & P(R|K_2) &= \frac{20}{150} = \frac{2}{15} \\ P(H|K_1) &= \frac{50}{100} = \frac{1}{2} & P(H|K_2) &= \frac{80}{150} = \frac{8}{15} \\ P(S|K_1) &= \frac{0}{100} = 0 & P(S|K_2) &= \frac{50}{150} = \frac{1}{3} \end{aligned}$$

Nu kan vi bruge sætning 1.4.4 til at finde sandsynligheden for at udtrække en rød kugle:

$$\begin{aligned} P(R) &= P(R|K_1)P(K_1) + P(R|K_2)P(K_2) \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{15} \cdot \frac{1}{2} = \frac{19}{60} = 0.3167. \end{aligned}$$

Det er naturligvis lige så let at udregne sandsynlighederne $P(H)$ og $P(S)$, men disse regnestykker overlades til den interesserede læser. \square

Følgende “omvendingsformel” for betingede sandsynligheder kan også være nyttig.

Sætning 1.4.6 *For vilkårlige hændelser A og B , som opfylder, at $P(A) > 0$ og $P(B) > 0$, gælder*

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A). \quad (1.4.4)$$

Bevis: Resultatet følger af definitionen af betinget sandsynlighed:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(B)}P(B|A).$$

\square

Eksempel 1.4.5 (fortsat). Lad os fortsætte diskussionen af kugler i kasser. Vi forstiller os nu, at en person uset af os udfører det omtalte eksperiment, men at vedkommende kun fortæller os farven på den udtrukne kugle, ikke hvilken kasse der blev valgt. Da giver det mening at spørge om den betingede sandsynlighed for, at kasse nr. 1 blev valgt, givet at kuglen er rød? Formel (1.4.4) giver, at

$$P(K_1|R) = \frac{P(K_1)}{P(R)}P(R|K_1) = \frac{1/2}{19/60} \cdot \frac{1}{2} = \frac{15}{19} = 0.7895.$$

Denne problemstilling illustrerer et væsentligt aspekt af begrebet betinget sandsynlighed. Uden kendskab til forsøgets udfald ville vi naturligvis sige, at sandsynligheden for at kuglen kommer fra kasse nr. 1 er $1/2$.

Men når vi modtager den ekstra information, at kuglen er rød, ændres denne vurdering, fordi en rød kugle er et mere sandsynligt resultat af en udtrækning fra kasse 1 end fra kasse 2. Man kan sige, at kuglens farve indeholder information om, hvilken kasse den kommer fra. En betinget sandsynlighed er en sandsynlighed, der er udregnet under hensyn til den information, der ligger i, at den hændelse, vi betinger med, er indtruffet. \square

Ved at kombinere de to foregående sætninger opnår man følgende nyttige regneregler, som er kendt under navnet *Bayes' formel*.

Sætning 1.4.7 *Lad A_1, A_2, \dots, A_n være indbyrdes disjunkte hændelser, som opfylder, at $E = A_1 \cup A_2 \cup \dots \cup A_n$, og at $P(A_i) > 0$ for $i = 1, \dots, n$. For en hændelse B med $P(B) > 0$ gælder for ethvert $k = 1, \dots, n$, at*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{j=1}^n P(B|A_j)P(A_j)}. \quad (1.4.5)$$

Bevis: Ifølge (1.4.4) er

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)},$$

og ved at benytte (1.4.3) i nævneren, opnås (1.4.5). \square

Bemærk, at udregningen i anden halvdel af Eksempel 1.4.5 også kan ses som en anvendelse af Bayes' formel, da vi jo netop i eksemplets første del havde beregnet $P(R)$ v.h.a. (1.4.3).

Lad os slutte dette afsnit med at studere, hvordan den betingede sandsynlighed $P(B|A)$ opfører sig som funktion af B .

Sætning 1.4.8 *Lad A være en hændelse, så $P(A) > 0$. Hændelsen A holdes fast. Da er funktionen*

$$B \mapsto P(B|A), \quad (1.4.6)$$

hvor B er en vilkårlig hændelse, et sandsynlighedsmål.

Bevis: Da $P(B|A) \in [0, 1]$, skal vi bare eftervise (1.3.2) og (1.3.3), hvoraf den første hurtigt kan klares, da $P(E|A) = P(E \cap A)/P(A) = P(A)/P(A) = 1$. For at vise (1.3.3) betragter vi to disjunkte hændelser B og C . Da også $B \cap A$ og $C \cap A$ er disjunkte, er

$$\begin{aligned} P(B \cup C|A) &= \frac{P(A \cap (B \cup C))}{P(A)} = \frac{P((A \cap B) \cup (A \cap C))}{P(A)} \\ &= \frac{P(A \cap B) + P(A \cap C)}{P(A)} = P(B|A) + P(C|A). \end{aligned}$$

□

Sandsynlighedsmålet givet ved (1.4.6) kaldes *den betingede fordeling givet A*. Et sandsynlighedsmål kaldes somme tider en sandsynlighedsfordeling; især når udfaldsrummet er \mathbb{R} eller \mathbb{R}^n .

1.5 Stokastisk uafhængighed

Lad A og B være to hændelser med $P(B) > 0$. I almindelighed er $P(A|B) \neq P(A)$, idet oplysningen om, at B er indtruffet, vil ændre sandsynligheden for A . Vi så et eksempel på dette i Eksempel 1.4.5. I de tilfælde, hvor oplysningen om, at B er indtruffet ikke har indflydelse på sandsynligheden for at A indtræffer, d.v.s. hvis

$$P(A|B) = P(A), \tag{1.5.1}$$

siger vi, at A er uafhængig af B . Hvis vi f.eks. slår plat og krone, og hvis vi lader A være hændelsen, at vi får krone i første slag, og B hændelsen, at vi får krone i andet slag, så er det klart, at A og B må være uafhængige. Vi ser da også, at

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B).$$

Ved at indsætte definitionen af den betingede sandsynlighed, ses det at (1.5.1) er ækvivalent med $P(A \cap B) = P(A) \cdot P(B)$. Denne ligning har den fordel, at man ikke behøver at antage, at $P(B) > 0$, samt at den er symmetrisk i A og B . Derfor defineres uafhængighed af to hændelser på følgende måde.

Definition 1.5.1 *To hændelser A og B siges at være stokastisk uafhængige, hvis*

$$P(A \cap B) = P(A) \cdot P(B). \quad (1.5.2)$$

Når misforståelser er udelukket, siger man som regel bare, at A og B er uafhængige.

Eksempel 1.2.1 (fortsat). Betragt endnu en gang kast med to terninger. Udfaldsrummet er $E = E_1 \times E_2$, hvor $E_1 = E_2 = \{1, 2, 3, 4, 5, 6\}$, og sandsynlighedsfunktionen er konstant (lig $1/36$). Lad os antage at den første terning er rød og den anden sort, så vi kan skelne dem. Hændelser, der kun vedrører antallet af øjne på den røde terning, har formen $A_1 \times E_2$, hvor $A_1 \subseteq E_1$, mens hændelser, der kun vedrører den sorte terning har formen $E_1 \times A_2$, hvor $A_2 \subseteq E_2$. For eksempel er $\{5\} \times E_2 = \{(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6)\}$ hændelsen, at første terning viser 5. To hændelser af typen $B_1 = A_1 \times E_2$ og $B_2 = E_1 \times A_2$, hvor $A_1 \subseteq E_1$ og $A_2 \subseteq E_2$, er uafhængige, som man let ser:

$$\begin{aligned} P(B_1 \cap B_2) &= P(A_1 \times A_2) = \frac{\#(A_1 \times A_2)}{36} \\ &= \frac{\#A_1 \cdot \#A_2}{36} = \frac{\#(A_1 \times E_2)}{36} \cdot \frac{\#(E_1 \times A_2)}{36} = P(B_1) \cdot P(B_2). \end{aligned}$$

Som sædvanlig betegner $\#M$ antallet af elementer i mængden M . Vi har altså vist, at den sandsynlighedsteoretiske model, som vi tidligere opstillede for kast med to terninger, medfører, at de to terninger er uafhængige. Hvis det ikke havde været tilfældet, måtte der enten være noget galt med definitionen af stokastisk uafhængighed eller med vores model. Heldigvis er begge OK.

Bemærk, at ovenstående ræsonnement kan gennemføres for ethvert endeligt sandsynlighedsfelt, hvor udfaldsrummet er et produkt af to mængder, og hvor sandsynlighedsfunktionen er konstant. Prøv f.eks. igen at overveje den beregning vedrørende møntkast, som blev givet lige før Definition 1.5.1 .

□

Eksempel 1.5.2 Et kort udtages tilfældigt fra et spil kort. Lad A , B og C være hændelserne

$$\begin{aligned} A &= \{ \text{vi trækker en klør} \} \\ B &= \{ \text{vi trækker et es} \} \\ C &= \{ \text{vi trækker en sort konge} \}. \end{aligned}$$

Intuitivt vil man nok forvente, at A og B er uafhængige, da hændelsen at trække et es ikke synes at have indflydelse på, om dette er en klør eller ikke. Det er da også rigtigt:

$$P(A) = \frac{1}{4}, \quad P(B) = \frac{1}{13} \quad \text{og} \quad P(A \cap B) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = P(A) \cdot P(B).$$

Derimod vil vi nok forvente, at information om, at vi har trukket en klør, vil forøge sandsynligheden for at trække en sort konge. Også dette kan bekræftes:

$$P(C) = \frac{1}{26}, \quad \text{så} \quad P(A) \cdot P(C) = \frac{1}{104}, \quad \text{mens} \quad P(A \cap C) = \frac{1}{52},$$

så A og C er ikke uafhængige. Vi ser videre, at

$$P(C|A) = \frac{1/52}{1/4} = \frac{1}{13} > P(C).$$

Hændelserne B og C er klart ikke uafhængige, da de er disjunkte.

Betragt nu hændelserne

$$\begin{aligned} D_1 &= \{ \text{vi trækker en klør eller en hjerter} \} \\ D_2 &= \{ \text{vi trækker et rødt kort} \}. \end{aligned}$$

Her er det vel ikke ganske klart om D_1 og D_2 er uafhængige, men det er de:

$$P(D_1) = \frac{1}{2}, \quad P(D_2) = \frac{1}{2} \quad \text{og} \quad P(D_1 \cap D_2) = \frac{1}{4} = P(D_1) \cdot P(D_2).$$

□

Eksempel 1.5.3 En mønt kastes n gange. I dette tilfælde er udfaldsrummet $E = \{P, K\}^n$, og sandsynlighedsfunktionen tillægger alle 2^n udfald samme sandsynlighed, nemlig 2^{-n} . Lad A være hændelsen “både plat og krone forekommer” og B hændelsen “der er højst én plat”. Er A og B uafhængige? Lad os først finde $P(A)$. Dette er en meget stor hændelse, men den komplementære hændelse består kun af to elementer: $E \setminus A = \{KK \cdots K, PP \cdots P\}$ (med en forhåbentlig indlysende notation). Derfor er

$$P(A) = 1 - P(E \setminus A) = 1 - 2 \cdot 2^{-n}.$$

Da $B = \{KK \cdots K, PK \cdots K, KPK \cdots K, \dots, KK \cdots KP\}$, som har $n + 1$ elementer, er

$$P(B) = (n + 1) \cdot 2^{-n}.$$

Endelig er $A \cap B$ hændelsen “der er præcis en plat”, d.v.s. $A \cap B = \{PK \cdots K, KPK \cdots K, \dots, KK \cdots KP\}$, som har n elementer, så

$$P(A \cap B) = n \cdot 2^{-n}.$$

Da altså

$$P(A) \cdot P(B) = (n + 1)(1 - 2^{-n+1})2^{-n},$$

kommer vi til den lidt overraskende konklusion, at A og B er uafhængige, hvis $n = 3$, mens de ikke er uafhængige, når $n = 2, 4, 5, 6, \dots$. Man skal være varsom med hurtige konklusioner i sandsynlighedsregning. \square

Vi vil nu se på, hvordan man definerer uafhængighed af mere end to hændelser. Følgende eksempel illustrerer, at man skal passe lidt på.

Eksempel 1.5.2 (fortsat). Igen udtages et kort tilfældigt fra et spil kort. Lad D_1 og D_2 være som tidligere, og definer

$$D_3 = \{ \text{vi trækker en spar eller en hjerter} \}.$$

At D_3 og D_2 er uafhængige, vises på helt samme måde, som vi viste, at D_1 og D_2 er uafhængige. At også D_1 og D_3 er uafhængige, ses igen på samme måde. Vi har altså, at de tre hændelser D_1 , D_2 og D_3 parvist

er uafhængige, men da $D_1 \cap D_3 \subseteq D_2$, kan $D_1 \cap D_3$ og D_2 ikke være uafhængige. Vi ser da også, at

$$P((D_1 \cap D_3) \cap D_2) = P(\{\text{vi trækker en hjerter}\}) = \frac{1}{4},$$

mens

$$P(D_1 \cap D_3) \cdot P(D_2) = P(D_1) \cdot P(D_3) \cdot P(D_2) = \frac{1}{8}.$$

Det er altså ikke rimeligt at sige, at hændelserne D_1 , D_2 og D_3 er uafhængige.

□

Ovenstående eksempel viser, at det for tre hændelser A , B og C ikke i almindelighed gælder, at ligningerne

$$P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C), P(B \cap C) = P(B)P(C)$$

medfører at

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Den sidste ligning må medtages i definitionen af stokastisk uafhængighed af A , B og C . Vi har følgende generelle definition.

Definition 1.5.4 *De n hændelser A_1, \dots, A_n ($n \geq 2$) siges at være indbyrdes stokastisk uafhængige, hvis der for enhver delmængde $\{i_1, \dots, i_k\}$ af $\{1, \dots, n\}$ gælder at*

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Hvis n er stor er der tale om ganske mange ligninger. Som regel udelader man ordet stokastisk, og siger bare, at A_1, \dots, A_n er indbyrdes uafhængige.

Eksempel 1.5.3 (fortsat). Da vi før diskuterede møntkast, sagde vi, at udfaldsrummet er $E = \{P, K\}^n$, og sandsynlighedsfunktionen er konstant lig 2^{-n} , hvilket jo også er yderst rimeligt. Nu kan vi også argumentere på en anden måde. I stedet for fra starten at antage, at sandsynlighedsfunktionen er konstant på E , kan vi nemlig antage at møntkastene er indbyrdes uafhængige i den forstand at, hvis A_1, \dots, A_n er hændelser, hvor

A_i kun vedrører udfaldet af det *ite* møntkast, så er disse hændelser indbyrdes uafhængige. Hvis vi yderligere antager, at der for hvert møntkast gælder, at udfaldene plat og krone har samme sandsynlighed, følger det, at sandsynligheden for ethvert udfald i E er 2^{-n} . Lad os nøjes med at betragte $n = 5$. Med en forhåbentlig indlysende (men unægteligt noget kritisabel) notation er $P(PKPKK) = P(P)P(K)P(P)P(K)P(K) = 2^{-5}$.

□

Sætning 1.5.5 *Lad A , B og C være indbyrdes uafhængige hændelser. Da gælder følgende:*

- 1) $A \setminus B$ og C er uafhængige,
- 2) $A \cap B$ og C er uafhængige,
- 3) $A \cup B$ og C er uafhængige,
- 4) $E \setminus A$, B og C er indbyrdes uafhængige.

Bevis: Vi viser kun 1): $P((A \setminus B) \cap C) = P((A \cap C) \setminus (A \cap B \cap C)) = P(A \cap C) - P(A \cap B \cap C) = P(A)P(C) - P(A \cap B)P(C) = P(A \setminus B)P(C)$. De øvrige resultater er også lette at vise. De vises i opgave 1.25.

□

Der findes naturligvis også en version af Sætning 1.5.5 for n indbyrdes uafhængige hændelser.

Det viser sig, at det ret enkle begreb uafhængighed i mange tilfælde ikke er tilstrækkeligt til at beskrive mere komplekse sammenhænge mellem fænomener. I sådanne situationer har begrebet *betinget uafhængighed* vist sig særdeles nyttigt.

Definition 1.5.6 *Lad A , B og C være tre hændelser. Da siges A og B at være betinget uafhængige givet C såfremt*

$$P(A \cap B | C) = P(A | C) \cdot P(B | C). \quad (1.5.3)$$

Begrebet betinget uafhængighed er relativt kompliceret og ikke helt let at få en god intuition for. Grundig overvejelse af det følgende eksempel og af Eksempel 1.5.8 burde imidlertid hjælpe på forståelsen.

Eksempel 1.4.5 (fortsat). Igen står der to kasser på et bord, hvoraf den ene indeholder 50 røde og 50 hvide kugler, mens den anden indeholder

20 røde, 80 hvide og 50 sorte kugler. Igen vælger en person (uset af os) en kasse ved tilfældig lodtrækning, udtager derefter tilfældigt en kugle fra kassen, og oplyser os kun om kuglens farve. Vi bemærker os farven, hvorefter personen lægger kuglen tilbage i sin kasse og derefter igen tilfældigt trækker en kugle fra den samme kasse. Vi oplyses også om den anden kugles farve. Betragt hændelserne

$$\begin{aligned} A &= \{\text{første kugle er rød}\} \\ B &= \{\text{anden kugle er rød}\} \\ K_1 &= \{\text{kasse nummer et blev valgt}\} \\ K_2 &= \{\text{kasse nummer to blev valgt.}\} \end{aligned}$$

Når personen først har valgt f.eks. kasse nummer et, er der jo bare tale om to identiske og *uafhængige* forsøg, som hvert består i at trække en kugle. Derfor er hændelserne A og B betinget uafhængige givet K_1 , og

$$P(A \cap B | K_1) = P(A | K_1) \cdot P(B | K_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Hændelserne A og B er selvfølgelig også betinget uafhængige givet K_2 . Derimod er A og B *ikke* uafhængige. Lad os indse dette. Vi viste tidligere, at $P(A) = 19/60$ (den gang hed den samme hændelse R). Da den anden udtrækning på ingen måde adskiller sig fra den første, er også $P(B) = 19/60$, så $P(A)P(B) = 361/3600$. Formel (1.4.3) giver, at

$$\begin{aligned} P(A \cap B) &= P(A \cap B | K_1)P(K_1) + P(A \cap B | K_2)P(K_2) \\ &= P(A | K_1)P(B | K_1)P(K_1) + P(A | K_2)P(B | K_2)P(K_2) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{15} \cdot \frac{2}{15} \cdot \frac{1}{2} = \frac{482}{3600} > \frac{361}{3600} = P(A)P(B). \end{aligned}$$

Dette kommer ikke som en overraskelse. Vi så tidligere, at den oplysning, at den første kugle er rød, indeholder information om, hvilken kasse vi har valgt, og dermed også om sandsynligheden for at den anden kugle er rød.

□

Vi kan naturligvis også definere betinget uafhængighed af mange hændelser.

Definition 1.5.7 Lad A_1, \dots, A_n ($n \geq 2$) og C være hændelser. Da siges A_1, \dots, A_n at være betinget uafhængige givet C , hvis der for enhver delmængde $\{i_1, \dots, i_k\}$ af $\{1, \dots, n\}$ gælder at

$$P(A_{i_1} \cap \dots \cap A_{i_k} | C) = P(A_{i_1} | C) \cdots P(A_{i_k} | C). \quad (1.5.4)$$

Eksempel 1.5.8 Et skadesforsikringselskab har N forsikringstagere. Vi nummererer disse $1, 2, \dots, N$. Det antages, at sandsynligheden for at forsikringstager nr. k har en skade i løbet af et år er p_k ($k = 1, \dots, N$). Disse sandsynligheder antages at være konstante fra år til år. Nu udtager selskabet tilfældigt en forsikringstager for at følge ham i en årrække. Hver forsikringstager har samme sandsynlighed $1/N$ for at blive udtrukket. Den udvalgte forsikringstager følges i m år, hvor det noteres i hvilke år vedkommende har uheld.

Lad A_j betegne hændelsen, at den tilfældigt udvalgte forsikringstager har et uheld i det j te år ($j = 1, \dots, m$), og lad B_k være hændelsen, at den k te forsikringstager er blevet udvalgt ($k = 1, \dots, N$). Da er altså

$$\begin{aligned} P(B_k) &= 1/N, \text{ for } k = 1, \dots, N \text{ og} \\ P(A_j | B_k) &= p_k, \text{ for } k = 1, \dots, N, \quad j = 1, \dots, m. \end{aligned}$$

Først vil vi finde $P(A_1 \cap \dots \cap A_m)$. Da B_1, \dots, B_N er disjunkte og $E = B_1 \cup \dots \cup B_N$, har vi ifølge Sætning 1.4.4, at

$$P(A_1 \cap \dots \cap A_m) = \sum_{k=1}^N P(A_1 \cap \dots \cap A_m | B_k) P(B_k).$$

For at komme videre antages det yderligere, at det faktum, at forsikringstager nr. k har haft et uheld i et år, ikke influerer på sandsynligheden for, at han får uheld et andet år. Vi antager altså, at hændelserne A_1, \dots, A_m er betinget uafhængige givet B_k , så

$$P(A_1 \cap \dots \cap A_m | B_k) = P(A_1 | B_k) \cdots P(A_m | B_k) = p_k^m.$$

Dermed er

$$P(A_1 \cap \dots \cap A_m) = \frac{1}{N} \sum_{k=1}^N p_k^m. \quad (1.5.5)$$

Specielt fås for $m = 1$, at

$$P(A_1) = \frac{1}{N} \sum_{k=1}^N p_k.$$

Denne sandsynlighed er altså gennemsnittet af p_k -erne. Lad os betegne denne gennemsnitlige sandsynlighed for uheld med \bar{p} , d.v.s.

$$\bar{p} = \frac{1}{N} \sum_{k=1}^N p_k.$$

Der gælder naturligvis også, at $P(A_2) = \dots = P(A_m) = \bar{p}$. For $m = 2$ er (1.5.5)

$$P(A_1 \cap A_2) = \frac{1}{N} \sum_{k=1}^N p_k^2,$$

hvorfor den betingede sandsynlighed for A_2 givet A_1 er

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{1}{\bar{p}N} \sum_{k=1}^N p_k^2.$$

Lad os nu sammenligne sandsynligheden $P(A_2) = \bar{p}$ med den betingede sandsynlighed $P(A_2|A_1)$:

$$\begin{aligned} P(A_2|A_1) - P(A_2) &= \bar{p}^{-1} \left(\frac{1}{N} \sum_{k=1}^N p_k^2 - \bar{p}^2 \right) \\ &= \bar{p}^{-1} \left(\frac{1}{N} \sum_{k=1}^N p_k^2 + \bar{p}^2 - 2\bar{p}^2 \right) \\ &= \bar{p}^{-1} \left(\frac{1}{N} \sum_{k=1}^N p_k^2 + \bar{p}^2 - 2\bar{p} \frac{1}{N} \sum_{k=1}^N p_k \right) \\ &= \bar{p}^{-1} \frac{1}{N} \sum_{k=1}^N (p_k^2 + \bar{p}^2 - 2\bar{p}p_k) \\ &= \bar{p}^{-1} \frac{1}{N} \sum_{k=1}^N (p_k - \bar{p})^2 \geq 0. \end{aligned}$$

Vi ser, at $P(A_2|A_1) = P(A_2)$ hvis og kun hvis alle forsikringstagere er lige uheldige, altså hvis $p_k = \bar{p}$ for alle $k = 1, \dots, N$. Eller med andre

ord: Hvis nogle er uheldigere end andre, så er A_1 og A_2 ikke uafhængige hændelser. Hvis den udvalgte forsikringstager har et uheld det første år, forøges hans sandsynlighed for at få et uheld det næste år. Dette fænomen synes måske at modsige vores antagelse om, at uheld eller ej i forskellige år ikke influerer på hinanden, men der er faktisk tale om helt det samme fænomen, som vi så, da vi trak to kugler fra en tilfældig kasse i Eksempel 1.4.5 ovenfor. Hvis den tilfældigt udvalgte forsikringstager har et uheld første år, så tyder denne information på at vedkommende hører til blandt de mindre heldige.

□

1.6 Sammenfatning

I dette kapitel har vi defineret en sandsynlighedsmodel helt generelt. Mængden af mulige udfald kaldes udfaldsrummet, og en delmængde af udfaldsrummet kaldes en hændelse. Sandsynligheden for en hændelse, d.v.s. sandsynligheden for et udfald i en delmængde af udfaldsrummet, angives ved hjælp af et sandsynlighedsmål. Et sandsynlighedsmål har værdier i $[0, 1]$ og skal opfylde to simple og naturlige betingelser. Ud fra disse udledte vi nogle vigtige regneregler for sandsynlighedsmål. Derefter indførte vi to begreber, som er særdeles nyttige, når man i praksis skal opbygge sandsynlighedsteoretiske modeller: Betingede sandsynligheder og stokastisk uafhængighed. Vi har også været inde på de vigtigste regneregler for betingede sandsynligheder.

1.7 Opgaver

- 1.1 En mønt kastes tre gange, og vi interesserer os for antallet af gange man får krone. Opstil en sandsynlighedsteoretisk model, d.v.s. angiv de mulige udfald og den tilsvarende sandsynlighedsfunktion. Beregn dernæst sandsynlighederne for følgende hændelser

{i alle kast viser mønten krone}
{i mindst ét kast viser mønten krone}
{i præcis ét kast viser mønten krone}.

- 1.2 Tre personer udvælges tilfældigt. Vi interesserer os for hvor mange af disse, der er født på en søndag. Opstil en sandsynlighedsteoretisk model.
- 1.3 Der slås et slag med to terninger. Hvad er sandsynligheden for følgende hændelser? (a) Summen af øjnene er større end eller lig 10. (b) De to terninger viser det samme. (c) Der er mindst én sekser. (d) Der er netop én sekser.
- 1.4 Hvad er sandsynligheden for ved et slag med tre terninger at summen af øjnene er 10?
- 1.5 To personer vælges tilfældigt. (a) Hvad er sandsynligheden for, at de har samme fødselsdag? (b) Hvad er sandsynligheden for at de har forskellige fødselsdage?
- 1.6 En terning kastes to gange. Find sandsynligheden for hændelsen, at der forekommer mindst én sekser. Find sandsynligheden for at der forekommer enten mindst én sekser eller mindst én toer.
- 1.7 En mønt kastes 10 gange. Hvad er sandsynligheden for hændelsen, at plat forekommer mindst to gange?
- 1.8 Syv personer vælges tilfældigt. (a) Hvad er sandsynligheden for at de alle er født på en søndag? (b) Hvad er sandsynligheden for at de alle er født på den samme ugedag? (c) Hvad er sandsynligheden for at de er født på hver sin ugedag?
- 1.9 Et spil kort blandes, og tretten kort trækkes. Hvad er sandsynligheden for hverken at få billedkort eller esser?
- 1.10 Lad A og B være hændelser, for hvilke $P(A) = 0.6$, $P(B) = 0.5$ og $P(A \cup B) = 0.9$. Find sandsynlighederne for følgende hændelser $A \cap B$, $E \setminus A$, $E \setminus B$, $(E \setminus A) \cap (E \setminus B)$ og $(E \setminus A) \cup (E \setminus B)$. Her betegner E som sædvanligt udfaldsrummet.
- 1.11 Et tal trækkes tilfældigt i intervallet mellem nul og en (se Eksempel 1.3.1). Hvad er sandsynligheden for, at første decimal efter kommaet er et lige tal? Vink: Benyt (1.3.14).

- 1.12 Hvad er i et slag med fem terninger sandsynligheden for at få mindst én sekser? Vink: det er lettere at udregne sandsynligheden for den komplementære hændelse.
- 1.13 Lad E være en vilkårlig mængde, og lad P være et sandsynligheds mål på E . Vis at der for tre vilkårlige delmængder af E gælder, at

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C).$$

Vink: Benyt 1.3.11 nogle gange.

- 1.14 Hvad er i et slag med tre terninger sandsynligheden for at mindst to terninger viser det samme? Vink: Lad A_{12} betegne den hændelse at terning nr. 1 og 2 viser det samme, og definer A_{13} og A_{23} tilsvarende. Benyt nu opgave 1.13.
- 1.15 Den franske adelsmand Chevalier de Méré, som levede i det syttende århundrede, elskede hazardspil. Desværre havde han problemer med sandsynlighedsregningen. Pascal skrev 29. juli 1654 om ham, at “de Méré er en intelligent mand, men han er ikke matematiker, og det er, som De ved, en stor mangel”. De Méré havde fået den lyse idé, at sandsynligheden for at få mindst én sekser i fire slag med en terning skulle være $4 \times \frac{1}{6} = \frac{2}{3}$. Han spillede nu i overensstemmelse med denne beregning, men da han tabte alle sine penge, indså han, at han havde taget fejl. Han fandt derfor sine pistoler frem, pudsede dem og skød sig. Hvori lå hans fejltagelse, og hvad er det korrekte resultat?
De Méré mente også, at sandsynligheden for at få mindst én dobbelt sekser (Sonnez) i 24 kast med to terninger var den samme som sandsynligheden for at få mindst én sekser i fire slag med en terning. Er det rigtigt?
- 1.16 Vis formel (1.3.14).
- 1.17 Betragt kast med to terninger. Vi antager, at den ene terning er hvid og den anden sort.

- 1) Hvad er den betingede sandsynlighed for, at summen af øjnene er 12, givet at summen er mindst 11?
 - 2) Hvad er den betingede sandsynlighed for, at de to terninger viser det samme, givet at summen er 7?
 - 3) Hvad er den betingede sandsynlighed for, at den hvide terning viser 3, givet at den sorte terning viser 5?
 - 4) Hvad er den betingede sandsynlighed for, at terningen med det mindste antal øjne viser 2, givet at terningen med det største antal øjne højst viser 5?
- 1.18 En mønt kastes 10 gange.
- (1) Hvad er sandsynligheden for at få krone den tiende gang, givet at de ni første kast giver plat?
 - (2) Hvad er sandsynligheden for at få krone den tiende gang, givet at netop ni af de ti kast giver plat?
- 1.19 Betragt igen eksemplet med de 25 børn og deres fødselsdage (Eksempel 1.2.2). Hvad er sandsynligheden for, at der ikke er to børn med samme fødselsdag, givet at ingen af børnene er født i januar?
- 1.20 En kasse indeholder 10 røde og 2 hvide kugler. Kugler trækkes tilfældigt uden tilbagelægning, indtil man får en hvid kugle. Hvad er sandsynligheden for, at der er 5 kugler tilbage i kassen efter dette?
- 1.21 En mønt kastes tre gange. Hver gang mønten viser krone kastes en terning. Hvad er sandsynligheden for at få mindst én sekser i dette spil? Vink: Lad A_n , ($n = 0, 1, 2, 3$) være hændelsen, at mønten viser krone netop n gange, og benyt Sætning 1.4.4.
- 1.22 En terning kastes. Dernæst kastes en mønt det antal gange, som terningen viser. Hvad er sandsynligheden for, at få præcis én krone.
- 1.23 Et gartneri dyrker stedmoderplanter. Sandsynligheden for anlæg for blå blomster er 0.7, for gule blomster 0.2 og for hvide blomster 0.1. Sandsynligheden for at en plante kommer i "groning" er 0.95 for planter med blå blomster, mens denne sandsynlighed er 0.9 både for planter med gule blomster og for planter med hvide blomster.

- (1) Hvad er sandsynligheden for at en tilfældigt udvalgt plante kommer i groning?
- (2) Hvad er sandsynligheden for, at en plante, der er kommet i groning, får gule blomster?

1.24 Betragt igen kast med to terninger. Vi antager, at den ene terning er hvid og den anden sort. Hvilke par blandt følgende hændelser er uafhængige?

$$\begin{aligned} A &= \{\text{den hvide terning viser } 4\} \\ B &= \{\text{den sorte terning viser } 1\} \\ C &= \{\text{terningen med det højeste antal øjne viser } 4\} \\ D &= \{\text{summen af øjnene er } 5\} \\ F &= \{\text{summen af øjnene er } 7\} \end{aligned}$$

Overvej, om det stemmer overens med din intuition.

1.25 Vis påstandene 2) – 4) i Sætning 1.5.5.

1.26 Lad P være et sandsynlighedsmål på mængden E , og lad A være en delmængde af E . Vis at A og E er uafhængige hændelser.

1.27 Vis, at hvis en hændelse A er uafhængig af sig selv, så er $P(A)$ lig nul eller en.

1.28 En terning kastes. Lad A være hændelsen, at vi får 1, 2 eller 3, og lad B være hændelsen, at vi får 1 eller 4. Vis, at A og B er uafhængige.

1.29 Et tal trækkes tilfældigt i intervallet mellem nul og en (se Eksempel 1.3.1). Betragt hændelserne $A = [0, 0.5]$, $B = [0.1, 0.7]$ og $C = [0.4, 0.9]$. Undersøg, om A og B er uafhængige, om A og C er uafhængige og om B og C er uafhængige.

1.30 Lad A , B og C være hændelser. Antag, at A og B er uafhængige, samt at A og C er uafhængige. Kan man deraf slutte, at A og $B \cup C$ er uafhængige?

1.31 En terning kastes to gange. Vis, at hændelserne

$$A = \{\text{ulige antal \u00f8jne i f\u00f8rste kast}\}$$

$$B = \{\text{ulige antal \u00f8jne i andet kast}\}$$

$$C = \{\text{ulige sum}\}.$$

parvist er uafh\u00e6ngige, men at de tre h\u00e6ndelser ikke er uafh\u00e6ngige.

1.32 Antag, at A_1, \dots, A_n er uafh\u00e6ngige h\u00e6ndelser. Vis, at $E \setminus A_1, A_2, \dots, A_n$ er uafh\u00e6ngige h\u00e6ndelser, og slut af dette resultat, at $E \setminus A_1, \dots, E \setminus A_n$ er uafh\u00e6ngige. Benyt til slut det viste til at bevise, at $P(A_1 \cup \dots \cup A_n) = 1$ hvis og kun hvis $P(A_i) = 1$ for mindst et i .

Kapitel 2

Stokastiske variable

Oftentimes we are not really interested in the large basic outcome space E . Instead, the interest is concentrated on certain aspects of the outcome, which can be expressed by numbers. This is what we saw in Chapter 1. For example, if we toss a coin n times, and we are only interested in the number of heads and tails, then we are actually free to ignore the 2^n sequences of n heads and tails, if possible. In fact, it turns out that, if we, as we will do later, consider the coin being tossed infinitely many times, in this chapter we will introduce the concept of a stochastic variable, which formalizes the very common situation, that the aspect of a stochastic phenomenon, which we are interested in, can be expressed by a number or even by a vector in \mathbb{R}^n . When one first has decided to work with stochastic variables, it is a very useful tool for the construction and analysis of mathematical models for random phenomena.

2.1 Stokastiske variable og deres fordeling

We assume, that a basic outcome space E with a corresponding probability measure P is given. Then a *stochastic variable* is a function from E into the real numbers \mathbb{R} . Stochastic variables are denoted as usual by capital letters, e.g. X .

Eksempel 1.5.3 (fortsat). Antag at vi kaster en mønt n gange. Vi lader udfaldsrummet E og sandsynligheds målet være som tidligere. Et udfald

kan betegnes med (R_1, \dots, R_n) , hvor hvert R_i enten er P eller K . Hvis vi kun er interesseret i antallet af gange vi får udfaldet krone, er følgende stokastiske variable den rette at betragte

$$X(R_1, \dots, R_n) = \#\{j | R_j = K\}. \quad (2.1.1)$$

□

En stokastisk variabel afhænger altså af udfaldet e på det basale udfaldsrum E . Når vi først er kommet godt i gang, vil vi imidlertid konsekvent udelade argumentet e , og skrive X for værdien af X i stedet for $X(e)$. Hvis A er en delmængde af \mathbb{R} , vil vi også skrive $\{X \in A\}$, når vi egentlig mener hændelsen $\{e \in E | X(e) \in A\}$ (d.v.s. originalmængden eller Urbilledet af A ved funktionen X). Dette gør vi ikke kun for at undgå en kluntet notation, men også for at understrege, at vi som regel kan glemme alt om det basale udfaldsrum E . Det intuitive indhold i begrebet "stokastisk variabel" er *ikke* en funktion defineret på et basalt udfaldsrum, men derimod en reel variabel, der er tilfældig, således at forstå at vi ikke på forhånd ved, hvad dens værdi er, men vi ved, hvad sandsynligheden er for, at den antager denne eller hinde værdi.

Vi kan beskrive den tilfældige variation af X i \mathbb{R} ved for enhver delmængde A af \mathbb{R} at angive sandsynligheden for at X ligger i A , d.v.s. ved at angive

$$P_X(A) = P(X \in A) \quad (2.1.2)$$

Bemærk, at vi yderligere har forenklet notationen ved at skrive $P(X \in A)$ i stedet for $P(\{X \in A\})$. Funktionen P_X fra delmængderne af \mathbb{R} ind i $[0, 1]$ kaldes *fordelingen* af X . Den er et sandsynlighedsmål på \mathbb{R} . For at indse det skal vi checke, at P_X opfylder (1.3.2) og (1.3.3). Da $\{X \in \mathbb{R}\} = X^{-1}(\mathbb{R}) = E$, er $P_X(\mathbb{R}) = P(E) = 1$, og hvis A og B er disjunkte, er $\{X \in A\}$ og $\{X \in B\}$ også disjunkte og opfylder, at

$$\{X \in A \cup B\} = \{X \in A\} \cup \{X \in B\}$$

(overvej hvorfor). Derfor er $P_X(A \cup B) = P(\{X \in A\} \cup \{X \in B\}) = P(\{X \in A\}) + P(\{X \in B\}) = P_X(A) + P_X(B)$. Her har vi naturligvis brugt at P opfylder (1.3.2) og (1.3.3).

Mange gange kan man umiddelbart angive X s fordeling uden tanke på noget basalt udfaldsrum E . I sådanne tilfælde, er det bedst at glemme

alt om, hvad E er, og vi vil komme med udsagn som “lad X være en stokastisk variabel med fordeling ...”. Hvis man vil have formalismen på plads, kan man i sådanne situationer definere $E = \mathbb{R}$ og lade X være den identiske afbildning. Til andre tider har man derimod brug for E for at kunne argumentere for, hvad fordelingen af X er. Også når man betragter flere stokastiske variable, er det nyttigt at have dem defineret på et fælles bagvedliggende E , når man skal studere sammenhængen mellem dem.

Eksempel 1.5.3 (fortsat). Vi kaster en mønt n gange, og lader X være den stokastiske variabel givet ved (2.1.1), som angiver antallet af kast med udfaldet krone. Det er klart at X må tilhøre mængden $\{0, 1, \dots, n\}$. Vi kan derfor angive X s fordeling ved at angive sandsynlighederne $P(X = k)$ for $k = 0, 1, \dots, n$. Da alle udfald i E har samme sandsynlighed, skal vi for at finde $P(X = k)$ bare finde ud af, hvor mange af disse der har udfaldet krone præcis k gange. Dette er et kombinatorisk spørgsmål, som er behandlet i Appendix B. Det er nemlig antallet af måder, vi kan udtage en delmængde med k elementer af en mængde med n elementer. Svaret er binomialkoefficienten $\binom{n}{k}$, så X s fordeling er givet ved

$$P(X = k) = \binom{n}{k} 2^{-n}, \quad k = 0, 1, \dots, n.$$

□

Lad os lige give nogle flere meget enkle eksempler på stokastiske variable.

Eksempel 1.2.1 (fortsat). Vi kaster endnu en gang to terninger, den ene rød og den anden sort. Udfaldsrummet kan skrives som $E = \{(r, s) | r, s = 1, \dots, 6\}$. Her følger nogle eksempler på stokastiske variable.

$R(r, s)$	$= r$	antal øjne på den røde terning
$S(r, s)$	$= s$	antal øjne på den sorte terning
$X(r, s)$	$= r + s$	summen af øjnene
$Y(r, s)$	$= r \wedge s$	det mindste antal øjne
$Z(r, s)$	$= r \vee s$	det største antal øjne

Her og senere betegner $x \wedge y$ og $x \vee y$ henholdsvis det mindste og det største af de to reelle tal x og y . D.v.s. $x \vee y = \max\{x, y\}$ og $x \wedge y = \min\{x, y\}$.

Det er let at angive fordelingen af disse stokastiske variable. Således kan R kun antage værdierne $1, 2, \dots, 6$, og hver af disse antages med sandsynligheden $1/6$. Det kunne vi naturligvis godt have fundet ud af uden at gå omkring E .

Bemærk at det bagvedliggende udfaldsrum E ikke er entydigt. Hvis vi kun er interesserede i R , kunne vi have valgt $\{1, 2, \dots, 6\}$ med en konstant sandsynlighedsfunktion som vores bagvedliggende udfaldsrum, men det er lige så rigtigt at benytte E . Vi kunne såmænd også have benyttet et endnu større udfaldsrum, f.eks. med plads til endnu nogle terningkast, et par møntkast samt udtrækning af et kort, og udstyret med et passende sandsynlighedsmål, uden at dette havde ændret fordelingen af R . Det basale udfaldsrum kan generelt sagtens gøres større, end hvad man i den konkrete situation har brug for. Det skal bare altid være mindst stort nok til at omfatte alt, hvad der skal bruges til at studere ens konkrete problem.

Den stokastiske variable X forudsætter i modsætning til R og S at vi har hele E (eller en endnu større mængde) som basalt udfaldsrum. Denne stokastiske variable kan antage værdierne $2, 3, \dots, 12$ med sandsynligheder, som blev givet i Eksempel 1.2.3. Man kunne selvfølgelig bare have specificeret disse sandsynligheder, men det ville være lidt kunstigt. Det ville i hvert fald ikke være klart, hvor de kom fra. Hvis man vil studere spillet mellem R og X , har man også brug for E . Læseren opfordres til selv at overveje fordelingerne af Y og Z .

□

Hvis X er en stokastisk variabel, og t er en funktion fra \mathbb{R} ind i \mathbb{R} , så er $Y = t(X)$ også en stokastisk variabel. Dette er klart, for da X er en funktion fra det underliggende udfaldsrum E ind i \mathbb{R} , og da Y er den sammensatte funktion $t \circ X$, er Y også en funktion fra E ind i \mathbb{R} . Den stokastiske variable Y har selvfølgelig sin egen fordeling P_Y , der imidlertid kan findes ud fra fordelingen af X , som vi tidligere betegnede med P_X . Der gælder nemlig for $A \subseteq \mathbb{R}$, at

$$\begin{aligned} P_Y(A) &= P(Y \in A) = P(t(X) \in A) \\ &= P(X \in t^{-1}(A)) = P_X(t^{-1}(A)), \end{aligned} \tag{2.1.3}$$

hvor $t^{-1}(A)$ betegner originalmængden af A ved t , d.v.s. $t^{-1}(A) = \{x \in \mathbb{R} | t(x) \in A\}$. Sagt med ord, består $t^{-1}(A)$ af de reelle tal, som afbildes ind i A ved t . Det skal understreges, at t altså ikke behøver at være injektiv, således at den har en invers funktion. Notationen kan forvirre, da den inverse funktion, hvis den eksisterer, traditionelt betegnes med t^{-1} . I de særligt pæne tilfælde, hvor den inverse funktion eksisterer, er $t^{-1}(A)$ naturligvis blot billedet af A ved funktionen t^{-1} , så notationen er konsistent. Resultatet (2.1.3) kaldes *transformationssætningen for fordelinger*. Man siger at Y er en transformation af X ved transformationen t .

2.2 Fordelingsfunktionen

Lad X være en stokastisk variabel. I reglen er det ikke praktisk at skulle angive $P(X \in A)$ for alle mulige delmængder A af \mathbb{R} . Vi har derfor brug for lettere måder at specificere fordelinger på. En generelt anvendelig fremgangsmåde er at angive *fordelingsfunktionen* for X . Denne er defineret ved

$$F(x) = P(X \leq x), \quad \text{for } x \in \mathbb{R}. \quad (2.2.1)$$

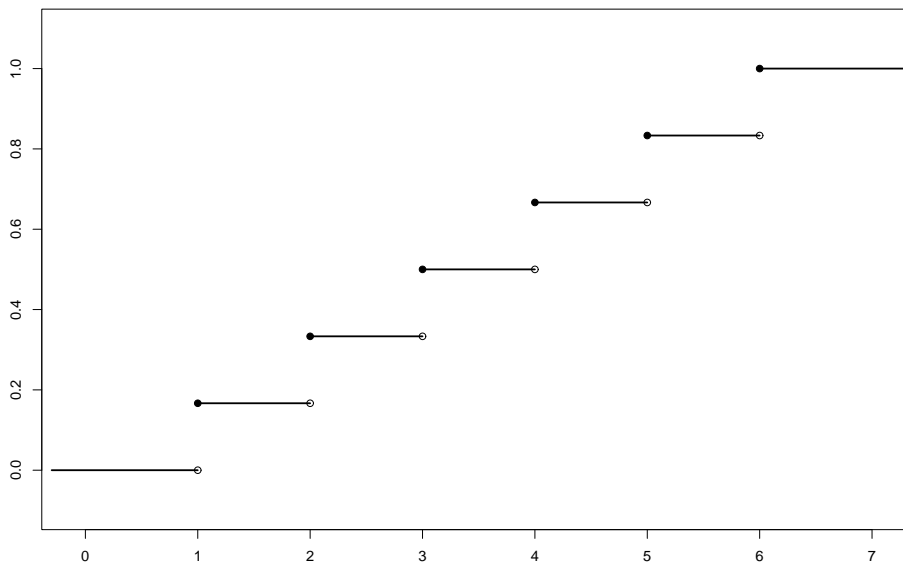
Fordelingsfunktionen er en funktion fra \mathbb{R} ind i $[0, 1]$. Hvis vi angiver fordelingsfunktionens værdi for alle x , har vi altså sandsynligheden for at X tilhører ethvert interval af formen $(-\infty, x]$. Dette er faktisk nok til at specificere hele fordelingen, men det har vi ikke forudsætninger for at vise for generelle fordelinger. Det gør ikke så meget, for vi vender tilbage til dette spørgsmål senere for de to vigtige typer af fordelinger, som vi især vil beskæftige os med i dette kursus. Lad os se på et par typiske eksempler.

Eksempel 2.2.1 Lad X være den stokastiske variable, som angiver antallet af øjne ved kast med en terning. Da X antager værdierne $\{1, 2, 3, 4,$

5, 6}, hver med sandsynlighed $1/6$, er fordelingsfunktionen for X givet ved

$$F(x) = \begin{cases} 0 & \text{hvis } x < 1 \\ \frac{1}{6} & \text{hvis } 1 \leq x < 2 \\ \frac{2}{6} & \text{hvis } 2 \leq x < 3 \\ \frac{3}{6} & \text{hvis } 3 \leq x < 4 \\ \frac{4}{6} & \text{hvis } 4 \leq x < 5 \\ \frac{5}{6} & \text{hvis } 5 \leq x < 6 \\ 1 & \text{hvis } 6 \leq x \end{cases}$$

Fordelingfunktionen vokser i spring fra 0 til 1. Den er kontinuert bortset fra springpunkterne, hvor den er kontinuert fra højre, men ikke fra venstre. Da $P(X = i) = F(i) - F(i - 1)$ for $i = 1, \dots, 6$ er det her klart, at F fuldstændigt specificerer fordelingen af X .

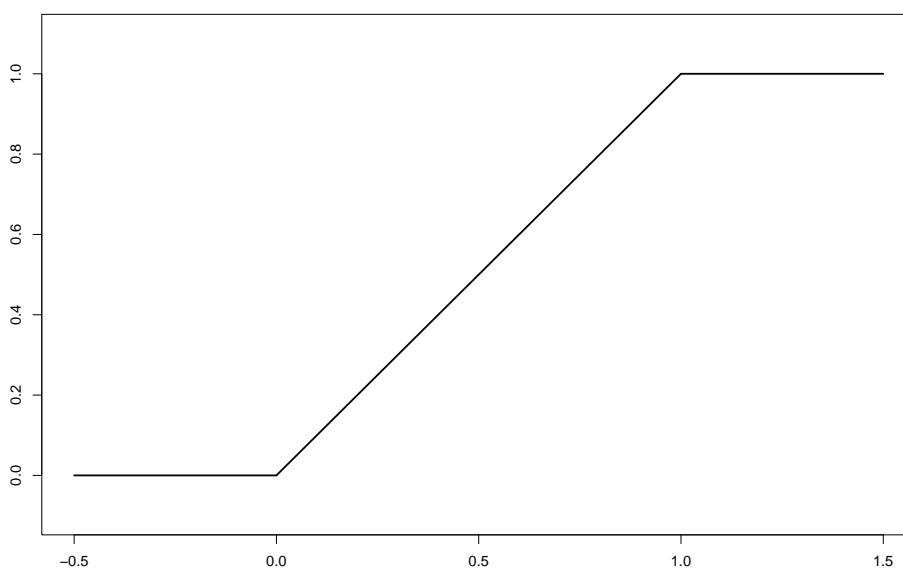


Figur 2.2.1: Fordelingsfunktionen for en stokastisk variabel, der angiver antallet af øjne ved kast med en terning.

Fordelingen af X kaldes ligefordelingen på mængden $\{1, 2, 3, 4, 5, 6\}$. Mere generelt kaldes en fordeling på en endelig delmængde M af \mathbb{R} eller \mathbb{R}^n med konstant sandsynlighedsfunktion for *ligefordelingen* på M , og en stokastisk variabel med denne fordeling siges at være ligefordelt på M . \square

Eksempel 1.3.1 (fortsat). Lad os udtrække et tilfældigt tal mellem nul og en, som forklaret i Afsnit 1.3, og lad X være den stokastiske variable, som angiver det udtrukne tal. Hvis $x \in [0, 1]$, er $P(X \in (-\infty, x]) = P(X \in [0, x]) = x$. Derfor er fordelingsfunktionen for X

$$F(x) = \begin{cases} 0 & \text{hvis } x < 0 \\ x & \text{hvis } 0 \leq x \leq 1 \\ 1 & \text{hvis } x > 1. \end{cases} \quad (2.2.2)$$



Figur 2.2.2: Fordelingsfunktionen for en stokastisk variabel, hvis værdi er et tilfældigt tal mellem nul og en.

Denne funktion vokser kontinuert fra 0 til 1. Den er strengt voksende i intervallet $[0, 1]$. \square

I begge disse eksempler vokser fordelingsfunktionen svagt fra 0 til 1. Dette er generelle egenskaber ved en fordelingsfunktion. Lad X være en stokastisk variabel med fordelingsfunktion F . At F er *svagt voksende* (også somme tider kaldet ikke-aftagende) er let at se, for hvis $x \leq y$, så er $(-\infty, x] \subseteq (-\infty, y]$, hvilket ifølge (1.3.8) medfører at

$$F(x) = P_X((-\infty, x]) \leq P_X((-\infty, y]) = F(y).$$

Fordelingsfunktionen F opfylder også, at

$$\lim_{n \rightarrow \infty} F(n) = 1 \tag{2.2.3}$$

og

$$\lim_{n \rightarrow \infty} F(-n) = 0, \tag{2.2.4}$$

men det vil vi ikke vise generelt. Vi skal senere se, at disse grænseresultater holder for to vigtige typer af fordelinger.

Det kan omvendt bevises, at enhver funktion F fra \mathbb{R} ind i $[0, 1]$, som er svagt voksende, og som opfylder (2.2.3) og (2.2.4) bestemmer et sandsynlighedsmål på \mathbb{R} (eller en delmængde af \mathbb{R}). Det vil vi dog ikke bevise.

Eksempel 2.2.2 I dette eksempel skal vi se, hvordan vi kan bruge fordelingsfunktionen til at argumentere os frem til, hvad fordelingen af en bestemt stokastisk variabel er. Vi betragter udsendelse af α -partikler fra et radioaktivt materiale. Lad X være den stokastiske variable, som angiver ventetiden mellem udsendelsen af to på hinanden følgende α -partikler. Her er det bedst ikke at tænke for meget over, hvad det bagvedliggende udfaldsrum er. Det er givetvis meget komplekst.

I hvert fald er $X > 0$. Endvidere må det være rimeligt at antage, at atomerne i det radioaktive materiale ikke kan huske, hvor lang tid der er gået, siden den sidste α -partikel blev udsendt. Det kan matematisk formuleres således:

$$P(X > t + s | X > t) = P(X > s) \tag{2.2.5}$$

for alle $t, s \geq 0$. Sagt med ord: Hvis vi ved, at der ikke er udsendt nogen α -partikel i tidsintervallet $[0, t]$, så er sandsynligheden for, at der

ikke udsendes nogen i de næste s tidsenheder (altså i $[t, t + s]$), lig sandsynligheden for, at der ikke udsendes nogen α -partikel i tidsintervallet $[0, s]$. Bemærk, at (2.2.5) kun giver mening for alle $t > 0$, hvis vi antager, at $P(X > t) > 0$ for alle $t > 0$. Det er jo rimeligt nok, så det antager vi. Lad F betegne fordelingsfunktionen for X , og definer $G(x) = 1 - F(x) = P(X > x)$. Da

$$P(X > t + s | X > t) = \frac{P(X > t + s)}{P(X > t)},$$

kan (2.2.5) skrives som

$$G(t + s) = \frac{G(t + s)}{G(t)},$$

så

$$G(t + s) = G(t)G(s)$$

for alle $t, s \geq 0$. Sæt nu $L(x) = \log(G(x))$, hvilket kan gøres, da $G(x) > 0$. Vi har da, at

$$L(t + s) = L(t) + L(s)$$

for alle $t, s \geq 0$, og da $G(0) = P(X > 0) = 1$, er $L(0) = 0$, så for ethvert $t > 0$ er

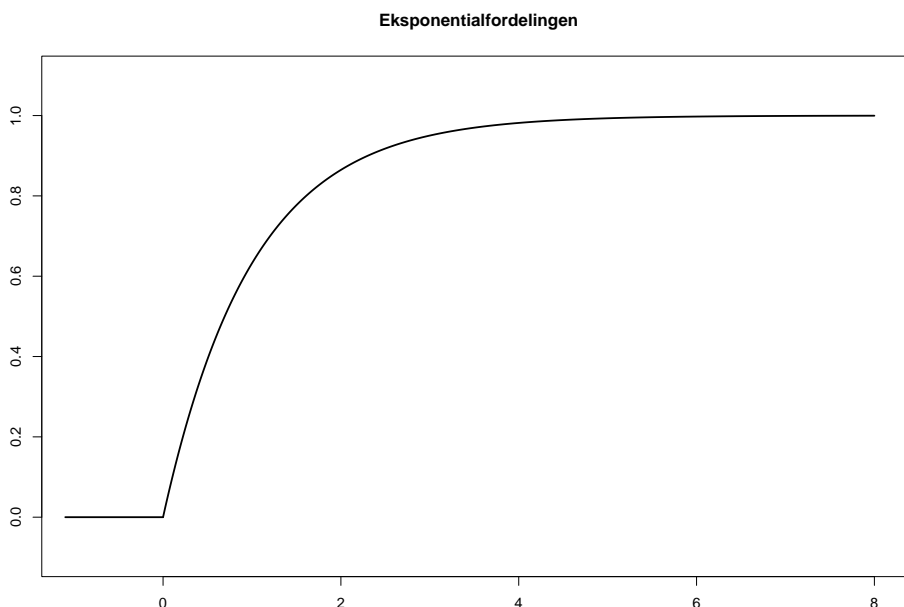
$$\frac{L(t + h) - L(t)}{h} = \frac{L(h) - L(0)}{h} \quad (2.2.6)$$

for alle $h > 0$. Hvis vi antager, at L er differentiabel i 0 med differentialkvotient $-\lambda$ (da L er aftagende, er $\lambda \geq 0$), finder vi ved at lade h gå mod nul i (2.2.6), at L er differentiabel for alle $t > 0$ med $L'(t) = -\lambda$. Derfor er

$$L(t) = -\lambda t + c$$

for en eller anden konstant c . Da imidlertid $L(0) = 0$, er $c = 0$, så $L(t) = -\lambda t$ for $t > 0$. Vi kan nu regne tilbage til F , idet $F(x) = 1 - \exp(L(x))$. Fordelingsfunktionen for X er altså

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{hvis } x > 0 \\ 0 & \text{hvis } x \leq 0. \end{cases}$$



Figur 2.2.3: Ekspontialfordelingens fordelingsfunktion.

Fordelingen med denne fordelingsfunktion kaldes *ekspontialfordelingen* med parameter λ . Læg mærke til, at vi i udledningen af ekspontialfordelingen benyttede at en stokastisk variabel, som beskriver en ventetid med denne fordeling, ikke har nogen hukommelse om, hvor længe vi har ventet. Man siger somme tider lidt løst, at ekspontialfordelingen ikke har nogen hukommelse.

□

Der findes også en *transformationssætning for fordelingsfunktioner*, som for strengt monotone funktioner t giver sammenhængen mellem fordelingsfunktionerne for de stokastiske variable X og $Y = t(X)$, som vi betegner med F_X og F_Y . Der mindes om, at $F_X(x) = P(X \leq x)$ og $F_Y(x) = P(Y \leq x)$. Antag, at $P(X \in I) = 1$ for et interval $I \subseteq \mathbb{R}$. Intervallet I kan være endeligt eller uendeligt og lukket, åbent eller halvåbent. Om t antages det, at den er en strengt monoton kontinuert funktion fra

I ind i \mathbb{R} . Da er $J = t(I)$ et interval. Definer venstre (nedre) endepunkt v og højre (øvre) endepunkt h for intervallet J ved

$$v = \inf J \quad \text{og} \quad h = \sup J.$$

Eventuelt kan v være $-\infty$, ligesom h eventuelt kan være ∞ . Med t^{-1} betegnes den inverse funktion til t , som eksisterer og er defineret på J , da t er antaget at være strengt monoton. Det er klart, at $P(Y \in J) = 1$.

Sætning 2.2.3 *Hvis t er strengt voksende, er*

$$F_Y(y) = \begin{cases} 0 & \text{hvis } y < v \\ F_X(t^{-1}(y)) & \text{hvis } y \in J \\ 1 & \text{hvis } y > h. \end{cases} \quad (2.2.7)$$

Hvis t er strengt aftagende, er

$$F_Y(y) = \begin{cases} 0 & \text{hvis } y < v \\ 1 - F_X(t^{-1}(y)) + P(X = t^{-1}(y)) & \text{hvis } y \in J \\ 1 & \text{hvis } y > h. \end{cases} \quad (2.2.8)$$

Bevis: Hvis t er voksende, er

$$F_Y(y) = P(Y \leq y) = P(t(X) \leq y) = P(X \leq t^{-1}(y)) = F_X(t^{-1}(y))$$

for $y \in J$. Hvis t er aftagende, er

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(t(X) \leq y) = P(X \geq t^{-1}(y)) \\ &= 1 - P(X < t^{-1}(y)) = 1 - [F_X(t^{-1}(y)) - P(X = t^{-1}(y))] \end{aligned}$$

for $y \in J$. I den sidste beregning har vi benyttet (1.3.9) og (1.3.7). □

Det er her egentlig vigtigere at mærke sig bevisteknikken end selve resultatet. Den er ofte nyttig. Sætning 2.2.3 gælder også, når t ikke er kontinuert, men i så fald er J ikke noget interval, og F_Y er konstant udenfor J .

Eksempel 2.2.4 Lad os igen udtrække et tilfældigt tal mellem nul og en, som forklaret i Afsnit 1.3, og lad X være den stokastiske variable, som angiver det udtrukne tal. Antag nu, at vi tegner et kvadrat med sidelængde X . Da er arealet af dette kvadrat en stokastisk variabel, nemlig $Y = X^2$. Fordelingsfunktionen for Y for $y \in [0, 1]$ findes let:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y},$$

idet vi bruger (2.2.2). Da $Y \in [0, 1]$, er $F_Y(y) = 0$ for $y < 0$, og $F_Y(y) = 1$ for $y > 1$. Disse resultater kunne vi naturligvis umiddelbart have skrevet op v.h.a. (2.2.7) og (2.2.2). □

2.3 Fler-dimensionale stokastiske variable

Ofte har man brug for at studere flere stokastiske variable på én gang. Vi skal derfor i dette afsnit behandle n -dimensionale stokastiske vektorer af formen $X = (X_1, \dots, X_n)$, hvor X_1, \dots, X_n er sædvanlige endimensionale stokastiske variable, som alle er definerede på det samme bagvedliggende udfaldsrum E . Således er X altså en funktion fra E ind i \mathbb{R}^n . Dette bekymrer vi os i reglen ikke meget om, men det er dog vigtigt at huske, at der kan være sammenhænge mellem de enkelte koordinater af X , da de jo afhænger af det samme basale udfald $e \in E$. Somme tider kaldes en n -dimensional stokastisk vektor også for en n -dimensional stokastisk variabel. Når vi taler om flere stokastiske variable på én gang, vil det *altid* være underforstået, at de er defineret på det samme bagvedliggende udfaldsrum E .

Eksempel 1.2.1 (fortsat). Ved kast med en rød og en sort terning er det basale udfaldsrum $E = \{(r, s) | r, s = 1, \dots, 6\}$. Betragt de stokastiske variable $R(r, s) = r$, $S(r, s) = s$ og $T(r, s) = 7 - r$. De er alle ligefordelte på $\{1, 2, 3, 4, 5, 6\}$, og man kunne måske fristes til at tro, at der ikke er den store forskel på fordelingen af de to stokastiske vektorer (R, S) og (R, T) , men det er der. F.eks. er

$$P(R \leq 4, S \leq 4) = \frac{16}{36} = \frac{4}{9},$$

mens

$$P(R \leq 4, T \leq 4) = P(R \in \{3, 4\}) = \frac{1}{3}.$$

(vis disse udsagn). Her bruger vi den forenklede notation $P(R \leq 4, S \leq 4)$ i stedet for at skrive $P(\{R \leq 4\} \cap \{S \leq 4\})$ eller det, der er værre. Det vil vi som regel gøre, også ved højere-dimensionale stokastiske vektorer. \square

Ovenstående eksempel viser, at to stokastiske vektorer (X_1, X_2) og (Y_1, Y_2) sagtens kan have den egenskab, at X_1 og Y_1 har samme fordeling, og X_2 og Y_2 har samme fordeling, men at der findes en delmængde A af \mathbb{R}^2 , så

$$P((X_1, X_2) \in A) \neq P((Y_1, Y_2) \in A).$$

Det betyder, at sandsynligheden $P((X_1, X_2) \in A)$ ikke kan beregnes ud fra fordelingerne af X_1 og X_2 . Dette er baggrunden for, at vi definerer den *simultane fordeling* for den stokastiske vektor $X = (X_1, \dots, X_n)$ som sandsynligheds målet på \mathbb{R}^n givet ved

$$P_X(A) = P((X_1, \dots, X_n) \in A),$$

hvor A er en delmængde af \mathbb{R}^n . At $P_X(A)$ er et sandsynligheds mål indses på samme måde som ved en en-dimensional stokastisk variabel.

De n fordelinger af de enkelte koordinater X_1, \dots, X_n , betragtet separat som en-dimensionale stokastiske variable, kaldes for de *marginale fordelinger* af X_1, \dots, X_n for at understrege, at de ikke er hele sandheden om den stokastiske vektor $X = (X_1, \dots, X_n)$.

Eksempel 1.2.1 (fortsat). Betragt igen de stokastiske variable R , S og T . Den simultane fordeling af (R, S) er en ligefordeling på mængden $\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, d.v.s. alle elementer i mængden har samme sandsynlighed. Derimod er den simultane fordeling af (R, T) en ligefordeling på mængden $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Alle de marginale fordelinger er, som tidligere nævnt, ligefordelinger på mængden $\{1, 2, 3, 4, 5, 6\}$. \square

Man kan også definere en *simultan fordelingsfunktion* for en n -dimensional stokastisk vektor. Det er funktionen fra \mathbb{R}^n ind i $[0, 1]$ givet ved

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n). \quad (2.3.1)$$

De n fordelingsfunktioner af de enkelte koordinater X_1, \dots, X_n , betragtet separat som en-dimensionale stokastiske variable, kaldes for de *marginale fordelingsfunktioner*.

Fordelingsfunktioner for stokastiske vektorer benyttes især i teoretiske overvejelser, og de vil ikke blive brugt meget i disse noter. Lad os her slutte med at bemærke, at man ligesom for en-dimensionale stokastiske variable kan vise, at den simultane fordelingsfunktion for en stokastisk vektor X bestemmer den simultane fordeling af X entydigt.

2.4 Uafhængige stokastiske variable

Vi har tidligere diskuteret begrebet uafhængighed af hændelser. Nu vil vi også definere, hvad vi skal forstå ved uafhængige stokastiske variable.

Definition 2.4.1 De stokastiske variable X_1, \dots, X_n siges at være uafhængige, hvis

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n) \quad (2.4.1)$$

for alle A_1, \dots, A_n , hvor hver A_i er en delmængde af \mathbb{R} .

Vi siger også somme tider at X_1, \dots, X_n er stokastisk uafhængige, når det er nødvendigt for at undgå misforståelser. En anden måde at udtrykke definitionen på er, at X_1, \dots, X_n er uafhængige, hvis hændelserne $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ er indbyrdes uafhængige for alle A_1, \dots, A_n , hvor hver A_i er en delmængde af \mathbb{R} (overvej at dette er samme definition som Definition 2.4.1). Specielt er to stokastiske variable X og Y uafhængige, hvis hændelserne $\{X \in A\}$ og $\{Y \in B\}$ er uafhængige for ethvert valg af delmængderne A og B af \mathbb{R} .

Eksempel 1.2.1 (fortsat). Lad R , S og T være defineret som tidligere. Bemærk, at for alle delmængder A_1 og A_2 af $M = \{1, 2, 3, 4, 5, 6\}$ er

$$P(R \in A_1, S \in A_2) = \frac{\#A_1 \cdot \#A_2}{36} = P(R \in A_1)P(S \in A_2)$$

Heraf følger det, at det for ethvert par B_1 og B_2 af vilkårlige delmængder af \mathbb{R} gælder, at

$$\begin{aligned} P(R \in B_1, S \in B_2) &= P(R \in B_1 \cap M, S \in B_2 \cap M) \\ &= P(R \in B_1 \cap M)P(S \in B_2 \cap M) \\ &= P(R \in B_1)P(S \in B_2), \end{aligned}$$

da $B_1 \cap M$ og $B_2 \cap M$ er delmængder af M . Altså er R og S uafhængige. Derimod har vi tidligere set, at

$$P(R \leq 4, T \leq 4) = \frac{1}{3},$$

hvilket er forskelligt fra $P(R \leq 4)P(T \leq 4) = \frac{4}{9}$. Derfor er R og T ikke uafhængige. Da $T = 7 - R$, kommer dette ikke som nogen stor overraskelse.

□

Sætning 2.4.2 *To stokastiske variable X_1 og X_2 er stokastisk uafhængige hvis og kun hvis det for ethvert par A_1 og A_2 af delmængder af \mathbb{R} med $P(X_2 \in A_2) > 0$ gælder, at*

$$P(X_1 \in A_1 | X_2 \in A_2) = P(X_1 \in A_1). \quad (2.4.2)$$

Bevis: Dette er en enkel konsekvens af definitionen af betinget sandsynlighed og af (2.4.1). Hvis nemlig X_1 og X_2 er uafhængige, er

$$P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2), \quad (2.4.3)$$

og (2.4.2) følger ved at dividere med $P(X_2 \in A_2)$. Hvis omvendt (2.4.2) gælder, følger (2.4.3) ved at gange på begge sider med $P(X_2 \in A_2)$. Definitionen af uafhængighed kræver også, at (2.4.3) holder, når $P(X_2 \in A_2) = 0$. Men i dette tilfælde er det klart, at (2.4.3) er opfyldt, da der står nul på begge sider af lighedstegnet.

□

Sætning 2.4.2 siger, at uanset hvad vi får at vide om X_2 , så vil dette ikke ændre vore forventninger til udfaldet af X_1 , når X_1 og X_2 er uafhængige.

Sætning 2.4.3 Lad X_1, \dots, X_n være uafhængige stokastiske variable. Da gælder følgende:

- 1) Hvis $\varphi_i, i = 1, \dots, n$ er funktioner fra \mathbb{R} ind i \mathbb{R} , er de stokastiske variable $\varphi_1(X_1), \dots, \varphi_n(X_n)$ uafhængige.
- 2) Hvis $k < n$ og ψ er en funktion fra \mathbb{R}^{n-k} ind i \mathbb{R} , er de stokastiske variable $X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n)$ uafhængige.
- 3) Lad k og ψ være som i 2), og lad φ være en funktion fra \mathbb{R}^k ind i \mathbb{R} . Da er de stokastiske variable $\varphi(X_1, \dots, X_k)$ og $\psi(X_{k+1}, \dots, X_n)$ uafhængige.

Bevis: Både 1) og 3) følger af 2), men da vi ikke vil vise 2) generelt, viser vi alligevel 1). Hvis A_1, \dots, A_n er delmængder af \mathbb{R} , er

$$\begin{aligned} P(\varphi_1(X_1) \in A_1, \dots, \varphi_n(X_n) \in A_n) &= P(X_1 \in \varphi_1^{-1}(A_1), \dots, X_n \in \varphi_n^{-1}(A_n)) \\ &= P(X_1 \in \varphi_1^{-1}(A_1)) \cdots P(X_n \in \varphi_n^{-1}(A_n)) \\ &= P(\varphi_1(X_1) \in A_1) \cdots P(\varphi_n(X_n) \in A_n). \end{aligned}$$

Dermed er 1) vist.

Vi vil her kun vise 2), når hver af de stokastiske variable X_1, \dots, X_n kun kan antage endeligt mange værdier. Lad M være den endelige delmængde af \mathbb{R}^{n-k} , i hvilken den stokastiske vektor (X_{k+1}, \dots, X_n) antager sine værdier, og lad A_1, \dots, A_k, A være delmængder af \mathbb{R} . Da består $\psi^{-1}(A) \cap M$ kun af endeligt mange elementer. Det vil sige, at $\psi^{-1}(A) \cap M$ kan skrives på formen $\{(x_{k+1}^{(j)}, \dots, x_n^{(j)}) | j = 1, \dots, N\}$. Derfor får vi ved at anvende (1.3.14) to gange, at

$$\begin{aligned} &P(X_1 \in A_1, \dots, X_k \in A_k, \psi(X_{k+1}, \dots, X_n) \in A) \\ &= P(X_1 \in A_1, \dots, X_k \in A_k, (X_{k+1}, \dots, X_n) \in \psi^{-1}(A)) \\ &= \sum_{j=1}^N P(X_1 \in A_1, \dots, X_k \in A_k, (X_{k+1}, \dots, X_n) = (x_{k+1}^{(j)}, \dots, x_n^{(j)})) \\ &= \sum_{j=1}^N P(X_1 \in A_1) \cdots P(X_k \in A_k) P((X_{k+1}, \dots, X_n) = (x_{k+1}^{(j)}, \dots, x_n^{(j)})) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) \sum_{j=1}^N P((X_{k+1}, \dots, X_n) = (x_{k+1}^{(j)}, \dots, x_n^{(j)})) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) P((X_{k+1}, \dots, X_n) \in \psi^{-1}(A)) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) P(\psi(X_{k+1}, \dots, X_n) \in A) \end{aligned}$$

Dermed er 2) vist.

□

Vi vil senere give et bevis for punkt 2) i Sætning 2.4.3 for de to vigtigste typer af fordelinger.

2.5 Sammenfatning

I dette kapitel har vi indført begreberne stokastisk variabel og stokastisk vektor, som formaliserer den meget almindelige situation, at vi ikke er interesserede i hele udfaldsrummet, men kun i et tal eller eventuelt en vektor, som afhænger af udfaldet. Det er meget vigtigt at blive god til at regne med stokastiske variable og deres fordeling. Fordelingen er et sandsynlighedsmål på \mathbb{R} eller \mathbb{R}^n , som beskriver den tilfældige variation af den stokastiske variable eller den stokastiske vektor. Et vigtigt hjælpemiddel til at beskrive og behandle fordelingen af en stokastisk variabel er fordelingsfunktionen. Vi har blandt andet bevist en transformationssætning for fordelingsfunktioner. Hvis vi kender fordelingsfunktionen for en stokastisk variabel X , fortæller transformationssætningen os, hvad fordelingen er af den stokastiske variable $t(X)$, som fremkommer ved at anvende funktionen t på udfaldet af X .

En stokastisk vektor i \mathbb{R}^n kan opfattes som n stokastiske variable. Det er naturligt at spørge, om der er en sammenhæng mellem disse stokastiske variables tilfældige variation. En sådan sammenhæng beskrives ved den simultane fordeling af de n stokastiske variable. Den tilfældige variation af en enkelt af de n stokastiske variable set isoleret beskrives ved den marginale fordeling af denne stokastiske variable. Normalt er de marginale fordelinger ikke nok til at beskrive den simultane variation af alle n stokastiske variable. Hvis de n stokastiske variable er uafhængige, er der dog en simpel sammenhæng givet ved (2.4.1) mellem den simultane fordeling og de n marginale fordelinger. Derfor spiller antagelser om uafhængighed ofte en vigtig rolle ved opbygningen af stokastiske modeller.

2.6 Opgaver

- 2.1 Betragt kast med to terninger. Hvad er fordelingerne af de to stokastiske variable, som angiver antallet af øjne på henholdsvis terningen med det mindste antal øjne og terningen med det største antal øjne? Dette er de to stokastiske variable, som i Eksempel 1.2.1 i Afsnit 2.1 blev kaldt Y og Z .
- 2.2 Tre personer udvælges tilfældigt. Hvad er fordelingen af antallet blandt disse, som er født på en søndag?
- 2.3 Lad X være en stokastisk variabel, hvis fordeling er givet ved

$$\begin{aligned} P(X = 1) &= 0.12 & P(X = 2) &= 0.08 & P(X = 3) &= 0.20 \\ P(X = 4) &= 0.11 & P(X = 5) &= 0.19 & P(X = 6) &= 0.14 \\ P(X = 7) &= 0.06 & P(X = 8) &= 0.10, \end{aligned}$$

mens $P(X = x) = 0$, når $x \notin \{1, 2, 3, 4, 5, 6, 7, 8\}$. Lad t være en funktion fra \mathbb{R} ind i \mathbb{R} , som opfylder, at

$$\begin{aligned} t(1) &= t(2) = t(3) = 1 \\ t(4) &= t(5) = 2 \\ t(6) &= t(7) = 3 \\ t(8) &= 4. \end{aligned}$$

Hvordan t er defineret udenfor mængden $\{1, 2, 3, 4, 5, 6, 7, 8\}$ er ligegyldigt. Definer en ny stokastisk variabel ved $Y = t(X)$. Hvad er fordelingen af Y ? Tegn fordelingsfunktionen for Y .

- 2.4 Lad X være en stokastisk variabel, som kun kan antage værdierne 1, 2 og 3, og hvis fordeling er givet ved $P(X = 1) = P(X = 2) = P(X = 3) = \frac{1}{3}$. Tegn fordelingsfunktionen for X og for den stokastiske variable $Y = 1/X$.
- 2.5 Lad X_1 og X_2 være stokastiske variable, som begge kun kan antage værdierne 0 og 1. Lad deres marginale fordelinger være givet ved $P(X_1 = 0) = 0.4$, $P(X_1 = 1) = 0.6$, $P(X_2 = 0) = 0.3$ og $P(X_2 = 1) = 0.7$. Antag først at den simultane fordeling af den

stokastiske vektor (X_1, X_2) er givet ved $P((X_1, X_2) = (0, 0)) = 0.12$, $P((X_1, X_2) = (1, 0)) = 0.18$, $P((X_1, X_2) = (0, 1)) = 0.28$ og $P((X_1, X_2) = (1, 1)) = 0.42$, og undersøg, om X_1 og X_2 under denne antagelse er uafhængige. Antag dernæst at den simultane fordeling af (X_1, X_2) er givet ved $P((X_1, X_2) = (0, 0)) = 0.15$, $P((X_1, X_2) = (1, 0)) = 0.15$, $P((X_1, X_2) = (0, 1)) = 0.25$ og $P((X_1, X_2) = (1, 1)) = 0.45$, og undersøg også for denne simultane fordeling, om X_1 og X_2 er uafhængige. Gør til slut rede for, at begge simultane fordelinger er i overensstemmelse med de angivne marginale fordelinger for X_1 og X_2 .

- 2.6 Lad M være en endelig delmængde af \mathbb{R} og lad X være en stokastisk variabel, som er ligefordelt på M (d.v.s. X antager kun værdier i M , og $P(X = x)$ er den samme for alle $x \in M$). Lad endvidere t være en funktion fra \mathbb{R} in i \mathbb{R} , og definer en ny stokastisk variabel ved $Y = t(X)$. Hvilke betingelser skal t opfylde, for at Y er ligefordelt på $t(M)$?
- 2.7 Lad X være en stokastisk variabel, som kun antager værdier i intervallet $[0, 1]$, og hvis fordeling er givet ved det sandsynlighedsmål, der blev studeret i Eksempel 1.3.3. Find fordelingsfunktionen for X . Definer en ny stokastisk variabel ved $Y = X^2$. Hvad er sandsynligheden for, at Y tilhører intervallet $[y_1, y_2]$, hvor $0 < y_1 < y_2 < 1$? Find fordelingsfunktionen for Y .
- 2.8 Antag at et tal udtrækkes tilfældigt mellem nul og en (se Eksempel 1.3.1), og lad X være den stokastiske variable, som angiver det udtrukne tal. Da er fordelingsfunktionen for X givet ved (2.2.2). Definer en ny stokastisk variabel ved $Y = -\log(X)/\lambda$, hvor $\lambda > 0$. I hvilket interval antager Y sine værdier? Hvad er fordelingsfunktionen for Y ? Sammenlign resultatet med Eksempel 2.2.2.
- 2.9 Betragt kast med to terninger. Hvad er den simultane fordeling for den to-dimensionale stokastiske vektor, hvis første koordinat, Y , angiver antallet af øjne på terningen med det mindste antal øjne, mens den anden koordinat, Z , angiver antallet af øjne på terningen med det største antal øjne? Er Y og Z uafhængige? (De to stokastiske variable Y og Z , blev defineret i Eksempel 1.2.1).

- 2.10 Lad X være en stokastisk variabel og lad t være en funktion fra \mathbb{R} ind i \mathbb{R} . Definer en ny stokastisk variabel ved $Y = t(X)$, og antag, at X og Y er stokastisk uafhængige. Vis, at Y så er stokastisk uafhængig af sig selv, og at der derfor for enhver hændelse A gælder, at $P(Y \in A) = P(Y \in A)^2$. Slut heraf, at der findes et a , så fordelingsfunktionen for Y har formen

$$F(y) = \begin{cases} 1 & \text{for } y \geq a \\ 0 & \text{for } y < a. \end{cases}$$

Vink: Husk fordelingsfunktionens egenskaber, ikke mindst (2.2.3) og (2.2.4). Det følger af opgaven, at $t(X)$ med sandsynlighed 1 er konstant lig a .

- 2.11 Vis, at 2) i Sætning 2.4.3 medfører 1) og 3) i samme sætning.

Kapitel 3

Fordelinger på endelige mængder

I dette kapitel vil vi studere stokastiske variable, som kun kan antage endeligt mange værdier, og deres fordeling. Ved i første omgang at nøjes med at diskutere sådanne stokastiske variable, undgår vi nogle matematiske komplikationer. Vi kan derfor indføre en række vigtige begreber uden at det væsentlige drukner i tekniske detaljer. I de følgende kapitler bliver vi så nødt til at tage hånd om disse tekniske problemer. I indeværende kapitel vil vi også indføre og behandle nogle vigtige fordelinger, som ofte forekommer i praksis.

Lad os begynde med nogle ikke særligt dybsindige betragtninger om et forhold, som det er vigtigt at gøre sig klart. Lad P være et sandsynlighedsmål på \mathbb{R}^n og lad M være en delmængde af \mathbb{R}^n . Vi siger, at P er *koncentreret* på M , hvis $P(M) = 1$. Tilsvarende siges en stokastisk variabel eller vektor X at være koncentreret på M , hvis dens fordeling er koncentreret på M , altså hvis $P(X \in M) = 1$. Hvis P er koncentreret på M , er det nok at angive sandsynligheden for alle delmængder af M . For en vilkårlig $B \subseteq \mathbb{R}^n$ er nemlig $P(B) = P(B \setminus M) + P(B \cap M) = P(B \cap M)$, da $B \setminus M \subseteq \mathbb{R}^n \setminus M$ og dermed $P(B \setminus M) \leq P(\mathbb{R}^n \setminus M) = 1 - P(M) = 0$. Det er let at vise, at hvis vi kun betragter P som en funktion af delmængderne af M , så er P et sandsynlighedsmål på M (overvej). Omvendt kan man udvide et sandsynlighedsmål P på M til et sandsynlighedsmål på hele \mathbb{R}^n ved for en vilkårlig delmængde $B \subseteq \mathbb{R}^n$ at definere $P(B) = P(B \cap M)$.

Det er ikke svært at se, at den således definerede funktion er et sandsynlighedsmål (bevis det). Alt efter hvad der er mest praktisk i en given situation, kan vi derfor frit opfatte et sandsynlighedsmål, som er koncentreret på M , enten som et sandsynlighedsmål på \mathbb{R}^n eller som et sandsynlighedsmål på M .

3.1 Sandsynlighedsfunktioner

Som nævnt i indledningen skal vi i dette kapitel beskæftige os med sandsynlighedsmål, som er koncentreret på en endelig delmængde M af \mathbb{R} eller \mathbb{R}^n . Det er somme tider nyttigt at nummerere M s elementer, altså at skrive M på formen $M = \{x_1, \dots, x_k\}$. Som omtalt i afsnit 1.2 er der til et sandsynlighedsmål, som er koncentreret på den endelige mængde M , knyttet en såkaldt sandsynlighedsfunktion.

Definition 3.1.1 *En funktion p fra en endelig mængde M ind i intervallet $[0, 1]$, som opfylder, at*

$$\sum_{x \in M} p(x) = 1 \quad (3.1.1)$$

kaldes en sandsynlighedsfunktion på M .

Med $\sum_{x \in M} p(x)$ menes summen over alle de værdier af $p(x)$, som opnås, når x gennemløber M , d.v.s. $\sum_{i=1}^k p(x_i)$.

Sætning 3.1.2 *Der er en entydig korrespondance mellem sandsynlighedsfunktionerne på M og sandsynlighedsmålene, som er koncentreret på M . Således svarer der til enhver sandsynlighedsfunktion p et sandsynlighedsmål P givet ved*

$$P(A) = \sum_{x \in A} p(x) \quad (3.1.2)$$

for enhver delmængde A af M . Ligeledes svarer der til ethvert sandsynlighedsmål P en sandsynlighedsfunktion p givet ved

$$p(x) = P(\{x\}) \quad (3.1.3)$$

for ethvert $x \in M$.

Bevis: Lad p være en sandsynlighedsfunktion på M og definer P ved (3.1.2). Vi skal da vise, at P er et sandsynlighedsmål på M . At der for enhver delmængde A af M gælder, at $0 \leq P(A) \leq 1$, er ikke svært at se, og at $P(M) = 1$ følger af (3.1.1). For to disjunkte delmængder af M , A og B , er

$$P(A \cup B) = \sum_{x \in A \cup B} p(x) = \sum_{x \in A} p(x) + \sum_{x \in B} p(x) = P(A) + P(B).$$

Dermed er det vist, at P er et sandsynlighedsmål på M .

Hvis vi omvendt ved (3.1.3) definerer en funktion p ud fra et sandsynlighedsmål P på M , skal det vises, at p er en sandsynlighedsfunktion på M . At $p(x) \in [0, 1]$ er klart, da P er et sandsynlighedsmål. Da $M = \{x_1\} \cup \dots \cup \{x_k\}$, følger det af (1.3.14) at

$$\sum_{i=1}^k p(x_i) = \sum_{i=1}^k P(\{x_i\}) = P(M) = 1.$$

Altså er p en sandsynlighedsfunktion på M . Hvis vi ud fra denne sandsynlighedsfunktion p konstruerer et sandsynlighedsmål ved (3.1.2), er det klart, at vi genfinder det sandsynlighedsmål, som vi startede med. \square

Bemærk, at vi også kan skrive (3.1.2) på følgende måde

$$P(A) = \sum_{i=1}^k p(x_i) 1_A(x_i),$$

hvor vi har benyttet indikatorfunktionen for A , se (A.1.15).

Der er grund til igen at understrege, at sandsynlighedsfunktionens værdi i et punkt x netop er lig den sandsynlighed, som det tilsvarende sandsynlighedsmål tillægger hændelse $\{x\}$, altså sandsynligheden for udfaldet x . Denne sandsynlighed kaldes også punktsandsynligheden i x . Hvis man har lyst, kan man godt definere p udenfor M ved at sætte $p(x) = 0$ for $x \notin M$.

Vi har allerede set flere eksempler på sandsynlighedsmål på endelige mængder. Det enkleste eksempel er den såkaldte *ligefordeling* på den endelige mængde M , hvor sandsynlighedsmassen er fordelt jævnt ud over M , d.v.s. den tilsvarende sandsynlighedsfunktion er konstant på M . Dens konstante værdi må, for at (3.1.1) kan opfyldes, være $1/\#M$.

Meget ofte kaldes et sandsynlighedsmaal på \mathbb{R} eller \mathbb{R}^n en sandsynlighedsfordeling eller blot en fordeling. Det vil vi også gøre i det følgende.

3.2 Binomialfordelingen

I dette afsnit betragter vi situationen, hvor et forsøg med to mulige udfald gentages n gange uafhængigt af hinanden. Vi er specielt interesserede i fordelingen af antallet af gange det ene af de to udfald forekommer. Fordelingen af dette antal kaldes en *binomialfordeling*. Et eksempel er antal krone i n kast med en mønt. Et andet er antal defekte produkter i en stikprøve på n fra en løbende produktion. Der vil blive givet flere eksempler i kurssets statistikdel. Her kommer en præcis definition af binomialfordelingen.

Definition 3.2.1 *Lad X_1, X_2, \dots, X_n være uafhængige stokastiske variable med værdier i $\{0, 1\}$, som alle har samme fordeling givet ved*

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p$$

for et givet tal p mellem 0 og 1. Fordelingen af summen $S = X_1 + X_2 + \dots + X_n$ kaldes da binomialfordelingen med antalsparameter n og sandsynlighedsparameter p .

Det er klart, at binomialfordelingen er koncentreret på mængden $\{0, 1, \dots, n\}$.

Eksempel 3.2.2 En mønt kastes 20 gange. Antallet af kast med udfaldet krone er binomialfordelt med antalsparameter 20 og sandsynlighedsparameter $1/2$.

□

For at finde sandsynlighedsfunktionen for binomialfordelingen, finder vi først sandsynlighedsfunktionen for den stokastiske vektor (X_1, \dots, X_n) , som er koncentreret på mængden $\{0, 1\}^n$. For $x_i \in \{0, 1\}$ er

$$P(X_i = x_i) = p^{x_i}(1 - p)^{1 - x_i},$$

hvor $i = 1, \dots, n$, og da X_1, \dots, X_n er uafhængige, er

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= p^{x_1 + \cdots + x_n} (1 - p)^{n - (x_1 + \cdots + x_n)} \end{aligned}$$

Lad nu t være funktionen givet ved $t(x_1, \dots, x_n) = x_1 + \cdots + x_n$. Vi skal da finde fordelingen af den stokastiske variable $S = t(X_1, \dots, X_n)$. For $s \in \{0, \dots, n\}$ er

$$\begin{aligned} P(S = s) &= P((X_1, \dots, X_n) \in t^{-1}(\{s\})) \\ &= \sum_{(x_1, \dots, x_n) \in t^{-1}(\{s\})} P((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{(x_1, \dots, x_n) \in t^{-1}(\{s\})} p^s (1 - p)^{n-s} \\ &= \#t^{-1}(\{s\}) p^s (1 - p)^{n-s}, \end{aligned}$$

hvor vi har benyttet (3.1.2) (eller om man vil (1.3.14)), samt at

$$t^{-1}(\{s\}) = \{(x_1, \dots, x_n) \in \{0, 1\}^n \mid x_1 + \cdots + x_n = s\}.$$

Som sædvanlig betegnes antallet af elementer i mængden $t^{-1}(\{s\})$ med $\#t^{-1}(\{s\})$. Vi mangler derfor blot at finde dette antal. Da

$$t^{-1}(\{s\}) = \{(x_1, \dots, x_n) \in \{0, 1\}^n \mid \text{netop } s \text{ af } x_i\text{-erne er lig } 1\},$$

er $\#t^{-1}(\{s\})$ lig antallet af delmængder med s elementer af en mængde med n elementer, men det er netop binomialkoefficienten $\binom{n}{s} = n! / (s!(n-s)!)$, se Appendix B. Vi har dermed bevist følgende sætning.

Sætning 3.2.3 *Binomialfordelingens sandsynlighedsfunktion er givet ved*

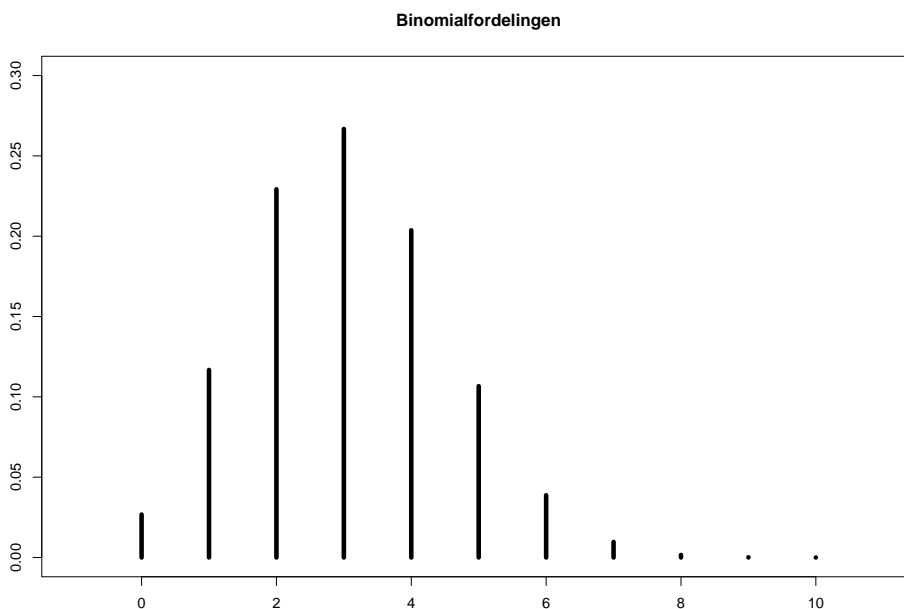
$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad (3.2.1)$$

for $x \in \{0, 1, \dots, n\}$.

Da S er koncentreret på $\{0, 1, \dots, n\}$, må $\sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = 1$, hvilket er i overensstemmelse med binomialformlen, se Appendix B.

Eksempel 3.2.4 Sandsynligheden for at få en hånd på tretten kort uden esser er $(52 - 4)^{13} / 52^{13} = 0.3038$ (jfr. opgave 1.9). Hvad er sandsynligheden for, at denne kedelige hændelse for en bestemt spiller indtræffer præcis 3 gange i løbet af 10 spil? Da antallet er binomialfordelt med antalsparameter 10 og sandsynlighedsparameter 0.3038, er svaret

$$\binom{10}{3} \times 0.3038^3 \times (1 - 0.3038)^{10-3} = 0.2667.$$



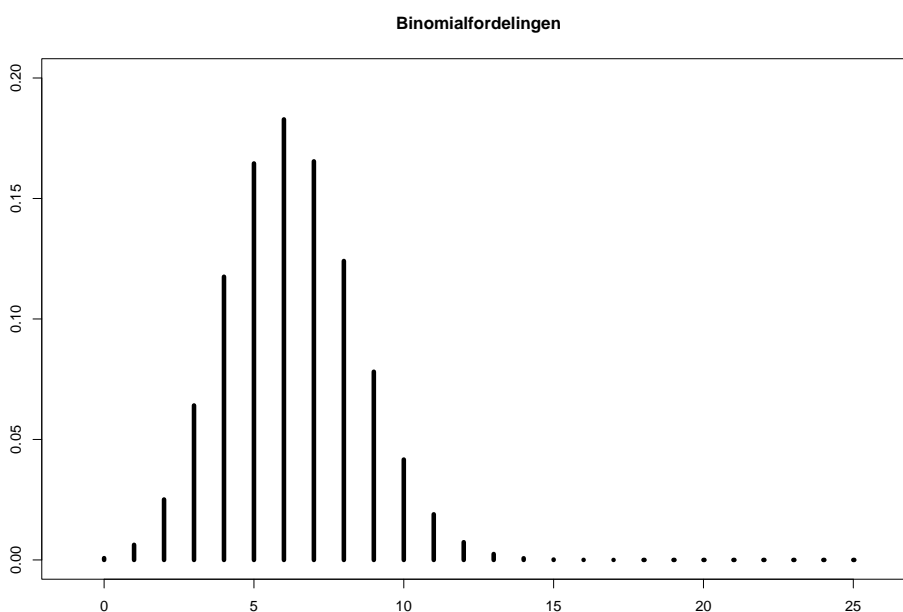
Figur 3.2.1: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 10 og sandsynlighedsparameter 0.3038. I Eksempel 3.2.4 er det fordelingen af antal es-løse hænder i 10 spil.

□

Eksempel 3.2.5 En person sår 25 planter, som hver med sandsynlighed $\frac{1}{4}$ får blå blomster og med sandsynlighed $\frac{3}{4}$ får gule blomster. Da er

antallet af planter med blå blomster binomialfordelt med antalsparameter 25 og sandsynlighedsparameter $\frac{1}{4}$, og sandsynligheden for at 10 planter får blå blomster er

$$\binom{25}{10} \times 3^{15} \times 4^{-25} = 0.0417.$$



Figur 3.2.2: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 25 og sandsynlighedsparameter $\frac{1}{4}$. I Eksempel 3.2.5 er det fordelingen af antal planter med blå blomster.

□

Sætning 3.2.6 *Lad S_1 og S_2 være uafhængige stokastiske variable, hvor S_i er binomialfordelt med antalsparameter n_i og sandsynlighedsparameter p , $i = 1, 2$. Da er $S_1 + S_2$ binomialfordelt med antalsparameter $n_1 + n_2$ og sandsynlighedsparameter p .*

Bevis: Se Opgave 3.5.

□

Det er vigtigt at gøre sig klart, at de to binomialfordelinger i Sætning 3.2.6 har samme sandsynlighedsparameter. Hvis det ikke er tilfældet, er summen ikke binomialfordelt. Overvej hvorfor.

3.3 Den hypergeometriske fordeling

Den model, som blev studeret i forrige afsnit dækker den situation, hvor man udtager et antal stikprøver, som alle har samme sandsynlighed for de to mulige udfald. Af og til kommer man ud for, at stikprøver udtrækkes af små populationer og at en person eller genstand, som er blevet udtrukket én gang ikke igen kan udtrækkes. Da vil det faktisk, at der allerede er udtrukket et vist antal stikprøver, påvirke fordelingen af den næste udtrækning. Denne type situation, som kaldes *stikprøveudtagning uden tilbagelægning*, kan formuleres ved hjælp af farvede kugler på følgende måde.

En kasse indeholder N kugler, af hvilke R er røde og $H = N - R$ er hvide. En stikprøve på n kugler udtages. Vi skal studere fordelingen af antallet af røde kugler i denne stikprøve. Denne fordeling kaldes *den hypergeometriske fordeling* med parametre N , R og n . Bemærk, at hvis vi n gange havde trukket en kugle og derefter lagt den tilbage i kassen før vi trak igen, så ville antallet af røde kugler blandt n udtrukne have været binomialfordelt med antalsparameter n og sandsynlighedsparameter R/N (overvej hvorfor denne påstand er rigtig). Dette kaldes stikprøveudtagning med tilbagelægning. Pointen i dette afsnit er at undersøge effekten af ikke at lægge de udtrukne kugler tilbage i kassen. I resten af afsnittet betragtes altså situationen uden tilbagelægning.

Lad X betegne antallet af røde kugler blandt de n udtrukne kugler. Da er X en stokastisk variabel som er koncentreret på mængden $\{0, 1, \dots, n\}$. Det er ikke altid, at X kan antage alle disse værdier. Det kan X kun når $n \leq R$ og $n \leq H$. Hvis for eksempel $n > R$, er der ikke tilstrækkeligt mange røde kugler i kassen til, at X kan antage værdien n .

Lad os nu finde sandsynlighedsfunktionen for den hypergeometriske fordeling. Sandsynligheden for hændelsen $\{X = x\}$ er nul, hvis $x > R$ eller hvis $n - x$ (antal hvide kugler i stikprøven) er større end $N - R$. Vi vil derfor antage, at $(n - (N - R)) \vee 0 \leq x \leq R$. Da rækkefølgen af de udtagne kugler ikke spiller nogen rolle, kan vi opfatte stikprøven som

en delmængde med n elementer af mængden $\{1, 2, \dots, N\}$ af kugler. Vi kan også antage, at de første R kugler, $\{1, \dots, R\}$, er de røde, mens de sidste $N - R$ kugler, $\{R + 1, \dots, N\}$ er de hvide. Vi ser da, at der er $\binom{N}{n}$ mulige udfald, som alle er lige sandsynlige. Vi har altså en ligefordeling på n -delmængderne af $\{1, 2, \dots, N\}$. Sandsynligheden for hændelsen $\{X = x\}$ er derfor forholdet mellem antallet af gunstige n -delmængder (altså delmængder med netop x røde og $n - x$ hvide kugler), og antallet af mulige n -delmængder, $\binom{N}{n}$. For at tælle hvor mange af delmængderne, der har netop x røde og $n - x$ hvide kugler, bemærker vi at en gunstig n -delmængde er beskrevet ved en x -delmængde af $\{1, \dots, R\}$ og en $(n - x)$ -delmængde af $\{R + 1, \dots, N\}$. Antallet af sådanne par af delmængder er åbenbart $\binom{R}{x} \times \binom{N-R}{n-x}$. Vi har derfor vist følgende sætning.

Sætning 3.3.1 *Den hypergeometriske fordelings punktsandsynligheder er givet ved*

$$P(X = x) = \frac{\binom{R}{x} \binom{N-R}{n-x}}{\binom{N}{n}} \quad (3.3.1)$$

for $(n - (N - R)) \vee 0 \leq x \leq R$, mens $P(X = x)$ er nul for alle andre værdier af x .

Sandsynlighed $P(X = x)$ kan også skrives på formen

$$\begin{aligned} P(X = x) &= \frac{R^{(x)}(N - R)^{(n-x)}n!}{x!(n - x)!N^{(n)}} \\ &= \binom{n}{x} \frac{R^{(x)}(N - R)^{(n-x)}}{N^{(n)}}, \end{aligned}$$

som gør sammenligning med binomialfordelingen lettere og involverer lidt færre faktorer.

Eksempel 3.3.2 15 kort trækkes tilfældigt fra et almindeligt spil kort. Hvad er sandsynligheden for at netop 7 af dem er spar? Svaret er punktsandsynligheden i punktet 7 for den hypergeometriske fordeling med parametrene $N = 52$, $R = 13$ og $n = 15$, altså

$$\frac{\binom{13}{7} \binom{39}{8}}{\binom{52}{15}} = \binom{15}{7} \frac{13^{(7)} \times 39^{(8)}}{52^{(15)}} = 0.0236.$$

□

Eksempel 3.3.3 En kasse indeholder 1000 røde og 2000 hvide kugler. 5 kugler udtages. Hvad er sandsynligheden for at netop 2 af disse er røde? Svaret er

$$\binom{5}{2} \frac{1000^{(2)} 2000^{(3)}}{3000^{(5)}} = 10 \times \frac{(1000 \times 999) \times (2000 \times 1999 \times 1998)}{3000 \times 2999 \times 2998 \times 2997 \times 2996} = 0.3295.$$

Det er ret oplagt, at man i næstsidste udtryk kan erstatte 999 med 1000, 1999 med 2000 osv. uden at ændre ret meget på resultatet. Hvis vi gør det, opnår vi følgende approksimative resultat

$$P(X = 2) \simeq 10 \times \frac{1000^2 2000^3}{3000^5} = 10 \times \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = 0.3292,$$

som netop er sandsynligheden for udfaldet $S = 2$, når S er binomialfordelt med antalsparameter 5 og sandsynlighedsparameter $1/3$. Dette er præcis det svar, vi ville få, hvis udtagningen af de 5 kugler var foregået med tilbagelægning. Intuitivt følger denne approksimation af, at når der er tilstrækkeligt mange kugler i kassen i forhold til det antal, der udtrækkes, kan det ikke gøre megen forskel, om vi lægger de få udtrukne kugler tilbage igen eller ej. \square

Eksempel 3.3.3 illustrerer følgende generelle resultat.

Sætning 3.3.4 (Approksimation med en binomialfordeling). *Lad X være hypergeometrisk fordelt med parametre N , R og n . Hvis stikprøvestørrelsen n holdes fast, mens parametrene N og R går imod uendelig på en sådan måde, at R/N konvergerer mod et tal $p \in (0, 1)$, vil*

$$P(X = x) \rightarrow \binom{n}{x} p^x (1 - p)^{n-x}.$$

Bevis: Vi finder efter en simpel omskrivning, at

$$P(X = x) = \binom{n}{x} \times \left[\left(\frac{R}{N}\right) \left(\frac{R-1}{N-1}\right) \cdots \left(\frac{R-x+1}{N-x+1}\right) \right] \\ \times \left[\left(\frac{N-R}{N-x}\right) \left(\frac{N-R-1}{N-x-1}\right) \cdots \left(\frac{N-R-(n-x)+1}{N-n+1}\right) \right].$$

Da

$$\begin{aligned} & \left(\frac{R}{N}\right) \left(\frac{R-1}{N-1}\right) \cdots \left(\frac{R-x+1}{N-x+1}\right) \\ &= \left(\frac{R}{N}\right)^x \left[\left(\frac{1-R^{-1}}{1-N^{-1}}\right) \left(\frac{1-2R^{-1}}{1-2N^{-1}}\right) \cdots \left(\frac{1-(x-1)R^{-1}}{1-(x-1)N^{-1}}\right) \right] \end{aligned}$$

og

$$\begin{aligned} & \left(\frac{N-R}{N-x}\right) \left(\frac{N-R-1}{N-x-1}\right) \cdots \left(\frac{N-R-(n-x)+1}{N-n+1}\right) = \left(\frac{N-R}{N}\right)^{n-x} \\ & \times \left[\left(\frac{1}{1-xN^{-1}}\right) \left(\frac{1-(N-R)^{-1}}{1-(x+1)N^{-1}}\right) \cdots \left(\frac{1-(n-x-1)(N-R)^{-1}}{1-(n-1)N^{-1}}\right) \right], \end{aligned}$$

ses det, at under grænseovergangen vil den første af de kantede parenteser i udtrykket for $P(X=x)$ gå imod p^x , medens den anden kantede parentes konvergerer mod $(1-p)^{n-x}$. Dette viser sætningen. \square

3.4 Transformation af fordelinger

Hvis p er sandsynlighedsfunktion for en stokastisk vektor, som er koncentreret på en endelig mængde, er der naturligvis relationer mellem den simultane sandsynlighedsfunktion p og sandsynlighedsfunktionerne for de tilsvarende marginale fordelinger. Følgende sætning giver disse relationer for en to-dimensional fordeling. Der gælder et helt tilsvarende resultat for højere-dimensionale fordelinger, som den interesserede studerende let selv kan formulere og bevise.

Sætning 3.4.1 *Lad (X_1, X_2) være en to-dimensional stokastisk vektor, hvor X_i er koncentreret på den endelige mængde T_i , $i = 1, 2$, og lad $p : T_1 \times T_2 \mapsto [0, 1]$ være sandsynlighedsfunktionen for den simultane fordeling af (X_1, X_2) . Da er sandsynlighedsfunktionerne for de marginale fordelinger af X_1 og X_2 , $p_1 : T_1 \mapsto [0, 1]$ og $p_2 : T_2 \mapsto [0, 1]$, givet ved*

$$p_1(x_1) = \sum_{x_2 \in T_2} p(x_1, x_2) \quad (3.4.1)$$

og

$$p_2(x_2) = \sum_{x_1 \in T_1} p(x_1, x_2). \quad (3.4.2)$$

Bevis: Vi viser kun (3.4.1), da de to resultater er helt symmetriske. For $x \in T_1$ følger det af (3.1.2), at

$$\begin{aligned} P(X_1 = x) &= P((X_1, X_2) \in \{x\} \times T_2) \\ &= \sum_{(x_1, x_2) \in \{x\} \times T_2} p(x_1, x_2) = \sum_{x_2 \in T_2} p(x, x_2). \end{aligned}$$

□

Der gælder følgende transformationssætning for fordelinger på endelige mængder, som Sætning 3.4.1 er et specialtilfælde af.

Sætning 3.4.2 *Lad (X_1, \dots, X_n) være en n -dimensional stokastisk vektor koncentreret på den endelige mængde $T \subseteq \mathbb{R}^n$. Betegn den simultane sandsynlighedsfunktion med p , og lad ψ være en afbildning fra T ind i \mathbb{R}^k . Da er sandsynlighedsfunktionen for den stokastiske vektor $Y = \psi(X_1, \dots, X_n)$ givet ved*

$$q(y) = \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) \quad \text{hvis } y \in \psi(T). \quad (3.4.3)$$

Udenfor $\psi(T)$ har q værdien nul.

Bevis: For $y \in \psi(T)$ er

$$\begin{aligned} P(Y = y) &= P((X_1, \dots, X_n) \in \psi^{-1}(\{y\})) \\ &= \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n). \end{aligned}$$

□

I beviset for Sætning 3.2.3 forkom faktisk en anvendelse af Sætning 3.4.2. Prøv at gå gennem dette bevis igen. Her følger endnu to eksempler.

Eksempel 3.4.3 Lad X være binomialfordelt med antalsparameter n og sandsynlighedsparameter p . Da er $Y = X/c$, hvor c er et positivt reelt tal, koncentreret på mængden $T = \{0, 1/c, \dots, n/c\}$ med sandsynlighedsfunktion

$$q(y) = \binom{n}{cy} p^{cy} (1-p)^{n-cy} \text{ for } y \in T.$$

Videre er $Z = X^2$ koncentreret på mængden $\tilde{T} = \{i^2 | i = 1, \dots, n\}$ med sandsynlighedsfunktion

$$q(z) = \binom{n}{\sqrt{z}} p^{\sqrt{z}} (1-p)^{n-\sqrt{z}} \text{ for } z \in \tilde{T}.$$

□

Den næste vigtige sætning følger også af transformationssætningen Sætning 3.4.2.

Sætning 3.4.4 Lad (X_1, X_2) være en to-dimensional stokastisk vektor, hvor X_i 's fordeling er koncentreret på $\{0, 1, \dots, n_i\}$ ($n_i \in \mathbb{N}$), $i = 1, 2$. Lad endvidere p være sandsynlighedsfunktionen for den simultane fordeling af (X_1, X_2) . Da er den stokastiske variable $Y = X_1 + X_2$ koncentreret på mængden $\{0, 1, \dots, n_1 + n_2\}$, og dens sandsynlighedsfunktion er givet ved

$$q(y) = \sum_{j=0}^y p(j, y-j), \quad (3.4.4)$$

når $y \in \{0, 1, \dots, n_1 + n_2\}$.

Bevis: Resultatet følger af Sætning 3.4.2 med $\psi(x_1, x_2) = x_1 + x_2$, idet

$$\psi^{-1}(\{y\}) = \{(0, y), (1, y-1), \dots, (y-1, 1)(y, 0)\}.$$

□

3.5 Polynomialfordelingen

Polynomialfordelingen, som også kaldes multinomialfordelingen, er den naturlige generalisering af binomialfordelingen til situationer, hvor det enkelte forsøg har mere end to mulige udfald.

Definition 3.5.1 Lad X_1, X_2, \dots, X_n være uafhængige stokastiske variable med værdier i $\{1, 2, \dots, k\}$, og antag at de alle har samme fordeling givet ved

$$P(X_i = j) = p_j$$

($j = 1, \dots, k$), hvor p_1, \dots, p_k er givne tal i $[0, 1]$, som opfylder at $p_1 + \dots + p_k = 1$. Definer den stokastiske vektor (S_1, \dots, S_k) ved

$$S_j = \#\{i | X_i = j\}.$$

Fordelingen af (S_1, \dots, S_k) kaldes en polynomialfordeling af orden k med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_k .

Det er ikke svært at indse, at den stokastiske vektor kun antager værdier i følgende delmængde af \mathbb{N}_0^k

$$D_k(n) = \{(s_1, \dots, s_k) \mid s_i \in \{0, \dots, n\}, i = 1, \dots, k, s_1 + \dots + s_k = n\}. \quad (3.5.1)$$

Det er også praktisk at have et navn til variationsområdet af sandsynlighedsparametrene. Derfor defineres mængden

$$\Delta_k = \{(p_1, \dots, p_k) \mid p_i \in [0, 1], i = 1, \dots, k, p_1 + \dots + p_k = 1\}. \quad (3.5.2)$$

Bemærk, at vi for $k = 2$ har

$$(S_1, S_2) = (S_1, n - S_1),$$

hvor S_1 er antal gange hændelsen $X_i = 1$ er indtruffet. Derfor er S_1 binomialfordelt med antalsparameter n og sandsynlighedsparameter p_1 . Polynomialfordelingen af orden 2 er således essentielt en binomialfordeling.

Eksempel 3.5.2 En terning kastes 20 gange. Lad S_i være antallet af kast, hvor udfaldet blev i øjne, $i = 1, \dots, 6$. Hvis de 20 kast antages at være uafhængige af hinanden, er den stokastiske vektor $(S_1, S_2, S_3, S_4, S_5, S_6)$ polynomialfordelt med antalsparameter 20 og sandsynlighedsparametre $1/6, 1/6, 1/6, 1/6, 1/6, 1/6$.

□

Eksempel 3.5.3 I 1865 udførte Mendel nogle berømte forsøg med ærteplanter. Ved at krydse to ærtestammer opnåede han nogle ærter, som skønt de så ens ud, var af fire mulige typer. Når de blev dyrket fremkom der nemlig ærteplanter med ærter, hvis form kunne være rund eller kantet, og hvis farve kunne være gul eller grøn. Ifølge Mendels arvelighedslære er sandsynligheden for, at en plante har ærter med en given form og farve som angivet i følgende tabel.

Runde		Kantede	
Gule	Grønne	Gule	Grønne
9/16	3/16	3/16	1/16

Mendel dyrkede 556 ærter, hvis typer kan betragtes som uafhængige. Hvis S_i er antallet af ærter blandt de dyrkede, som viste sig at være af den i te type, er den stokastiske vektor (S_1, S_2, S_3, S_4) polynomialfordelt med antalsparameter 556 og sandsynlighedsparametre $9/16, 3/16, 3/16, 1/16$. I Mendels forsøg blev udfaldet af den stokastiske vektor $(315, 108, 101, 32)$. Man kan med metoder fra statistikken undersøge, om dette udfald er i overensstemmelse med Mendels teori.

□

For at opskrive polynomialfordelingens sandsynlighedsfunktion, er det nødvendigt at benytte *polynomialkoefficienterne*

$$\binom{n}{s_1, \dots, s_k} = \frac{n!}{s_1! \cdots s_k!},$$

som omtales nærmere i Appendiks B. Disse koefficienter kaldes også for multinomialkoefficienterne. Bemærk, at vi for $k = 2$ genfinder binomialkoefficienterne, under den lidt ændrede betegnelse

$$\binom{n}{s} = \binom{n}{s, n-s}$$

Sætning 3.5.4 *Polynomialfordelingens sandsynlighedsfunktion er givet ved*

$$p(s_1, \dots, s_k) = \binom{n}{s_1, \dots, s_k} p_1^{s_1} \cdots p_k^{s_k} \quad (3.5.3)$$

for $(s_1, \dots, s_k) \in D_k(n)$.

Bevis: Sandsynlighedsfunktionen for den stokastiske vektor (X_1, \dots, X_n) er

$$P(X_1 = x_1, \dots, X_n = x_n) = p_1^{\#\{i|x_i=1\}} \cdots p_k^{\#\{i|x_i=k\}},$$

hvor $x_i \in \{1, \dots, k\}$. Vi definerer en funktion t fra mængden af værdier af (X_1, \dots, X_n) , altså $\{1, \dots, k\}^n$ ind i $D_k(n)$ ved

$$t(x_1, \dots, x_n) = (\#\{i|x_i = 1\}, \dots, \#\{i|x_i = k\}).$$

Lad endvidere $A_{(s_1, \dots, s_k)}$ betegne originalmængden for $\{(s_1, \dots, s_k)\}$ ved t , dvs.

$$\begin{aligned} A_{(s_1, \dots, s_k)} &= t^{-1}(\{(s_1, \dots, s_k)\}) \\ &= \{(x_1, \dots, x_n) \in \{1, \dots, k\}^n \mid \#\{i|x_i = j\} = s_j, j = 1, \dots, k\}. \end{aligned}$$

Da $(S_1, \dots, S_k) = t(X_1, \dots, X_n)$, følger det af Sætning 3.4.2, at for $(s_1, \dots, s_k) \in D_k(n)$ er

$$\begin{aligned} P((S_1, \dots, S_k) = (s_1, \dots, s_k)) &= \sum_{(x_1, \dots, x_n) \in A_{(s_1, \dots, s_k)}} P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{(x_1, \dots, x_n) \in A_{(s_1, \dots, s_k)}} p_1^{s_1} \cdots p_k^{s_k} \\ &= \#A_{(s_1, \dots, s_k)} p_1^{s_1} \cdots p_k^{s_k} \\ &= \binom{n}{s_1, \dots, s_k} p_1^{s_1} \cdots p_k^{s_k}. \end{aligned}$$

Det sidste lighedstegn følger, fordi antallet af elementer i $A_{(s_1, \dots, s_k)}$ er lig antallet af måder, hvorpå vi kan udvælge de k delmængder $\{i \mid x_i = j\}$

($j = 1, \dots, k$) af $\{1, \dots, n\}$, således at der er s_j elementer i $\{i \mid x_i = j\}$ for ethvert $j = 1, \dots, k$. Ifølge Appendix B er

$$\#A_{(s_1, \dots, s_k)} = \binom{n}{s_1, \dots, s_k}.$$

□

Følgende sætning kan ret let vises ud fra definitionen af polynomialfordelingen. Overvej resultatets intuitive indhold.

Sætning 3.5.5 *Antag, at (S_1, \dots, S_k) er polynomialfordelt med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_k . For en delmængde I af $\{1, 2, \dots, k\}$ er $S = \sum_{i \in I} S_i$ binomialfordelt med antalsparameter n og sandsynlighedsparameter $\sum_{i \in I} p_i$. Lad mere generelt I_1, \dots, I_m være disjunkte delmængder af $\{1, 2, \dots, k\}$, som opfylder, at $I_1 \cup \dots \cup I_m = \{1, 2, \dots, k\}$, og definer*

$$T_j = \sum_{i \in I_j} S_i \quad \text{og} \quad q_j = \sum_{i \in I_j} p_i,$$

$j = 1, \dots, m$. Da er (T_1, \dots, T_m) polynomialfordelt med antalsparameter n og sandsynlighedsparametre q_1, \dots, q_m .

Bevis: Se Opgave 3.12.

□

Bemærk, at hvis vi lader mængden I bestå af et enkelt element, ser vi, at S_i er binomialfordelt med antalsparameter n og sandsynlighedsparameter p_i .

Eksempel 3.5.2 (fortsat). Som tidligere er S_i antallet af kast, hvor udfaldet blev i øjne, $i = 1, \dots, 6$. Vi definerer en ny stokastisk vektor (T_1, T_2, T_3) ved $T_1 = S_1 + S_2$, $T_2 = S_3 + S_4$ og $T_3 = S_5 + S_6$, hvor altså for eksempel T_1 er antal kast med højst to øjne. Da er (T_1, T_2, T_3) ifølge Sætning 3.5.5 polynomialfordelt med antalsparameter n og sandsynlighedsparametre $1/3, 1/3, 1/3$. Sætningen fortæller os også, at antallet af kast med et lige antal øjne, $S_2 + S_4 + S_6$, er binomialfordelt med antalsparameter n og sandsynlighedsparameter $1/2$. Formentlig er ingen af disse resultater nogen stor overraskelse.

□

Eksempel 3.5.3 (fortsat). I Mendels forsøg er antallet af gule ærter, $S_1 + S_3$, binomialfordelt med antalsparameter 556 og sandsynlighedsparameter $3/4$. Antallet af kantede ærter, $S_3 + S_4$, er binomialfordelt med antalsparameter 556 og sandsynlighedsparameter $1/4$. Her har vi igen benyttet Sætning 3.5.5.

□

3.6 Uafhængige stokastiske variable

Det er ret let at se på sandsynlighedsfunktionen for en n -dimensional stokastisk vektor, hvorvidt de n koordinater er uafhængige stokastiske variable.

Sætning 3.6.1 *Antag, at (X_1, \dots, X_n) er en n -dimensional stokastisk vektor, hvor X_i er koncentreret på den endelige mængde T_i , $i = 1, \dots, n$. Definer $T = T_1 \times \dots \times T_n$, lad $p : T \mapsto [0, 1]$ være sandsynlighedsfunktionen for den simultane fordeling af (X_1, \dots, X_n) , og lad $p_i : T_i \mapsto [0, 1]$, være sandsynlighedsfunktionen for den marginale fordeling af X_i , $i = 1, \dots, n$. Da er de følgende tre udsagn ækvivalente:*

1) X_1, \dots, X_n er stokastisk uafhængige,

2) For alle $(x_1, \dots, x_n) \in T$ er

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n), \quad (3.6.1)$$

3) Der findes n ikke-negative reelle funktioner g_i , $i = 1, \dots, n$, så

$$p(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) \quad (3.6.2)$$

for alle $(x_1, \dots, x_n) \in T$.

Bevis: Vi viser kun sætningen i tilfældet $n = 2$. Det er tilstrækkeligt til at give ideen i beviset. Det generelle bevis er helt tilsvarende, men notationsmæssigt temmelig besværligt.

At 1) medfører 2) følger af definitionen af uafhængighed af stokastiske variable (Definition 2.4.1) ved at vælge $A_1 = \{x_1\}$ og $A_2 = \{x_2\}$. At 2) medfører 3) er klart.

Lad os nu antage, at 3) holder og vise, at det medfører 2). Ifølge Sætning 3.4.1 er

$$p_1(x_1) = \sum_{x_2 \in T_2} p(x_1, x_2) = g_1(x_1) \sum_{x_2 \in T_2} g_2(x_2),$$

og

$$p_2(x_2) = \sum_{x_1 \in T_1} p(x_1, x_2) = g_2(x_2) \sum_{x_1 \in T_1} g_1(x_1).$$

Da

$$\begin{aligned} 1 = \sum_{x_1 \in T_1} p_1(x_1) &= \sum_{x_1 \in T_1} \left(g_1(x_1) \sum_{x_2 \in T_2} g_2(x_2) \right) \\ &= \left(\sum_{x_1 \in T_1} g_1(x_1) \right) \left(\sum_{x_2 \in T_2} g_2(x_2) \right), \end{aligned}$$

følger det, at

$$p_1(x_1)p_2(x_2) = g_1(x_1)g_2(x_2) = p(x_1, x_2),$$

hvilket netop er (3.6.1).

Endelig skal vi til slut antage, at 2) er opfyldt, og vise at X_1 og X_2 så er uafhængige. Lad derfor A_1 og A_2 være vilkårlige delmængder af de reelle tal, og definer $B_1 = A_1 \cap T_1$ og $B_2 = A_2 \cap T_2$. Vi har da, at

$$\begin{aligned} &P(X_1 \in A_1, X_2 \in A_2) \\ &= P((X_1, X_2) \in B_1 \times B_2) = \sum_{(x_1, x_2) \in B_1 \times B_2} p(x_1, x_2) \\ &= \sum_{(x_1, x_2) \in B_1 \times B_2} p_1(x_1)p_2(x_2) = \sum_{x_1 \in B_1} \left(\sum_{x_2 \in B_2} p_1(x_1)p_2(x_2) \right) \\ &= \sum_{x_1 \in B_1} p_1(x_1) \left(\sum_{x_2 \in B_2} p_2(x_2) \right) = \sum_{x_1 \in B_1} p_1(x_1)P(X_2 \in B_2) \\ &= P(X_1 \in B_1)P(X_2 \in B_2) = P(X_1 \in A_1)P(X_2 \in A_2). \end{aligned}$$

D.v.s. X_1 og X_2 er uafhængige, jfr. Definition 2.4.1. □

Læg mærke til, at vi undervejs har vist, at $p_1(x_1) = cg_1(x_1)$ og $p_2(x_2) = c^{-1}g_2(x_2)$, hvor $c = \sum_{x_2 \in T_2} g_2(x_2)$. Funktionerne g_1 og g_2 er altså proportionale med henholdsvis p_1 og p_2 . De mangler bare at blive normerede, så de bliver til sandsynlighedsfunktioner.

Det er også vigtigt at bemærke følgende konsekvens af Sætning 3.6.1. Vi kan vælge mængderne T_i således at $p_i(x_i) > 0$ for alle $x_i \in T_i$. Lad os for overskuelighedens skyld holde os til tilfældet $n = 2$. Hvis X_1 og X_2 er uafhængige, er (3.6.1) opfyldt for alle $(x_1, x_2) \in T_1 \times T_2$. Dette medfører på grund af den måde, hvorpå vi lige har valgt mængderne T_1 og T_2 , at $p(x_1, x_2) > 0$ for alle $(x_1, x_2) \in T_1 \times T_2$. Vi ser altså, at hvis X_1 og X_2 er uafhængige, er den stokastiske vektor (X_1, X_2) nødvendigvis koncentreret på en produktmængde. Det er derfor somme tider let at konstatere, hvis to stokastiske variable på endelige mængder *ikke* er uafhængige: *Hvis mængden*

$$\{(x_1, x_2) \mid p(x_1, x_2) > 0\}$$

ikke er en produktmængde, kan X_1 og X_2 ikke være uafhængige. Her betegner $p(x_1, x_2)$ som tidligere den simultane sandsynlighedsfunktion for (X_1, X_2) . Prøv igen at kaste et blik på Eksempel 1.2.1 i Afsnit 2.4.

Eksempel 3.6.2 Antag at den stokastiske vektor (S_1, \dots, S_k) er polynomialfordelt med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_k . Da polynomialfordelingen er koncentreret på $D_k(n) = \{(x_1, \dots, x_k) \mid x_i \in \{0, \dots, n\}, i = 1, \dots, k, x_1 + \dots + x_k = n\}$, som tydeligvis ikke er nogen produktmængde, og da $p(x_1, \dots, x_k) > 0$ for $(x_1, \dots, x_k) \in D_k(n)$, kan de stokastiske variable S_1, \dots, S_k ikke være uafhængige. Da $S_1 + \dots + S_k = n$ er dette ikke nogen overraskelse. □

Man har meget ofte brug for at finde fordelingen af summen af to eller flere uafhængige stokastiske variable. Derfor er der grund til at give følgende specialtilfælde af Sætning 3.4.4 for uafhængige stokastiske variable.

Korollar 3.6.3 *Lad X_1 og X_2 være to uafhængige stokastiske variable, hvor X_i s fordeling er koncentreret på $\{0, 1, \dots, n_i\}$ ($n_i \in \mathbb{N}$) og har*

sandsynlighedsfunktion p_i , $i = 1, 2$. Da er den stokastiske variable $Y = X_1 + X_2$ koncentreret på mængden $\{0, 1, \dots, n_1 + n_2\}$, og dens sandsynlighedsfunktion er givet ved

$$q(y) = \sum_{j=0}^y p_1(j)p_2(y-j) \quad (3.6.3)$$

når $y \in \{0, 1, \dots, n_1 + n_2\}$.

Sandsynlighedsfunktionen q givet ved (3.6.3) kaldes *foldningen* af de to sandsynlighedsfunktioner p_1 og p_2 . Man kalder derfor ofte fordelingen af $X_1 + X_2$ foldningen af de to marginale fordelinger.

Lad os slutte dette afsnit med det i mange statistiske analyser vigtige begreb *betinget uafhængighed af to stokastiske variable givet en tredje*.

Definition 3.6.4 Lad X_1 , X_2 og X_3 være tre stokastiske variable, som er koncentreret på de endelige mængder T_1, T_2 og T_3 . Da siges X_1 og X_2 at være betinget uafhængige givet X_3 , hvis der for alle $(x_1, x_2, x_3) \in T_1 \times T_2 \times T_3$, hvor $P(X_3 = x_3) > 0$ gælder, at

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2 | X_3 = x_3) = \\ P(X_1 = x_1 | X_3 = x_3)P(X_2 = x_2 | X_3 = x_3). \end{aligned} \quad (3.6.4)$$

Eksempel 1.4.5 (fortsat). Sidst vi betragtede de to kasser på et bord, hvoraf den ene indeholder 50 røde og 50 hvide kugler, mens den anden indeholder 20 røde, 80 hvide og 50 sorte kugler, trak vi to kugler. Vi betragter igen helt samme situation og giver farverne rød, hvid og sort numrene 1, 2 og 3. Lad X_1 være den stokastiske variable, som angiver nummeret på farven af den først trukne kugle, og lad X_2 angive nummeret på farven af den anden kugle. Lad endelig X_3 angive kassenummeret. Da er X_1 og X_2 betinget uafhængige givet X_3 , mens vi har set, at X_1 og X_2 ikke er uafhængige. □

Man kan tilsvarende lave et eksempel på betinget uafhængige stokastiske variable ved at definere passende stokastiske variable i Eksempel 1.5.8.

Sætning 3.6.5 Lad X_1, X_2 og X_3 være stokastiske variable, hvor X_i er koncentreret på den endelige mængde T_i , $i = 1, 2, 3$. Betegn sandsynlighedsfunktionen for den simultane fordeling af (X_1, X_2, X_3) med p , den simultane sandsynlighedsfunktion for (X_i, X_j) med p_{ij} , og sandsynlighedsfunktionen for X_3 med p_3 . Da er følgende tre udsagn ækvivalente:

1) X_1 og X_2 er betinget uafhængige givet X_3 ,

2) For alle $(x_1, x_2, x_3) \in T$, hvor $T = T_1 \times T_2 \times T_3$, gælder, at

$$p(x_1, x_2, x_3)p_3(x_3) = p_{13}(x_1, x_3)p_{23}(x_2, x_3), \quad (3.6.5)$$

3) Der findes to ikke-negative reelle funktioner g og h , så

$$p(x_1, x_2, x_3) = g(x_1, x_3)h(x_2, x_3) \quad (3.6.6)$$

for alle $(x_1, x_2, x_3) \in T$.

Bevis: At 1) og 2) er ækvivalente, følger umiddelbart af Definition 3.6.4 og definitionen af betinget sandsynlighed. Man skal lige huske, at når $p_3(x_3) = 0$, er begge sider af (3.6.5) altid lig nul, så der er ikke noget at vise i den situation. At 2) medfører 3) er heller ikke svært at se: Hvis $p_3(x_3) > 0$, skal man bare dividere (3.6.5) igennem med $p_3(x_3)$, og f.eks. definere $g(x_1, x_3) = p_{13}(x_1, x_3)$ og $h(x_2, x_3) = p_{23}(x_2, x_3)/p_3(x_3)$. Når $p_3(x_3) = 0$, er også $p(x_1, x_2, x_3) = 0$, så i den situation kan man sætte $g(x_1, x_3) = 0$ og dernæst definere $h(x_2, x_3)$ ganske som man vil.

Endelig skal vi vise, at 3) medfører 2). En version af Sætning 3.4.1 for tre-dimensionale diskrete stokastiske vektorer giver, at (3.6.6) medfører, at

$$p_3(x_3) = \sum_{x_1 \in T_1} \sum_{x_2 \in T_2} p(x_1, x_2, x_3) = \left(\sum_{x_1 \in T_1} g(x_1, x_3) \right) \left(\sum_{x_2 \in T_2} h(x_2, x_3) \right),$$

og

$$\begin{aligned} p_{13}(x_1, x_3) &= \sum_{x_2 \in T_2} p(x_1, x_2, x_3) = g(x_1, x_3) \sum_{x_2 \in T_2} h(x_2, x_3), \\ p_{23}(x_2, x_3) &= \sum_{x_1 \in T_1} p(x_1, x_2, x_3) = h(x_2, x_3) \sum_{x_1 \in T_1} g(x_1, x_3), \end{aligned}$$

hvilket sammenholdt med (3.6.6) viser, at (3.6.5) holder. \square

3.7 Middelværdi og varians

I dette afsnit skal vi definere to størrelser, som giver en meget summarisk beskrivelse af en given sandsynlighedsfordeling. Her betragter vi kun fordelinger, der er koncentreret på en endelig delmængde af \mathbb{R} . Størrelserne kan også defineres for mere generelle stokastiske variable, men det vil vi vende tilbage til senere. Den første størrelse, middelværdien, angiver, hvor på den reelle akse hovedparten af sandsynlighedsmassen ligger. Den anden, variansen, er et mål for, hvor meget sandsynlighedsmassen er spredt ud.

Lad X_1, \dots, X_N være uafhængige stokastiske variable, som alle er koncentreret på den endelige mængde $\{a_1, \dots, a_k\}$, ($a_i \in \mathbb{R}$) og som alle har samme fordeling med sandsynlighedsfunktion p . Hvis N er stor, er ifølge sandsynlighedsregningens frekvensfortolkning (som vi skal bevise i Kapitel 7),

$$\frac{\#\{i | X_i = a_j\}}{N} \simeq P(X_1 = a_j) = p(a_j).$$

Dermed er

$$\begin{aligned} & \frac{X_1 + \dots + X_N}{N} \\ &= \frac{a_1 \cdot \#\{i | X_i = a_1\} + a_2 \cdot \#\{i | X_i = a_2\} + \dots + a_k \cdot \#\{i | X_i = a_k\}}{N} \\ &\simeq a_1 \cdot p(a_1) + a_2 \cdot p(a_2) + \dots + a_k \cdot p(a_k). \end{aligned}$$

Når N er stor, er det gennemsnitlige udfald altså nær ved den sidste størrelse. Denne overvejelse motiverer følgende definition.

Definition 3.7.1 *Lad X være en stokastisk variabel, der er koncentreret på den endelige mængde $T = \{a_i : i = 1, \dots, k\} \subseteq \mathbb{R}$, og som har sandsynlighedsfunktion p . Da definerer vi middelværdien af X som*

$$E(X) = \sum_{i=1}^k a_i p(a_i). \quad (3.7.1)$$

Af og til udelader man parenteserne og skriver EX i stedet for $E(X)$, når dette kan ske uden fare for misforståelser. Betegnelsen “E” kan

henføres til det engelske ord *expectation* eller til det tyske ord *Erwartungswert*. Bemærk, at middelværdi i disse noter skrives ikke-kursiveret E for at undgå forveksling med udfaldsrummet E .

Læg mærke til, at middelværdien af en stokastisk variabel X kun afhænger af fordelingen af X . Den er altså egentlig en egenskab ved fordelingen, og vi omtaler også $E(X)$ som fordelingsens middelværdi. Vi skal senere i dette afsnit finde middelværdierne af de fordelinger, som vi tidligere i dette kapitel har indført.

Eksempel 3.7.2 Lad X være en stokastisk variabel, som er koncentreret på mængden $\{0, 1\}$, hvis fordeling er givet ved $P(X = 1) = p$ og $P(X = 0) = 1 - p$. Da er

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p. \quad (3.7.2)$$

Bemærk, at hvis Y er en vilkårlig stokastisk variabel, og hvis 1_A er indikatorfunktionen for en vilkårlig delmængde af \mathbb{R} , er den stokastiske variable $1_A(Y)$ koncentreret på $\{0, 1\}$, og $P(1_A(Y) = 1) = P(Y \in A)$. Derfor er

$$E(1_A(Y)) = P(Y \in A).$$

Man kan således udtrykke sandsynligheden for en hændelse ved en middelværdi. Opgave 3.17 giver et eksempel på en anvendelse af denne observation.

Man kan opfatte fordelingen af X som en binomialfordeling med antalsparameter 1 og sandsynlighedsparameter p . Hvis S er binomialfordelt med antalsparameter 2 og sandsynlighedsparameter p , er

$$E(S) = 0 \cdot (1 - p)^2 + 1 \cdot 2p(1 - p) + 2 \cdot p^2 = 2p.$$

Man kan godt beregne middelværdien af en vilkårlig binomialfordeling ved at indsætte i (3.7.1), men vi kan om nogle få sider finde denne middelværdi på en langt lettere måde. Vi venter derfor lidt med beregningen. \square

Vi skal nu se, at man kan beregne middelværdien af en transformeret stokastisk variabel på en enkel og naturlig måde.

Sætning 3.7.3 *Lad X være en n -dimensional stokastisk vektor, som er koncentreret på den endelige mængde $T \subseteq \mathbb{R}^n$, og som har sandsynlighedsfunktion p . Lad endvidere ψ være en funktion fra T ind i \mathbb{R} . Da har den stokastiske variable $\psi(X)$ middelværdien*

$$E(\psi(X)) = \sum_{(x_1, \dots, x_n) \in T} \psi(x_1, \dots, x_n) p(x_1, \dots, x_n). \quad (3.7.3)$$

Bevis: Ifølge Sætning 3.4.2 har fordelingen af $Y = \psi(X)$ sandsynlighedsfunktionen

$$q(y) = \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n)$$

og er koncentreret på den endelige mængde $\psi(T)$. Ifølge Definition 3.7.1 er middelværdien af den stokastiske variable $Y = \psi(X)$

$$\begin{aligned} E(Y) &= \sum_{y \in \psi(T)} y q(y) \\ &= \sum_{y \in \psi(T)} y \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) \\ &= \sum_{y \in \psi(T)} \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} \psi(x_1, \dots, x_n) p(x_1, \dots, x_n) \\ &= \sum_{(x_1, \dots, x_n) \in T} \psi(x_1, \dots, x_n) p(x_1, \dots, x_n). \end{aligned}$$

□

Eksempel 3.7.4 Lad (X_1, X_2) være den stokastiske vektor, som angiver udfaldet af et kast med to terninger. Definer en stokastisk variabel ved $Z = X_1 \vee X_2$, d.v.s. antallet af øjne på den terning, som viser det største antal øjne. Vi har nu to måder at beregne denne middelværdi på. Den ene er at benytte definitionen (3.7.1). For at gøre det, skal vi først finde sandsynlighedsfunktionen for Z . Det er ikke vanskeligt at se, at den er givet ved følgende tabel.

z	1	2	3	4	5	6
$p(z)$	1/36	1/12	5/36	7/36	1/4	11/36

Ifølge (3.7.1) er

$$E(Z) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{1}{4} + 6 \cdot \frac{11}{36} = 4.472.$$

Hvis vi ikke ønsker at bekymre os om sandsynlighedsfunktionen for Z , kan vi benytte Sætning 3.7.3 og beregne $E(Z)$ ved hjælp af sandsynlighedsfunktionen for den oprindelige stokastiske vektor (X_1, X_2) . Ifølge (3.7.3) er

$$E(Z) = \sum_{i=1}^6 \sum_{j=1}^6 (i \vee j) \frac{1}{36} = 4.472.$$

□

Følgende sætninger er lette anvendelser af Sætning 3.7.3.

Sætning 3.7.5 *Lad X være en stokastisk variabel på en endelig mængde, og lad a og b være vilkårlige reelle tal. Da er*

$$E(a + bX) = a + bE(X) \quad (3.7.4)$$

Bevis: Betegn den mængde, som X er koncentreret på med $\{x_1, \dots, x_k\}$, og lad p være sandsynlighedsfunktionen for X . Ifølge (3.7.3) kan middelværdien af $a + bX$ beregnes ved

$$\sum_{i=1}^k (a + bx_i)p(x_i) = a \sum_{i=1}^k p(x_i) + b \sum_{i=1}^k x_i p(x_i) = a + bE(X).$$

□

Sætning 3.7.6 *Lad (X_1, X_2) være en to-dimensional stokastisk vektor på en endelig mængde, som opfylder, at $X_1 \leq X_2$. Da er $E(X_1) \leq E(X_2)$. Specielt gælder for en stokastisk variabel X , at*

$$|E(X)| \leq E(|X|). \quad (3.7.5)$$

Bevis: Lad T være den endelige mængde, på hvilken (X_1, X_2) er koncentreret, og lad p betegne sandsynlighedsfunktionen for (X_1, X_2) . Udsagnet at $X_1 \leq X_2$ betyder, at $x_1 \leq x_2$ for alle $(x_1, x_2) \in T$. Derfor er

$$E(X_1) = \sum_{(x_1, x_2) \in T} x_1 p(x_1, x_2) \leq \sum_{(x_1, x_2) \in T} x_2 p(x_1, x_2) = E(X_2),$$

hvor vi endnu en gang har benyttet Sætning 3.7.3. Sætningens anden påstand følger af det netop viste resultat, da $-|X| \leq X \leq |X|$, således at $-\mathbb{E}(|X|) \leq \mathbb{E}(X) \leq \mathbb{E}(|X|)$. Her benytter vi (3.7.4) til at slutte, at $\mathbb{E}(-|X|) = -\mathbb{E}(|X|)$.

□

Sætning 3.7.7 *Lad X_1, X_2, \dots, X_n være stokastiske variable på endelige mængder. Da er*

$$\mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n). \quad (3.7.6)$$

Hvis det yderligere antages, at de stokastiske variable X_1, X_2, \dots, X_n er uafhængige, er

$$\mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_n) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \dots \cdot \mathbb{E}(X_n). \quad (3.7.7)$$

Bevis: Vi kan nøjes med at bevise sætningen for $n = 2$, da det generelle resultat følger ved induktion. Sæt $T = T_1 \times T_2$, hvor T_i er den endelige mængde, på hvilken X_i er koncentreret, og betegn sandsynlighedsfunktionen for den stokastiske vektor (X_1, X_2) med p . Ifølge Sætning 3.7.3 er

$$\mathbb{E}(X_i) = \sum_{(x_1, x_2) \in T} x_i p(x_1, x_2), \quad i = 1, 2,$$

og

$$\begin{aligned} \mathbb{E}(X_1 + X_2) &= \sum_{(x_1, x_2) \in T} (x_1 + x_2) p(x_1, x_2) \\ &= \sum_{(x_1, x_2) \in T} x_1 p(x_1, x_2) + \sum_{(x_1, x_2) \in T} x_2 p(x_1, x_2) \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2). \end{aligned}$$

For at vise sætningens sidste påstand antages det nu, at X_1 og X_2 er uafhængige, så $p(x_1, x_2) = p_1(x_1)p_2(x_2)$, hvor p_i er sandsynlighedsfunktionen for X_i , $i = 1, 2$. Ifølge (3.7.3) er

$$\mathbb{E}(X_1 X_2) = \sum_{(x_1, x_2) \in T} x_1 x_2 p(x_1, x_2)$$

$$\begin{aligned}
&= \sum_{x_2 \in T_2} \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) x_2 p_2(x_2) \right) = \sum_{x_2 \in T_2} x_2 p_2(x_2) \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) \right) \\
&= \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) \right) \left(\sum_{x_2 \in T_2} x_2 p_2(x_2) \right) = E(X_1)E(X_2).
\end{aligned}$$

□

Eksempel 3.7.8 Vi vil nu finde *binomialfordelingens middelværdi*. Husk, at binomialfordelingen med antalsparameter n og sandsynlighedsparameter p er defineret som fordelingen af summen $S = X_1 + X_2 + \dots + X_n$, hvor X_1, X_2, \dots, X_n er uafhængige stokastiske variable med værdier i $\{0, 1\}$, som alle har samme fordeling givet ved $P(X_i = 1) = p$ og $P(X_i = 0) = 1 - p$. Ifølge Eksempel 3.7.2 er $E(X_i) = p$, så ifølge (3.7.6) er

$$E(S) = E(X_1) + \dots + E(X_n) = np.$$

Når man tænker på binomialfordelingens definition, er udtrykket for middelværdien ikke overraskende. Det forventede antal “succes’er” er naturligvis antallet af forsøg ganget med sandsynligheden for “succes” i det enkelte forsøg, jvf. sandsynlighedsregningens frekvensfortolkning.

□

Eksempel 3.7.9 Middelværdien i den *hypergeometriske fordeling* kan beregnes ved følgende argument. Lad kuglerne være nummereret $1, \dots, N$, hvor de røde kugler har numrene $1, \dots, R$. Lad A_i betegne den hændelse, at kugle nr. i er blandt de n udtrukne, og definer N stokastiske variable 1_{A_i} , $i = 1, \dots, N$ ved at $1_{A_i} = 1$, hvis kugle nr. i er blandt de udtrukne, og $1_{A_i} = 0$, hvis dette ikke er tilfældet. Ifølge Eksempel 3.7.2 er $E(1_{A_i}) = P(A_i)$. Da $1_{A_1} + \dots + 1_{A_N} = n$, og da alle kugler har samme sandsynlighed for at blive udtrukket, følger det af (3.7.4) og (3.7.6) at

$$n = E(1_{A_1} + \dots + 1_{A_N}) = E(1_{A_1}) + \dots + E(1_{A_N}) = NE(1_{A_1}),$$

dvs.

$$E(1_{A_1}) = \frac{n}{N}.$$

Hvis X er antallet af røde kugler i stikprøven, er

$$E(X) = E(1_{A_1} + \dots + 1_{A_R}) = E(1_{A_1}) + \dots + E(1_{A_R}) = R \frac{n}{N} = n \frac{R}{N}.$$

□

Vi kommer nu til *variansen*, som er et mål for, hvor meget sandsynlighedsmassen spreder sig omkring middelværdien.

Definition 3.7.10 *Lad X være en stokastisk variabel på en endelig mængde. Da defineres variansen af X som*

$$\text{Var}(X) = \text{E}([X - \text{E}(X)]^2). \quad (3.7.8)$$

Hvis X er koncentreret på mængden T og har sandsynlighedsfunktion p , kan variansen ifølge Sætning 3.7.3 beregnes ved formlen

$$\text{Var}(X) = \sum_{x \in T} (x - \text{E}(X))^2 p(x).$$

Da

$$[X - \text{E}(X)]^2 = X^2 + \text{E}(X)^2 - 2X\text{E}(X),$$

følger den nyttige formel for variansen af X

$$\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2, \quad (3.7.9)$$

som oftest er den letteste måde at beregne variansen på.

Kvadratrodten af variansen kaldes *spredningen* eller *standardafvigelsen*. Spredningen er et mere naturligt mål for en fordelings bredde end variansen. Dette ses tydeligt af følgende resultat. Lad X være en stokastisk variabel, og definer en ny stokastisk variabel ved $a + bX$, hvor a og b er reelle tal. Da er

$$\text{Var}(a + bX) = b^2 \text{Var}(X) \quad (3.7.10)$$

og dermed

$$\sqrt{\text{Var}(a + bX)} = |b| \sqrt{\text{Var}(X)}.$$

Ligning (3.7.10) indses let:

$$\begin{aligned} \text{Var}(a + bX) &= \text{E}([a + bX - \text{E}(a + bX)]^2) \\ &= \text{E}(b^2[X - \text{E}(X)]^2) = b^2 \text{Var}(X). \end{aligned}$$

Eksempel 3.7.11 Lad X være en stokastisk variabel, hvis fordeling er givet ved $P(X = 1) = p$ og $P(X = 0) = 1 - p$. Da er $X^2 = X$, så ifølge (3.7.2) er

$$E(X^2) = E(X) = p,$$

således at (3.7.9) giver

$$\text{Var}(X) = p - p^2 = p(1 - p). \quad (3.7.11)$$

Hvis S er binomialfordelt med antalsparameter 2 og sandsynlighedsparameter p , er

$$E(S^2) = 0 \cdot (1 - p)^2 + 1 \cdot 2p(1 - p) + 4 \cdot p^2 = 2p + 2p^2,$$

så

$$\text{Var}(S) = 2p + 2p^2 - 4p^2 = 2p(1 - p).$$

Man kan godt beregne variansen af en vilkårlig binomialfordeling ved at beregne middelværdien af $E(X(X - 1))$, men da vi i næste afsnit udleder en betydeligt lettere måde at beregne denne varians på, udskyder vi dette problem lidt.

Definer endelig for $a > 0$ en stokastisk variabel ved $Y = aX$. Det er klart, at $P(Y = a) = P(X = 1) = p$ og $P(Y = 0) = P(X = 0) = 1 - p$, således at Y 's tilfældige variation om sin middelværdi $E(Y) = aE(X) = ap$ vokser med a . Vi finder da også, at

$$\text{Var}(Y) = a^2 \text{Var}(X) = a^2 p(1 - p),$$

og at spredningen er $a\sqrt{p(1 - p)}$.

□

Lad os kort behandle den ekstreme situation, hvor variansen er nul. Det følgende resultat vil næppe overraske nogen.

Sætning 3.7.12 *For en stokastisk variabel X på en endelig mængde er $\text{Var}(X) = 0$ hvis og kun hvis der findes et reelt tal c , så $P(X = c) = 1$.*

Bevis: Lad p betegne sandsynlighedsfunktionen for X . Hvis $P(X = c) = 1$, er

$$p(x) = \begin{cases} 1 & \text{hvis } x = c \\ 0 & \text{hvis } x \neq c, \end{cases}$$

og da $E(X) = c$, er

$$\text{Var}(X) = (E(X) - E(X))^2 p(E(X)) = 0.$$

Antag omvendt, at $\text{Var}(X) = 0$, og lad T være den endelige mængde, hvorpå X er koncentreret. Da

$$\sum_{x \in T} (x - E(X))^2 p(x) = 0$$

medfører, at alle led i summen er lig nul, er $p(x) = 0$, når $x \neq E(X)$. Dette viser sætningen. □

Definition 3.7.13 *En stokastisk variabel X siges at være udartet, hvis der findes et $c \in \mathbb{R}$, så $P(X = c) = 1$. Fordelingen af X kaldes da den udartede fordeling i punktet c .*

Vi kan altså kort sige, at en fordeling har varians nul hvis og kun hvis den er udartet i et punkt c , som altså er dens middelværdi.

3.8 Kovarians og korrelation

Vi kommer nu til en størrelse, korrelationen, som giver en summarisk beskrivelse af samspillet mellem to stokastiske variable, altså af aspekter ved den simultane fordeling af (X, Y) . Først skal vi definere kovariansen, som spiller en lignende rolle, men er sværere at fortolke. Hvis X og Y har tendens til at følges ad, er kovariansen og korrelationen positive tal. Hvis derimod X og Y opfører sig modsat i den forstand, at X og $-Y$ har tendens til at følges ad, er kovariansen og korrelationen negative. Hvis der ikke er nogen sammenhæng mellem de to stokastiske variables variation er de to størrelser begge lig nul. Eksempel 3.8.2 nedenfor giver en simpel illustration af disse forhold.

Definition 3.8.1 *For to stokastiske variable på endelige mængder defineres kovariansen mellem X og Y ved*

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]. \quad (3.8.1)$$

Ved at regne på højresiden af (3.8.1), er det er ikke vanskeligt at indse, at vi har følgende formel, som ofte er nyttig til beregning af kovariansen

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \quad (3.8.2)$$

Bemærk, at

$$\text{Cov}(X, X) = \text{Var}(X).$$

Lad X , Y og Z være tre stokastiske variable, og lad a , b , c og d være reelle tal. Følgende regneregler for kovarians vises lige så let som (3.7.10) ved at regne efter.

$$\text{Cov}(a + bX, c + dY) = bd \text{Cov}(X, Y) \quad (3.8.3)$$

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z) \quad (3.8.4)$$

$$\text{Cov}(X, Y) = \text{Cov}(Y, X) \quad (3.8.5)$$

Eksempel 3.8.2 Antag at (X, Y) er koncentreret på mængden

$$\{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$$

med punktsandsynlighederne $P((X, Y) = (1, 1)) = P((X, Y) = (-1, -1)) = \frac{1}{3}$ og $P((X, Y) = (1, -1)) = P((X, Y) = (-1, 1)) = \frac{1}{6}$. Da er $E(X) = E(Y) = 0$ og $P(XY = 1) = \frac{2}{3}$ og $P(XY = -1) = \frac{1}{3}$, så $\text{Cov}(X, Y) = E(XY) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$.

Vi kan mere generelt antage, at den simultane fordeling af (X, Y) er givet ved $P((X, Y) = (1, -1)) = P((X, Y) = (-1, 1)) = \frac{1}{m}$ og $P((X, Y) = (1, 1)) = P((X, Y) = (-1, -1)) = \frac{1}{2} - \frac{1}{m}$, hvor m er et helt tal større end eller lig to. Jo større m er, jo større er sandsynligheden for, at $X = Y$. Denne sandsynlighed er jo lig med $1 - \frac{2}{m}$. Man kan altså konkludere, at jo større m er, jo større tendens har X og Y til at følges ad. Lad os nu beregne deres kovarians. Også for den mere generelle model er $E(X) = E(Y) = 0$, mens $P(XY = 1) = 1 - \frac{2}{m}$ og $P(XY = -1) = \frac{2}{m}$, så $\text{Cov}(X, Y) = E(XY) = 1 - \frac{4}{m}$. Vi ser, at kovariansen er negativ for $m = 2$ og $m = 3$, mens den er lig nul for $m = 4$ og positiv for m større end eller lig 5. Kovariansen vokser tydeligvis, når m vokser.

Definer til slut en ny stokastisk variabel ved $T = X + aY$, hvor a er et reelt tal. Lad os antage, at $m = 4$, så $\text{Cov}(X, Y) = 0$. Det er vist klart, at hvis $a > 0$, har T og Y tendens til at følges, mens de for $a < 0$ har tendens til at variere modsat. Hvis $a = 0$, er der ingen sammenhæng mellem den stokastiske variation af T og Y . I overensstemmelse med disse overvejelser finder vi ved at benytte (3.8.3) og (3.8.4), at

$$\text{Cov}(T, Y) = \text{Cov}(X, Y) + \text{Cov}(aY, Y) = a\text{Var}(Y).$$

□

Sætning 3.8.3 Hvis X og Y er uafhængige, er $\text{Cov}(X, Y) = 0$.

Bevis: Ifølge (3.7.7) er $E(XY) = E(X)E(Y)$, når X og Y er uafhængige, hvorfor sætningens konklusion kan sluttet af (3.8.2).

□

Den omvendte implikation gælder *ikke*. Selv om $\text{Cov}(X, Y) = 0$, kan de to stokastiske variable sagtens være særdeles afhængige, som følgende eksempel viser.

Eksempel 3.8.4 Antag, at (X, Y) er ligefordelt på $\{(0, 1), (1, 0), (0, -1), (-1, 0)\}$. Da $E(X) = E(Y) = 0$, og da $XY = 0$ (og dermed specielt $E(XY) = 0$), er $\text{Cov}(X, Y) = 0$. Da (X, Y) ikke lever på en produktmængde, kan de ikke være uafhængige. Man kan også bemærke, at hvis man ved at $Y = 1$, så må X være nul. Der er altså en kraftig afhængighed. Hvis man har behov for yderligere bevis, er $P(X = 1, Y = 1) = 0$, mens $P(X = 1)P(Y = 1) = \frac{1}{16}$.

□

Kovariansen spiller en vigtig rolle, når man beregner variansen af en sum af stokastiske variable. Vi ser først, at

$$\begin{aligned} \text{Var}(X_1 + X_2) & & (3.8.6) \\ &= E\left(\left(X_1 + X_2 - (E(X_1) + E(X_2))\right)^2\right) \\ &= E\left(\left(X_1 - E(X_1)\right)^2 + \left(X_2 - E(X_2)\right)^2 + 2\left(X_1 - E(X_1)\right)\left(X_2 - E(X_2)\right)\right) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2). \end{aligned}$$

Det er ikke svært at gennemføre det induktionsargument, som viser, at for n stokastiske variable X_1, \dots, X_n er

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} \text{Cov}(X_i, X_j). \quad (3.8.7)$$

Hvis man ikke bryder sig om induktionsargumenter, kan formelen også vises direkte i analogi med (3.8.6).

Eksempel 3.8.5 Vi vil i dette eksempel finde variansen af den hypergeometriske fordeling v.h.a. (3.8.7). Lad notationen være som i Eksempel 3.7.9. Specielt er A_i den hændelse, at kugle nr. i er blandt de n udtrukne, og den stokastiske variable 1_{A_i} er lig en, hvis kugle nr. i er blandt de udtrukne, og $1_{A_i} = 0$ ellers. Da er $X = 1_{A_1} + \dots + 1_{A_R}$ hypergeometrisk fordelt. Fra Eksempel 3.7.9 ved vi, at $E(1_{A_i}) = \frac{n}{N}$, og da $1_{A_i}^2 = 1_{A_i}$ er også $E(1_{A_i}^2) = \frac{n}{N}$. Det vil sige, at

$$\text{Var}(1_{A_i}) = \frac{n}{N} \left(1 - \frac{n}{N}\right).$$

For $i \neq j$ er

$$\begin{aligned} E(1_{A_i} 1_{A_j}) &= E(1_{A_i \cap A_j}) = P(A_i \cap A_j) \\ &= P(A_j)P(A_i|A_j) = \frac{n}{N} \cdot \frac{n-1}{N-1}. \end{aligned}$$

Vi har brugt at $P(A_i) = E(1_{A_i}) = \frac{n}{N}$ ifølge Eksempel 3.7.2. At $P(A_i|A_j) = (n-1)/(N-1)$ følger af, at når vi ved, at kugle nr. j er blevet udtrukket, er situationen som den var uden denne oplysning bortset fra, at der nu kun skal udtrækkes $n-1$ kugler fra en population af $N-1$ kugler. Vi finder, at for $i \neq j$ er

$$\text{Cov}(1_{A_i}, 1_{A_j}) = \frac{n}{N} \cdot \frac{n-1}{N-1} - \left(\frac{n}{N}\right)^2 = \frac{n(n-N)}{N^2(N-1)},$$

så ifølge (3.8.7) er

$$\begin{aligned} \text{Var}(X) &= R \frac{n}{N} \left(1 - \frac{n}{N}\right) + 2 \left[\frac{1}{2} R(R-1)\right] \frac{n(n-N)}{N^2(N-1)} \\ &= n \frac{R}{N} \left(1 - \frac{R}{N}\right) \frac{N-n}{N-1}. \end{aligned}$$

Vi har her brugt at, der er $\frac{1}{2}R(R-1)$ led i dobbeltsummen i (3.8.7). \square

Årsagen til, at kovariansen mellem to stokastiske variable X og Y er stor, kan være, at X eller Y varierer meget, mens der ikke er den store afhængighed mellem dem. Betragt for eksempel de stokastiske variable aX og Y , hvor a er et positivt tal. Ifølge (3.8.3) er $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$, så vi kan ved at vælge a passende gøre kovariansen mellem aX og Y vilkårligt stor eller lille, selv om sammenhængen mellem aX og Y egentlig er den samme som sammenhængen mellem X og Y . Hvis for eksempel X er en længde, ser vi, at vi kan gøre kovariansen 10 gange større ved at måle X i millimeter i stedet for centimeter. Det er ikke heldigt. Et bedre mål for samspelet mellem X og Y er derfor korrelationen mellem X og Y , som ikke afhænger af de enheder, hvormed man måler X og Y .

Definition 3.8.6 For to stokastiske variable X og Y , som opfylder at $\text{Var}(X) > 0$ og $\text{Var}(Y) > 0$, defineres korrelationen mellem X og Y ved

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \quad (3.8.8)$$

Korrelationen kan fortolkes som kovariansen mellem de to *standardiserede* variable

$$X_0 = \frac{X - \text{E}(X)}{\sqrt{\text{Var}(X)}}, \quad Y_0 = \frac{Y - \text{E}(Y)}{\sqrt{\text{Var}(Y)}},$$

idet

$$\text{Cov}(X_0, Y_0) = \text{E}(X_0 Y_0) = \frac{\text{E}((X - \text{E}(X))(Y - \text{E}(Y)))}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \text{corr}(X, Y).$$

De stokastiske variable X_0 og Y_0 er standardiserede, så de har middelværdi nul og varians en. Korrelationen er således uafhængig af valget af nulpunkt og enhed for de skalaer, hvorpå X og Y måles. Kovariansen er også uafhængig af valget af nulpunkter, men ikke af valget af enheder.

Eksempel 3.8.2 (fortsat). For vilkårligt $m (\geq 2)$ er $P(X = 1) = P(X = -1) = P(Y = 1) = P(Y = -1) = \frac{1}{2}$, således at $\text{Var}(X) = \text{Var}(Y) = E(Y^2) = \frac{1}{2} + \frac{1}{2} = 1$. Derfor er $\text{corr}(X, Y) = \text{Cov}(X, Y) = 1 - \frac{4}{m}$.

For $m = 4$ er $\text{Cov}(X, aY) = a\text{Cov}(X, Y) = 0$, hvorfor vi ifølge (3.8.6) har, at $\text{Var}(T) = \text{Var}(X) + a^2\text{Var}(Y) = 1 + a^2$. Dermed er $\text{corr}(T, Y) = \text{Cov}(T, Y)/\sqrt{1 + a^2} = a/\sqrt{1 + a^2}$.

□

Hvis to stokastiske variable X og Y opfylder, at $\text{corr}(X, Y) = 0$, siges de at være *ukorrelerede*. Da $\text{corr}(X, Y) = 0$ naturligvis er ensbetydende med $\text{Cov}(X, Y) = 0$, kan man udtrykke resultatet af Sætning 3.8.3 således: Hvis X og Y er uafhængige, er de også ukorrelerede. Eksempel 3.8.4 viste, at X og Y godt kan være ukorrelerede uden at være uafhængige. Man skal sædvanligvis være varsom med at fortolke korrelationen mellem to stokastiske variable. Vi skal senere se, at når den simultane fordeling af (X, Y) er en såkaldt normalfordeling, har man mere fast grund under fødderne.

Andre vigtige egenskaber ved korrelationen fremgår af følgende sætning.

Sætning 3.8.7

(a) $-1 \leq \text{corr}(X, Y) \leq 1$.

(b) $\text{corr}(X, Y) = \pm 1$ hvis og kun hvis der for passende $\alpha, \beta \in \mathbb{R}$ gælder, at $Y = \alpha + \beta X$. Værdien 1 af korrelationen svarer til $\beta > 0$, værdien -1 til $\beta < 0$.

Bevis: (a) følger af, at

$$|\text{corr}(X, Y)| = |E(X_0 Y_0)| \leq E|X_0 Y_0| \leq E\left(\frac{1}{2}X_0^2 + \frac{1}{2}Y_0^2\right) = 1,$$

hvor X_0 og Y_0 er de ovenfor indførte standardiserede variable, og hvor vi har benyttet (3.7.5). Vi har også brugt, at der for vilkårlige reelle tal x og y gælder, at $-\frac{1}{2}(x^2 + y^2) \leq xy \leq \frac{1}{2}(x^2 + y^2)$. Disse to uligheder følger let af, at $(x + y)^2 \geq 0$ og $(x - y)^2 \geq 0$.

Antag, at $\text{corr}(X, Y) = 1$. Da er $\text{Cov}(X_0, Y_0) = 1$, så variansen af differensen $X_0 - Y_0$ er

$$\begin{aligned} \text{Var}(X_0 - Y_0) &= E\left((X_0 - Y_0)^2\right) \\ &= E(X_0^2) + E(Y_0^2) - 2E(X_0 Y_0) = 1 + 1 - 2 = 0. \end{aligned}$$

Ifølge Sætning 3.7.12 findes der en konstant c , så $X_0 - Y_0 = c$. Da middelværdien af $X_0 - Y_0$ er lig nul, må $c = 0$, dvs. $X_0 = Y_0$. Dette er ækvivalent med at der eksisterer $\alpha \in \mathbb{R}$ og $\beta > 0$, så $Y = \alpha + \beta X$. Tilsvarende ses, at hvis $\text{corr}(X, Y) = -1$, vil $X_0 + Y_0$ have varians 0, hvorefter følger at $X_0 = -Y_0$, hvilket er ækvivalent med eksistensen af en tilsvarende lineær relation mellem X og Y med $\beta < 0$.

□

Se igen på de korrelationer, som vi ovenfor beregnede i Eksempel 3.8.2, og overvej dem i lyset af Sætning 3.8.7. Betragt blandt andet tilfældet $m = 2$.

Sætning 3.8.8 *Lad X_1, \dots, X_n være parvis ukorrelerede reelle stokastiske variable. Da er*

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n). \quad (3.8.9)$$

Bevis: At sætningen følger umiddelbart af (3.8.7), da $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$.

□

Additionsreglen (3.8.9) gælder specielt, hvis X_1, \dots, X_n er *uafhængige*. Denne regneregul er en af grundene til, at varians og spredning må anses for mere naturlige mål for en stokastisk variabels variation end så mange andre, der er blevet foreslået; f. eks. $E(|Y - EY|)$.

Eksempel 3.8.9 Ved at benytte (3.8.9) kan vi let beregne binomialfordelingens varians. Binomialfordelingen med antalsparameter n og sandsynlighedsparameter p er fordelingen af summen $S = X_1 + X_2 + \dots + X_n$, hvor X_1, X_2, \dots, X_n er uafhængige stokastiske variable med værdier i $\{0, 1\}$, som alle har samme fordeling givet ved $P(X_i = 1) = p$ og $P(X_i = 0) = 1 - p$. Ifølge Eksempel 3.7.11 er $\text{Var}(X_i) = p(1 - p)$, så ifølge (3.8.9) er

$$\text{Var}(S) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = np(1 - p). \quad (3.8.10)$$

Binomialfordelingen med antalsparameter 100 og sandsynlighedsparameter 0.5, som giver positiv sandsynlighed til alle hele tal mellem nul og 100, har middelværdi $100 \times 0.5 = 50$, og altså varians $100 \times 0.5 \times 0.5 = 25$ og spredning $\sqrt{25} = 5$.

Prøv at sammenligne variansen (3.8.10) med resultatet i Eksempel 3.8.5, og overvej om forskellen kan forstås intuitivt.

□

Eksempel 3.8.10 Lad Y_1, Y_2, \dots være en følge af uafhængige stokastiske variable med samme fordeling (af vilkårlig type). Lad A være en delmængde af \mathbb{R} , så $1 > P(Y_1 \in A) > 0$. Da er den stokastiske variable

$$X_n = 1_A(Y_1) + \dots + 1_A(Y_n)$$

binomialfordelt med antalsparameter n og sandsynlighedsparameter $P(Y_1 \in A)$. Vi kan tænke på Y_i -erne som udfaldene af en række uafhængige forsøg. Den relative frekvens af forsøg med udfald i mængden A blandt de første n forsøg er X_n/n . Denne har middelværdi $P(Y_1 \in A)$ og varians $P(Y_1 \in A)(1 - P(Y_1 \in A))/n$. Variansen aftager altså, når antallet af forsøg, n , vokser, så vi ser, at fordelingen af den relative frekvens bliver mere og mere koncentreret omkring $P(Y_1 \in A)$, når n går mod uendelig, hvilket underbygger sandsynlighedsregningens frekvensfortolkning. Vi skal i Kapitel 7 give et mere præcist udsagn.

□

Eksempel 3.8.11 Lad (S_1, \dots, S_k) være polynomialfordelt med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_k . Da er $E(S_i) = np_i$, $\text{Var}(S_i) = np_i(1 - p_i)$,

$$\text{Cov}(S_i, S_j) = -np_i p_j, \quad \text{når } i \neq j,$$

og

$$\text{corr}(S_i, S_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}},$$

se opgave 3.25.

□

Eksempel 3.8.12 Lad $(x_1, y_1), \dots, (x_n, y_n)$ være givne talpar. Vi kan tænke på n par af samhørende målinger, f. eks. højde og vægt for n personer. Ved den *empiriske fordeling* af disse observationer forstås fordelingen på mængden bestående af de n givne talpar, som fremkommer ved at

tildele hvert observationspar sandsynligheden $\frac{1}{n}$. Det skal forstås således, at hvis to eller flere af disse talpar er sammenfaldende, så får det tilsvarende punkt i \mathbb{R}^2 sandsynlighed lig antal sammenfaldende observationer divideret med n . En stokastisk variabel (X, Y) med denne fordeling kan opfattes som en observation, der er tilfældigt udvalgt blandt de n givne. For en sådan stokastisk variabel gælder, at

$$\begin{aligned} E(X) = \bar{x} &= \frac{1}{n}(x_1 + \dots + x_n), \\ E(Y) = \bar{y} &= \frac{1}{n}(y_1 + \dots + y_n), \\ \text{Var}(X) &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \\ \text{Var}(Y) &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2, \\ \text{Cov}(X, Y) &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \\ \text{corr}(X, Y) &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \end{aligned}$$

Her er \bar{x} (den empiriske middelværdi) og $\frac{1}{n} \sum (x_i - \bar{x})^2$ (den empiriske varians) velkendte størrelser fra deskriptiv statistik, knyttet til den endimensionale empiriske fordeling af x 'erne (bortset fra, at man sædvanligvis dividerer med $n - 1$ i stedet for n ved udregning af den empiriske varians, se opgave 3.26). Størrelsen $\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$ er kendt som den *empiriske kovarians* (idet man dog også her plejer at normere med $n - 1$, ikke n). Den sidste størrelse $\text{corr}(X, Y)$ kaldes naturligvis den *empiriske korrelation*. Fortolkningen af denne er, at hvis der på nær små afvigelser gælder $y_i \approx \alpha + \beta x_i$, $\beta > 0$, svarende til at de to-dimensionale observationer grupperer sig omkring en linie med positiv hældning, så vil den empiriske korrelation være nær ved 1. Tilsvarende vil en empirisk korrelation være nær -1 , når observationerne grupperer sig omkring en linie med negativ hældning. Hvis den empiriske korrelation er nær 0, er der ingen afhængighed af denne simple lineære art mellem x 'er og y 'er. Præcis hvor nær 0, den empiriske korrelation skal være, for at dette kan konkluderes, siger ovenstående ikke noget om. En tommelfingerregel,

som ikke kan begrundes her, går ud på, at hvis den empiriske korrelation ligger inden for grænserne $\pm 2/\sqrt{n}$, så er der ikke tegn på en sådan sammenhæng.

□

3.9 Sammenfatning

I Kapitel 3 har vi studeret sandsynlighedsmål på endelige mængder og i den forbindelse indført en række centrale begreber og vist nogle vigtige resultater. Vi har indført sandsynlighedsfunktionen, som er en naturlig måde at specificere et sandsynlighedsmål på en endelig mængde. Vi har set, hvordan vi i situationer, hvor vi kender sandsynlighedsfunktionen for en stokastisk variabel eller vektor X , kan finde sandsynlighedsfunktionen for den stokastiske variable, der fremkommer ved at anvende en funktion på X . Et vigtigt specialtilfælde er, at vi kan finde den marginale fordeling af en af koordinaterne i en stokastisk vektor ud fra den simultane sandsynlighedsfunktion. Et lige så vigtigt specialtilfælde er, at vi kan finde sandsynlighedsfunktionen for en sum af to stokastiske variable. Videre har vi lært, hvordan man på en simultan sandsynlighedsfunktion for en stokastisk vektor ser, om koordinaterne er uafhængige stokastiske variable.

Vi har også indført nogle begreber, som giver en mere summarisk beskrivelse af den stokastiske variation af en stokastisk variabel eller vektor. Middelværdien er et tal, som angiver en typisk værdi af en stokastisk variabel. Den angiver altså en slags centrum for fordelingen. Variansen er et tal, der angiver, hvor meget en stokastisk variabel varierer. Jo større variansen er, jo mere er fordelingsmassen spredt ud. Hvis variansen er nul, ligger hele sandsynlighedsmassen således i det punkt, der er angivet ved middelværdien. Endvidere er kovariansen og korrelationen tal, der for to stokastiske variable indikerer, om der er en sammenhæng mellem deres tilfældige variation. Korrelationen er lettest at fortolke, da den ikke afhænger af de anvendte måleenheder, mens kovariansen spiller en vigtig rolle ved beregningen af variansen af en sum af stokastiske variable.

Endelig har vi indført nogle ofte anvendte fordelinger på endelige mængder, nemlig binomialfordelingen, polynomialfordelingen og den hypergeometriske fordeling.

3.10 Opgaver

3.1 Lad N være et positivt helt tal og definer funktionen

$$p(i) = c2^i \quad \text{for } i \in \{1, 2, \dots, N\}.$$

Hvad skal værdien af c være, for at p er en sandsynlighedsfunktion på mængden $\{1, 2, \dots, N\}$? Vink: Brug formel (C.0.1) på side 293.

3.2 Et panel af colasmagere på fem personer skal forsøge at skelne mellem to colamærker, som vi kan betegne med C og P . De 5 colasmagere får hver serveret et glas af den samme slags cola, idet man ved kast med en ærlig mønt har valgt cola P eller cola C . Antag, at hver colasmager har sandsynlighed p for at bestemme colamærket korrekt. Det viser sig, at 4 ud af 5 gættede på cola P , mens 1 gættede på cola C . Hvad er den betingede sandsynlighed, givet dette resultat, for at det var cola C , der blev serveret. Vink: Bayes' formel.

3.3 Lad S være binomialfordelt med parametre n og p .

(a) Vis at

$$P(S = s + 1) = \frac{n - s}{s + 1} \cdot \frac{p}{1 - p} \cdot P(S = s).$$

(b) Vis at

$$P(S = s + 1) > P(S = s) \Leftrightarrow s < (n + 1)p - 1$$

$$P(S = s + 1) = P(S = s) \Leftrightarrow s = (n + 1)p - 1$$

$$P(S = s + 1) < P(S = s) \Leftrightarrow s > (n + 1)p - 1$$

(c) Vis, at hvis $(n + 1)p$ ikke er et heltal, så har binomialfordelingens sandsynlighedsfunktion entydigt maksimum i punktet $s = [(n + 1)p]$, dvs. dens værdi i dette punkt er større end værdien i ethvert andet punkt. Her betegner $[x]$ heltalsdelen af x , dvs. det største hele tal, som er mindre end eller lig x .

Hvad sker der hvis $(n + 1)p$ er et heltal?

- 3.4 Fem terninger kastes. Hvad er sandsynligheden for at få
- netop tre seksere?
 - mindst tre seksere?
 - tre (men ikke fire eller fem) ens?
 - tre, fire eller fem ens?
- 3.5 Bevis Sætning 3.2.6, f. eks. ved at gå tilbage til binomialfordelingens definition som fordelingen af en sum af uafhængige 0-1-variable (se Definition 3.2.1).

- 3.6 Lad X_1 og X_2 være uafhængige stokastiske variable, som er binomialfordelte med parametre henholdsvis (n_1, p) og (n_2, p) . Sæt $Y = (X_1 + X_2)/(n_1 + n_2)$.
- Find mængden F af mulige værdier af Y .
 - Find $q(y) = P(Y = y)$ for $y \in F$.

- 3.7 Lad $p_{N,R,n} : \{0, 1, \dots, n\} \rightarrow [0, 1]$ betegne sandsynlighedsfunktionen for den hypergeometriske fordeling med parametre N, R og n . Vis relationerne

$$p_{N,R,n}(x) = p_{N,n,R}(x),$$

$$p_{N,R,n}(x) = p_{N,R,N-n}(R-x),$$

$$p_{N,R,n}(x) = p_{N,N-R,n}(n-x).$$

Disse relationer følger af formler, udledt i Afsnit 3.3, men de har også et intuitivt indhold, som bør overvejes.

- 3.8 13 kort udtages tilfældigt fra et almindeligt spil bestående af 52 kort.
- Hvad er sandsynligheden for at få mindst 2 esser ?
 - Hvad er sandsynligheden for at få mindst 2 esser, givet at man ikke får en eneste konge?
 - Hvad er den betingede fordeling af antallet af esser, givet at man ikke får en eneste konge? (opskriv de 5 punktsandsynligheder).
 - Er antallet af esser stokastisk uafhængigt af antallet af konger?

3.9 Formuler det til Sætning 3.4.1 svarende resultat for højere-dimensionale fordelinger, og diskuter, hvordan det kan vises. Overvej også at resultatet er et specialtilfælde af Sætning 3.4.2.

3.10 Vis ved at kombinere resultaterne i Sætning 3.2.6 og Korollar 3.6.3 at

$$\binom{n_1 + n_2}{i} = \sum_{j=0}^i \binom{n_1}{j} \binom{n_2}{i-j},$$

hvor n_1 og n_2 er vilkårlige positive hele tal, og i er et helt tal mellem 0 og $n_1 + n_2$. Prøv at give en kombinatorisk fortolkning af den fundne formel.

3.11 En terning kastes 12 gange.

(a) Hvad er sandsynligheden for netop at få 2 1'ere, 2 2'ere, \dots , 2 6'ere?

(b) Hvad er sandsynligheden for netop at få 2 1'ere og 2 6'ere?

3.12 Lad (S_1, \dots, S_5) være polynomialfordelt af orden 5 med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_5 . De følgende spørgsmål besvares lettest ved direkte anvendelse af polynomialfordelingens definition (Definition 3.5.1).

(a) Vis at S_1 er binomialfordelt med antalsparameter n og sandsynlighedsparameter p_1 .

(b) Vis at $S_1 + S_2$ er binomialfordelt med antalsparameter n og sandsynlighedsparameter $p_1 + p_2$.

(c) Vis at $(S_1 + S_2, S_3 + S_4, S_5)$ er polynomialfordelt af orden 3 med antalsparameter n og sandsynlighedsparametre $p_1 + p_2, p_3 + p_4$ og p_5 .

(d) Vis til slut Sætning 3.5.5, som (a), (b) og (c) er specialtilfælde af.

3.13 Lad X og Y være uafhængige stokastiske variable, som begge er ligefordelte på mængden $\{0, 1, \dots, N\}$. Find $P(X \geq Y)$ og $P(X = Y)$. Find dernæst sandsynlighedsfunktionerne for de stokastiske variable $X \vee Y, X \wedge Y$ og $|X - Y|$.

- 3.14 Lad (X_1, X_2) være en to-dimensional stokastisk vektor, og antag at sandsynlighedsfunktion $p(x_1, x_2)$ for den simultane fordeling af (X_1, X_2) er givet ved følgende tabel.

		x_1			
		-1	0	2	6
x_2	3	0	0	1/9	4/27
	1	2/9	0	1/9	1/9
	-2	1/9	1/27	1/27	1/9

Find sandsynligheden for følgende tre hændelser:

- X_1 er et lige tal,
 - $X_1 X_2$ er et ulige tal,
 - $X_2 > 0$ og $X_1 \geq 0$.
- 3.15 Lad (X_1, X_2) være en to-dimensional stokastisk vektor, og antag at sandsynlighedsfunktion $p(x_1, x_2)$ for den simultane fordeling af (X_1, X_2) er givet ved følgende tabel.

		x_1		
		-1	0	1
x_2	1	0.11	0.10	0.15
	0	0.14	0.12	0.14
	-1	0.09	0.08	0.07

- Find sandsynlighedsfunktionerne for de marginale fordelinger for X_1 og X_2 .
- Undersøg om X_1 og X_2 er uafhængige.
- Find sandsynlighedsfunktionen for den stokastiske variable $Z = X_1 X_2$.
- Definer en ny stokastisk vektor ved $(Y_1, Y_2) = (X_1^2, X_2^2)$. Find sandsynlighedsfunktion for den simultane fordeling af (Y_1, Y_2) , og undersøg om Y_1 og Y_2 er uafhængige.
- Beregn middelværdien af Z , Y_1 og Y_2 . Find dernæst middelværdi og varians af X_1 og X_2 samt $\text{Cov}(X_1, X_2)$ og $\text{corr}(X_1, X_2)$.

3.16 Lad X være en stokastisk variabel, hvis fordeling er givet ved

$$P(X = 1) = 1 - P(X = 0) = p,$$

og lad den stokastiske variable Y opfylde, at

$$\begin{aligned} P(Y = 1 | X = 0) &= 1 - P(Y = 0 | X = 0) = p_0 \\ P(Y = 1 | X = 1) &= 1 - P(Y = 0 | X = 1) = p_1, \end{aligned}$$

hvor $(p, p_0, p_1) \in (0, 1)^3$.

- Find fordelingen af Y .
- Find den simultane fordeling af den stokastiske vektor (X, Y) .
- Find en betingelse på (p_0, p_1) , som sikrer, at X og Y er uafhængige.
- Find en betingelse på (p, p_0, p_1) , som sikrer, at X og Y har samme marginale fordeling.

3.17 Additionsformlerne (1.3.11) og Opgave 1.13 angav formler for henholdsvis $P(A_1 \cup A_2)$ og $P(A_1 \cup A_2 \cup A_3)$, udtrykt ved sandsynligheder for fællesmængder mellem hændelserne A_i , $i = 1, 2, 3$. Vis den generelle formel

$$P(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

Formlen kan vises ved induktion efter n ; men grunden til, at vi har udskudt den til nu, er at den er nemmere at gennemskue, når man tænker på sandsynligheder som middelværdier af indikatorfunktioner. Hermed menes, at man til enhver hændelse A knytter en stokastisk variabel 1_A , som er lig en, når udfaldet er i A , og som ellers er lig nul. Ifølge Eksempel 3.7.2 er $E(1_A) = P(A)$. Begynd med at vise at

$$\begin{aligned} 1_{A_1 \cup \dots \cup A_n} &= 1 - (1 - 1_{A_1}) \dots (1 - 1_{A_n}) \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} 1_{A_{i_1}} \dots 1_{A_{i_k}}. \end{aligned}$$

- 3.18 Benyt Opgave 3.17 til at løse følgende klassiske problem, kaldet *rencontreproblemet*: n breve, nummereret $1, \dots, n$, lægges tilfældigt i n konvolutter, ligeledes nummereret $1, \dots, n$. Hvad er sandsynligheden for, at ikke et eneste brev kommer i konvolutten med samme nummer? Vink: Lad A_i betegne hændelse at det i 'te brev kommer i konvolut nr. i ; så er det ikke svært at beregne sandsynligheder for hændelser af formen $A_{i_1} \cap \dots \cap A_{i_k}$.

Det mest interessante ved dette problem er dets løsning. Den ønskede sandsynlighed viser sig at blive

$$1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots - (-1)^n \frac{1}{n!}\right),$$

som kan vises (eksponentialfunktionens rækkeudvikling) at konvergere mod $e^{-1} = 0.36788$ for $n \rightarrow \infty$. Konvergensten er meget hurtig. I praksis er løsningen uafhængig af n , når blot n er større end ca. 7.

- 3.19 Vis, at der for ethvert $a > 0$ gælder, at

$$P(|X| \geq a) \leq \frac{E|X|}{a}.$$

Denne ulighed er kendt som Markovs ulighed. Vink: Bemærk at $1_{[a, \infty)}(|X|) \leq \frac{|X|}{a}$.

- 3.20 Angiv middelværdi, varians og spredning for

(a) En stokastisk variabel som er ligefordelt på $\{1, 2, 3, 4, 5, 6\}$, f. eks. antal øjne ved et slag med en terning.

(b) Summen af øjnene i et slag med to terninger.

(c) En stokastisk variabel som er ligefordelt på $\{1, 2, \dots, n\}$. Vink: $1 + 2 + \dots + n = \frac{1}{2}n(n+1)$ og $1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(2n+1)(n+1)$.

- 3.21 Lad X være binomialfordelt med antalsparameter n og sandsynlighedsparameter p . Find $E(X(X-1))$ ved hjælp af Sætning 3.7.3, og brug resultatet til at give et alternativt bevis for, at $\text{Var}(X) = np(1-p)$.

3.22 For en stokastisk variabel Y defineres det *andet moment* omkring $\mu \in \mathbb{R}$ som $E((Y - \mu)^2)$. Vis at denne størrelse, som funktion af μ , antager sin mindste værdi for $\mu = EY$. Den mindste værdi er $\text{Var}(Y)$.

3.23 Vis, for en stokastisk variabel Y , at hvis

$$P(Y \in [a, b]) = 1,$$

så er

$$a \leq EY \leq b \quad \text{og} \quad \sqrt{\text{Var}(Y)} \leq \frac{b-a}{2}.$$

Vink: Angående den sidste ulighed, vis at $E\left(\left(Y - \frac{a+b}{2}\right)^2\right) \leq \left(\frac{b-a}{2}\right)^2$, og benyt Opgave 3.22. For hver af disse uligheder, hvornår gælder der lighedstegn?

3.24 En stokastisk variabel X har middelværdi 5 og varians 2. Hvad er $E(7 + 8X + X^2)$?

3.25 Vis påstandene i Eksempel 3.8.11. Vink: Ifølge Sætning 3.5.5 er S_i , S_j og $S_i + S_j$ binomialfordelte med antalsparameter n og sandsynlighedsparametre henholdsvis p_i , p_j og $p_i + p_j$. Udtryk $\text{Var}(S_i + S_j)$ ved varianserne af S_i og S_j og kovariansen mellem disse.

Hvornår er korrelationen mellem S_i og S_j lig -1?

3.26 Lad X_1, \dots, X_n være uafhængige, identisk fordelte stokastiske variable med middelværdi μ og varians σ^2 . Sæt $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, og definer den *empiriske varians* s^2 ved

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Vis at $E(s^2) = \sigma^2$.

3.27 Lad X_1 , X_2 og X_3 være identisk fordelte, uafhængige stokastiske variable med strengt positiv varians. Vis at

$$\text{corr}(X_1 + X_2, X_2 + X_3) = \frac{1}{2}.$$

3.28 Lad X være en stokastisk variabel, som antager værdier i mængden $\{-1, 0, 1\}$, og lad dens sandsynlighedsfunktion p være givet ved

$$p(-1) = 1/4, \quad p(0) = 1/2, \quad p(1) = 1/4.$$

Definer desuden $Y = X^2$

- (a) Bestem middelværdi og varians for X .
 - (b) Bestem sandsynlighedsfunktionen for Y , samt middelværdi og varians for Y .
 - (c) Vis at X og Y er ukorreleerede, d.v.s. $\text{Cov}(X, Y) = 0$.
 - (d) Bestem sandsynlighedsfunktionen for (X, Y) , og vis at X og Y *ikke* er uafhængige.
- 3.29 For en stokastisk vektor (X_1, \dots, X_k) defineres *kovariansmatricen* som $k \times k$ -matricen $V = \{v_{ij}\}$, hvor $v_{ii} = \text{Var}(X_i)$, $i = 1, \dots, k$, og hvor $v_{ij} = \text{Cov}(X_i, X_j)$, når $i \neq j$. Kovariansmatricen er altså symmetrisk ($v_{ij} = v_{ji}$).

(a) En symmetrisk $k \times k$ -matrix $V = \{v_{ij}\}$ kaldes positiv semi-definit, hvis

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j v_{ij} \geq 0$$

for alle $(a_1, \dots, a_k) \in \mathbb{R}^k$. Vis, at kovariansmatricen er positiv semi-definit. Vink: Beregn variansen af $a_1 X_1 + \dots + a_k X_k$.

(b) Determinanten af en positiv semi-definit matrix er ikke-negativ. Benyt denne kendsgerning til at give et alternativt bevis for (a) i Sætning 3.8.7.

(c) En symmetrisk $k \times k$ -matrix $V = \{v_{ij}\}$ kaldes positiv definit, hvis den er positiv semi-definit og desuden opfylder, at

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j v_{ij} = 0$$

hvis og kun hvis $a_1 = \dots = a_k = 0$. Antag, at den stokastiske vektor (X_1, \dots, X_k) opfylder, at $a_1 X_1 + \dots + a_k X_k$ er degenereret hvis og kun hvis vektoren $a_1 = \dots = a_k = 0$. Vis, at kovariansmatricen da er positiv definit. Overvej betingelsens geometriske fortolkning.

-
- 3.30 Antag, at X_1 og X_2 er uafhængige stokastiske variable, som begge er ligefordelte på $\{0, 1, \dots, N\}$. Find sandsynlighedsfunktionen for $X_1 + X_2$.

Kapitel 4

Generelle diskrete fordelinger

4.1 Diskrete fordelinger

I Kapitel 3 betragtede vi fordelinger, hvor sandsynlighedsmassen ligger på en endelig mængde. Dette er ikke altid naturligt, som følgende eksempel viser.

Eksempel 4.1.1 Et servicefirma har indrettet en web-site, som kunderne kan besøge for at få forskellige informationer og bestille tjenesteydelser. Erfaringen viser, at web-siten i gennemsnit får 180 besøg i timen, men at antallet af besøg varierer på tilfældig måde. For at vælge en passende computerkapacitet, som sikrer at kunderne kun sjældent oplever problemer, er det vigtigt at have en model for den stokastiske variation af antallet af besøg der begynder per minut. Det er ikke nok at vide, at der i gennemsnit begynder tre besøg per minut; man må også kende sandsynligheden for at væsentligt højere antal optræder.

Vi kan dele et minut op i små tidsintervaller af længde 1 sekund, og definere 60 stokastiske variable X_1, \dots, X_{60} ved

$$X_i = \begin{cases} 1 & \text{hvis et besøg begynder i } i\text{te tidsinterval} \\ 0 & \text{ellers.} \end{cases}$$

Hvis vi ignorerer muligheden for, at mere end ét besøg begynder i løbet af et af de små tidsintervaller, er det samlede antal besøg, som begynder i løbet af et minut, X , givet ved $X = X_1 + \dots + X_{60}$. Vi antager yderligere,

at de stokastiske variable X_1, \dots, X_{60} er uafhængige og identisk fordelte med $P(X_i = 1) = 3/60 = 1/20$. Det sidste valg er i overensstemmelse med sandsynlighedsregningens frekvensfortolkning. Af disse antagelser følger det, at den stokastiske variabel X er binomialfordelt med antalsparameter 60 og sandsynlighedsparameter $1/20$. Dette er imidlertid ikke nogen helt perfekt model, fordi vi valgte at ignorere muligheden for, at mere end ét besøg begynder i samme lille tidsinterval, og sandsynligheden for dette er nok ikke helt negligibel. Vi kan dog let forbedre modellen. Vi kan nemlig bare dele et minut op i delintervaller af længde $1/10$ sekund, og sætte sandsynligheden for at et besøg begynder i et delinterval til $1/200$. Sandsynligheden for at mere end et besøg begynder i løbet af den samme $1/10$ sekund er næppe særlig stor. I denne model er det samlede antal besøg per minut, X , binomialfordelt med antalsparameter 600 og sandsynlighedsparameter $1/200$.

Men vi har stadig lavet en approximation, så der er stadig basis for at forbedre modellen. For ethvert $n \in \mathbb{N}$ kan vi dele et minut op i n delintervaller af længde $1/n$ minut og sætte sandsynligheden for at et besøg begynder i et af intervallerne til $3/n$. Ved at argumentere som tidligere fås, at antallet af besøg per time er binomialfordelt med antalsparameter n og sandsynlighedsparameter $3/n$. For ethvert n laver vi stadig en lille approximation, som dog bliver mindre og mindre, når vi gør n større og større. Vi fornemmer, at den perfekte model kan opnås ved at lade n gå mod uendelig.

□

For at lave den grænseovergang, som synes at kunne give en tilfredsstillende model i Eksempel 4.1.1, har vi brug for følgende sætning.

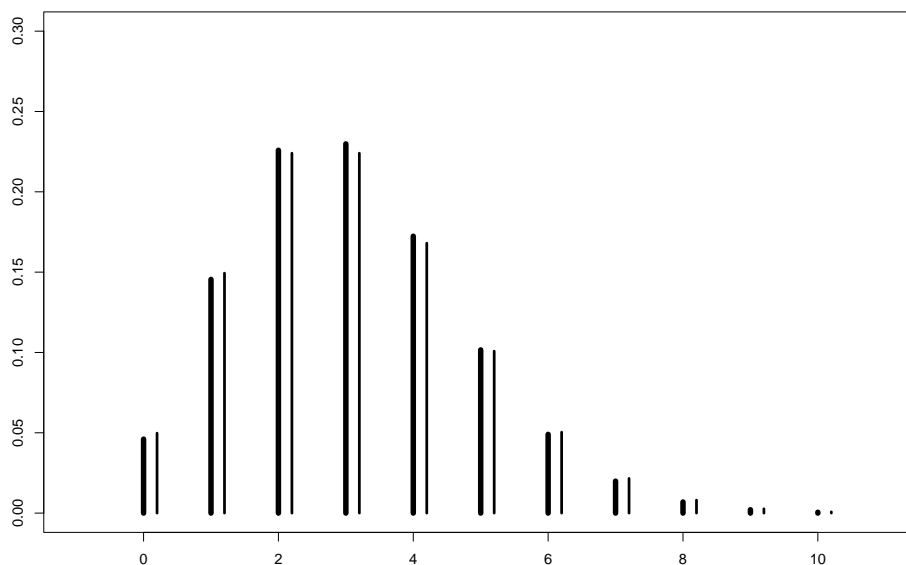
Sætning 4.1.2 *Lad $\{p_n\}$ være en følge af reelle tal i intervallet $(0, 1)$, som opfylder, at $np_n \rightarrow \lambda \in [0, \infty)$ for $n \rightarrow \infty$. Da gælder for ethvert $x \in \mathbb{N}_0$, at*

$$\lim_{n \rightarrow \infty} \binom{n}{x} p_n^x (1 - p_n)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}. \quad (4.1.1)$$

Bevis: Bemærk først, at da λ er et endeligt tal, medfører $np_n \rightarrow \lambda$, at $p_n \rightarrow 0$. Vi begynder med tilfældet $x > 0$ og laver følgende enkle

omskrivning

$$\begin{aligned} & \binom{n}{x} p_n^x (1 - p_n)^{n-x} \\ &= [1 - 1/n] \cdots [1 - (x-1)/n] (1 - p_n)^{-x} \frac{(np_n)^x}{x!} (1 - p_n)^n. \end{aligned}$$



Figur 4.1.1: Som en illustration af Sætning 4.1.2 er her sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 60 og sandsynlighedsparameter 1/20 tegnet som fede lodrette streger, mens grænseværdien givet ved (4.1.1) er angivet ved siden af med tyndere streger.

Det er klart, at

$$[1 - 1/n] \cdots [1 - (x-1)/n] (1 - p_n)^{-x} \rightarrow 1,$$

og at

$$(np_n)^x \rightarrow \lambda^x$$

når $n \rightarrow \infty$, men vi må se lidt nærmere på $(1 - p_n)^n$. Da den naturlige logaritme er differentiabel i 1 med differentialkoefficient 1, ser vi at

$$\log(1 - p_n)^n = -np_n \frac{\log(1) - \log(1 - p_n)}{p_n} \rightarrow -\lambda,$$

når $n \rightarrow \infty$, så $(1 - p_n)^n \rightarrow e^{-\lambda}$.

Tilfældet $x = 0$ er lettere. Her er binomialsandsynligheden lig $(1 - p_n)^n$, som vi lige har set konvergerer mod $e^{-\lambda}$. Altså gælder (4.1.1) også for $x = 0$.

□

Sætning 4.1.2 siger, at hvis X_n er binomialfordelt med antalsparameter n og sandsynlighedsparameter p_n , hvor p_n er som beskrevet i sætningen, så vil

$$P(X_n = x) \rightarrow \frac{\lambda^x}{x!} e^{-\lambda}$$

når $n \rightarrow \infty$. For et givet n er de mulige værdier for X_n mængden $\{0, 1, \dots, n\}$, som jo vokser op mod \mathbb{N}_0 , når $n \rightarrow \infty$. Man må derfor spørge sig, om man på en eller anden måde kan lade grænseværdien i (4.1.1) definere en sandsynlighedsfordeling på \mathbb{N}_0 . Med $p_n = 3/n$ (og dermed $\lambda = 3$) ville vi derved få en tilfredsstillende model i Eksempel 4.1.1. Det viser sig, at det kan man godt, men vi har endnu ikke behandlet fordelinger på en uendelig mængde som \mathbb{N}_0 , så det må vi lige have på plads først.

Vi vil definere sandsynlighedsmål på mængder, som kan skrives på formen

$$T = \{x_i \mid i \in \mathbb{N}\}. \quad (4.1.2)$$

Man kan kort sige, at man skal kunne tælle elementerne i T . Elementerne x_i behøver ikke være reelle tal; de kan for eksempel godt tilhøre \mathbb{R}^n . Vigtige eksempler på mængder af denne type er \mathbb{N} , \mathbb{Z} og \mathbb{N}^k . At \mathbb{Z} kan skrives på formen (4.1.2) ses på følgende måde. Hvis vi definerer

$$x_i = \begin{cases} \frac{i-1}{2} & \text{hvis } i \text{ er ulige} \\ -\frac{i}{2} & \text{hvis } i \text{ er lige,} \end{cases}$$

er $\mathbb{Z} = \{x_i \mid i \in \mathbb{N}\}$. At \mathbb{N}^k kan skrives på formen (4.1.2) er mindre oplagt, men vil blive vist i et matematikkursus. Man skal bare finde en

snedig måde at tælle elementerne i \mathbb{N}^k på. Vi nøjes her med at antyde, hvordan man kan tælle elementerne i \mathbb{N}^2 :

$$\begin{array}{ccccccc} & & & & & & \vdots \\ & & & & & & 10 & \vdots \\ & & & & & & 6 & 9 \\ & & & & & & 3 & 5 & 8 & \dots \\ & & & & & & 1 & 2 & 4 & 7 & \dots \end{array}$$

Ved en lignende argumentation kan det mere generelt vises, at hvis T_1 og T_2 er to mængder, som begge kan skrives på formen (4.1.2), så kan også produktmængden $T_1 \times T_2$ skrives på denne form. Mængder af formen (4.1.2) og endelige mængder kaldes under et *tællelige mængder*.

En sandsynlighedsfordeling, som er koncentreret på en tællelig mængde kaldes en *diskret fordeling*. Da vi i forrige kapitel gennemgik teorien for fordelinger, der er koncentreret på en endelig mængde, vil vi her kun eksplicit betragte det uendelige tilfælde, altså $T = \{x_i \mid i \in \mathbb{N}\}$. Det vil dog være klart, at alt, hvad der står i dette kapitel, også gælder for en fordeling, som er koncentreret på en endelig mængde $T = \{x_1, \dots, x_N\}$. I resten af dette kapitel betegner T en mængde på formen $T = \{x_i \mid i \in \mathbb{N}\}$.

Ligesom for en endelig mængde defineres en fordeling på en tællelig mængde ved hjælp af en sandsynlighedsfunktion.

Definition 4.1.3 *En funktion p fra en tællelig mængde $T = \{x_i \mid i \in \mathbb{N}\}$ ind i intervallet $[0, 1]$, som opfylder, at*

$$\sum_{i=1}^{\infty} p(x_i) = 1, \quad (4.1.3)$$

kaldes en sandsynlighedsfunktion på T .

Uendelige summer som den i (4.1.3) er behandlet i Appendix C. Den væsentlige forskel mellem dette kapitel og Kapitel 3 er, at man skal være lidt mere påpasselig med uendelige summer end med endelige summer.

Eksempel 4.1.4 Funktionen

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \mathbb{N}_0,$$

hvor $\lambda > 0$, er en sandsynlighedsfunktion på \mathbb{N}_0 . Det er klart, at $p(x) \geq 0$. Fra den matematiske analyse er det endvidere kendt (eller bliver det), at eksponentialfunktionen kan skrives som den uendelige række

$$\exp(y) = \sum_{n=0}^{\infty} \frac{y^n}{n!} = 1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \frac{y^4}{24} + \dots,$$

der er konvergent for alle reelle tal y . Heraf følger det, at

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = 1.$$

Så mangler vi bare at indse, at $p(x) \leq 1$, men det følger af, at $p(x) \leq \sum_{n=0}^{\infty} p(n) = 1$ for alle $x \in \mathbb{N}_0$, da alle led i summen er positive. \square

Eksempel 4.1.5 Funktionen

$$p(x) = (1 - \theta)^x \theta, \quad x \in \mathbb{N}_0,$$

hvor $\theta \in (0, 1)$, er en sandsynlighedsfunktion på \mathbb{N}_0 . Det er klart, at $p(x) \in [0, 1]$. Da

$$\sum_{i=0}^n (1 - \theta)^i = \frac{1 - (1 - \theta)^{n+1}}{1 - (1 - \theta)} = \frac{1 - (1 - \theta)^{n+1}}{\theta},$$

følger det, at

$$\sum_{i=0}^{\infty} (1 - \theta)^i \theta = \theta \lim_{n \rightarrow \infty} \sum_{i=0}^n (1 - \theta)^i = \theta \frac{1}{\theta} = 1,$$

dvs. $p(x)$ opfylder (4.1.3). \square

Enhver sandsynlighedsfunktion p på en tællelig mængde $T = \{x_i \mid i \in \mathbb{N}\}$ definerer et sandsynlighedsmål på T ved

$$P(A) = \sum_{x \in A} p(x) = \sum_{i=1}^{\infty} p(x_i) 1_A(x_i), \quad (4.1.4)$$

hvor A er en vilkårlig delmængde af T . I den sidste sum betegner 1_A indikatorfunktionen for mængden A , som er defineret ved (A.1.15). Der mindes om, at $1_A(x)$ er lig en, hvis x tilhører A , og at $1_A(x)$ ellers er lig nul. Summen i (4.1.4) har ofte uendelig mange led, men da $p(x_i)1_A(x_i) \leq p(x_i)$, følger det af (4.1.3), at summen konvergerer. Hvis et sandsynlighedsmål er givet ved (4.1.4), siger vi, at det har sandsynlighedsfunktion p .

Lad os nu vise, at P er et sandsynlighedsmål på T . Først bemærkes det, at $P(A) \geq 0$, da alle led i summen (4.1.4) er positive, og at $P(A) \leq \sum_{i=1}^{\infty} p(x_i) = 1$. Da

$$P(T) = \sum_{i=1}^{\infty} p(x_i)1_T(x_i) = 1,$$

er (1.3.2) opfyldt, og hvis A og B er disjunkte delmængder af T er

$$\begin{aligned} P(A \cup B) &= \sum_{i=1}^{\infty} p(x_i)1_{A \cup B}(x_i) = \sum_{i=1}^{\infty} p(x_i)[1_A(x_i) + 1_B(x_i)] \\ &= \sum_{i=1}^{\infty} p(x_i)1_A(x_i) + \sum_{i=1}^{\infty} p(x_i)1_B(x_i) = P(A) + P(B), \end{aligned}$$

så også (1.3.3) holder. Vi har her brugt, at $1_{A \cup B}(x) = 1_A(x) + 1_B(x)$ (overvej hvorfor dette er rigtigt).

En stokastisk variabel eller vektor, hvis fordeling er en diskret fordeling, kaldes en *diskret stokastisk variabel* eller *vektor*. For en diskret stokastisk variabel eller vektor X , hvis fordeling har sandsynlighedsfunktionen p (man siger kort, at X har sandsynlighedsfunktion p), gælder, at

$$P(X = x) = p(x)$$

for alle $x \in T$. Dette følger umiddelbart af (4.1.4) ved at sætte $A = \{x\}$. Man kalder derfor også $p(x)$ for *punktsandsynligheden* i x .

Der skal lige erindres om, at man godt kan opfatte en diskret fordeling på en tællelig mængde $T \subseteq \mathbb{R}^n$ som en fordeling på hele \mathbb{R}^n ved for en vilkårlig delmængde $A \subseteq \mathbb{R}^n$ at sætte $P(A) = P(A \cap T)$, se diskussionen i begyndelsen af Kapitel 3. For en diskret stokastisk variabel eller vektor X koncentreret på den tællelige mængde T med sandsynlighedsfunktion

p gælder ifølge (4.1.4), at

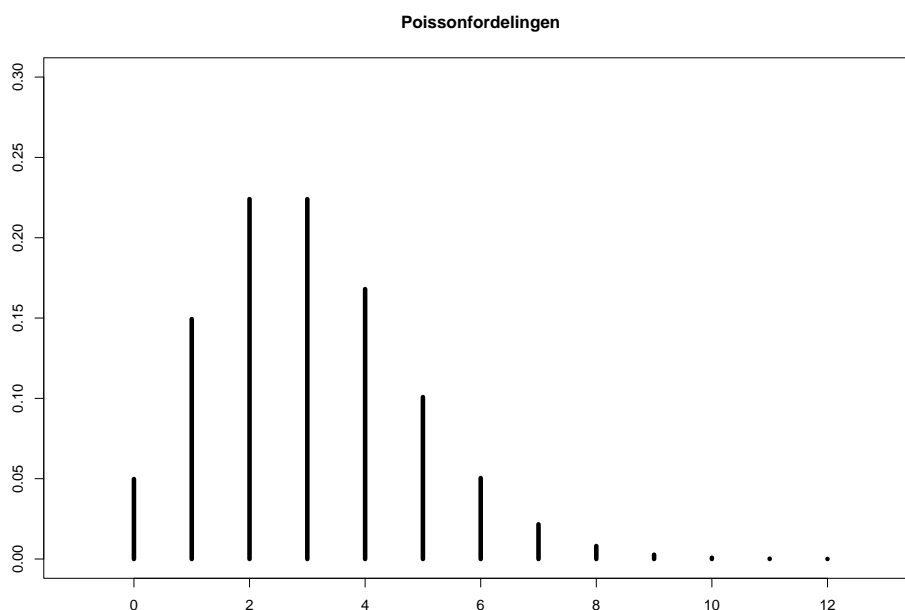
$$P(X \in A) = \sum_{x \in A \cap T} p(x) = \sum_{x \in A \cap T} P(X = x), \quad (4.1.5)$$

hvor A er en vilkårlig delmængde af \mathbb{R} eller \mathbb{R}^n .

Eksempel 4.1.4 (fortsat). Vi så ovenfor, at funktionen

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \mathbb{N}_0, \quad (4.1.6)$$

hvor $\lambda > 0$, er en sandsynlighedsfunktion på \mathbb{N}_0 . Den diskrete fordeling på \mathbb{N}_0 defineret ved sandsynlighedsfunktion (4.1.6) kaldes *Poissonfordelingen med parameter λ* . Eksempel 4.1.1 sammen med Sætning 4.1.2



Figur 4.1.2: Sandsynlighedsfunktionen for Poissonfordelingen med parameter 3.

viser, at Poissonfordelingen er en god model for antallet af tilfældigt indtræffende hændelser, som for eksempel besøg på en web-site. Der er

mange andre situationer, hvor man kan give argumenter, som er helt analoge til argumentet i Eksempel 4.1.1, for at et observeret antal er udfald af en Poissonfordelt stokastisk variabel. Eksempler på størrelser, som ofte kan antages at være Poissonfordelte, er antallet af skadesanmeldelser af en bestemt type til et forsikringselskab i en bestemt periode, antal trafikulykker per dag af en given type i et givet område, og antal fødsler per dag i et givet område. I nogle tilfælde er det ikke tiden, men derimod rummet, der deles op i små delmængder, når der skal argumenteres for at en størrelse er Poissonfordelt. Således er antallet af bakterier af en bestemt type i en vandprøve ofte Poissonfordelt.

□

Lad os lige kort betragte *fordelingsfunktionen* for en diskret fordeling. Vi nøjes med at betragte en diskret fordeling, som er koncentreret på \mathbb{N}_0 . Lad p betegne dens sandsynlighedsfunktion. Da er fordelingsfunktionen F for P givet ved

$$F(x) = \sum_{i=0}^{\infty} p(i) 1_{(-\infty, x]}(i) = \sum_{i=0}^{[x]} p(i) \quad \text{for } x \geq 0, \quad (4.1.7)$$

hvor $[x]$ betegner heltalsdelen af x (dvs. det største hele tal, der er mindre end eller lig x), mens $F(x) = 0$, når $x < 0$, og at

$$F(n) = \sum_{i=0}^n p(i) \rightarrow \sum_{i=0}^{\infty} p(i) = 1,$$

når $n \rightarrow \infty$. Funktionen F er konstant, når $x \in \mathbb{R} \setminus \mathbb{N}_0$, mens den har et spring opad af størrelse $p(i)$ i de $i \in \mathbb{N}_0$, for hvilke $p(i) > 0$. I springpunkterne er F kontinuert fra højre, men ikke fra venstre. Hermed menes, at $\lim_{n \rightarrow \infty} F(i + \frac{1}{n}) = F(i)$, mens $\lim_{n \rightarrow \infty} F(i - \frac{1}{n}) = F(i - 1) \neq F(i)$.

Bemærk, at vi kan finde sandsynlighedsfunktionen ud fra fordelingsfunktionen, idet

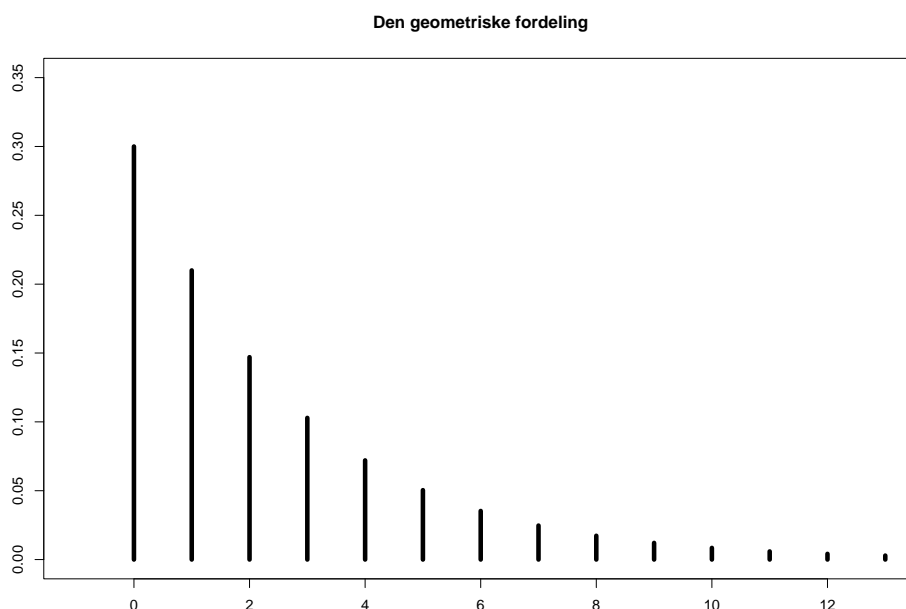
$$p(x) = F(x) - F(x - 1) \quad \text{for } x \in \mathbb{N}_0. \quad (4.1.8)$$

Da sandsynlighedsfunktionen bestemmer fordelingen, gælder det derfor også fordelingsfunktionen, som påstået i Afsnit 2.2.

Eksempel 4.1.5 (fortsat). Funktionen

$$p(x) = (1 - \theta)^x \theta, \quad x \in \mathbb{N}_0, \quad (4.1.9)$$

hvor $\theta \in (0, 1)$, er som vist ovenfor en sandsynlighedsfunktion på \mathbb{N}_0 . Den diskrete fordeling med sandsynlighedsfunktionen (4.1.9) kaldes *den geometriske fordeling* med parameter θ . Vi vil nu studere et typisk eksempel, hvor denne fordeling optræder.



Figur 4.1.3: Sandsynlighedsfunktionen for den geometriske fordeling med parameter 0.3.

Betragt et forsøg med to mulige udfald K og U , hvor sandsynligheden for udfaldet K er $\theta \in (0, 1)$. Uafhængige gentagelser af dette forsøg udføres, indtil første gang resultatet K opnås. Lad T være den stokastiske variable, som angiver ventetiden på udfaldet K . Hermed menes det antal forsøg med udfaldet U , der skal udføres inden udfaldet K opnås. Med andre ord er $T = t$, hvis de første t forsøg har udfaldet U , mens forsøg nummer $t + 1$ har udfaldet K . Den stokastiske variabel T antager kun

værdier i mængden \mathbb{N}_0 , og dens fordelingsfunktion har formen (4.1.7), som vi nu skal se.

Da $T \geq i$, præcis når de første i forsøg har udfaldet U , og da sandsynligheden for udfaldet U er $1 - \theta$, er fordelingsfunktionen for T givet ved

$$\begin{aligned} F(i) &= P(T \leq i) = 1 - P(T > i) \\ &= 1 - P(T \geq i + 1) = 1 - (1 - \theta)^{i+1} \end{aligned} \quad (4.1.10)$$

for ethvert $i \in \mathbb{N}_0$. Heraf følger det, at

$$\begin{aligned} P(T = i) &= P(T \leq i) - P(T \leq i - 1) \\ &= F(i) - F(i - 1) = (1 - \theta)^i \theta \end{aligned}$$

for alle $i \in \mathbb{N}_0$. Her er $p(i) = (1 - \theta)^i \theta$ netop sandsynlighedsfunktionen defineret ved (4.1.9). Det ses af (4.1.10), at

$$F(i) = \sum_{j=0}^i (1 - \theta)^j \theta,$$

altså, at F er af formen (4.1.7). Det vil sige, at T 's fordeling er den geometriske fordeling med parameter θ . □

4.2 Transformation af diskrete fordelinger

Hvis en diskret stokastisk vektor har sandsynlighedsfunktion p , er der, ligesom vi så det for fordelinger på endelige mængder, relationer mellem p og sandsynlighedsfunktionerne for de tilsvarende marginale fordelinger. Også her nøjes vi med at give disse relationer for en to-dimensional diskret fordeling, da den interesserede studerende let selv kan formulere og bevise resultatet for højere-dimensionale diskrete fordelinger.

Sætning 4.2.1 *Lad $p : T_1 \times T_2 \mapsto [0, 1]$ være sandsynlighedsfunktionen for den simultane fordeling af den to-dimensionale diskrete stokastiske vektor (X_1, X_2) , hvor X_i er koncentreret på den tællelige mængde T_i , $i = 1, 2$. Da har de marginale fordelinger af X_1 og X_2 sandsynlighedsfunktionerne $p_1 : T_1 \mapsto [0, 1]$ og $p_2 : T_2 \mapsto [0, 1]$ givet ved*

$$p_1(x_1) = \sum_{x_2 \in T_2} p(x_1, x_2) \quad (4.2.1)$$

og

$$p_2(x_2) = \sum_{x_1 \in T_1} p(x_1, x_2). \quad (4.2.2)$$

Bevis: Vi viser kun (4.2.1), da de to resultater er helt symmetriske. At p_1 er en sandsynlighedsfunktion på T_1 ses af, at $\sum_{x_1 \in T_1} p_1(x_1) = \sum_{x_1 \in T_1} \sum_{x_2 \in T_2} p(x_1, x_2) = 1$, da p er en sandsynlighedsfunktion på $T_1 \times T_2$. Lad A være en delmængde af T_1 . Af (4.1.4) følger det, at

$$\begin{aligned} P(X_1 \in A) &= P((X_1, X_2) \in A \times T_2) \\ &= \sum_{x_1 \in A} \sum_{x_2 \in T_2} p(x_1, x_2) = \sum_{x_1 \in A} p_1(x_1). \end{aligned}$$

Dette betyder netop, at p_1 er sandsynlighedsfunktionen for den marginale fordeling af X_1 , igen ifølge (4.1.4). □

Der gælder følgende transformationssætning for diskrete fordelinger, som Sætning 4.2.1 er et specialtilfælde af. Sætning 3.4.2 er selvfølgelig også et specialtilfælde.

Sætning 4.2.2 *Lad (X_1, \dots, X_n) være en n -dimensional diskret stokastisk vektor koncentreret på den tællelige mængde $T \subseteq \mathbb{R}^n$ med simultan sandsynlighedsfunktion p , og lad ψ være en afbildning fra T ind i \mathbb{R}^k . Da er $Y = \psi(X_1, \dots, X_n)$ en diskret stokastisk vektor koncentreret på $\psi(T)$, hvis sandsynlighedsfunktion er givet ved*

$$q(y) = \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) \quad \text{for } y \in \psi(T). \quad (4.2.3)$$

Bevis: At q er en sandsynlighedsfunktion følger af, at $\sum_{y \in \psi(T)} q(y) = \sum_{y \in \psi(T)} \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) = \sum_{(x_1, \dots, x_n) \in T} p(x_1, \dots, x_n) = 1$. Lad nu A være en delmængde af $\psi(T)$. Da er

$$\begin{aligned} P(Y \in A) &= P((X_1, \dots, X_n) \in \psi^{-1}(A)) \\ &= \sum_{(x_1, \dots, x_n) \in \psi^{-1}(A)} p(x_1, \dots, x_n) \end{aligned}$$

$$\begin{aligned}
&= \sum_{y \in A} \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) \\
&= \sum_{y \in A} q(y).
\end{aligned}$$

□

Eksempel 4.2.3 Lad X være Poisson fordelt med parameter μ . Da er $Y = X/c$, hvor c er et positivt reelt tal, en diskret stokastisk variabel koncentreret på mængden $T = \{i/c \mid i \in \mathbb{N}_0\}$ med sandsynlighedsfunktion

$$q(y) = \frac{\mu^{cy}}{(cy)!} e^{-\mu} \text{ for } y \in T.$$

Videre er $Z = X^2$ en diskret stokastisk variabel koncentreret på mængden $\tilde{T} = \{i^2 \mid i \in \mathbb{N}_0\}$ med sandsynlighedsfunktion

$$q(z) = \frac{\mu^{\sqrt{z}}}{(\sqrt{z})!} e^{-\mu} \text{ for } z \in \tilde{T}.$$

□

Med præcis samme bevis som i tilfældet med stokastiske variable koncentreret på endelige mængder vises det, at den næste sætning følger af transformationssætningen 4.2.2.

Sætning 4.2.4 Lad (X_1, X_2) være en to-dimensional stokastisk vektor, hvis simultane fordeling har sandsynlighedsfunktionen p . Antag endvidere, at X_i s fordeling er koncentreret på \mathbb{N}_0 , $i = 1, 2$. Da er sandsynlighedsfunktionen for den stokastiske variable $Y = X_1 + X_2$ givet ved

$$q(y) = \sum_{j=0}^y p(j, y-j), \quad (4.2.4)$$

for $y \in \mathbb{N}_0$.

4.3 Uafhængige diskrete stokastiske variable

Følgende sætning, som giver betingelser for, at n diskrete stokastiske variable er *uafhængige*, vises på helt samme måde som Sætning 3.6.1, bortset fra at summerne her kan være uendelige.

Sætning 4.3.1 *Lad (X_1, \dots, X_n) være en n -dimensional diskret stokastisk vektor, hvor X_i er koncentreret på den tællelige mængde T_i , $i = 1, \dots, n$, og antag, at den simultane fordeling af (X_1, \dots, X_n) har sandsynlighedsfunktion $p : T \mapsto [0, 1]$, hvor $T = T_1 \times \dots \times T_n$. Lad endelig $p_i : T_i \mapsto [0, 1]$, være sandsynlighedsfunktionen for den marginale fordeling af X_i , $i = 1, \dots, n$. Da er følgende tre udsagn ækvivalente:*

1) X_1, \dots, X_n er stokastisk uafhængige,

2) For alle $(x_1, \dots, x_n) \in T$ er

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n), \quad (4.3.1)$$

3) Der findes n ikke-negative reelle funktioner g_i , $i = 1, \dots, n$, så

$$p(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) \quad (4.3.2)$$

for alle $(x_1, \dots, x_n) \in T$.

Også for generelle diskrete fordelinger er funktionerne g_1, \dots, g_{n-1} og g_n proportionale med henholdsvis p_1, \dots, p_{n-1} og p_n , ligesom der (med samme argumenter som i Kapitel 3) gælder, at *Hvis mængden*

$$\{(x_1, x_2) | p(x_1, x_2) > 0\}$$

ikke er en produktmængde, kan X_1 og X_2 ikke være uafhængige. Her betegner $p(x_1, x_2)$ som tidligere den simultane sandsynlighedsfunktion for (X_1, X_2) . Det er derfor i den beskrevne situation let at konstatere, hvis to generelle diskrete stokastiske variable *ikke* er uafhængige.

I Kapitel 2 blev det lovet at bevise 2) i Sætning 2.4.3 for diskrete fordelinger. Ideen i beviset for diskrete fordelinger er præcis den samme som i beviset i Kapitel 2, men lad os alligevel for fuldstændighedens skyld vise følgende sætning.

Sætning 4.3.2 *Lad (X_1, \dots, X_n) være en n -dimensional diskret stokastisk vektor med sandsynlighedsfunktion p , og antag at de stokastiske variable X_1, \dots, X_n er uafhængige. Hvis ψ er en funktion fra \mathbb{R}^{n-k} ind i \mathbb{R} , hvor $k < n$, er de $k+1$ stokastiske variable $X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n)$ uafhængige.*

Bevis: Vi viser, at sandsynlighedsfunktionen for den diskrete stokastiske vektor $(X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n))$ spalter op i et produkt af formen (4.3.2). Lad T betegne den tællelige mængde, hvorpå vektoren (X_{k+1}, \dots, X_n) er koncentreret. Ifølge Sætning 4.2.2 er den simultane sandsynlighedsfunktion for $(X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n))$

$$\begin{aligned}
 & P(X_1 = x_1, \dots, X_k = x_k, \psi(X_{k+1}, \dots, X_n) = x) \\
 &= \sum_{(x_{k+1}, \dots, x_n) \in \psi^{-1}(\{x\}) \cap T} P(X_1 = x_1, \dots, X_k = x_k, X_{k+1} = x_{k+1}, \dots, X_n = x_n) \\
 &= \sum_{(x_{k+1}, \dots, x_n) \in \psi^{-1}(\{x\}) \cap T} P(X_1 = x_1) \cdots P(X_k = x_k) P(X_{k+1} = x_{k+1}, \dots, X_n = x_n) \\
 &= P(X_1 = x_1) \cdots P(X_k = x_k) \sum_{(x_{k+1}, \dots, x_n) \in \psi^{-1}(\{x\}) \cap T} P(X_{k+1} = x_{k+1}, \dots, X_n = x_n) \\
 &= p_1(x_1) \cdots p_k(x_k) h(x),
 \end{aligned}$$

hvor $h(x)$ er lig med summen i næstsidste linie. Den simultane sandsynlighedsfunktion spalter altså op som ønsket, således at de stokastiske variable $X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n)$ ifølge Sætning 4.3.1 er uafhængige. \square

Sætning 2.4.3 indeholder endnu to resultater, som imidlertid følger af 2), så disse [d.v.s. 1) og 3)] er nu også vist for diskrete stokastiske variable.

Man har meget ofte brug for at finde fordelingen af summen af to eller flere uafhængige stokastiske variable på \mathbb{N}_0 , hvilket begrundes, at vi giver følgende specialtilfælde af Sætning 4.2.4 for uafhængige stokastiske variable.

Sætning 4.3.3 *Lad (X_1, X_2) være en to-dimensional diskret stokastisk vektor. Antag, at X_1 og X_2 er uafhængige, og at de er koncentreret på \mathbb{N}_0 med sandsynlighedsfunktioner p_1 og p_2 . Definer en ny stokastisk variabel ved $Y = X_1 + X_2$. Da er sandsynlighedsfunktionen for den stokastiske variable $Y = X_1 + X_2$ givet ved*

$$q(y) = \sum_{j=0}^y p_1(j) p_2(y-j) \quad (4.3.3)$$

for $y \in \mathbb{N}_0$.

Sandsynlighedsfunktionen q givet ved (4.3.3) kaldes, som tidligere nævnt, *foldningen* af de to sandsynlighedsfunktioner p_1 og p_2 , ligesom man kalder fordelingen af $X_1 + X_2$ foldningen af de to marginale fordelinger.

Sætning 4.3.4 (Poissonfordelingens foldningsegenskab). *Lad X_1, \dots, X_n være uafhængige, Poissonfordelte stokastiske variable med parametre $\lambda_1, \dots, \lambda_n$. Da er $X_1 + \dots + X_n$ Poissonfordelt med parameter $\lambda_1 + \dots + \lambda_n$.*

Bevis: Vi kan nøjes med at vise sætningen for $n = 2$, da det generelle resultat da følger ved induktion (idet $X_1 + \dots + X_n = [X_1 + \dots + X_{n-1}] + X_n$). For $n = 2$ følger resultatet af (4.3.3). For $y \in \mathbb{N}_0$ er

$$\begin{aligned} P(X_1 + X_2 = y) &= \sum_{j=0}^y \frac{\lambda_1^j}{j!} e^{-\lambda_1} \frac{\lambda_2^{(y-j)}}{(y-j)!} e^{-\lambda_2} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{y!} \sum_{j=0}^y \binom{y}{j} \lambda_1^j \lambda_2^{(y-j)} \\ &= \frac{(\lambda_1 + \lambda_2)^y}{y!} e^{-(\lambda_1 + \lambda_2)}, \end{aligned}$$

hvor det sidste lighedstegn følger af binomialformlen; se (B.3.3) i Appendix B.

□

Alt, hvad der stod i forrige kapitel om *betinget uafhængighed*, holder også for generelle diskrete stokastiske variable. For god ordens skyld kommer det hele igen i den diskrete version.

Definition 4.3.5 *Lad X_1, X_2 og X_3 være tre diskrete stokastiske variable, som er koncentreret på de tællelige mængder T_1, T_2 og T_3 . Da siges X_1 og X_2 at være betinget uafhængige givet X_3 , hvis der for alle (x_1, x_2, x_3) i $T_1 \times T_2 \times T_3$, for hvilke $P(X_3 = x_3) > 0$ gælder, at*

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2 | X_3 = x_3) &= \\ P(X_1 = x_1 | X_3 = x_3) P(X_2 = x_2 | X_3 = x_3). \end{aligned} \tag{4.3.4}$$

Beviset for følgende sætning er identisk med beviset for Sætning 3.6.5

Sætning 4.3.6 *Lad (X_1, X_2, X_3) være en tre-dimensional diskret stokastisk vektor med simultan sandsynlighedsfunktion p . Antag, at X_i er koncentreret på den tællelige mængde T_i , $i = 1, 2, 3$. Betegn den simultane sandsynlighedsfunktion for (X_i, X_j) med p_{ij} , og sandsynlighedsfunktionen for X_3 med p_3 . Da er følgende tre udsagn ækvivalente:*

- 1) X_1 og X_2 er betinget uafhængige givet X_3 ,
- 2) For alle $(x_1, x_2, x_3) \in T_1 \times T_2 \times T_3$, gælder, at

$$p(x_1, x_2, x_3)p_3(x_3) = p_{13}(x_1, x_3)p_{23}(x_2, x_3), \quad (4.3.5)$$

- 3) Der findes to ikke-negative reelle funktioner g og h , så

$$p(x_1, x_2, x_3) = g(x_1, x_3)h(x_2, x_3) \quad (4.3.6)$$

for alle $(x_1, x_2, x_3) \in T_1 \times T_2 \times T_3$.

4.4 Middelværdi og varians

Middelværdi og varians kan for mange diskrete fordelinger defineres på en måde som er helt analog til definitionerne for fordelinger på endelige mængder. Det er dog ikke muligt at definere en middelværdi og en varians for alle diskrete fordelinger. Problemet er, at der indgår uendelige summer i definitionerne, og disse kan eventuelt divergere. Vi må derfor passe lidt på med summerne, hvilket er den eneste forskel mellem dette afsnit og afsnit 3.7.

Definition 4.4.1 *Lad X være en diskret stokastisk variabel, som er koncentreret på den tællelige mængde $T = \{x_i \mid i \in \mathbb{N}\}$ ($x_i \in \mathbb{R}$), og som har sandsynlighedsfunktion p . Da siges X at have middelværdi, hvis*

$$\sum_{i=1}^{\infty} |x_i|p(x_i) < \infty, \quad (4.4.1)$$

og vi definerer i så fald middelværdien af X som

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i). \quad (4.4.2)$$

Hvis (4.4.1) ikke er opfyldt, siger man, at X ikke har middelværdi, eller at middelværdien ikke er defineret. Når vi kræver, at (4.4.1) er opfyldt, er det for at udelukke den ubehagelige situation, hvor

$$\sum_{x_i \geq 0} x_i p(x_i) = \infty \quad \text{og} \quad \sum_{x_i < 0} x_i p(x_i) = -\infty.$$

Hvis $T \subseteq \mathbb{R}_+$, kan dette ikke ske, hvorfor man ofte i dette tilfælde tillader sig at skrive $E(X) = \infty$, når (4.4.1) ikke er opfyldt.

Lad os lige nævne et tilfælde, hvor man altid kan være sikker på at middelværdien eksisterer. Antag, at X er koncentreret på en begrænset mængde, d.v.s. at der findes et $c > 0$, så $T \subseteq [-c, c]$ (og dermed $P(X \in [-c, c]) = 1$). En stokastisk variabel med denne egenskab kaldes *begrænset*. Når $T \subseteq [-c, c]$, er

$$\sum_{i=1}^{\infty} |x_i| p(x_i) \leq \sum_{i=1}^{\infty} c p(x_i) = c \sum_{i=1}^{\infty} p(x_i) = c < \infty,$$

så X har middelværdi. Bemærk, at en fordeling på en endelig mængde er et specialtilfælde.

Eksempel 4.1.4 (fortsat). Middelværdien i Poissonfordelingen findes ved følgende udregning:

$$\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} \frac{\lambda^{(i-1)}}{(i-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda,$$

hvor $j = i - 1$, og hvor vi bruger, at summen i næstsidste udtryk er summen af Poissonfordelingens punktsandsynligheder, som jo er lig med en. Parameteren i Poissonfordelingen er altså lig med dens middelværdi. \square

Eksempel 4.1.5 (fortsat). Middelværdien i den geometriske fordeling kan beregnes på følgende måde. Lad X være en stokastisk variabel, der er geometrisk fordelt med parameter θ . Da er

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} i\theta(1-\theta)^i = (1-\theta) \sum_{i=1}^{\infty} i\theta(1-\theta)^{i-1} \\ &= (1-\theta) \left(\sum_{i=1}^{\infty} (i-1)\theta(1-\theta)^{i-1} + \sum_{i=1}^{\infty} \theta(1-\theta)^{i-1} \right) \\ &= (1-\theta)(E(X) + 1), \end{aligned}$$

hvoraf det følger, at

$$E(X) = \frac{1-\theta}{\theta}.$$

Man bør naturligvis først overveje, om middelværdien er endelig, for ellers er ovenstående udledning fejlagtig. At summen konvergerer følger af, at $(1-\theta)^i$ går eksponentielt hurtigt mod nul, mens i kun går lineært hurtigt mod uendelig.

Det forventede antal plat før første krone i en serie af møntkast er altså $\frac{1-0.5}{0.5} = 1$, og det forventede antal slag før første sekser i en serie af terningkast er $\frac{1-1/6}{1/6} = 5$. Begge dele lyder jo troværdigt. \square

Man kan beregne middelværdien af en transformeret diskret stokastisk variabel på samme enkle måde som for fordelinger på endelige mængder.

Sætning 4.4.2 *Lad X være en diskret n -dimensional stokastisk vektor, som er koncentreret på den tællelige mængde $T \subseteq \mathbb{R}^n$, og som har sandsynlighedsfunktion p . Lad endvidere ψ være en funktion fra T ind i \mathbb{R} . Da har den stokastiske variable $\psi(X)$ middelværdi, hvis og kun hvis*

$$\sum_{(x_1, \dots, x_n) \in T} |\psi(x_1, \dots, x_n)| p(x_1, \dots, x_n) < \infty, \quad (4.4.3)$$

og i så fald er

$$E(\psi(X)) = \sum_{(x_1, \dots, x_n) \in T} \psi(x_1, \dots, x_n) p(x_1, \dots, x_n). \quad (4.4.4)$$

Bevis: Ifølge Sætning 4.2.2 har fordelingen af $\psi(X)$ sandsynlighedsfunktion

$$q(y) = \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n),$$

$y \in \psi(T)$. Ifølge Definition 4.4.1 har den stokastiske variabel $\psi(X)$ middelværdi, netop når dens sandsynlighedsfunktion opfylder (4.4.1), d.v.s. når

$$\sum_{y \in \psi(T)} |y|q(y) < \infty.$$

Dette er det samme som (4.4.3), da

$$\begin{aligned} \sum_{y \in \psi(T)} |y|q(y) &= \sum_{y \in \psi(T)} |y| \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} p(x_1, \dots, x_n) \\ &= \sum_{y \in \psi(T)} \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} |y|p(x_1, \dots, x_n) \\ &= \sum_{y \in \psi(T)} \sum_{(x_1, \dots, x_n) \in \psi^{-1}(\{y\})} |\psi(x_1, \dots, x_n)|p(x_1, \dots, x_n) \\ &= \sum_{(x_1, \dots, x_n) \in T} |\psi(x_1, \dots, x_n)|p(x_1, \dots, x_n). \end{aligned}$$

Dermed er sætningens første påstand vist. Den anden følger ved samme beregning uden numerisk tegn om y , dvs. helt samme beregning som i beviset for Sætning 3.7.3. □

Også beviset for den følgende sætning er lig beviset for den tilsvarende sætning for fordelinger på endelige mængder, Sætning 3.7.5, når man først har sikret sig, at rækken er konvergent.

Sætning 4.4.3 *Lad X være en diskrete stokastisk variabel, som har middelværdi, og lad a og b være vilkårlige reelle tal. Da har også den stokastiske variable $a + bX$ middelværdi, og*

$$E(a + bX) = a + bE(X) \tag{4.4.5}$$

Bevis: Lad $T = \{x_i \mid i \in \mathbb{N}\}$ være den tællelige mængde, på hvilken X er koncentreret, og lad p være sandsynlighedsfunktionen for X . Da X har middelværdi, er

$$\sum_{i=1}^{\infty} |x_i|p(x_i) < \infty,$$

og da $|a + bx_i| \leq |a| + |b||x_i|$, medfører dette, at

$$\sum_{i=1}^{\infty} |a + bx_i|p(x_i) < \infty.$$

Ved hjælp af Sætning 4.4.2 kan vi derfor slutte, at $a+bX$ har middelværdi, som kan beregnes ved

$$\sum_{i=1}^{\infty} (a + bx_i)p(x_i) = a + b \sum_{i=1}^{\infty} x_i p(x_i) = a + bE(X).$$

□

Sætning 4.4.4 *Lad (X_1, X_2) være en to-dimensional diskret stokastisk vektor, som opfylder, at $X_1 \leq X_2$, samt at X_1 og X_2 har middelværdi. Da er $E(X_1) \leq E(X_2)$. Specielt gælder for en stokastisk variabel X , som har middelværdi, at*

$$|E(X)| \leq E(|X|). \quad (4.4.6)$$

Beviset er nærmest identisk med beviset for Sætning 3.7.6 og udelades derfor.

Sætning 4.4.5 *Lad (X_1, X_2) være en diskret to-dimensional stokastisk vektor, som opfylder, at X_2 har middelværdi, samt at $|X_1| \leq |X_2|$. Da har også X_1 middelværdi.*

Bevis: Lad T være den tællelige mængde, på hvilken (X_1, X_2) er koncentreret, og lad p betegne sandsynlighedsfunktionen for (X_1, X_2) . Da X_2 har middelværdi, har vi ifølge Sætning 4.4.2, at

$$\sum_{(x_1, x_2) \in T} |x_2|p(x_1, x_2) < \infty.$$

Udsagnet at $|X_1| \leq |X_2|$ betyder, at $|x_1| \leq |x_2|$ for alle $(x_1, x_2) \in T$. Derfor er

$$\sum_{(x_1, x_2) \in T} |x_1|p(x_1, x_2) \leq \sum_{(x_1, x_2) \in T} |x_2|p(x_1, x_2) < \infty,$$

hvilket, igen ifølge Sætning 4.4.2, viser at X_1 har middelværdi.

□

Sætning 4.4.6 *Lad (X_1, X_2, \dots, X_n) være en diskret stokastisk vektor, hvor X_i har middelværdi for alle $i = 1, \dots, n$. Da har den stokastiske variable $X_1 + X_2 + \dots + X_n$ middelværdi, og*

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n). \quad (4.4.7)$$

Hvis det yderligere antages, at de stokastiske variable X_1, X_2, \dots, X_n er uafhængige, har også den stokastiske variable $X_1 \cdot X_2 \cdot \dots \cdot X_n$ middelværdi, og

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n). \quad (4.4.8)$$

Bevis: Vi kan nøjes med at bevise sætningen for $n = 2$, da det generelle resultat følger ved induktion. Som sædvanlig betegner T den tællelige mængde, på hvilken (X_1, X_2) er koncentreret, mens sandsynlighedsfunktionen for (X_1, X_2) betegnes med p , og sandsynlighedsfunktionen for X_i , $i = 1, 2$, med p_i . Da X_i har middelværdi, er ifølge Sætning 4.4.2

$$\sum_{(x_1, x_2) \in T} |x_i| p(x_1, x_2) < \infty, \quad i = 1, 2.$$

Derfor medfører trekantsuligheden $|x_1 + x_2| \leq |x_1| + |x_2|$ at

$$\sum_{(x_1, x_2) \in T} |x_1 + x_2| p(x_1, x_2) < \infty,$$

hvorfor den stokastiske variable $X_1 + X_2$ ifølge Sætning 4.4.2 har middelværdi, som kan beregnes ved

$$\begin{aligned} \sum_{(x_1, x_2) \in T} (x_1 + x_2) p(x_1, x_2) &= \sum_{(x_1, x_2) \in T} x_1 p(x_1, x_2) + \sum_{(x_1, x_2) \in T} x_2 p(x_1, x_2) \\ &= E(X_1) + E(X_2). \end{aligned}$$

For at vise den sidste påstand, antages det nu, at X_1 og X_2 er uafhængige. Dermed er $p(x_1, x_2) = p_1(x_1)p_2(x_2)$, og vi kan antage, at $T = T_1 \times T_2$, hvor T_1 og T_2 er tællelige delmængder af \mathbb{R} , jfr. Sætning 4.3.1 og bemærkningerne efter denne sætning. Derfor er

$$\sum_{(x_1, x_2) \in T} |x_1 x_2| p(x_1, x_2) = \sum_{x_2 \in T_2} \left(\sum_{x_1 \in T_1} |x_1| p_1(x_1) |x_2| p_2(x_2) \right)$$

$$\begin{aligned}
&= \sum_{x_2 \in T_2} |x_2| p_2(x_2) \left(\sum_{x_1 \in T_1} |x_1| p_1(x_1) \right) \\
&= \left(\sum_{x_1 \in T_1} |x_1| p_1(x_1) \right) \left(\sum_{x_2 \in T_2} |x_2| p_2(x_2) \right),
\end{aligned}$$

og de to summer i det sidste udtryk er endelige, da X_1 og X_2 begge har middelværdi. Igen kan vi bruge Sætning 4.4.2 til at slutte, at den stokastiske variable $X_1 X_2$ har middelværdi, som kan beregnes ved

$$\begin{aligned}
\mathbb{E}(X_1 X_2) &= \sum_{(x_1, x_2) \in T} x_1 x_2 p(x_1, x_2) \\
&= \sum_{x_2 \in T_2} \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) x_2 p_2(x_2) \right) = \sum_{x_2 \in T_2} x_2 p_2(x_2) \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) \right) \\
&= \left(\sum_{x_1 \in T_1} x_1 p_1(x_1) \right) \left(\sum_{x_2 \in T_2} x_2 p_2(x_2) \right) = \mathbb{E}(X_1) \mathbb{E}(X_2).
\end{aligned}$$

□

For at sikre, at *variansen* af en diskret fordeling eksisterer, må vi kræve mere end for middelværdien.

Definition 4.4.7 *Lad X være en diskret stokastisk variabel, som er koncentreret på den tællelige mængde $T = \{x_i \mid i \in \mathbb{N}\}$, og som har sandsynlighedsfunktion p . Da siges X at have varians, hvis*

$$\sum_{i=1}^{\infty} x_i^2 p(x_i) < \infty, \quad (4.4.9)$$

og vi definerer i så fald variansen af X ved

$$\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2). \quad (4.4.10)$$

Først må vi lige overveje, at variansen er veldefineret. Da $|X| \leq X^2 + 1$, medfører (4.4.9) ifølge Sætning 4.4.5, at X har middelværdi, og da

$$[X - \mathbb{E}(X)]^2 = X^2 + \mathbb{E}(X)^2 - 2X\mathbb{E}(X)$$

sikrer Sætning 4.4.6, at den stokastiske variable $[X - E(X)]^2$ har middelværdi. Betingelsen (4.4.9) kan kort skrives som $E(X^2) < \infty$.

Det er klart, at formlerne (3.7.9) og (3.7.10) også gælder for generelle diskrete fordelinger. For disse kaldes kvadratroden af variansen naturligvis også *spredningen* eller *standardafvigelsen*.

Eksempel 4.1.4 (fortsat). Poissonfordelingens varians kan udregnes på følgende måde. Lad X være Poissonfordelt med parameter λ . Da er

$$\begin{aligned} E(X(X-1)) &= \sum_{i=0}^{\infty} i(i-1) \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} e^{-\lambda} = \lambda^2 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda^2, \end{aligned}$$

hvor $j = i - 2$. Heraf følger det, at

$$\text{Var}(X) = E(X(X-1)) + EX - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Bemærk, at for en Poissonfordeling er variansen lig middelværdien. \square

Eksempel 4.1.5 (fortsat). Lad os nu finde variansen af en stokastisk variabel X , der er geometrisk fordelt med parameter θ .

$$\begin{aligned} E(X^2) &= \sum_{i=0}^{\infty} i^2 \theta (1-\theta)^i = (1-\theta) \sum_{i=1}^{\infty} i^2 \theta (1-\theta)^{i-1} \\ &= (1-\theta) \left(\sum_{i=1}^{\infty} (i-1)^2 \theta (1-\theta)^{i-1} + \sum_{i=1}^{\infty} \theta (1-\theta)^{i-1} \right. \\ &\quad \left. + 2 \sum_{i=1}^{\infty} (i-1) \theta (1-\theta)^{i-1} \right) \\ &= (1-\theta) (E(X^2) + 1 + 2E(X)). \end{aligned}$$

Da vi i afsnit 4.4 så, at $E(X) = (1-\theta)/\theta$, er

$$E(X^2) = \frac{1-\theta}{\theta} + \frac{2(1-\theta)^2}{\theta^2},$$

således at

$$\text{Var}(X) = \frac{1 - \theta}{\theta} + \frac{(1 - \theta)^2}{\theta^2} = \frac{1 - \theta}{\theta^2}.$$

At rækken $\sum_{i=0}^{\infty} i^2 \theta (1 - \theta)^i$ er konvergent, kan man overbevise sig om på omtrent samme måde som i forbindelse med middelværdien af en geometrisk fordeling.

□

Sætning 4.4.8 *For en diskret stokastisk variabel X , som har varians, er $\text{Var}(X) = 0$ hvis og kun hvis $P(X = E(X)) = 1$.*

Beviset for denne sætning udelades, da det er omtrent identisk med beviset for Sætning 3.7.12.

4.5 Kovarians og korrelation

For at definere kovariansen mellem to diskrete stokastiske variable X og Y , må man sikre sig, at middelværdien i (3.8.1) eksisterer. Antag, at X og Y begge har varians, d.v.s. at

$$E(X^2) < \infty \quad \text{og} \quad E(Y^2) < \infty. \quad (4.5.1)$$

Vi så i forrige afsnit, at dette medfører, at middelværdien af X og Y også eksisterer. Da

$$(X - E(X))(Y - E(Y)) = XY - XE(Y) - YE(X) + E(X)E(Y)$$

og

$$|XY| \leq \frac{1}{2}(X^2 + Y^2),$$

ser vi af Sætningerne 4.4.5 og 4.4.6, at betingelsen (4.5.1) også sikrer, at den stokastiske variable $(X - E(X))(Y - E(Y))$ har middelværdi, således at vi kan definere kovariansen ved (3.8.1) og dermed korrelationen ved (3.8.8). Det er klart, at vi stadig kan beregne kovariansen ved (3.8.2), samt at regnereglerne (3.8.3) – (3.8.5) gælder for tre diskrete stokastiske, forudsat at de alle har varians.

En gennemgang af beviserne i afsnit 3.8 viser, at alt hvad der står i det afsnit om varianser, kovarianser og korrelationer også holder for

generelle diskrete stokastiske variable, hvis blot det forudsættes, at alle indgående stokastiske variable har varians, altså opfylder (4.4.9). Der er derfor ingen grund til at bruge papir på gentage alle disse resultater med denne betingelse tilføjet.

4.6 Betingede fordelinger

Lad P være en diskret fordeling, som er koncentreret på den tællelige mængde $T \subseteq \mathbb{R}^n$ med sandsynlighedsfunktion p . Vi så i Sætning 1.4.8, at funktionen $B \mapsto P(B|A)$ for enhver fast delmængde $A \subseteq T$ med $P(A) > 0$ er et sandsynlighedsmål, den betingede fordeling givet A . Den betingede fordeling er også diskret. Den er koncentreret på A og har en sandsynlighedsfunktion, som er givet ved

$$p(x|A) = p(x)1_A(x)/P(A). \quad (4.6.1)$$

Lad os indse dette. At (4.6.1) er en sandsynlighedsfunktion på A er klart:

$$\sum_{x \in A} p(x|A) = \frac{1}{P(A)} \sum_{x \in A} p(x) = 1.$$

Endvidere er

$$\sum_{x \in B} p(x|A) = \frac{1}{P(A)} \sum_{x \in A \cap B} p(x) = \frac{P(A \cap B)}{P(A)} = P(B|A).$$

Vi har dermed vist, at (4.6.1) er sandsynlighedsfunktion for den betingede fordeling givet A .

Betragt for eksempel en to-dimensional diskret stokastisk vektor (X_1, X_2) med sandsynlighedsfunktion $p(x_1, x_2)$. Antag, at X_1 er koncentreret på den tællelige mængde T_1 . Da er den betingede fordeling af (X_1, X_2) givet $X_2 = a$, hvor $P(X_2 = a) > 0$, koncentreret på $T_1 \times \{a\}$, så man kan opfatte den som en fordeling på T_1 (a er jo fast) med sandsynlighedsfunktion

$$p(x_1|X_2 = a) = \frac{P(X_1 = x_1, X_2 = a)}{P(X_2 = a)} = \frac{p(x_1, a)}{p_2(a)} \quad \text{for } x_1 \in T_1, \quad (4.6.2)$$

Her er p_2 sandsynlighedsfunktionen for den marginale fordeling af X_2 .

Som et andet eksempel, betragt en stokastisk vektor (X_1, X_2) med simultan sandsynlighedsfunktion P , hvor både X_1 og X_2 er koncentreret på \mathbb{N}_0 . Da er den betingede fordeling af (X_1, X_2) givet $X_1 + X_2 = y$ ($y \in \mathbb{N}_0$) koncentreret på mængden

$$M = \{(x_1, x_2) | x_1 + x_2 = y\} = \{(0, y), (1, y-1), \dots, (y-1, 1), (y, 0)\}$$

med sandsynlighedsfunktionen

$$p(x_1, x_2 | X_1 + X_2 = y) = \frac{p(x_1, x_2)}{\sum_{i=0}^y p(i, y-i)}, \quad (4.6.3)$$

for $(x_1, x_2) \in M$. Nævneren $\sum_{i=0}^y p(i, y-i)$ er sandsynlighedsfunktionen for $X_1 + X_2$ i punktet y , som er udtrykt ved p under brug af Sætning 4.2.4.

Eksempel 4.6.1 Lad X_1 og X_2 være to uafhængige stokastiske variable, som begge er geometrisk fordelte med parameter $\theta \in (0, 1)$. Ifølge opgave 4.9 er sandsynlighedsfunktionen for $Y = X_1 + X_2$ givet ved $q(y) = \theta^2(1-\theta)^y(y+1)$, $y \in \mathbb{N}_0$, så den betingede fordeling af (X_1, X_2) givet $Y = y$ har sandsynlighedsfunktionen

$$p(x_1, x_2 | Y = y) = \frac{\theta(1-\theta)^{x_1} \theta(1-\theta)^{x_2}}{\theta^2(1-\theta)^y(y+1)} = \frac{1}{y+1},$$

for $(x_1, x_2) \in M$. Vi ser, at den betingede fordeling af (X_1, X_2) givet $Y = y$ er ligefordelingen på M . Da der er en entydig forbindelse mellem $(x_1, x_2) \in M$ og x_1 , kan vi konkludere (jfr. (4.2.1)), at den betingede fordeling af X_1 givet $Y = y$ er en ligefordeling på $\{0, 1, \dots, y\}$. □

4.7 Sammenfatning

I dette kapitel har vi beskæftiget os med generelle diskrete fordelinger, d.v.s. sandsynlighedsmål, der er defineret på tællelige mængder. I lighed med fordelingerne i Kapitel 3, kan en generel diskret fordeling mest naturligt specificeres ved hjælp af en sandsynlighedsfunktion. I det hele

taget er resultater og definitioner i Kapitel 4 en gentagelse af, hvad vi så i Kapitel 3, dog med den vigtige forskel, at vi for generelle diskrete fordelinger i mange tilfælde må sikre os, at de indgående uendelige summer konvergerer. Noget nyt i forhold til forrige kapitel er, at vi for en diskret fordeling betragtede den betingede fordeling givet en delmængde af de mulige udfald og beviste, at denne betingede fordeling også er diskret og kan specificeres ved en sandsynlighedsfunktion.

Sidst men ikke mindst blev den vigtige Poissonfordeling og den geometriske fordeling indført og studeret.

4.8 Opgaver

- 4.1 (a) Vis, for X Poissonfordelt med parameter λ , rekursionsformlen

$$P(X = x + 1) = \frac{\lambda}{x + 1}P(X = x).$$

(b) Vis, at Poissonfordelingens sandsynlighedsfunktion har et entydigt maksimum hvis og kun hvis λ ikke er et positivt heltal, og at dette maksimum i så fald antages i punktet $x = [\lambda]$ (hvor $[\lambda]$ betegner heltalsdelen af λ , dvs. det største hele tal som er mindre end eller lig λ).

- 4.2 Lad X_1 og X_2 være uafhængige Poisson fordelte stokastiske variable med parametre λ_1 og λ_2 . Sæt $Y = (X_1 + X_2)/2$.

(a) Find mængden F af mulige værdier af Y .
(b) Find $q(y) = P(Y = y)$ for $y \in F$.

- 4.3 Lad X_1 og X_2 være uafhængige Poisson fordelte stokastiske variable med parametre αm_1 og αm_2 . Sæt $Y = (X_1 + X_2)/(m_1 + m_2)$.

(a) Find mængden F af mulige værdier af Y .
(b) Find $q(y) = P(Y = y)$ for $y \in F$.

- 4.4 A står ved en lidet trafikeret vej og håber på at fange en ledig taxa. Der kommer i gennemsnit én hver halve time. Vi antager derfor, at antal taxaer pr. minut er Poissonfordelt med parameter $1/30$, og at disse antal er uafhængige fra minut til minut.

- (a) Hvad er sandsynligheden for at A må vente mere end 1/2 time?
- (b) Hvad er sandsynligheden for at A må vente mere end 1 1/2 time?
- (c) Hvad er sandsynligheden for at A får en taxa allerede før der er gået 10 minutter?
- (d) Vis, at ventetiden, afrundet nedad til helt minuttal, er geometrisk fordelt med parameter $p = 1 - e^{-1/30} \simeq 1/30$.
- 4.5 Antag at sandsynligheden for at en sikring er defekt er 0.03. Benyt Sætning 4.1.2 til at beregne en approksimation til sandsynligheden for, at der i en pakke med 100 sikringer højst er 2 defekte.
- 4.6 En terning kastes indtil den første sekser opnås. Lad X betegne antallet af kast før en sekser opnås. Hvad er sandsynligheden for at $X < 6$? Hvad er den største værdi af $i \in \mathbb{N}$ for hvilken $P(X > i) \geq \frac{1}{2}$?
- 4.7 Antag, at den stokastiske variable X er geometrisk fordelt med parameter p . Find sandsynlighedsfunktionerne for de stokastiske variable $Y = X^2$ og $Z = X + 3$.
- 4.8 Antag, at X er geometrisk fordelt med parameter p , og definer en ny stokastisk variabel ved $Y = X \vee M$, hvor $M \in \mathbb{N}$. Find sandsynlighedsfunktionen for Y .
- 4.9 Lad X_1 og X_2 være uafhængige stokastiske variable, som er geometrisk fordelte med samme parameter θ . Vis, at sandsynlighedsfunktionen for summen $Y = X_1 + X_2$ er

$$q(y) = \theta^2(1 - \theta)^y(y + 1).$$

Fordelingen med denne sandsynlighedsfunktion kaldes den negative binomialfordeling med antalsparameter 2 og sandsynlighedsparameter θ .

- 4.10 Lad X_1 og X_2 være uafhængige stokastiske variable, som er geometrisk fordelte med parametre θ_1 og θ_2 . Find $P(X_1 \geq X_2)$ og $P(X_1 = X_2)$.

- 4.11 Lad X være en diskret stokastisk variabel, som er koncentreret på mængden $M = \{\frac{n}{2} | n = 0, 1, 2, \dots\}$. Antag, at X har sandsynlighedsfunktionen

$$p(x) = \theta^{2x}(1 - \theta), \quad x \in M,$$

hvor $\theta \in (0, 1)$.

- (a) Vis, at $2X$ er geometrisk fordelt med parameter $1 - \theta$.
 (b) Find $E(X)$ og $\text{Var}(X)$.
 (c) Vis, at $P(X \in \mathbb{N}_0) = \frac{1}{1+\theta}$.
- 4.12 Lad (X, Y) være en diskret stokastisk vektor i \mathbb{R}^2 med sandsynlighedsfunktion

$$p(x, y) = \begin{cases} \frac{\alpha^y e^{-2\alpha}}{x!(y-x)!} & \text{når } (x, y) \in M \\ 0 & \text{ellers,} \end{cases}$$

hvor $M = \{(x, y) \in \mathbb{N}_0^2 | x \leq y\}$.

- (a) Vis, at Y er Poissonfordelt med parameter 2α . Vink: Ifølge binomialformlen (B.3.3) med $x = y = 1$ er

$$\sum_{x=0}^y \binom{y}{x} = 2^y.$$

Definer en ny stokastisk variabel Z ved $Z = Y - X$.

- (b) Vis, at X og Z er stokastisk uafhængige, og at de begge er Poissonfordelte med parameter α .
- 4.13 Lad X være en stokastisk variabel på \mathbb{N}_0 . Vis at

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

- 4.14 Lad X være Poissonfordelt med parameter λ . Hvad er $E(2^X)$?
 Hvad er $E((1 + X)^{-1})$?

4.15 (a) Vis, at

$$p(n) = \frac{1}{n(n+1)}, \quad n \in \mathbb{N},$$

er en sandsynlighedsfunktion. Vink: Benyt, at $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$.

(b) Undersøg, om den tilsvarende fordeling har middelværdi.

4.16 Lad S_1 og S_2 være uafhængige, binomialfordelte stokastiske variable med antalsparametre n_1 og n_2 og med samme sandsynlighedsparameter p . Vis, at den betingede fordeling af S_1 givet $S_1 + S_2 = s$, er den hypergeometriske fordeling med parametre $N = n_1 + n_2$, $R = n_1$ og $n = s$.

4.17 Lad (S_1, \dots, S_5) være polynomialfordelt af orden 5 med antalsparameter n og sandsynlighedsparametre p_1, \dots, p_5 .

(a) Vis, at den betingede fordeling af (S_1, S_2, S_3) givet $S_1 + S_2 + S_3 = m$ ($m \in \{0, 1, \dots, n\}$) er en polynomialfordeling af orden 3 med antalsparameter m og sandsynlighedsparametre $\frac{p_1}{p_1+p_2+p_3}$, $\frac{p_2}{p_1+p_2+p_3}$, $\frac{p_3}{p_1+p_2+p_3}$.

(b) Formuler det generelle resultat, som (a) er et specialtilfælde af.

4.18 Lad X_1 og X_2 være uafhængige Poissonfordelte stokastiske variable med parametre λ_1 og λ_2 . Vis at den betingede fordeling af X_1 givet $X_1 + X_2 = n$ ($n \in \mathbb{N}$), er binomialfordelingen med antalsparameter n og sandsynlighedsparameter $\frac{\lambda_1}{\lambda_1 + \lambda_2}$.

4.19 Lad X_1, \dots, X_k være uafhængige Poissonfordelte stokastiske variable med parametre $\lambda_1, \dots, \lambda_k$, og definer $Z = X_1 + \dots + X_k$. Hvad er den betingede fordeling af (X_1, \dots, X_k) givet $Z = n$ ($n \in \mathbb{N}$). (Her bliver den betingede fordeling en polynomialfordeling). Hvad er $P(X_1 = x_1 | Z = n)$?

Kapitel 5

En-dimensionale kontinuerte fordelinger

I mange situationer er det naturligt at benytte en sandsynlighedsteoretisk model, hvor alle tal i et interval eller på hele den reelle akse er mulige udfald. Eksempler er længdemålinger med målefejl, vægten af en tilfældigt udvalgt person eller ventetiden mellem to tilfældige begivenheder, for eksempel anmeldelsen af skader til et forsikringselskab. I sådanne situationer har man behov for modeller, hvor sandsynlighedsmassen er fordelt ud over alle reelle tal i et interval og ikke koncentreret på en endelig eller tællelig delmængde, som den er for diskrete fordelinger. Sådanne fordelinger, der kaldes kontinuerte fordelinger, er emnet for dette kapitel.

5.1 Kontinuerte fordelinger og tætheder

Vi så to eksempler på kontinuerte fordelinger i Kapitel 1. I Eksempel 1.3.1 udledte vi en model for udtrækning af et tilfældigt tal i intervallet $[0, 1]$. Sandsynligheden for et udfald i et interval $I \subseteq [0, 1]$ er i dette tilfælde længden af intervallet $|I|$. Vi kan skrive dette som

$$P(I) = \int_0^1 1_I(x) dx,$$

hvor 1_I betegner indikatorfunktionen for I , se (1.3.5). Her er sandsynlighedsmassen fordelt jævnt ud over intervallet $[0, 1]$. I Eksempel 1.3.3

var

$$P(I) = \int_0^1 1_I(x) 3x^2 dx.$$

Her er sandsynlighedsmassen også fordelt ud over hele $[0, 1]$, men den er ikke delt jævnt ud. Dens fordeling er bestemt af funktionen $p(x) = 3x^2$, som kaldes sandsynlighedstætheden. I Eksempel 1.3.1 er sandsynlighedstætheden $p(x) = 1_{[0,1]}(x)$. Lad os se på endnu et eksempel.

Eksempel 5.1.1 Betragt tilfældigt ankommende begivenheder. Det kan for eksempel være besøg på en web-site eller ankomst af kunder til et kasseapparat. Som diskuteret i Eksempel 4.1.1, er det en god model for tilfældigt ankommende begivenheder at antage, at antallet af begivenheder i tidsintervallet $[0, t]$, hvor $t > 0$, er Poissonfordelt med parameter λt (med λ passende valgt; i Eksempel 4.1.1 var $\lambda = 3$ et godt valg). Lad T være den stokastiske variable, som angiver tidspunktet for den første begivenhed efter tid 0. Da er hændelsen $T > t$ lig med hændelsen, at antallet af hændelser i tidsintervallet $[0, t]$ er lig nul. Sandsynligheden for den sidste hændelse er $e^{-\lambda t}$, så

$$P(T > t) = e^{-\lambda t}.$$

Heraf ses, at for $t_2 > t_1 > 0$ er

$$P(T \in (t_1, t_2]) = P(T > t_1) - P(T > t_2) = e^{-\lambda t_1} - e^{-\lambda t_2},$$

eller

$$P(T \in (t_1, t_2]) = \int_{t_1}^{t_2} \lambda e^{-\lambda x} dx.$$

Her er sandsynlighedsmassen fordelt ud over hele den positive halvakse $(0, \infty)$ med sandsynlighedstætheden $p(x) = \lambda e^{-\lambda x}$. Bemærk, at

$$\begin{aligned} \frac{P(T \in (x, x+h])}{h} &= \frac{e^{-\lambda x} - e^{-\lambda(x+h)}}{h} \\ &= \lambda e^{-\lambda x} \frac{1 - e^{-\lambda h}}{\lambda h} \rightarrow \lambda e^{-\lambda x} \end{aligned}$$

for $h \rightarrow 0$. For små værdier af h er altså $P(T \in (x, x+h]) \simeq p(x)h$. □

Definition 5.1.2 En funktion p fra et interval $I \subseteq \mathbb{R}$ ind i $[0, \infty)$, som opfylder, at

$$\int_I p(x)dx = 1, \quad (5.1.1)$$

kaldes en sandsynlighedstæthed på I .

Det er naturligvis underforstået, at p er sådan, at integralet i (5.1.1) eksisterer. Information om integration over reelle intervaller er samlet i Appendiks D. En sandsynlighedstæthed kaldes somme tider en tæthedsfunktion eller blot en tæthed. Enhver sandsynlighedstæthed p på et interval $I \subseteq \mathbb{R}$ definerer et sandsynlighedsmål P på I ved

$$P(A) = \int_I 1_A(x)p(x)dx = \int_A p(x)dx, \quad (5.1.2)$$

hvor A er en vilkårlig delmængde af I (eller i hvert fald en delmængde, for hvilken man kan give mening til integralet), og hvor 1_A er indikatorfunktionen for A , se (A.1.15). Bemærk analogien til (4.1.4), som definerer en diskret fordeling. Den eneste forskel er, at summen i (4.1.4) her er erstattet med et integral og sandsynlighedsfunktionen med en sandsynlighedstæthed.

Lad os overbevise os om, at P er et sandsynlighedsmål på I . Hvis A er en delmængde af I , er for alle $x \in I$

$$0 \leq 1_A(x)p(x) \leq p(x),$$

hvoraf det følger ved integration over I at $0 \leq P(A) \leq 1$. Endvidere er ifølge (5.1.1)

$$P(I) = \int_I p(x)dx = 1,$$

så (1.3.2) er opfyldt. Lad endelig A og B være disjunkte delmængder af I . Da

$$1_{A \cup B}(x)p(x) = 1_A(x)p(x) + 1_B(x)p(x),$$

følger det, at

$$\begin{aligned} P(A \cup B) &= \int_I 1_{A \cup B}(x)p(x)dx \\ &= \int_I 1_A(x)p(x)dx + \int_I 1_B(x)p(x)dx = P(A) + P(B), \end{aligned}$$

således at (1.3.3) er opfyldt. Vi har dermed godtgjort, at P defineret ved (5.1.2) er et sandsynlighedsmål på I .

Et sandsynlighedsmål, der er defineret ved (5.1.2) ud fra en sandsynlighedstæthed p kaldes en *kontinuert fordeling*, og man siger, at det har sandsynlighedstæthed p . Som diskuteret i begyndelsen af Kapitel 3 kan man opfatte P som en fordeling på hele den reelle akse ved for $B \subseteq \mathbb{R}$ at sætte $P(B) = P(B \cap I)$. Dette svarer til at definere $P(B)$ ved (5.1.2), hvor $p(x)$ er sat lig nul for $x \notin I$. Man definerer ofte på denne måde $p(x)$ for alle $x \in \mathbb{R}$, selv om fordelingen er koncentreret på et mindre interval. En stokastisk variabel, hvis fordeling er kontinuert, kaldes en *kontinuert stokastisk variabel*. Hvis X er en kontinuert stokastisk variabel, hvis fordeling har sandsynlighedstæthed p , er

$$P(X \in A) = \int_{-\infty}^{\infty} 1_A(x)p(x)dx$$

for enhver delmængde $A \subseteq \mathbb{R}$. Bemærk, at der specielt gælder, at

$$P(X = a) = \int_{-\infty}^{\infty} 1_{\{a\}}(x)p(x)dx = 0$$

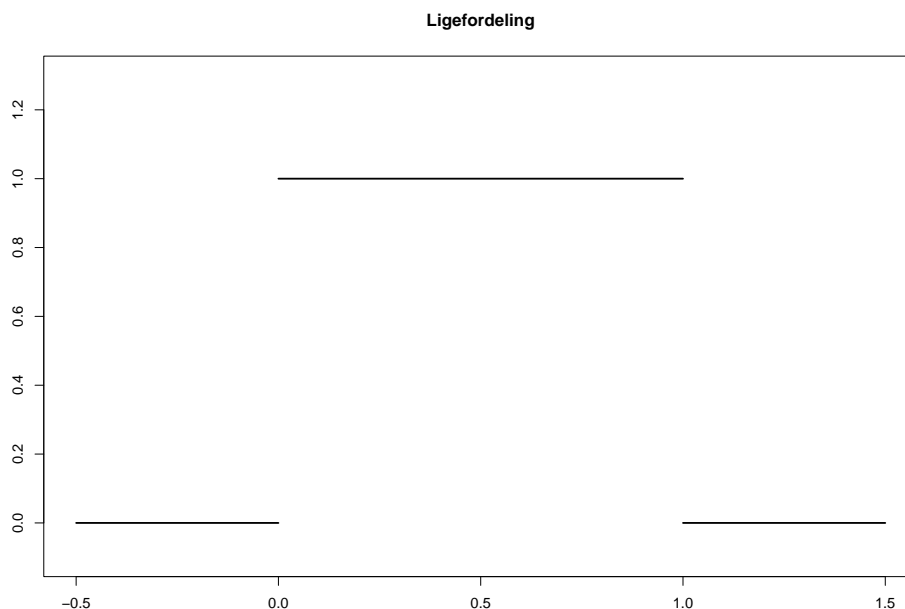
for ethvert $a \in \mathbb{R}$. Bemærk også, at man kan ændre værdien af p i et endeligt antal punkter uden at ændre den tilsvarende sandsynlighedsfordeling. En bestemt kontinuert fordeling kan altså godt have flere sandsynlighedstætheder.

Eksempel 5.1.3 For et begrænset interval $I = [a, b]$, hvor $a < b$, er funktionen

$$p(x) = \frac{1_{[a,b]}(x)}{b-a} \tag{5.1.3}$$

en sandsynlighedstæthed på I . Den tilsvarende kontinuerte fordeling kaldes *ligefordelingen* på $[a, b]$. Den kaldes også den rektangulære fordeling eller den uniforme fordeling på $[a, b]$. Denne fordeling svarer til at trække et tal tilfældigt i intervallet $[a, b]$. Man kan på tilsvarende måde definere en ligefordeling på enhver delmængde af \mathbb{R} , som kan tilskrives en endelig længde, f.eks. (a, b) eller $(a, b) \cup [c, d]$, hvor $a < b < c < d$.

□



Figur 5.1.1: Sandsynlighedstætheden for ligefordelingen på $[0, 1]$.

Eksempel 5.1.1 (fortsat). Funktionen

$$\lambda e^{-\lambda x} \quad x > 0 \quad (5.1.4)$$

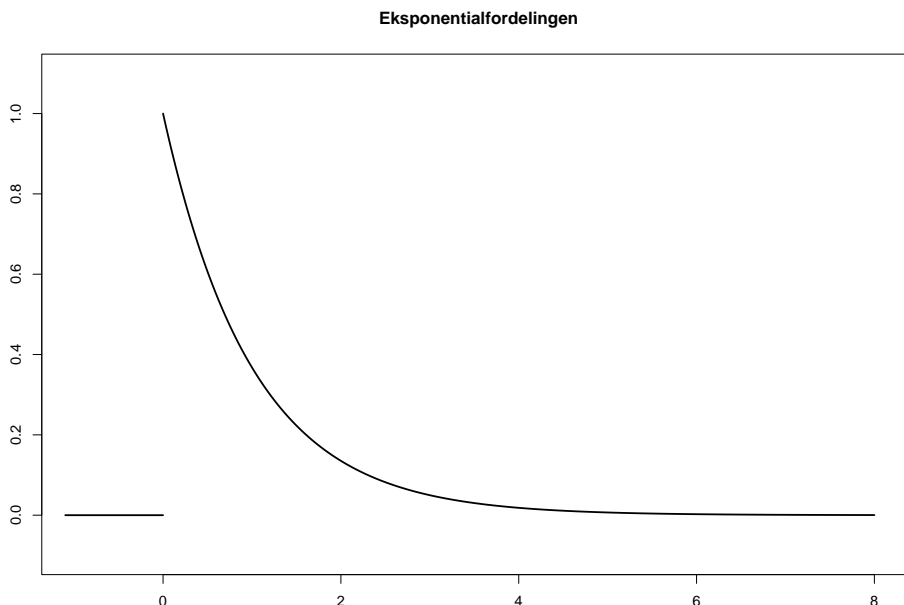
er en sandsynlighedstæthed på intervallet $(0, \infty)$. Den tilsvarende kontinuerte fordeling kaldes *eksponentialfordelingen* med parameter λ . Vi kan altså sige, at T er eksponentialfordelt med parameter λ . Når $\lambda = 1$ kaldes fordelingen standard eksponentialfordelingen eller den normerede eksponentialfordeling.

□

Eksempel 5.1.4 Funktionen

$$p(x) = \beta x^{\beta-1}, \quad x \in (0, 1), \quad (5.1.5)$$

hvor $\beta > 0$, er en sandsynlighedstæthed på $(0, 1)$ (overvej lige tilfældet $\beta \in (0, 1)$ en ekstra gang). Sandsynlighedstætheden i Eksempel 1.3.3



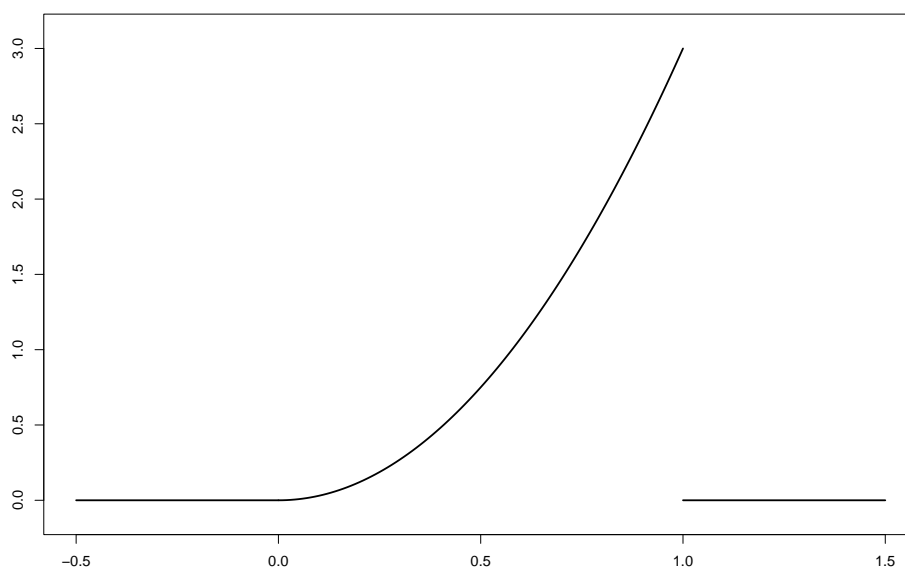
Figur 5.1.2: Sandsynlighedstætheden for eksponentialfordelingen med parameter $\lambda = 1$.

er af denne type med $\beta = 3$, mens ligefordelingen på $(0, 1)$ opnås for $\beta = 1$. Den tilsvarende kontinuerte fordeling kaldes en *Beta-fordeling*. Vi betragter kun en særligt enkel type Beta-fordeling her. Den generelle Beta-fordeling har tæthed $cx^{\beta-1}(1-x)^{\alpha-1}$, $x \in (0, 1)$, hvor $\alpha > 0$ og $\beta > 0$, og hvor c er en konstant, som sikrer, at der er tale om en sandsynlighedstæthed. Vi vil her kun beskæftige os med den specielle form givet ved (5.1.5). Et udtryk for konstanten c vil dog blive fundet i beviset for Sætning 8.1.3, se også opgave 8.7.

□

Lad P være en kontinuert fordeling med sandsynlighedstæthed p . For små værdier af $\delta > 0$ er

$$P([x_0, x_0 + \delta]) = \int_{x_0}^{x_0 + \delta} p(x) dx \simeq p(x_0)\delta,$$



Figur 5.1.3: Sandsynlighedstætheden givet ved (5.1.5) med $\beta = 3$.

forudsat at funktionen p er kontinuert i punktet x_0 . Vi ser, at værdien af sandsynlighedstætheden p i dens kontinuitetspunkter kan fortolkes som sandsynlighedsmasse per måleenhed (f. eks. længdeenhed), hvilket jo stemmer godt overens med betegnelsen sandsynlighedstæthed.

Fordelingsfunktionen for en kontinuert fordeling på den reelle akse med sandsynlighedstæthed p er givet ved

$$F(x) = \int_{-\infty}^x p(y) dy. \quad (5.1.6)$$

Fordelingen kan godt være koncentreret på en delmængde M af \mathbb{R} , men dette tilfælde er jo dækket ved at sætte værdien af p lig nul udenfor M .

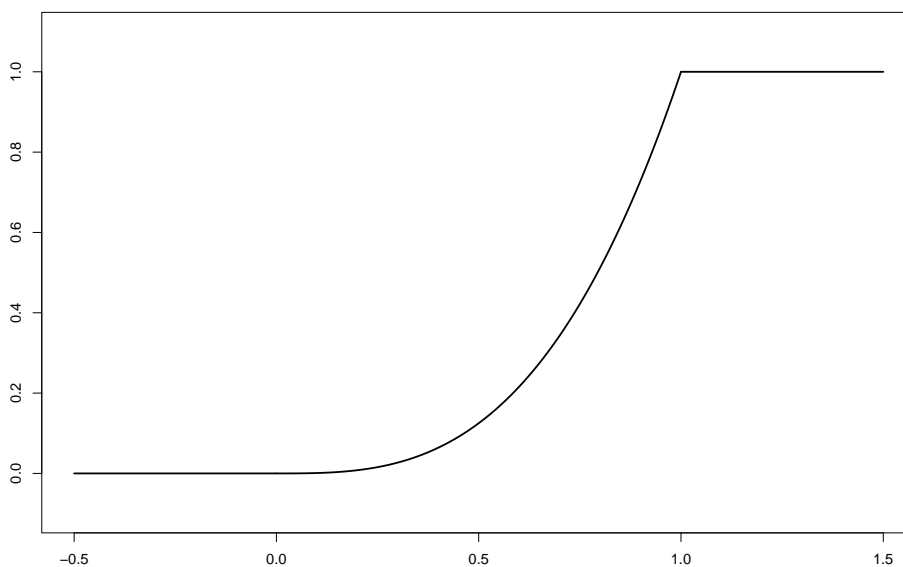
Eksempel 5.1.3 (fortsat). Fordelingsfunktionen for ligefordelingen på

intervallet (a, b) er

$$F(x) = \int_{-\infty}^x \frac{1_{(a,b)}(y)}{b-a} dy = \begin{cases} 0 & \text{for } x \leq a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x \geq b, \end{cases}$$

som er stykvis lineær. Se Figur 2.2.2.

□



Figur 5.1.4: Fordelingsfunktionen for fordelingen med sandsynlighedstæthed givet ved (5.1.5) med $\beta = 3$.

Eksempel 5.1.4 (fortsat). Fordelingsfunktionen for fordelingen med tæthed (5.1.5) er

$$F(x) = \int_{-\infty}^x \beta y^{\beta-1} 1_{(0,1)}(y) dy = \begin{cases} 0 & \text{for } x \leq 0 \\ x^\beta & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1. \end{cases}$$

□

Fordelingsfunktionen er en svagt voksende kontinuert funktion fra \mathbb{R} ind i $[0, 1]$, som opfylder, at

$$\lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} \int_{-\infty}^n p(y) dy = \int_{-\infty}^{\infty} p(y) dy = 1,$$

samt at

$$\begin{aligned} \lim_{n \rightarrow \infty} F(-n) &= \lim_{n \rightarrow \infty} P((-\infty, -n]) = \lim_{n \rightarrow \infty} [1 - P((-\infty, -n))] \\ &= 1 - \lim_{n \rightarrow \infty} \int_{-\infty}^{-n} p(y) dy = 1 - \int_{-\infty}^{\infty} p(y) dy = 0. \end{aligned}$$

Af (5.1.6) ses, at F er en stamfunktion til p . Derfor er F differentiabel i ethvert kontinuitetspunkt for tætheden p , og

$$F'(x) = p(x). \quad (5.1.7)$$

Vi ser, at i hvert fald for kontinuerte fordelinger med kontinuert sandsynlighedstæthed bestemmer fordelingsfunktionen sandsynlighedstætheden og dermed fordelingen. Dette resultat blev allerede nævnt i Afsnit 2.2.

Nu følger et par resultater, som er nyttige, hvis en fordelingsfunktion er givet, og man ud fra denne ønsker at afgøre, om den tilsvarende fordeling er kontinuert.

Sætning 5.1.5 *Antag, at en fordelingsfunktion F har formen*

$$F(x) = \int_{-\infty}^x f(y) dy, \quad (5.1.8)$$

hvor f er ikke-negativ. Da er fordelingen svarende til F kontinuert med sandsynlighedstæthed f .

Bevis: Da

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^n f(x) dx = \lim_{n \rightarrow \infty} F(n) = 1,$$

hvor vi har brugt (2.2.3), ses det, at f er en sandsynlighedstæthed på \mathbb{R} . Den tilsvarende fordelingsfunktion er F . Da der til enhver fordelingsfunktion svarer netop én fordeling, er den til F svarende fordeling altså kontinuert med tæthed f . □

Af hensyn til den næste sætning mindes der om, at en funktion f kaldes kontinuert differentiabel, hvis den er differentiabel og den afledede funktion f' er kontinuert.

Sætning 5.1.6 *Lad F være fordelingsfunktion for en fordeling, som er koncentreret på et interval (a, b) , d.v.s. $P((a, b)) = 1$. Eventuelt kan a være $-\infty$, ligesom b eventuelt kan være ∞ . Antag at F er kontinuert differentiabel på (a, b) . Da er den til F svarende fordeling kontinuert med sandsynlighedstæthed $p(x) = F'(x)$ for $x \in (a, b)$ og $p(x) = 0$ for $x \notin (a, b)$.*

Bevis: Definer en funktion p ved $p(x) = F'(x)$ for $x \in (a, b)$ og $p(x) = 0$ for $x \notin (a, b)$. Da en fordelingsfunktion iflg. afsnit 2.2 er svagt voksende, er $p(x) \geq 0$, og da (overvej tilfældet $-\infty < a$)

$$\int_{-n}^x p(y)dy = F(x) - F(-n)$$

for alle $n \in \mathbb{N}$ og for alle reelle $x \geq -n$, følger det af (2.2.4) ved at lade n gå mod uendelig, at F opfylder (5.1.8). Derfor følger sætningen af Sætning 5.1.5. □

Hvis a eller b er endelige tal, kan man definere $p(a)$ og $p(b)$ som man har lyst til.

Eksempel 5.1.1 (fortsat). I Eksempel 2.2.2 udledte vi fordelingsfunktionen

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{hvis } x > 0 \\ 0 & \text{hvis } x \leq 0 \end{cases}$$

for en fordeling, der er koncentreret på $(0, \infty)$. Da $F'(x) = \lambda e^{-\lambda x}$ for $x > 0$, er der tale om eksponentialfordelingen med parameter λ . □

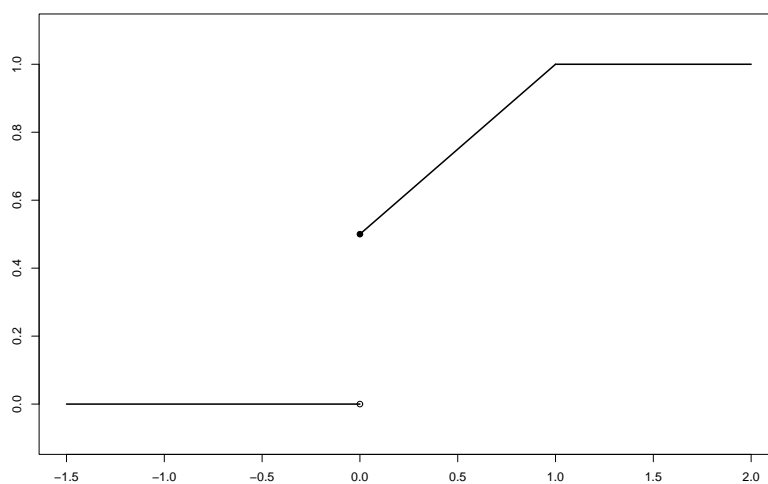
Man kan benytte fordelingsfunktionen til at beregne sandsynligheden for udfald i foreningsmængden af endeligt mange intervaller. Fremgangsmåden illustreres af følgende eksempel.

Eksempel 5.1.3 (fortsat). Lad den stokastiske variable X være ligefordelt på $[-1, 2]$. Da er

$$\begin{aligned} P\left(|X| \geq \frac{2}{3}\right) &= P\left(X \in \left[-1, -\frac{2}{3}\right] \cup \left[\frac{2}{3}, 2\right]\right) \\ &= \left(F\left(-\frac{2}{3}\right) - F(-1)\right) + \left(F(2) - F\left(\frac{2}{3}\right)\right) \\ &= \left(\frac{1}{9} - 0\right) + \left(1 - \frac{5}{9}\right) = \frac{5}{9}. \end{aligned}$$

Bemærk, at vi har brugt, at for eksempel $P\left(X = \frac{2}{3}\right) = 0$. □

Lad os slutte dette afsnit med at nævne, at der findes andre fordelinger på den reelle akse end de diskrete og de kontinuerte. Dels findes der nogle meget sære fordelinger, som er koncentreret på lige så sære delmængder af \mathbb{R} , dels findes der de mere medgørlige fordelinger af blandet type, som ikke sjældent optræder i praksis. Vi giver et eksempel på den sidste type, som vi derefter ikke vil beskæftige os yderligere med i disse noter.



Figur 5.1.5: Fordelingsfunktionen for den stokastiske variable Y i Eksempel 5.1.7.

Eksempel 5.1.7 Lad den stokastiske variable X være ligefordelt på $[-1, 1]$. Fordelingen af $Y = X \vee 0$ er af den nævnte blandede type. Halvdelen af sandsynlighedsmassen i denne fordeling er koncentreret i punktet 0, og den anden halvdel er jævnt fordelt over enhedsintervallet. Man kan naturligvis hverken opskrive en sandsynlighedsfunktion eller en sandsynlighedstæthed i dette tilfælde, men fordelingsfunktionen er veldefineret. For $y \in (0, 1]$ er $P(Y \leq y) = P(Y = 0) + P(0 < Y \leq y) = P(X \leq 0) + P(0 < X \leq y)$, så

$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ \frac{1}{2} + \frac{1}{2}y & \text{for } 0 \leq y \leq 1, \\ 1 & \text{for } 1 < y. \end{cases}$$

□

5.2 Middelværdi og varians

Også for mange kontinuerte fordelinger kan man definere størrelserne middelværdi og varians. Definitionen af en kontinuert fordeling ligner definitionen af en diskret fordeling bortset fra, at summen er erstattet med et integral og sandsynlighedsfunktionen med en sandsynlighedstæthed. På tilsvarende måde definerer vi middelværdi og varians for en kontinuert fordeling ved at erstatte summen med et integral og sandsynlighedsfunktionen med en sandsynlighedstæthed i definitionerne for en diskret fordeling. Da sandsynlighedstætheden angiver sandsynlighedsmasse per måleenhed, har middelværdi og varians for kontinuerte fordelinger samme fortolkning som for diskrete fordelinger. Resultater og beviser er helt parallelle til, hvad vi tidligere har set, men med summation erstattet med integration. Ligesom vi for diskrete fordelinger skulle sikre os, at de indgående summer konvergerer, må vi her passe lidt på med integralerne.

Definition 5.2.1 *Lad X være en kontinuert stokastisk variabel, som har sandsynlighedstæthed p . Da siges X at have middelværdi, hvis*

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty, \quad (5.2.1)$$

og vi definerer i så fald middelværdien af X som

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx < \infty. \quad (5.2.2)$$

Her er p defineret på hele \mathbb{R} . Hvis fordelingen af X er koncentreret på et delinterval I af den reelle akse, sætter vi som sædvanlig $p(x) = 0$, når $x \notin I$. I sådanne tilfælde, er det naturligt nok at integrere over intervallet I .

Hvis (5.2.1) ikke er opfyldt, siger man, at X ikke har middelværdi, eller at middelværdien ikke er defineret. Når vi kræver, at (5.2.1) er opfyldt, er det for at udelukke den ubehagelige situation, hvor

$$\int_{-\infty}^0 xp(x)dx = -\infty \quad \text{og} \quad \int_0^{\infty} xp(x)dx = \infty.$$

Hvis X er koncentreret på et interval $I \subseteq \mathbb{R}_+$, kan dette ikke ske, hvorfor man ofte i dette tilfælde tillader sig at skrive $E(X) = \infty$, når (5.2.1) ikke

er opfyldt. Vi kan i samme ånd generelt skrive betingelse (5.2.1) kort som $E(|X|) < \infty$.

Hvis X er en begrænset stokastisk variabel, kan man altid være sikker på at middelværdien eksisterer. Antag nemlig, at der findes et $c > 0$, så $P(X \in [-c, c]) = 1$. Da er

$$\int_{-\infty}^{\infty} |x|p(x)dx = \int_{-c}^c |x|p(x)dx \leq c \int_{-c}^c p(x)dx = c < \infty,$$

så X har middelværdi.

Eksempel 5.1.3 (fortsat). Middelværdien for ligefordelingen på $[a, b]$ er

$$\frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{1}{2}x^2 \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2},$$

altså midtpunktet af intervallet $[a, b]$, hvilket giver god mening. \square

Eksempel 5.1.1 (fortsat). Middelværdien for eksponentialfordelingen med parameter λ er λ^{-1} , da

$$\begin{aligned} \int_0^n x \lambda e^{-\lambda x} dx &= \left[x(-e^{-\lambda x}) \right]_0^n - \int_0^n (-e^{-\lambda x}) dx \\ &= -ne^{-\lambda n} + \lambda^{-1}(1 - e^{-\lambda n}) \rightarrow \lambda^{-1} \end{aligned}$$

for $n \rightarrow \infty$. \square

Eksempel 5.2.2 Funktionen $p(x) = \alpha x^{-(\alpha+1)}$ for $x > 1$, hvor $\alpha > 0$, er en sandsynlighedstæthed på $[1, \infty)$ (check dette). Den tilsvarende fordeling kaldes Pareto-fordelingen. Da

$$\int_1^n x \alpha x^{-(\alpha+1)} dx = \alpha \int_1^n x^{-\alpha} dx = \alpha \left[\frac{x^{-\alpha+1}}{-\alpha+1} \right]_1^n$$

for $\alpha \neq 1$, mens der for $\alpha = 1$ gælder, at

$$\int_1^n x x^{-2} dx = [\log(x)]_1^n,$$

ses det, at hvis $\alpha \leq 1$, har Pareto-fordelingen ikke middelværdi. For $\alpha > 1$ er middelværdien $\alpha/(\alpha-1)$. \square

Også for kontinuerte stokastiske variable kan man på enkel vis beregne middelværdien af en transformeret stokastisk variabel.

Sætning 5.2.3 *Lad X være en kontinuert stokastisk variabel, som er koncentreret på intervallet I , og som har sandsynlighedstæthed p . Lad endvidere t være en funktion fra I ind i \mathbb{R} . Da har den stokastiske variable $t(X)$ middelværdi hvis og kun hvis*

$$\int_I |t(x)|p(x)dx < \infty, \quad (5.2.3)$$

og i så fald er

$$E(t(X)) = \int_I t(x)p(x)dx. \quad (5.2.4)$$

Vi vil ikke vise denne sætning generelt. I Afsnit 5.4 vil den blive vist i det specialtilfælde, hvor afbildningen t er strengt monoton og kontinuert differentiabel. Bemærk, at vi ikke kan være sikre på, at fordelingen af $t(X)$ er kontinuert eller diskret, se Eksempel 5.1.7. Sætning 5.2.3 er i disse noter mest nyttig, når $t(X)$ har en fordeling af en type, for hvilken vi har defineret middelværdien, d.v.s. når $t(X)$ er kontinuert eller diskret. I andre tilfælde kan vi eventuelt tage (5.2.4) som definition af middelværdien.

Eksempel 5.1.4 (fortsat). Lad X være en stokastisk variabel, hvis fordeling har tæthed (5.1.5), og definer en ny stokastisk variabel ved $Y = 1/X$. Da

$$\int_0^1 x^{-1}\beta x^{\beta-1}dx = \beta \left[\frac{x^{\beta-1}}{\beta-1} \right]$$

for $\beta \neq 1$, og da

$$\int_0^1 x^{-1}dx = [\log(x)]_0^1,$$

ses det, at (5.2.3) kun er opfyldt, når $\beta > 1$. For $\beta \leq 1$ har Y ikke middelværdi. For $\beta > 1$ giver (5.2.4), at

$$E(Y) = \frac{\beta}{\beta-1}.$$

□

Sætning 5.2.4 *Lad X være en kontinuert stokastisk variabel, som er koncentreret på intervallet I . Lad t_1 og t_2 være funktioner fra I ind i \mathbb{R} , som begge opfylder (5.2.3). Da har den stokastiske variable $t_1(X) + t_2(X)$ middelværdi, og*

$$E(t_1(X) + t_2(X)) = E(t_1(X)) + E(t_2(X)). \quad (5.2.5)$$

Hvis det yderligere antages, at $t_1(x) \leq t_2(x)$ for alle $x \in I$, er

$$E(t_1(X)) \leq E(t_2(X)). \quad (5.2.6)$$

Bevis: Lad p betegne sandsynlighedstætheden for X . Da $|t_1(x) + t_2(x)| \leq |t_1(x)| + |t_2(x)|$, er

$$\int_I |t_1(x) + t_2(x)|p(x)dx \leq \int_I |t_1(x)|p(x)dx + \int_I |t_2(x)|p(x)dx < \infty.$$

Da funktionen $t_1(x) + t_2(x)$ altså opfylder (5.2.3), har $t_1(X) + t_2(X)$ ifølge Sætning 5.2.3 middelværdi givet ved

$$\begin{aligned} \int_I (t_1(x) + t_2(x))p(x)dx &= \int_I t_1(x)p(x)dx + \int_I t_2(x)p(x)dx \\ &= E(t_1(X)) + E(t_2(X)). \end{aligned}$$

Sætningens sidste påstand følger af en velkendt egenskab ved integralet. \square

Sætning 5.2.5 *Lad X være en kontinuert stokastisk variabel, som har middelværdi, og lad a og b være vilkårlige reelle tal. Da har også den stokastiske variable $a + bX$ middelværdi, og*

$$E(a + bX) = a + bE(X) \quad (5.2.7)$$

Bevis: Sætningen er en konsekvens af Sætning 5.2.4 med $t_1(x) = a$ og $t_2(x) = bx$. Lad p betegne sandsynlighedstætheden for X . Da X har middelværdi, er

$$\int_{-\infty}^{\infty} |bx|p(x)dx = |b| \int_{-\infty}^{\infty} |x|p(x)dx < \infty,$$

så t_2 opfylder (5.2.3). At t_1 opfylder (5.2.3) er klart. Ifølge Sætning 5.2.4 har $a + bX$ derfor middelværdi, som kan beregnes ved

$$\int_{-\infty}^{\infty} ap(x)dx + \int_{-\infty}^{\infty} bxp(x)dx = a \int_{-\infty}^{\infty} p(x)dx + b \int_{-\infty}^{\infty} xp(x)dx = a + bE(X).$$

□

Bemærk, at det følger af resultaterne i Sætningerne 5.2.4 og 5.2.5, at der som for diskrete fordelinger gælder, at $|E(X)| \leq E(|X|)$.

De øvrige resultater for middelværdi, som vi har vist for diskrete fordelinger, involverer mere end én stokastisk variabel. Da vi først vil definere kontinuerte stokastiske vektorer i Kapitel 6, må vi vente til denne definition er på plads, før vi kan vise de tilsvarende resultater om middelværdi for kontinuerte stokastiske variable.

Definition 5.2.6 *Lad X være en kontinuert stokastisk variabel, som har sandsynlighedstæthed p . Da siges X at have varians, hvis*

$$\int_{-\infty}^{\infty} x^2 p(x) dx < \infty, \quad (5.2.8)$$

og vi definerer i så fald variansen af X ved

$$\text{Var}(X) = E([X - E(X)]^2). \quad (5.2.9)$$

Bemærk, at betingelsen (5.2.8) ifølge Sætning 5.2.3 er ensbetydende med $E(X^2) < \infty$. Da $|x| \leq x^2 + 1$, medfører (5.2.8) at også (5.2.1) er opfyldt. D.v.s. at (5.2.8) også sikrer, at middelværdien eksisterer. At variansen er veldefineret indses nu på samme måde som for diskrete fordelinger (under anvendelse af Sætning 5.2.4).

Ved beregninger som i Kapitel 5 følger det af Sætningerne 5.2.4 og 5.2.5, at formlerne (3.7.9) og (3.7.10) også gælder for kontinuerte fordelinger. Som altid kaldes kvadratroden af variansen *spredningen* eller *standardafvigelsen*.

Eksempel 5.1.1 (fortsat). For en eksponentialfordelt stokastisk variabel X er

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ &= [x^2 (-e^{-\lambda x})]_0^{\infty} - \int_0^{\infty} 2x (-e^{-\lambda x}) dx \\ &= 0 + 2\lambda^{-1} E(X) = 2\lambda^{-2}, \end{aligned}$$

så eksponentialfordelingen har varians, og ifølge (3.7.9) er

$$\text{Var}(X) = E(X^2) - (E(X))^2 = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}.$$

Vi har her været lidt hurtigere end tidligere ved beregningen af integralet over intervallet $(0, \infty)$, men det er naturligvis stadig defineret som en grænseværdi. □

Eksempel 5.1.3 (fortsat). Variansen for ligefordelingen på $[a, b]$ er

$$\begin{aligned} \frac{1}{b-a} \int_a^b [x - (a+b)/2]^2 dx &= \frac{1}{b-a} \int_{-(b-a)/2}^{(b-a)/2} y^2 dy \\ &= \frac{2}{b-a} \int_0^{(b-a)/2} y^2 dy \\ &= \frac{2}{3(b-a)} \left(\frac{b-a}{2} \right)^3 = \frac{(b-a)^2}{12}. \end{aligned}$$

Spredningen er $(b-a)/(2\sqrt{3})$. □

Eksempel 5.2.2 (fortsat). For Pareto-fordelingen er

$$\int_1^\infty x^2 \alpha x^{-(\alpha+1)} dx = \alpha \int_1^\infty x^{-\alpha+1} dx = \alpha \left[\frac{x^{-\alpha+2}}{-\alpha+2} \right]_1^\infty$$

for $\alpha \neq 2$, og for $\alpha = 2$ fås

$$\int_1^\infty x^2 2x^{-3} dx = [\log(x)]_1^\infty = \infty.$$

Det ses derfor, at hvis $\alpha \leq 2$, har Pareto-fordelingen ikke varians. For $\alpha > 2$ er

$$\int_1^\infty x^2 \alpha x^{-(\alpha+1)} dx = \frac{\alpha}{\alpha-2},$$

så i det tilfælde er variansen ifølge (3.7.9) lig

$$\frac{\alpha}{\alpha-2} - \left(\frac{\alpha}{\alpha-1} \right)^2 = \frac{\alpha}{(\alpha-2)(\alpha-1)^2}.$$

□

Resultater om varians, som involverer flere kontinuerte stokastiske variable, og alt vedrørende kovarians og korrelation kan først behandles i kapitel 6.

5.3 Normalfordelingen

Den kontinuerte fordeling med sandsynlighedstæthed

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}, \quad (5.3.1)$$

kaldes *standard normalfordelingen*. Den kaldes også den normerede normalfordeling, Gauss-fordelingen og *u*-fordelingen.

Det er ikke ganske klart, at (5.3.1) er en sandsynlighedstæthed. At integralet af $\varphi(x)$ over \mathbb{R} er endeligt, kan ses på følgende måde

$$\begin{aligned} \int_{-n}^n e^{-x^2/2} dx &= 2 \int_0^n e^{-x^2/2} dx < 2 \left(1 + \int_1^n e^{-x^2/2} dx \right) \\ &< 2 + 2 \int_1^n e^{-x/2} dx = 2 + 4(e^{-\frac{1}{2}} - e^{-\frac{1}{2}n}). \end{aligned}$$

Da det sidste udtryk konvergerer, må også den voksende følge $\int_{-n}^n e^{-x^2/2} dx$ konvergere. Grænseværdien er lig integralet $\int_{-\infty}^{\infty} e^{-x^2/2} dx$. At integralet er lig $\sqrt{2\pi}$, således at φ er en sandsynlighedstæthed, vil vi ikke vise her. Det vil blive vist i Opgave 6.1 i Kapitel 6. Fordelingsfunktionen for standard normalfordelingen betegnes Φ , d.v.s.

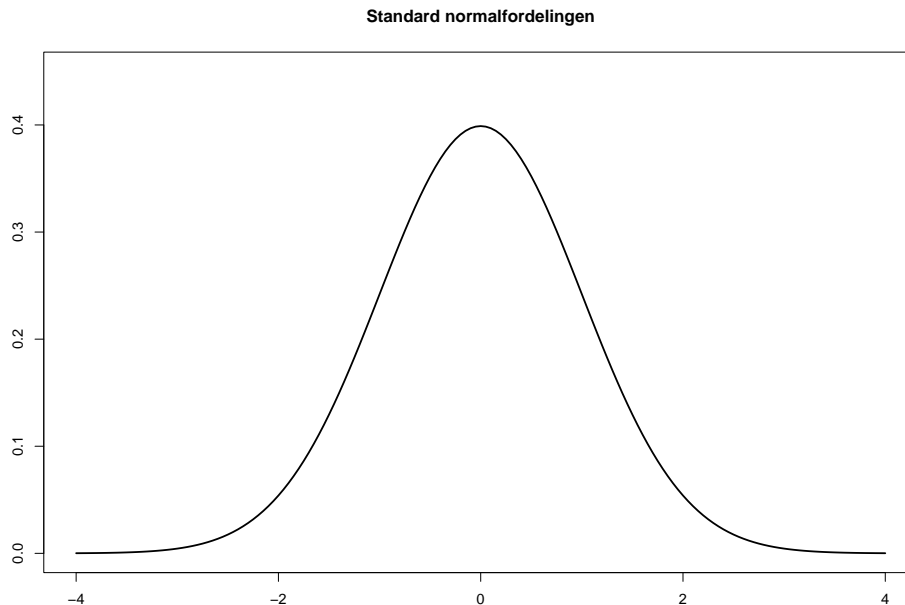
$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy. \quad (5.3.2)$$

Der findes ikke noget eksplicit udtryk for Φ ved simple funktioner, så værdierne af $\Phi(x)$ må findes enten ved tabelopslag eller ved hjælp af regnemaskine eller computer.

For standard normalfordelingen er $E(|X|^k) < \infty$ for alle $k \in \mathbb{N}$. Specielt eksisterer både middelværdi og varians. At

$$\int_{-\infty}^{\infty} |x|^k e^{-x^2/2} dx < \infty$$

kan for eksempel ses af, at man kan vise, at der findes et reelt tal $K > 0$, så $|x|^k e^{-x^2/2} < K e^{-x^2/4}$ for alle $x \in \mathbb{R}$. At $e^{-x^2/4}$ er integrabel, ses på samme måde, som vi brugte til at vise, at $e^{-x^2/2}$ er integrabel.



Figur 5.3.1: Standard normalfordelingens sandsynlighedstæthed.

Lad os så finde middelværdi og varians af standard normalfordelingen. Da φ er en lige funktion (d.v.s. $\varphi(-x) = \varphi(x)$), er middelværdien nul:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-x^2/2} dx &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^0 xe^{-x^2/2} dx + \int_0^{\infty} xe^{-x^2/2} dx \right] \\ &= \frac{1}{\sqrt{2\pi}} \left[- \int_0^{\infty} xe^{-x^2/2} dx + \int_0^{\infty} xe^{-x^2/2} dx \right] = 0. \end{aligned}$$

For at finde variansen bemærkes følgende. Differentiation af tætheden φ giver

$$\varphi'(x) = \frac{d}{dx} \left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) = -\frac{1}{\sqrt{2\pi}} xe^{-x^2/2},$$

og

$$\varphi''(x) = \frac{d}{dx} \left(-\frac{1}{\sqrt{2\pi}} xe^{-x^2/2} \right) = \frac{1}{\sqrt{2\pi}} (x^2 - 1)e^{-x^2/2}.$$

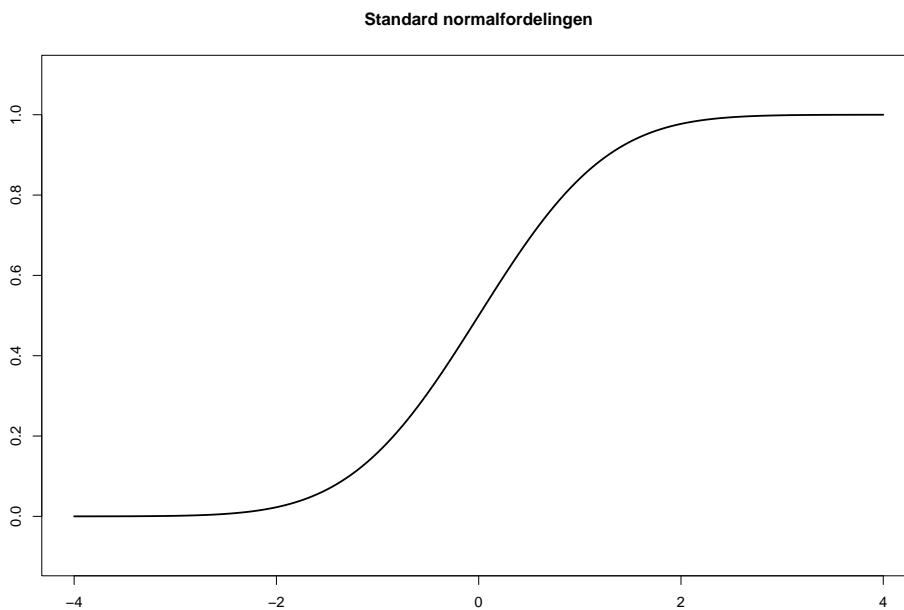
Da $\varphi'(n) \rightarrow 0$ for $n \rightarrow \pm\infty$, er

$$\begin{aligned} \int_{-\infty}^{\infty} (x^2 - 1) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx &= \int_{-\infty}^{\infty} \varphi''(x) dx = \lim_{n \rightarrow \infty} \int_{-n}^{+n} \varphi''(x) dx \\ &= \lim_{n \rightarrow \infty} (\varphi'(n) - \varphi'(-n)) = 0. \end{aligned}$$

Dermed er

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

hvor vi har brugt, at φ er en sandsynlighedstæthed. Da middelværdien er nul, er også variansen lig 1.



Figur 5.3.2: Standard normalfordelingens fordelingsfunktion.

Den generelle normalfordeling opnås ved at indføre en positionsparameter $\mu \in \mathbb{R}$ og en skalaparameter $\sigma > 0$. Hvis X er standard normalfordelt, er

$$Y = \mu + \sigma X \tag{5.3.3}$$

normalfordelt med middelværdi μ og varians σ^2 . At Y har denne middelværdi og varians ses af regneregler for middelværdi og varians, (5.2.7) og (3.7.10). Fordelingsfunktionen for normalfordelingen med middelværdi μ og varians σ^2 er

$$\begin{aligned} F(y) &= P(Y \leq y) = P(\mu + \sigma X \leq y) \\ &= P\left(X \leq \frac{y - \mu}{\sigma}\right) = \Phi\left(\frac{y - \mu}{\sigma}\right). \end{aligned} \quad (5.3.4)$$

Da Φ er kontinuert differentiabel, er F det også, og

$$F'(y) = \frac{1}{\sigma} \varphi\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right).$$

Ifølge Sætning 5.1.6 er sandsynlighedstætheden for normalfordelingen med middelværdi μ og varians σ^2 derfor lig

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right), \quad y \in \mathbb{R}. \quad (5.3.5)$$

Normalfordelingen er langt den vigtigste af de kontinuerte fordelinger. Dette skyldes især den *centrale grænseværdisætning*, som vi vil behandle i Kapitel 7. Dels kan denne sætning bruges til at argumentere for, at målefejl ofte er normalfordelte, dels er den baggrunden for, at normalfordelingen hyppigt optræder som approksimativ fordeling af estimatorer for parametre i statistiske modeller. Det skal også nævnes, at en række statistiske modeller, baseret på normalfordelingen, har en særligt simpel struktur, som har relation til n -dimensional Euklidisk geometri og lineær algebra. Disse *lineære normalfordelingsmodeller* er de vigtigste og mest anvendte blandt alle statistiske modeller. Vi skal se lidt på et simpelt eksempel i Afsnit 8.3.

5.4 Transformation af kontinuerte fordelinger på \mathbb{R}

Vi skal i dette afsnit behandle fordelingen af $t(X)$, hvor X er en kontinuert stokastisk variabel, og t er en reel funktion. Vi så i Eksempel 5.1.7,

at fordelingen af $t(X)$ ikke behøver at være kontinuert. Vi vil derfor lægge betingelser på funktionen t , som sikrer, at fordelingen af $t(X)$ er kontinuert.

Lad X være en kontinuert stokastisk variabel, som opfylder, at $P(X \in I) = 1$ for et interval $I \subseteq \mathbb{R}$. Intervallet I kan være endeligt eller uendeligt og lukket, åbent eller halvåbent. I det følgende betegner t en kontinuert, strengt monoton afbildning af I ind i \mathbb{R} . Da er $J = t(I)$ et interval. Definer venstre endepunkt v og højre endepunkt h for intervallet J ved

$$v = \inf J \quad \text{og} \quad h = \sup J.$$

Eventuelt kan v være $-\infty$, ligesom h eventuelt kan være ∞ . Med t^{-1} betegnes den inverse funktion til t , som eksisterer og er defineret på J , da t er antaget at være strengt monoton.

Definer nu en stokastisk variabel Y ved $Y = t(X)$. Det er klart, at $P(Y \in J) = 1$. Faktisk er $P(Y \in (v, h)) = 1$. Hvis nemlig f. eks. $v \in J$, er $P(Y = v) = P(X = t^{-1}(v)) = 0$.

Lad F_X betegne fordelingsfunktionen for X og F_Y fordelingsfunktionen for Y , d.v.s. $F_X(x) = P(X \leq x)$ og $F_Y(x) = P(Y \leq x)$. Lad videre a og b være højre og venstre endepunkt for intervallet I (dvs. $a = \inf I$ og $b = \sup I$). Der mindes om, at hvis X s sandsynlighedstæthed p er kontinuert på intervallet (a, b) , er F_X kontinuert differentiabel på (a, b) , og $F'_X(x) = p(x)$ for $x \in (a, b)$. Videre gælder ifølge Sætning 5.1.6, at hvis F_Y er kontinuert differentiabel på (v, h) , så er Y en kontinuert stokastisk variabel med sandsynlighedstæthed $q = F'_Y$.

Sætning 5.4.1 *Antag, at X er en kontinuert stokastisk variabel, som er koncentreret på intervallet I , og at X s sandsynlighedstæthed p er kontinuert på (a, b) . Lad den reelle funktion t være kontinuert differentiabel, og antag at $t'(x) \neq 0$ for alle $x \in (a, b)$. Da er $Y = t(X)$ en kontinuert stokastisk variabel med sandsynlighedstæthed q givet ved*

$$q(y) = \begin{cases} p(t^{-1}(y)) \left| \frac{d}{dy} t^{-1}(y) \right| & \text{hvis } y \in (v, h) \\ 0 & \text{hvis } y \notin (v, h). \end{cases} \quad (5.4.1)$$

Et alternativt udtryk for q er

$$q(y) = \begin{cases} p(t^{-1}(y))/|t'(t^{-1}(y))| & \text{hvis } y \in (v, h) \\ 0 & \text{hvis } y \notin (v, h). \end{cases} \quad (5.4.2)$$

Bevis: Bemærk først, at betingelsen, at $t'(x) \neq 0$ for alle $x \in (a, b)$, dels sikrer, at t er strengt monotont, dels at t^{-1} er kontinuert differentiabel med afledet

$$\frac{d}{dx}t^{-1}(x) = \frac{1}{t'(t^{-1}(x))}.$$

Dette viser, at de to udtryk for q er identiske, samt at $F_X(t^{-1}(x))$ er kontinuert differentiabel på (v, h) . Ifølge transformationssætningen for fordelingsfunktioner, Sætning 2.2.3, er fordelingsfunktionen for Y , $F_Y(y)$, enten lig $F_X(t^{-1}(y))$ eller lig $1 - F_X(t^{-1}(y))$ afhængigt af, om t er voksende eller aftagende. I begge tilfælde er F_Y altså kontinuert differentiabel på (v, h) , således at Y ifølge Sætning 5.1.6 er en kontinuert stokastisk variabel med sandsynlighedstæthed $q(y) = F'_Y(y)$ for $y \in (v, h)$. Hvis $t' > 0$, er

$$F'_Y(y) = \frac{d}{dy}F_X(t^{-1}(y)) = F'_X(t^{-1}(y)) \frac{d}{dy}t^{-1}(y) = \frac{p(t^{-1}(y))}{t'(t^{-1}(y))}$$

for alle $y \in (v, h)$. Tilsvarende er

$$F'_Y(y) = \frac{d}{dy}(1 - F_X(t^{-1}(y))) = -F'_X(t^{-1}(y)) \frac{d}{dy}t^{-1}(y) = \frac{p(t^{-1}(y))}{-t'(t^{-1}(y))}$$

for alle $y \in (v, h)$, hvis $t' < 0$. □

Man kan godt forstå resultatet i Sætning 5.4.1 intuitivt. Hvis nemlig $h > 0$ er et lille tal, er $t(x+h) \simeq t(x) + t'(x)h$, så hvis $y = t(x)$, og hvis t er voksende, er

$$\begin{aligned} P(Y \in [y, y + t'(x)h]) &\simeq P(t(X) \in [t(x), t(x+h)]) \\ &= P(X \in [x, x+h]) \simeq p(x)h = p(t^{-1}(y))h. \end{aligned}$$

Dermed er

$$q(y) \simeq \frac{P(Y \in [y, y + t'(x)h])}{t'(x)h} \simeq \frac{p(t^{-1}(y))}{t'(t^{-1}(y))}.$$

Den sandsynlighedsmasse, der oprindeligt var i intervallet $[x, x+h]$ flyttes ved t over i et interval, hvis længde er $t'(x)h$. Det er grunden til at $t'(x)$ optræder i formelen for q .

Eksempel 5.4.2 (Positions- og skalaparameter). Hvis X har sandsynlighedstæthed p vil $Y = a + bX$, hvor $b \neq 0$, have tæthed

$$q(y) = \frac{1}{|b|} p\left(\frac{y-a}{b}\right). \quad (5.4.3)$$

Fordelingen af Y siges her at være fremkommet af X 's fordeling ved at forsyne denne med *positionsparameteren* a og *skalaparameteren* b . Fordelingerne af X og Y siges at være af samme *type*. Når man taler om positions- og skalaparameter for en fordeling, er det altid med reference til en normeret fordeling af samme type. Hvis f. eks. Y er eksponentialfordelt med positionsparameter a og skalaparameter b betyder det, at Y har tæthed

$$q(y) = \frac{1}{|b|} e^{-(y-a)/b} 1_{(0,\infty)}((y-a)/b),$$

med reference til eksponentialfordelingen med parameter 1, der har tæthed e^{-x} på den positive halvakse. Bemærk at denne fordeling er koncentreret på intervallet (a, ∞) . Eksponentialfordelingen med parameter λ er altså den samme som eksponentialfordelingen med skalaparameter λ^{-1} . Bemærk også, at hvis X er standard normalfordelt, så er Y normalfordelt med middelværdi a og varians b^2 , og at (5.3.5) er i overensstemmelse med (5.4.3).

□

Eksempel 5.4.3 Lad X være eksponentialfordelt med parameter 1, og betragt den stokastiske variabel $Y = X^2$. Da funktionen $t(x) = x^2$ har den afledede $t'(x) = 2x$ og for $y > 0$ har den inverse $t^{-1}(y) = \sqrt{y}$, er tætheden for Y ifølge Sætningen 5.4.1

$$q(y) = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}, \quad y > 0.$$

□

Eksempel 5.4.4 Lad X være en kontinuert stokastisk variabel, der er koncentreret på det åbne interval I (det er ikke nogen indskrænkning at antage, at I er åbent). Antag at X s tæthed p er kontinuert på I og opfylder, at $p(x) > 0$ for alle $x \in I$. Lad F betegne fordelingsfunktionen for X , og sæt $t = F$. Da opfylder t betingelserne i Sætning 5.4.1, og da $F' = p$, ses det af (5.4.2), at $q(y) = p(F^{-1}(y))/p(F^{-1}(y)) = 1$ for $y \in (0, 1)$. Den stokastiske variable $F(X)$ er således ligefordelt på intervallet $(0, 1)$. Dette resultat bruges somme tider til at undersøge, om givne data kan antages at være observationer af en stokastisk variabel med fordelingsfunktion F .

Også den inverse funktion F^{-1} opfylder betingelserne i Sætning 5.4.1. Dette kan bruges til at konstruere en stokastisk variabel med fordeling F ud fra en stokastisk variabel R , som er ligefordelt på $(0, 1)$, ved

$$Z = F^{-1}(R).$$

Da den inverse til F^{-1} er F , og da $F' = p$, følger det af (5.4.1) at $q(z) = 1_{(0,1)}(F(z))p(z) = p(z)$ for $z \in I$. Altså er Z koncentreret på I med tæthed p . Det er let at generere uafhængige stokastiske variable R_1, \dots, R_n , der er ligefordelte på $(0, 1)$. Det kan enhver respektabel statistikprogrampakke og mange lommeregnerne gøre. Hvis man kan beregne $F^{-1}(R_i)$, enten ved at bruge et eksplicit udtryk eller ved en numerisk procedure, kan man derfor frembringe uafhængige udfald $Z_i = F^{-1}(R_i)$, $i = 1, \dots, n$, af en stokastisk variabel med fordelingsfunktion F . Dette har man jævnligt brug for i praksis. Det kaldes at simulere udfald af den stokastiske variable.

Fordelingsfunktionen for eksponentialfordelingen er $1 - e^{-\lambda x}$, $x > 0$. Derfor er $-\lambda^{-1} \log(1 - R)$ eksponentialfordelt med parameter λ . Det er ikke svært at indse (f. eks. ved hjælp af Sætning 5.4.1), at hvis R er ligefordelt på $(0, 1)$, så har $1 - R$ den samme fordeling. Derfor er også

$$-\lambda^{-1} \log(R)$$

eksponentialfordelt med parameter λ .

□

Eksempel 5.4.5 Lad os betragte udsendelse af α -partikler fra et radioaktivt materiale. Lidt fra det radioaktive materiale opsættes en skærm, som kan registrere, når α -partikler rammer den. Vi tænker os, at der er

tale om en ganske lille klump af materialet, som kan antages at ligge i et punkt. Vi ser på opstillingen ovenfra, så skærmen ses som et ret linie. For at lette forklaringen indlægges et retvinklet koordinatsystem, hvis y -akse følger skærmen. Vi kan antage, at det radioaktive materiale ligger i punktet $(-1, 0)$.

Den retning, som en partikel, der rammer skærmen, udsendes i, kan måles ved vinklen i forhold til x -aksen. Denne vinkel er en stokastisk variabel, som vi betegner med X . Da alle retninger er lige sandsynlige, er X ligefordelt på $(-\frac{\pi}{2}, \frac{\pi}{2})$. Vi vil nu finde fordelingen af det punkt på y -aksen, som rammes af α -partiklen; altså fordelingen af $Y = \tan(X)$.

Da sandsynlighedstætheden for X er $\frac{1}{\pi}1_{(-\frac{\pi}{2}, \frac{\pi}{2})}(x)$, og da den afledede af $\tan^{-1}(y)$ er $1/(1+y^2)$, følger det af (5.4.1), at sandsynlighedstætheden for Y er

$$q(y) = \frac{1}{\pi(1+y^2)}, \quad y \in \mathbb{R}. \quad (5.4.4)$$

Fordelingen med denne tæthed kaldes *Cauchy-fordelingen*.

Cauchy-fordelingen har ikke middelværdi (og derfor heller ikke varians). Dette følger af at

$$\begin{aligned} \int_{-n}^n |x| \frac{1}{\pi(1+x^2)} dx &\geq \frac{1}{\pi} \int_1^n \frac{x}{1+x^2} dx \\ &\geq \frac{1}{\pi} \int_1^n \frac{x}{2x^2} dx = \frac{1}{2\pi} \int_1^n \frac{1}{x} dx = \frac{1}{2\pi} \log(n). \end{aligned}$$

□

Vi kan nu **bevise Sætning 5.2.3** i det tilfælde, hvor den stokastiske variable X og transformationen t er som i Sætning 5.4.1. Vi nøjes med beviset, hvor t er voksende. Læseren kan let selv klare det aftagende tilfælde. Da sandsynlighedstætheden for fordelingen af $t(X)$ er givet ved (5.4.1), har $t(X)$ middelværdi, når følgende integral er endeligt:

$$\int_v^h |y| p(t^{-1}(y)) \frac{d}{dy} t^{-1}(y) dy = \int_a^b |t(x)| p(x) dx,$$

hvor vi har skiftet integrationsvariabel til $x = t^{-1}(y)$. Heraf ses, at $t(X)$ har middelværdi hvis og kun hvis (5.2.3) holder. At middelværdien kan beregnes ved (5.2.4) følger ved at foretage det samme variabelskift i integralet uden numerisk tegn.

Eksempel 5.4.6 Lad X være en stokastisk variabel, hvis mulige værdier er hele den reelle akse, og antag, at man har brug for at finde fordelingen af $Z = X^2$. I denne situation, som jævnligt forekommer i statistik, kan Sætning 5.4.1 ikke anvendes, da funktionen $x \mapsto x^2$ jo ikke er monoton på hele \mathbb{R} . Alligevel er det ikke vanskeligt at finde sandsynlighedstætheden for X^2 . Man skal blot, i lighed med beviset for Sætning 5.4.1, gå via fordelingsfunktionen.

Lad altså p betegne sandsynlighedstætheden for X . Vi antager, at denne er kontinuert på hele \mathbb{R} . Lad endvidere F_X og F_Z betegne fordelingsfunktionerne for henholdsvis X og Z . For ethvert $z > 0$ er

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X^2 \leq z) \\ &= P(-\sqrt{z} \leq X \leq \sqrt{z}) = F_X(\sqrt{z}) - F_X(-\sqrt{z}). \end{aligned}$$

Da $Z \geq 0$, og da F_Z er kontinuert differentiabel på $(0, \infty)$, er Z ifølge Sætning 5.1.6 en kontinuert stokastisk variabel med sandsynlighedstæthed

$$q(z) = F'_Z(z) = \frac{d}{dz} F_X(\sqrt{z}) - \frac{d}{dz} F_X(-\sqrt{z}) = \frac{p(\sqrt{z}) + p(-\sqrt{z})}{2\sqrt{z}}$$

for $z > 0$. For $z \leq 0$ er sandsynlighedstætheden naturligvis lig med nul.

Lad os til slut betragte det specialtilfælde, hvor X er standard normalfordelt. Da er $p(x) = e^{-x^2/2}/\sqrt{2\pi}$, så det er ikke vanskeligt at se, at tætheden for X^2 er

$$q(z) = \begin{cases} e^{-z/2}/\sqrt{2\pi z} & \text{hvis } z > 0 \\ 0 & \text{hvis } z \leq 0. \end{cases} \quad (5.4.5)$$

Denne fordeling kaldes χ^2 -fordelingen med en frihedsgrad. Vi vil vende tilbage til χ^2 -fordelingerne i Kapitel 8.

□

Man kommer af og til i den situation, at man gerne vil finde fordelingen af en ikke-monoton transformation af en kontinuert stokastisk variabel. Sætning 5.4.1 kan altså ikke anvendes, men man kan ofte løse problemet ved at gå via fordelingsfunktionen i lighed med, hvad der blev gjort i Eksempel 5.4.6.

For at minde om, at der findes mange typer af transformationer, hvor metoden i Sætning 5.4.1 og dens variationer, så som Eksempel 5.4.6, ikke kan anvendes, slutter vi med et eksempel på en transformation af en kontinuert stokastisk variabel, hvor man opnår en diskret fordeling.

Eksempel 5.4.7 Antag, at X er eksponentialfordelt med parameter λ , og definer en funktion $t : (0, \infty) \mapsto \mathbb{N}_0$ ved $t(x) = [x]$, hvor $[x]$ betegner heltalsdelen af x , altså det største hele tal, der er mindre end eller lig x . Den stokastiske variable $Y = t(X)$ er koncentreret på \mathbb{N}_0 . For $y \in \mathbb{N}_0$ er

$$\begin{aligned} P(Y = y) &= P([X] = y) = P(y \leq X < y + 1) \\ &= \int_y^{y+1} \lambda e^{-\lambda x} dx = (e^{-\lambda})^y (1 - e^{-\lambda}), \end{aligned}$$

d.v.s. Y er geometrisk fordelt med parameter $1 - e^{-\lambda}$. Når man tænker på, hvordan vi udledte eksponentialfordelingen i Eksempel 5.1.1 som ventetiden på en tilfældigt ankommende begivenhed (se også Eksempel 4.1.1), og sammenligner dette med definitionen af den geometriske fordeling, er resultatet ikke helt overraskende.

□

5.5 Sammenfatning

I Kapitel 5 har vi indført og studeret en-dimensionale kontinuerte fordelinger. Det er fordelinger, hvor sandsynligheden er fordelt så jævnt ud over den reelle akse eller et delinterval af denne, at fordelingen ikke kan beskrives ved en sandsynlighedsfunktion, som vi gjorde det for de diskrete fordelinger. I stedet benyttes en sandsynlighedstæthed: Sandsynligheden for at få et udfald i en delmængde A af \mathbb{R} beregnes som integralet af sandsynlighedstætheden over A . Sandsynlighedstætheden kan (i sine kontinuitetspunkter) fortolkes som sandsynlighed per længdeenhed. Definitioner og mange resultater for kontinuerte fordelinger minder om dem, vi har set for diskrete fordelinger, idet dog alle summer er erstattet med integraler. Det gælder ikke mindst resultaterne om middelværdi og varians.

Et centralt resultat i dette kapitel er transformationsætningen for en-dimensionale kontinuerte fordelinger, som anvendes til at beregne fordelingen af en funktion af en grundliggende stokastisk variabel, hvis fordeling man kender. Sætningen kan vises ved at udnytte, at man som regel kan finde sandsynlighedstætheden for en kontinuert fordeling ved at differentiere dens fordelingsfunktion.

Vi har også indført nogle nye fordelinger. De vigtigste er normalfordelingen, eksponentialfordelingen og ligefordelingen på et begrænset interval. Endvidere er vi stødt på Paretofordelingen og Cauchy fordelingen, som begge anvendes i situationer, hvor man har behov for en model, som tillægger (numerisk) meget store udfald en større sandsynlighed end de har for en normalfordeling eller en eksponentialfordeling. I de følgende opgaver dukker endnu et par fordelinger op: Laplacefordelingen, arcsinus fordelingen og den logaritmiske normalfordeling.

5.6 Opgaver

- 5.1 Antag at den stokastiske variable X er eksponentialfordelt med parameter λ . Find $P(X > x)$ for alle $x > 0$. Find for $\lambda = 1$ sandsynligheden $P(1 < X < 2)$.
- 5.2 Antag at X er en kontinuert stokastisk variabel på $(1, \infty)$ med sandsynlighedstæthed $p(x) = \alpha x^{-(\alpha+1)}$ for $x > 1$, hvor $\alpha > 0$ (se Eksempel 5.2.2). Find fordelingsfunktionen for X .
- 5.3 Lad X være en stokastisk variabel med fordelingsfunktion

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x/3 & \text{for } 0 < x \leq 1 \\ (2x - 1)/3 & \text{for } 1 < x \leq 2 \\ 1 & \text{for } x > 2. \end{cases}$$

Find $P(\frac{1}{2} < X < 1)$, $P(1 \leq X < \frac{3}{2})$ og $P(\frac{2}{3} < X \leq \frac{4}{3})$. Gør rede for, at X er kontinuert, og find sandsynlighedstætheden for X .

5.4 Definer en fordelingsfunktion F ved

$$F(x) = \frac{1}{2} + \frac{x}{2(|x| + 1)}, \quad x \in \mathbb{R}.$$

Gør rede for, at den tilsvarende fordeling er kontinuert og angiv sandsynlighedstætheden.

5.5 (a) Vis, at

$$p(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R},$$

er en sandsynlighedstæthed på \mathbb{R} . Den tilsvarende fordeling kaldes *Laplace-fordelingen* eller den tosidede eksponentialfordeling.

(b) Find fordelingsfunktionen.

(c) Vis, at fordelingen har middelværdi og varians, og find disse størrelser.

5.6 (a) Vis, at

$$p(x) = \frac{1}{\pi\sqrt{1-x^2}} \quad x \in (-1, 1),$$

er en sandsynlighedstæthed på intervallet $(-1, 1)$. Vink: Det er en god idé at skifte integrationsvariabel til v givet ved $x = \sin(v)$. Den tilsvarende fordeling kaldes *arcus-sinus fordelingen*.

(b) Find fordelingsfunktionen. Fordelingsfunktionen forklarer fordelingsnavn.

(c) Beregn fordelingsmiddelværdi og varians.

5.7 Vis, at fordelingen med sandsynlighedstæthed givet ved (5.1.5) har middelværdi $\beta/(\beta + 1)$. Hvad er variansen?

5.8 Lad X være exponentialfordelt med parameter λ . Hvad er

(a) $E(X^3)$?

(b) $E(\exp(-aX))$?

5.9 Sæt

$$I_n = \int_0^\infty x^n e^{-x} dx,$$

hvor $n \in \mathbb{N}_0$.

(a) Overvej, at integralet er endeligt.

(b) Vis, at $I_n = n!$ Vink: Vis, at $I_n = nI_{n-1}$.

Funktionen $p_n : [0, \infty) \mapsto \mathbb{R}$ givet ved $p_n(x) = \frac{1}{n!}x^n e^{-x}$ er således en sandsynlighedstæthed. Den tilsvarende fordeling kaldes en *Erlang fordeling* efter en dansk matematiker, se også Opgave 6.7. Denne fordeling er et specialtilfælde af Γ -fordelingerne, som vil blive behandlet grundigt i Kapitel 8.

(c) Udregn middelværdi og varians af fordelingen med sandsynlighedstæthed p_n .

5.10 Vis, at funktionen

$$p(x) = \begin{cases} 2(x+1)^{-3} & \text{hvis } x > 0 \\ 0 & \text{hvis } x \leq 0 \end{cases}$$

er en sandsynlighedstæthed, og find den tilsvarende fordelingsfunktion. Undersøg, om fordelingen har middelværdi og varians, og angiv i bekræftende fald disse størrelser.

5.11 Vis den sidste påstand i Sætning 5.2.4.

5.12 (a) Antag at X er standard normalfordelt. Vis, at $E(X^3) = 0$ og at $E(X^4) = 3$. Vink til $E(X^4)$: Ved hjælp af partiel integration kan integralet $\int_{-\infty}^{\infty} x^4 \varphi(x) dx$ udtrykkes ved $E(X^2)$.

(b) Antag at Y er normalfordelt med middelværdi μ og varians σ^2 . Bestem $E(Y^3)$ og $E(Y^4)$. Vink: Binomialformlen kan være nyttig.

5.13 Lad X være en kontinuert stokastisk variabel, som er koncentreret på et interval (a, b) , og antag at X 's sandsynlighedstæthed p er kontinuert på (a, b) . Find sandsynlighedstætheden for

(a) $\exp(X)$.

Antag nu, at $a \geq 0$, og find tætheden for fordelingen af

- (b) \sqrt{X}
- (c) $1/X$
- (d) X^2 .

- (e) Antag til slut, at X kan antage værdier på hele den reelle akse fraregnet nul. Hvad er da tætheden for fordelingen af $1/X$? Hvad er svaret, hvis også nul er en mulig værdi for X ?
- 5.14 Lad X være en kontinuert stokastisk variabel med sandsynlighedstæthed (5.1.5). Hvad er tætheden for fordelingen af $1/X$. Diskuter derefter, hvordan Eksempel 5.2.2 og den del af Eksempel 5.1.4, som står lige efter Sætning 5.2.3, hænger sammen.
- 5.15 Lad X være normalfordelt med middelværdi μ og varians σ^2 . Fordelingen af $Y = \exp(X)$ kaldes den *logaritmiske normalfordeling* med parametre (μ, σ^2) .
- (a) Find sandsynlighedstætheden for fordelingen af Y .
- (b) Vis, at hvis Y er logaritmisk normalfordelt, så er også βY logaritmisk normalfordelt for ethvert $\beta > 0$. En klasse af fordelinger med denne egenskab kaldes skalainvariant. Opskriv den formel, som viser, hvorledes den logaritmiske normalfordelings parametre ændrer sig ved en sådan skalatransformation.
- (c) Vis, at middelværdien i den logaritmiske normalfordeling med parametre $\mu = 0$ og $\sigma^2 = 1$ er $\sqrt{e} = 1.6487$.
- 5.16 (a) Vis, at fordelingsfunktionen for Cauchy-fordelingen er
- $$\frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x).$$
- Vink: Resultatet vises lettest ved at bruge den måde, som Cauchy-fordelingen blev udledt på i Eksempel 5.4.5.
- (b) Hvis X er Cauchy-fordelt, hvad er så fordelingen af X^2 og $1/X$?
- 5.17 Antag, at X er ligefordelt på intervallet $(0, 1)$. Find fordelingen af $Y = -\log(X)$.
- 5.18 Betragt en kontinuert fordeling på $I \subseteq \mathbb{R}$ (I kan selvfølgelig godt være hele \mathbb{R}) med sandsynlighedstæthed p , som opfylder, at $p(x) > 0$ for alle $x \in I$. Lad F være den tilsvarende fordelingsfunktion. For ethvert $\alpha \in (0, 1)$ defineres fordelingsens α -fraktile ved $F^{-1}(\alpha)$. Traditionelt angives α i procent, således at f. eks. 50%-fraktilen (som også kaldes *medianen*) er det tal x , for hvilket $F(x) = \frac{1}{2}$.

- (a) Find 5%-fraktilen, medianen og 95%-fraktilen i eksponentialfordelingen med parameter $\lambda = 1$.
- (b) Find 5%-fraktilen, medianen og 95%-fraktilen i normalfordelingen.
- (c) Find 5%-fraktilen, medianen og 95%-fraktilen i arcus-sinus fordelingen, som blev defineret i Opgave 5.6.
- (d) Find 5%-fraktilen, medianen og 95%-fraktilen i Cauchy-fordelingen. Vink: Se Opgave 5.16.
- (e) Hvilke regler gælder for fraktiler i en transformeret fordeling? Specielt, hvordan afhænger fraktilerne i en fordeling med positions- og skalaparameter af disse parametre?
- 5.19 Antag, at X s fordeling har sandsynlighedstæthed p på \mathbb{R} , og at p er en kontinuert funktion. Hvad er sandsynlighedstætheden for fordelingen af $|X|$?
- 5.20 Lad X være en kontinuert stokastisk variabel med kontinuert sandsynlighedstæthed p . Vis, at $Y = X^n$, hvor $n \in \mathbb{N}$, har sandsynlighedstæthed
- $$q(y) = \begin{cases} \frac{p(-y^{n^{-1}}) + p(y^{n^{-1}})}{ny^{1-n^{-1}}} & \text{for } y > 0 \\ 0 & \text{for } y \leq 0, \end{cases}$$
- når n er lige, og
- $$q(y) = \frac{p(y^{n^{-1}})}{n|y|^{1-n^{-1}}} \quad \text{for } y \neq 0,$$
- når n er ulige.
- 5.21 Lad X være ligefordelt i $[0, 2\pi]$, og sæt $Y = \cos(X)$. Vis, at Y er arcus-sinus fordelt (se Opgave 5.6).
- 5.22 Inge stiller sin cykel fra sig efter en længere tur. Hvad er fordelingen af baghjulsventilens højde over jorden.

5.23 Med $[x]$ betegnes som sædvanligt heltalsdelen af det reelle tal x , d.v.s. det største hele tal som er mindre end eller lig x . Lad X være ligefordelt på $(0, 1)$, og definer diskrete stokastiske variable X_1, X_2, \dots, X_n med værdier i $\{0, 1\}$ ved

$$\begin{aligned}X_1 &= [2X] \\X_2 &= [4X - 2X_1] \\X_3 &= [8X - 4X_1 - 2X_2] \\X_4 &= [16X - 8X_1 - 4X_2 - 2X_3] \\&\vdots\end{aligned}$$

Vis at X_1, X_2, \dots, X_n er uafhængige, ligefordelte på $\{0, 1\}$ Vink: Bemærk at der i totalssystemet gælder $X = 0.X_1X_2X_3\dots X_n$, på nær afrunding nedad til n betydende cifre. Bemærk at opgaven viser, at man kan simulere et vilkårligt antal møntkast ud fra én ligefordelt stokastisk variabel.

Kapitel 6

Fler-dimensionale kontinuerte fordelinger

I praksis er det sjældent nok at betragte en enkelt stokastisk variabel af gangen. For eksempel betragter man i statistik som regel mange observationer, der antages at være et udfald af hver sin stokastiske variable. Vi er derfor tvunget til at definere kontinuerte fordelinger på \mathbb{R}^n og kontinuerte stokastiske vektorer, hvilket teknisk er lidt sværere end fordelinger på \mathbb{R} , da vi bliver nødt til at integrere i \mathbb{R}^n .

6.1 Tætheder og fordelinger på \mathbb{R}^n

Som for kontinuerte fordelinger på \mathbb{R} er sandsynlighedstætheden et centralt begreb.

Definition 6.1.1 *En funktion p fra en delmængde $B \subseteq \mathbb{R}^n$ ind i $[0, \infty)$, som opfylder, at*

$$\int_B p(x_1, \dots, x_n) dx_1 \cdots dx_n = 1, \quad (6.1.1)$$

kaldes en sandsynlighedstæthed på B .

Integration i \mathbb{R}^n diskuteres i Appendix D.

Det til en sandsynlighedstæthed p hørende sandsynlighedsmål på B er givet ved

$$P(A) = \int_B 1_A(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

hvor $A \subseteq B$. At P er et sandsynlighedsmål vises på helt samme måde, som det i Kapitel 5 blev gjort for den reelle akse. En fordeling af denne type kaldes en *kontinuert fordeling*, og en stokastisk vektor (X_1, \dots, X_n) , hvis fordeling er kontinuert, kaldes en *kontinuert stokastisk vektor*. Hvis X s fordeling har sandsynlighedstæthed p er altså

$$P((X_1, \dots, X_n) \in A) = \int_B 1_A(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Som diskuteret tidligere, kan en fordeling, der er koncentreret på $B \subseteq \mathbb{R}^n$, opfattes som en fordeling på hele \mathbb{R}^n , ligesom en sandsynlighedstæthed på B kan defineres på hele \mathbb{R}^n ved at sætte dens værdi lig nul udenfor B .

Eksempel 6.1.2 En naturlig model for udtrækning af et tilfældigt punkt i $[0, 1] \times [0, 1]$ er sandsynlighedsmålet givet ved

$$P(A) = |A|$$

for $A \subseteq [0, 1] \times [0, 1]$, hvor $|A|$ betegner arealet af A . At dette er et sandsynlighedsmål, er ikke svært at vise. Det er det kontinuerte sandsynlighedsmål på $[0, 1] \times [0, 1]$ med sandsynlighedstætheden

$$p(x, y) = 1_{[0,1] \times [0,1]}(x, y).$$

Fordelingen kaldes *ligefordelingen* på $[0, 1] \times [0, 1]$. Vi kan generelt definere ligefordelingen på en begrænset delmængde B af \mathbb{R}^n som fordelingen med sandsynlighedstætheden

$$p(x_1, \dots, x_n) = \frac{1_B(x_1, \dots, x_n)}{|B|}, \quad (6.1.2)$$

hvor $|B|$ betegner det n -dimensionale volumen af B , se (D.2.6) i Appendix D.

□

Ligesom for en-dimensionale kontinuerte fordelinger, kan man i det fler-dimensionale tilfælde give en fortolkning af værdien af sandsynlighedstætheden i dens kontinuitetspunkter. Antag nemlig, at p er kontinuert i punktet (x_1, \dots, x_n) . Da kan man for ethvert givet $\epsilon > 0$ finde et $\delta > 0$, så

$$p(x_1, \dots, x_n) - \epsilon \leq p(\tilde{x}_1, \dots, \tilde{x}_n) \leq p(x_1, \dots, x_n) + \epsilon \quad (6.1.3)$$

for alle $(\tilde{x}_1, \dots, \tilde{x}_n) \in A$, hvor

$$A = [x_1, x_1 + \delta] \times \dots \times [x_n, x_n + \delta].$$

Ved overalt i (6.1.3) at gange med $1_A(\tilde{x}_1, \dots, \tilde{x}_n)$ og integrere med hensyn til $(\tilde{x}_1, \dots, \tilde{x}_n)$ finder vi, at

$$\begin{aligned} (p(x_1, \dots, x_n) - \epsilon)|A| &= \int_{\mathbb{R}^n} (p(x_1, \dots, x_n) - \epsilon) 1_A(\tilde{x}_1, \dots, \tilde{x}_n) d\tilde{x}_1 \cdots d\tilde{x}_n \\ &\leq \int_{\mathbb{R}^n} 1_A(\tilde{x}_1, \dots, \tilde{x}_n) p(\tilde{x}_1, \dots, \tilde{x}_n) d\tilde{x}_1 \cdots d\tilde{x}_n \\ &\leq \int_{\mathbb{R}^n} (p(x_1, \dots, x_n) + \epsilon) 1_A(\tilde{x}_1, \dots, \tilde{x}_n) d\tilde{x}_1 \cdots d\tilde{x}_n \\ &= (p(x_1, \dots, x_n) + \epsilon)|A|, \end{aligned}$$

hvor $|A|$ betegner det n -dimensionale volumen af A , se (D.2.6) Da $P(A) = \int_{\mathbb{R}^n} 1_A(\tilde{x}_1, \dots, \tilde{x}_n) p(\tilde{x}_1, \dots, \tilde{x}_n) d\tilde{x}_1 \cdots d\tilde{x}_n$, ser vi, at

$$p(x_1, \dots, x_n) \simeq P(A)/|A|. \quad (6.1.4)$$

Sandsynlighedstæthedens værdi i punktet (x_1, \dots, x_n) angiver altså, hvor megen sandsynlighedsmasse der nær dette punkt er per volumenenhed (når $n = 2$ er det per arealenhed).

Hvis man kender sandsynlighedstætheden for en kontinuert stokastisk vektor (X_1, \dots, X_n) , er det relativt let at finde sandsynlighedstætheden for den *marginale fordeling* af den stokastiske vektor, som fremkommer ved kun at betragte en delmængde af koordinaterne. For nemheds skyld kan vi nøjes med at finde fordelingen af (X_1, \dots, X_k) , hvor $k < n$. Lad os først betragte tilfældet $n = 2$ og $k = 1$, hvor det hele er lidt mere overskueligt. Lad derfor (X, Y) være en to-dimensional kontinuert stokastisk

vektor med sandsynlighedstæthed $p(x, y)$. For $A \subseteq \mathbb{R}$ er

$$\begin{aligned} P(X \in A) &= P((X, Y) \in A \times \mathbb{R}) \\ &= \int_{\mathbb{R}^2} 1_{A \times \mathbb{R}}(x, y) p(x, y) dx dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_{A \times \mathbb{R}}(x, y) p(x, y) dy \right) dx \\ &= \int_{\mathbb{R}} 1_A(x) \left(\int_{\mathbb{R}} p(x, y) dy \right) dx, \end{aligned}$$

hvoraf det ses, at X er en kontinuert stokastisk variabel med sandsynlighedstæthed

$$q(x) = \int_{\mathbb{R}} p(x, y) dy. \quad (6.1.5)$$

Man integrerer simpelthen med hensyn til den overflødige koordinat y .

Vi vender os derefter til den generelle situation.

Sætning 6.1.3 *Lad (X_1, \dots, X_n) være en kontinuert stokastisk vektor med sandsynlighedstæthed $p(x_1, \dots, x_n)$. Den stokastiske vektor (X_1, \dots, X_k) , hvor $k < n$, er da kontinuert med sandsynlighedstæthed*

$$q(x_1, \dots, x_k) = \int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_k, y_{k+1}, \dots, y_n) dy_{k+1} \cdots dy_n. \quad (6.1.6)$$

Bevis: For $A \subseteq \mathbb{R}^k$ er

$$\begin{aligned} P((X_1, \dots, X_k) \in A) &= P((X_1, \dots, X_n) \in A \times \mathbb{R}^{n-k}) \\ &= \int_{\mathbb{R}^n} 1_{A \times \mathbb{R}^{n-k}}(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^{n-k}} 1_{A \times \mathbb{R}^{n-k}}(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_{k+1} \cdots dx_n \right) dx_1 \cdots dx_k \\ &= \int_{\mathbb{R}^k} 1_A(x_1, \dots, x_k) \left(\int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_n) dx_{k+1} \cdots dx_n \right) dx_1 \cdots dx_k \\ &= \int_{\mathbb{R}^k} 1_A(x_1, \dots, x_k) q(x_1, \dots, x_k) dx_1 \cdots dx_k, \end{aligned}$$

hvilket viser, at (X_1, \dots, X_k) er en kontinuert stokastisk vektor med sandsynlighedstæthed q . □

Sagt med ord finder man altså tætheden for den marginale fordeling ved at integrere med hensyn til de koordinater, som man ikke er interesseret i.

Eksempel 6.1.4 Antag at (X, Y) er ligefordelt på trekanten

$$B = \{(x, y) \in [0, 1] \times [0, 1] \mid x \geq y\}.$$

Da arealet af B er $\frac{1}{2}$, er sandsynlighedstætheden for (X, Y) lig $2 \cdot 1_B(x, y)$, så sandsynlighedstætheden for fordelingen af X er

$$q(x) = \int_{\mathbb{R}} 2 \cdot 1_B(x, y) dy = 2 \int_0^x dy = 2x,$$

hvis $x \in [0, 1]$, mens $q(x)$ er lig nul ellers. □

6.2 Uafhængighed

Følgende sætning om, hvordan man kan se på sandsynlighedstætheden for en kontinuert stokastisk vektor, at de enkelte koordinater er uafhængige kontinuerte stokastiske variable, minder meget om det tilsvarende resultat for diskrete stokastiske variable.

Sætning 6.2.1 *Antag, at (X_1, \dots, X_n) er en n -dimensional kontinuert stokastisk vektor med sandsynlighedstæthed p . Betegn sandsynlighedstætheden for den marginale fordeling af X_i med p_i , $i = 1, \dots, n$. Da er følgende tre udsagn ækvivalente:*

1) X_1, \dots, X_n er stokastisk uafhængige,

2) For alle $(x_1, \dots, x_n) \in \mathbb{R}^n$ er

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n), \quad (6.2.1)$$

3) Der findes n ikke-negative reelle funktioner g_i , $i = 1, \dots, n$, så

$$p(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) \quad (6.2.2)$$

for alle $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Bevis: Vi viser kun sætningen i tilfældet $n = 2$. Det generelle bevis er helt tilsvarende, men notationsmæssigt mere besværligt.

Først vises, at 1) medfører 2). Hvis X_1 og X_2 er uafhængige, følger det af definitionen af uafhængighed af stokastiske variable (Definition 2.4.1), at vi for $A_i \subseteq \mathbb{R}$, $i = 1, 2$ har at

$$\begin{aligned} P((X_1, X_2) \in A_1 \times A_2) &= P(X_1 \in A_1)P(X_2 \in A_2) \\ &= \int_{\mathbb{R}} 1_{A_1}(x_1)p_1(x_1)dx_1 \int_{\mathbb{R}} 1_{A_2}(x_2)p_2(x_2)dx_2 \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_{A_1}(x_1)1_{A_2}(x_2)p_1(x_1)p_2(x_2)dx_2 \right) dx_1 \\ &= \int_{\mathbb{R}^2} 1_{A_1 \times A_2}(x_1, x_2)p_1(x_1)p_2(x_2)dx_1dx_2, \end{aligned}$$

hvoraf det ses, at sandsynlighedstætheden for (X_1, X_2) er $p_1(x_1)p_2(x_2)$.

For at vise, at 2) medfører 1), antager vi, at $p(x_1, x_2) = p_1(x_1)p_2(x_2)$. For $A_i \subseteq \mathbb{R}$, $i = 1, 2$, er da

$$P((X_1, X_2) \in A_1 \times A_2) = \int_{\mathbb{R}^2} 1_{A_1 \times A_2}(x_1, x_2)p_1(x_1)p_2(x_2)dx_1dx_2,$$

så ved ovenstående regnestykke i omvendt rækkefølge ses det, at

$$P((X_1, X_2) \in A_1 \times A_2) = P(X_1 \in A_1)P(X_2 \in A_2).$$

Altså er X_1 og X_2 er uafhængige, jfr. Definition 2.4.1.

At 2) medfører 3) er klart.

Lad os nu antage, at 3) holder og vise, at det medfører 2). Ifølge Sætning 6.1.3 er

$$p_1(x_1) = \int_{\mathbb{R}} p(x_1, x_2)dx_2 = g_1(x_1) \int_{\mathbb{R}} g_2(x_2)dx_2,$$

og

$$p_2(x_2) = \int_{\mathbb{R}} p(x_1, x_2)dx_1 = g_2(x_2) \int_{\mathbb{R}} g_1(x_1)dx_1.$$

Da

$$\begin{aligned} 1 = \int_{\mathbb{R}} p_1(x_1)dx_1 &= \int_{\mathbb{R}} \left(g_1(x_1) \int_{\mathbb{R}} g_2(x_2)dx_2 \right) dx_1 \\ &= \left(\int_{\mathbb{R}} g_1(x_1)dx_1 \right) \left(\int_{\mathbb{R}} g_2(x_2)dx_2 \right), \end{aligned}$$

følger det, at

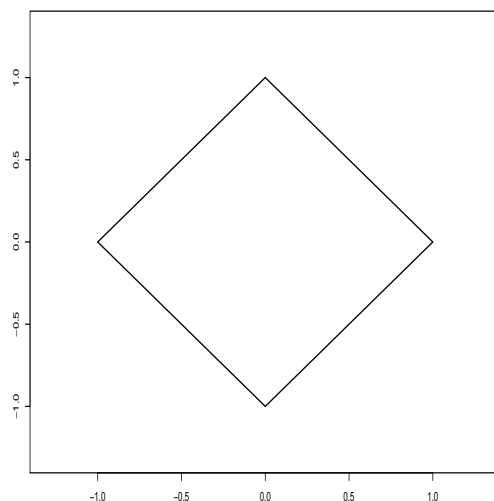
$$\begin{aligned} p_1(x_1)p_2(x_2) &= g_1(x_1)g_2(x_2) \left(\int_{\mathbb{R}} g_2(y_2)dy_2 \right) \left(\int_{\mathbb{R}} g_1(y_1)dy_1 \right) \\ &= g_1(x_1)g_2(x_2) = p(x_1, x_2), \end{aligned}$$

hvilket netop er (6.2.1). □

Egentlig skulle man i beviset for at 1) medfører 2) betragte en vilkårlig delmængde A af \mathbb{R}^2 og ikke kun en produktmængde $A_1 \times A_2$. På et senere kursus i sandsynlighedsregning vil det imidlertid blive vist, at det faktisk er tilstrækkeligt at betragte produktmængder.

I analogi med, hvad der gælder for diskrete fordelinger, er funktionerne g_1, \dots, g_{n-1} og g_n proportionale med henholdsvis p_1, \dots, p_{n-1} og p_n . Ligeledes gælder der (med argumenter som i Kapitel 3), at *hvis vi ikke kan finde en produktmængde $T_1 \times T_2$, således at (X_1, X_2) er koncentreret på $T_1 \times T_2$, og således at $p(x_1, x_2) > 0$ for alle $(x_1, x_2) \in T_1 \times T_2$, så kan X_1 og X_2 ikke være uafhængige.*

Eksempel 6.2.2 Antag, at (X_1, X_2) er ligefordelt på mængden $K = \{(x_1, x_2) \mid |x_1| + |x_2| \leq 1\}$. Da K ikke er nogen produktmængde, kan vi



Figur 6.2.1: Mængden K .

ved hjælp af ovenstående observation konkludere, at X_1 og X_2 ikke er uafhængige. Dette er også intuitivt klart: Hvis vi for eksempel ved, at $X_1 = \frac{1}{2}$, så kan vi konkludere, at $-\frac{1}{2} \leq X_2 \leq \frac{1}{2}$. Lad os - til overflod - også bevise, at den simultane sandsynlighedstæthed ikke kan skrives som et produkt af de to marginale tætheder.

Da $|K| = 2$, er den simultane sandsynlighedstæthed $p(x_1, x_2) = 1_K(x_1, x_2)/2$. Sandsynlighedstætheden for X_1 er ifølge (6.1.6)

$$p_1(x_1) = \frac{1}{2} \int_{\mathbb{R}} 1_K(x_1, x_2) dx_2 = \frac{1}{2} \int_{|x_1|-1}^{1-|x_1|} 1 dx_2 = 1 - |x_1|$$

for $x_1 \in [-1, 1]$. På tilsvarende måde fås, at sandsynlighedstætheden for X_2 er $p_2(x_2) = 1 - |x_2|$ for $x_2 \in [-1, 1]$. Det er klart, at $1_K(x_1, x_2)/2 \neq (1 - |x_1|)1_{[-1,1]}(x_1)(1 - |x_2|)1_{[-1,1]}(x_2)$.

□

I Kapitel 2 blev det lovet at bevise 2) i Sætning 2.4.3 for kontinuerte fordelinger. Lad os gøre det.

Sætning 6.2.3 *Lad (X_1, \dots, X_n) være en n -dimensional kontinuert stokastisk vektor, og antag at de stokastiske variable X_1, \dots, X_n er uafhængige. Hvis ψ er en funktion fra \mathbb{R}^{n-k} ind i \mathbb{R} , hvor $k < n$, så er de $k+1$ stokastiske variable $X_1, \dots, X_k, \psi(X_{k+1}, \dots, X_n)$ uafhængige.*

Bevis: Lad A_1, \dots, A_k, A være delmængder af \mathbb{R} , lad p betegne sandsynlighedstætheden for den stokastiske vektor (X_1, \dots, X_n) , og lad p_i betegne sandsynlighedstætheden for X_i , $i = 1, \dots, n$. Ifølge Sætning 6.2.1 er $p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n)$, så

$$\begin{aligned} & P(X_1 \in A_1, \dots, X_k \in A_k, \psi(X_{k+1}, \dots, X_n) \in A) \\ &= P(X_1 \in A_1, \dots, X_k \in A_k, (X_{k+1}, \dots, X_n) \in \psi^{-1}(A)) \\ &= \int_{\mathbb{R}^n} 1_{A_1 \times \dots \times A_k \times \psi^{-1}(A)}(x_1, \dots, x_n) p_1(x_1) \cdots p_n(x_n) dx_1 \cdots dx_n \\ &= \left(\int_{\mathbb{R}} 1_{A_1}(x_1) p_1(x_1) dx_1 \right) \cdots \left(\int_{\mathbb{R}} 1_{A_k}(x_k) p_k(x_k) dx_k \right) \\ &\quad \left(\int_{\mathbb{R}^{n-k}} 1_{\psi^{-1}(A)}(x_{k+1}, \dots, x_n) p_{k+1}(x_{k+1}) \cdots p_n(x_n) dx_{k+1} \cdots dx_n \right) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) P((X_{k+1}, \dots, X_n) \in \psi^{-1}(A)) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) P(\psi(X_{k+1}, \dots, X_n) \in A), \end{aligned}$$

således, at de $k + 1$ stokastiske variable er uafhængige. □

Sætning 2.4.3 indeholder endnu to resultater, som imidlertid følger af 2), så disse [d.v.s. 1) og 3)] er nu også vist for kontinuerte stokastiske variable.

6.3 Transformation af kontinuerte fordelinger på \mathbb{R}^n

I dette afsnit behandles nogle ofte forekommende transformationer af kontinuerte stokastiske vektorer.

6.3.1 Reelle transformationer

Først betragter vi transformationer, som resulterer i en reel stokastisk variabel. Det drejer sig for eksempel om summen eller produktet af to stokastiske variable.

Sætning 6.3.1 *Lad (X, Y) være en to-dimensionel kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , og definer en ny stokastisk variabel ved $Z = X + Y$. Da er Z en kontinuert stokastisk variabel med sandsynlighedstæthed*

$$q(z) = \int_{-\infty}^{\infty} p(x, z - x) dx. \quad (6.3.1)$$

Bevis: Fordelingsfunktionen for Z er givet ved

$$\begin{aligned} F(z) &= P(X + Y \leq z) = P((X, Y) \in \{(x, y) | y \leq z - x\}) \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} p(x, y) dy \right) dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^z p(x, u - x) du \right) dx \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} p(x, u - x) dx \right) du = \int_{-\infty}^z q(u) du. \end{aligned}$$

Vi har foretaget substitutionen $u = y + x$ i det inderste integral og dernæst ombyttet integrationsordenen. Da fordelingsfunktionen for Z opfylder (5.1.8), følger sætningen af Sætning 5.1.5. □

Det næste resultat, som vedrører fordelingen af summen af to uafhængige stokastiske variable, følger af Sætning 6.3.1 ved at benytte at den simultane sandsynlighedstæthed i denne situation er produktet af de to marginale sandsynlighedstætheder, jfr. Sætning 6.2.1.

Korollar 6.3.2 *Lad X og Y være to uafhængige kontinuerte stokastiske variable med marginale tætheder p_1 og p_2 , og definer en ny stokastisk variabel ved $Z = X + Y$. Da er Z en kontinuert stokastisk variabel med sandsynlighedstæthed*

$$q(z) = \int_{-\infty}^{\infty} p_1(x)p_2(z-x)dx. \quad (6.3.2)$$

Funktionen q givet ved (6.3.2) kaldes *foldningen* af de to funktioner p_1 og p_2 . Man kalder derfor fordelingen af $X + Y$ foldningen af de to marginale fordelinger.

Eksempel 6.3.3 *Lad X og Y være to uafhængige stokastiske variable, som begge er eksponentialfordelte med parameter λ . Da sandsynlighedstætheden for eksponentialfordelingen er $p(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x)$, er sandsynlighedstætheden q for $X + Y$ ifølge (6.3.2) givet ved*

$$q(z) = \int_0^z \lambda^2 e^{-\lambda(x+(z-x))} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

Denne fordeling er et eksempel på en Γ -fordeling, en fordeling som vi skal studere nærmere i Kapitel 8.

□

Det følgende generelle resultat kan blandt andet bruges til at finde fordelingen af produktet af eller kvotienten mellem to stokastiske variable ved at vælge funktionen f passende.

Sætning 6.3.4 *Lad (X, Y) være en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , og definer en ny stokastisk variabel ved $Z = Y/f(X)$, hvor f er en strengt positiv reel funktion. Da er Z en kontinuert stokastisk variabel med sandsynlighedstæthed*

$$q(z) = \int_{-\infty}^{\infty} p(x, f(x)z) f(x) dx. \quad (6.3.3)$$

Bevis: Fordelingsfunktionen for Z er givet ved

$$\begin{aligned} F(z) &= P(Y/f(X) \leq z) = P((X, Y) \in \{(x, y) | y \leq zf(x)\}) \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{zf(x)} p(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^z p(x, f(x)u) f(x) du \right) dx = \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} p(x, f(x)u) f(x) dx \right) du = \int_{-\infty}^z q(u) du. \end{aligned}$$

Vi har foretaget substitutionen $y = f(x)u$ i det inderste integral og dernæst ombyttet integrationsordenen. Da fordelingsfunktionen for Z opfylder (5.1.8), følger sætningen af Sætning 5.1.5. □

Bemærk, at hvis X er koncentreret på en mængde M , er det nok, at $f(x) > 0$ for $x \in M$.

De følgende tre vigtige specialtilfælde vises ved at sætte f lig henholdsvis $1/x$, x og \sqrt{x} .

Korollar 6.3.5 *Lad (X, Y) være en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , hvor $X > 0$. Da er den stokastiske variable $Z = X \cdot Y$ kontinuert med sandsynlighedstæthed*

$$q(z) = \int_0^{\infty} \frac{p(x, z/x)}{x} dx. \quad (6.3.4)$$

Korollar 6.3.6 *Lad (X, Y) være en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , hvor $X > 0$. Da er den stokastiske variable $Z = Y/X$ kontinuert med sandsynlighedstæthed*

$$q(z) = \int_0^{\infty} p(x, zx) x dx. \quad (6.3.5)$$

Korollar 6.3.7 *Lad (X, Y) være en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , hvor $X > 0$. Da er den stokastiske variable $Z = Y/\sqrt{X}$ kontinuert med sandsynlighedstæthed*

$$q(z) = \int_0^{\infty} p(x, z\sqrt{x}) \sqrt{x} dx. \quad (6.3.6)$$

Eksempel 6.3.8 Lad X og Y være to uafhængige stokastiske variable, som begge er eksponentialfordelte med parameter 1. Da er deres simultane sandsynlighedstæthed $p(x, y) = e^{-(x+y)}$, så sandsynlighedstætheden q for Y/X er ifølge (6.3.5) givet ved

$$q(z) = \int_0^\infty e^{-x(1+z)} x dx = \frac{1}{(1+z)^2} \int_0^\infty ye^{-y} dy = \frac{1}{(1+z)^2}, \quad (6.3.7)$$

når $z > 0$. Vi har her sat $y = x(1+z)$ og brugt, at $\int_0^\infty ye^{-y} dy$ er middelværdien af en standard eksponentialfordeling, som vi tidligere har set er lig en. Når $z \leq 0$ er $q(z) = 0$. Fordelingen med sandsynlighedstæthed givet ved (6.3.7) er et specialtilfælde af den såkaldte F-fordeling, som vi skal vende tilbage til i Kapitel 8. □

Vi har her kun behandlet nogle få, men vigtige, eksempler på reelle transformationer af kontinuerte stokastiske vektorer. Vi slutter med et eksempel på en transformation, som ikke er dækket af ovenstående resultater, men hvor man alligevel kan komme igennem via fordelingsfunktionen.

Eksempel 6.3.9 Antag, at X_1, \dots, X_n er uafhængige stokastiske variable, som alle er ligefordelte på $[0, 1]$. Lad $X_{(1)}$ være den stokastiske variable, som angiver det mindste udfald blandt de n variable, altså

$$X_{(1)} = X_1 \wedge X_2 \wedge \dots \wedge X_n = \min\{X_1, \dots, X_n\}.$$

Fordelingsfunktionen for $X_{(1)}$ er for $x \in (0, 1)$ givet ved

$$\begin{aligned} F_1(x) = P(X_{(1)} \leq x) &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - P(X_1 > x) \cdots P(X_n > x) = 1 - (1-x)^n. \end{aligned}$$

Da F_1 er kontinuert differentiabel på $(0, 1)$ er sandsynlighedstætheden for $X_{(1)}$ (Sætning 5.1.6)

$$f_1(x) = F_1'(x) = n(1-x)^{n-1}, \quad x \in (0, 1).$$

Vi kan også klare fordelingen af den k te mindste blandt X_1, \dots, X_n ($1 \leq k \leq n$). Lad os betegne denne stokastiske variable med $X_{(k)}$. Dens fordelingsfunktion er for $x \in (0, 1)$ givet ved

$$\begin{aligned} F_k(x) &= P(\text{mindst } k \text{ af de } n \text{ variable er mindre end eller lig } x) \\ &= P(\#\{i \mid X_i \leq x\} \geq k). \end{aligned}$$

Da $P(X_i \leq x) = x$ for $x \in (0, 1)$, og da

$$\#\{i \mid X_i \leq x\} = \sum_{i=1}^n 1_{[0,x]}(X_i),$$

er $\#\{i \mid X_i \leq x\}$ binomialfordelt med antalsparameter n og sandsynlighedsparameter x . Derfor er

$$F_k(x) = \sum_{i=k}^n \binom{n}{i} x^i (1-x)^{n-i}.$$

Også F_k er kontinuert differentiabel på $(0, 1)$, så sandsynlighedstætheden for $X_{(k)}$ er for $x \in (0, 1)$

$$\begin{aligned} f_k(x) &= F_k'(x) = \sum_{i=k}^n \binom{n}{i} \left[ix^{i-1}(1-x)^{n-i} - (n-i)x^i(1-x)^{n-i-1} \right] \\ &= \binom{n}{k} kx^{k-1}(1-x)^{n-k} \\ &\quad + \sum_{i=k+1}^n \binom{n}{i} ix^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n-1} \binom{n}{i} (n-i)x^i(1-x)^{n-(i+1)} \\ &= \binom{n}{k} kx^{k-1}(1-x)^{n-k}, \end{aligned}$$

idet en omskrivning viser, at de to sidste summer er identiske. □

6.3.2 Lineær transformation af en kontinuert fordeling

Vi skal her betragte lineære transformationer af kontinuerte stokastiske vektorer. Sådanne transformationer er vigtige i den statistiske teori for

normalfordelte observationer. Vi begynder med en to-dimensional stokastisk vektor (X_1, X_2) , d.v.s. med transformationer af formen

$$t(X_1, X_2) = A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

hvor A er en 2×2 -matrix.

Sætning 6.3.10 *Lad (X_1, X_2) være en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , og definer nye stokastiske variable ved*

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 \\ Y_2 &= a_{21}X_1 + a_{22}X_2, \end{aligned}$$

hvor a_{11}, a_{12}, a_{21} og a_{22} er reelle tal, som opfylder, at determinanten af matricen

$$A = \begin{Bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Bmatrix}$$

er forskellig fra nul (d.v.s. $\det(A) = a_{11}a_{22} - a_{12}a_{21} \neq 0$). Da er (Y_1, Y_2) en to-dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed

$$q(y_1, y_2) = p(t^{-1}(y_1, y_2))|\det(A^{-1})|, \quad (6.3.8)$$

hvor

$$t^{-1}(y_1, y_2) = A^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Et alternativt udtryk for q er

$$q(y_1, y_2) = \frac{p(t^{-1}(y_1, y_2))}{|\det(A)|}. \quad (6.3.9)$$

“Bevis”: Nedenstående er ikke noget helt korrekt bevis, men giver ikke desto mindre et rigtigt godt indtryk af, hvorfor sætningen er rigtig. Følgende notation vil blive brugt:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad \text{og} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

hvor $Y = AX$, $y = Ax$ og $x = A^{-1}y$.

Vi begynder med at minde om approximationen

$$q(y_1, y_2) \simeq \frac{P(Y \in K_h)}{h^2},$$

hvor $K_h = [y_1, y_1 + h] \times [y_2, y_2 + h]$ er et lille kvadrat med det ene hjørne i punktet y , cf. (6.1.4). Da $\det(A) \neq 0$, har A en invers, så

$$P(Y \in K_h) = P(AX \in K_h) = P(X \in A^{-1}K_h) \simeq p(A^{-1}y)|A^{-1}K_h|,$$

hvor $A^{-1}K_h = \{A^{-1}z | z \in K_h\}$, og hvor $|A^{-1}K_h|$ betegner arealet af $A^{-1}K_h$. Vi har her brugt, at $A^{-1}K_h$ er et lille parallelogram, hvis ene hjørne er punktet $A^{-1}y$. De andre hjørner er

$$A^{-1}y + A^{-1} \begin{pmatrix} h \\ 0 \end{pmatrix}, \quad A^{-1}y + A^{-1} \begin{pmatrix} 0 \\ h \end{pmatrix} \quad \text{og} \quad A^{-1}y + A^{-1} \begin{pmatrix} h \\ h \end{pmatrix},$$

så $A^{-1}K_h$ er udspændt af vektorerne

$$A^{-1} \begin{pmatrix} h \\ 0 \end{pmatrix} = \begin{pmatrix} b_{11}h \\ b_{21}h \end{pmatrix} \quad \text{og} \quad A^{-1} \begin{pmatrix} 0 \\ h \end{pmatrix} = \begin{pmatrix} b_{12}h \\ b_{22}h \end{pmatrix},$$

hvor vi benytter notationen

$$A^{-1} = \begin{Bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Bmatrix}.$$

Da arealet af parallelogrammet udspændt af to vektorer u og v er lig den numeriske værdi af determinanten af matricen, hvis to søjler består af vektorerne u og v , er

$$|A^{-1}K_h| = |b_{11}hb_{22}h - b_{12}hb_{21}h| = |b_{11}b_{22} - b_{12}b_{21}|h^2 = |\det(A^{-1})|h^2.$$

Sammenfattende har vi altså, at

$$q(y_1, y_2) \simeq p(A^{-1}y)|\det(A^{-1})|.$$

Denne approximation kan gøres vilkårligt nøjagtig ved at vælge h tilstrækkeligt lille, men da den ikke afhænger af h , må den være eksakt. Det

alternative udtryk (6.3.9) for q opnås nu ved at bemærke, at $\det(A^{-1}) = [\det(A)]^{-1}$.

□

Resultatet vedrørende arealet af parallelogrammet udspændt af to vektorer u og v er let at vise. Her er en mulig fremgangsmåde: Hvis v betragtes som parallelogrammets grundlinie, er højden lig længden af projektionen af u på en vektor, som står vinkelret på v . Resultatet følger nu ved at gange højden med længden af grundlinien.

Der gælder et tilsvarende resultat for lineære transformationer af en n -dimensional kontinuert stokastisk vektor (X_1, \dots, X_n) , d.v.s. transformationer af formen

$$t(X_1, \dots, X_n) = A \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

hvor A er en $n \times n$ -matrix. I dette resultat benyttes det, at man også kan knytte en determinant til en $n \times n$ -matrix, og at en $n \times n$ -matrix har en invers matrix, når determinanten er forskellig fra nul. Det generelle resultat kan sandsynliggøres på samme måde som Sætning 6.3.10. Vi benytter notationen

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Sætning 6.3.11 *Lad X være en n -dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed p , og definer en ny stokastisk vektor ved*

$$Y = AX,$$

hvor A er en $n \times n$ -matrix, som opfylder $\det(A) \neq 0$. Da er Y en n -dimensional kontinuert stokastisk vektor med simultan sandsynlighedstæthed

$$q(y) = p(A^{-1}y)|\det(A^{-1})|. \quad (6.3.10)$$

Et alternativt udtryk for q er

$$q(y) = \frac{p(A^{-1}y)}{|\det(A)|}. \quad (6.3.11)$$

Hvis Y er en n -dimensional kontinuert stokastisk vektor med sandsynlighedstæthed p , så har den stokastiske vektor $Z = Y + \mu$, hvor μ er en (ikke-stokastisk) vektor i \mathbb{R}^n , sandsynlighedstætheden

$$q(z) = p(z - \mu). \quad (6.3.12)$$

For $K_h = [z_1, z_1 + h] \times [z_2, z_2 + h] \times \cdots \times [z_n, z_n + h]$, hvor $h > 0$, er nemlig

$$P(Z \in K_h) = P(Y + \mu \in K_h) = P(Y \in (K_h - \mu)) \simeq p(z - \mu)h^n.$$

Her betegner $K_h - \mu$ mængden $\{u - \mu | u \in K_h\}$, altså en n -dimensional kasse med det ene hjørne i punktet $z - \mu$, som har volumen h^n . Ved at kombinere (6.3.11) og (6.3.12), ser vi, at sandsynlighedstætheden for $V = AX + \mu$ er

$$q(v) = \frac{p(A^{-1}(v - \mu))}{|\det(A)|}. \quad (6.3.13)$$

Vi slutter dette afsnit med at bevise følgende vigtige sætning om fordelingen af en sum af uafhængige normalfordelte stokastiske variable. Vi kunne faktisk godt have vist sætningen i næste afsnit ved hjælp af (6.3.2). Det kræver imidlertid en del kedeligt regneri. Beviset her er lidt mere elegant og giver iøvrigt en forsmag på et bevis, vi skal studere i Kapitel 8.

Sætning 6.3.12 *Lad X_1, \dots, X_n være uafhængige stokastiske variable, hvor X_i er normalfordelt med middelværdi μ_i og varians σ_i^2 , $i = 1, \dots, n$. Da er $X_1 + \cdots + X_n$ normalfordelt med middelværdi $\mu_1 + \cdots + \mu_n$ og varians $\sigma_1^2 + \cdots + \sigma_n^2$.*

Bevis: Først bemærker vi, at det er nok at bevise sætningen for $n = 2$, da det generelle tilfælde følger ved induktion.

Dernæst vil vi betragte to uafhængige standard normalfordelte stokastiske variable U_1 og U_2 . Deres simultane sandsynlighedstæthed er

$$p(u_1, u_2) = \varphi(u_1)\varphi(u_2) = \frac{1}{2\pi}e^{-(u_1^2+u_2^2)/2}.$$

Som sædvanlig betegner φ sandsynlighedstætheden for standard normalfordelingen. Vi definerer to nye stokastiske variable V_1 og V_2 ved

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix},$$

hvor det antages, at $\alpha^2 + \beta^2 = 1$. Matricen

$$A = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$$

har determinant 1 og svarer til en rotation af planen. Da

$$A^{-1} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix},$$

er sandsynlighedstætheden for (V_1, V_2) ifølge (6.3.9)

$$\begin{aligned} q(v_1, v_2) &= p(\alpha v_1 - \beta v_2, \beta v_1 + \alpha v_2) \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}[(\alpha v_1 - \beta v_2)^2 + (\beta v_1 + \alpha v_2)^2]} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(v_1^2 + v_2^2)} = \varphi(v_1)\varphi(v_2). \end{aligned}$$

Vi konkluderer derfor ved hjælp af Sætning 6.2.1, at V_1 og V_2 er uafhængige, og at de begge er standard normalfordelte. Specielt gælder, at hvis $\alpha^2 + \beta^2 = 1$, så er $\alpha U_1 + \beta U_2$ standard normalfordelt.

Vi kan nu vende tilbage til X_1 og X_2 . Vi definerer de standardiserede variable

$$U_1 = \frac{X_1 - \mu_1}{\sigma_1} \quad \text{og} \quad U_2 = \frac{X_2 - \mu_2}{\sigma_2},$$

og skriver $X_1 + X_2$ ved hjælp af disse:

$$\begin{aligned} X_1 + X_2 &= (\mu_1 + \sigma_1 U_1) + (\mu_2 + \sigma_2 U_2) \\ &= (\mu_1 + \mu_2) + (\sigma_1 U_1 + \sigma_2 U_2) \\ &= (\mu_1 + \mu_2) + \sqrt{\sigma_1^2 + \sigma_2^2} U, \end{aligned}$$

hvor

$$U = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} U_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} U_2.$$

Da U_1 og U_2 er uafhængige og standard normalfordelte, er U også standard normalfordelt, hvorfor sætningen følger. □

I næste afsnit skal vi se, at formlerne (3.7.6) og (3.8.9) vedrørende middelværdi og varians af en sum af stokastiske variable også gælder for kontinuerte stokastiske variable. Det er derfor ikke overraskende, at middelværdien og variansen af $X_1 + \dots + X_n$ er som angivet i Sætningen 6.3.12. Det interessante er, at summen er normalfordelt. Hvis man kan huske formlerne (3.7.6) og (3.8.9), og det bør man kunne, er det ikke svært at huske Sætning 6.3.12.

6.4 Middelværdi, varians og kovarians

Middelværdien af en transformation af en stokastisk vektor kan beregnes på en måde, der helt svarer til, hvad vi tidligere har set for diskrete stokastiske variable og en-dimensionale kontinuerte stokastiske variable.

Sætning 6.4.1 *Lad X være en n -dimensional kontinuert stokastisk vektor med sandsynlighedstæthed p . Lad endvidere ψ være en funktion fra \mathbb{R}^n ind i \mathbb{R} . Da har den stokastiske variable $\psi(X)$ middelværdi hvis og kun hvis*

$$\int_{\mathbb{R}^n} |\psi(x_1, \dots, x_n)| p(x_1, \dots, x_n) dx_1 \cdots dx_n < \infty, \quad (6.4.1)$$

og i så fald er

$$E(\psi(X)) = \int_{\mathbb{R}^n} \psi(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n. \quad (6.4.2)$$

Vi kan ikke på dette kursus bevise Sætning 6.4.1. Som i det en-dimensionale tilfælde kan vi ikke være sikre på, at fordelingen af $\psi(X)$ er af en type, for hvilken vi har defineret middelværdien. Sætningen er derfor mest nyttig for os, når fordelingen af $\psi(X)$ er kontinuert eller diskret. For i det mindste at antyde, at Sætning 6.4.1 er korrekt, vil vi se på tilfældet $\psi = 1_B$, hvor $B \subseteq \mathbb{R}^n$:

$$\begin{aligned} \int_{\mathbb{R}^n} 1_B(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n &= \\ P((X_1, \dots, X_n) \in B) &= E(1_B(X_1, \dots, X_n)), \end{aligned}$$

idet $1_B(X_1, \dots, X_n)$ er en diskret stokastisk variabel, som kun antager værdierne 0 og 1, se Eksempel 3.7.2. Selv om dette er et meget enkelt eksempel, er det faktisk første skridt i beviset for (6.4.2).

Ved hjælp af Sætning 6.4.1 kan vi nu vise de samme resultater, som vi før har vist for diskrete stokastiske variable ved at bruge Sætning 4.4.2. Beviserne er helt analoge, blot er summer erstattet med integraler og sandsynlighedsfunktioner med sandsynlighedstætheder. Vi nøjes med at bevise den næste sætning for at vise, hvordan man gør. Resten overlades trygt til læseren.

Sætning 6.4.2 *Lad X_1, X_2, \dots, X_n være kontinuerte stokastiske variable, som alle har middelværdi. Da har den stokastiske variable $X_1 + X_2 + \dots + X_n$ middelværdi, og*

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n). \quad (6.4.3)$$

Hvis det yderligere antages, at de stokastiske variable X_1, X_2, \dots, X_n er uafhængige, har også $X_1 \cdot X_2 \cdot \dots \cdot X_n$ middelværdi, og

$$E(X_1 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n). \quad (6.4.4)$$

Bevis: Vi kan nøjes med at bevise sætningen for $n = 2$, da det generelle resultat følger ved induktion.

Betegn sandsynlighedstætheden for den stokastiske vektor (X_1, X_2) med p og sandsynlighedstætheden for X_i , $i = 1, 2$, med p_i . Da X_i har middelværdi, er ifølge Sætning 6.4.1

$$\int_{\mathbb{R}^2} |x_i| p(x_1, x_2) dx_1 dx_2 < \infty, \quad i = 1, 2.$$

Derfor medfører trekantsuligheden $|x_1 + x_2| \leq |x_1| + |x_2|$ at

$$\begin{aligned} \int_{\mathbb{R}^2} |x_1 + x_2| p(x_1, x_2) dx_1 dx_2 &\leq \\ \int_{\mathbb{R}^2} |x_1| p(x_1, x_2) dx_1 dx_2 + \int_{\mathbb{R}^2} |x_2| p(x_1, x_2) dx_1 dx_2 &< \infty, \end{aligned}$$

hvorfor den stokastiske variable $X_1 + X_2$ ifølge Sætning 6.4.1 har middelværdi, som kan beregnes ved

$$\begin{aligned} &\int_{\mathbb{R}^2} (x_1 + x_2) p(x_1, x_2) dx_1 dx_2 \\ &= \int_{\mathbb{R}^2} x_1 p(x_1, x_2) dx_1 dx_2 + \int_{\mathbb{R}^2} x_2 p(x_1, x_2) dx_1 dx_2 \\ &= E(X_1) + E(X_2). \end{aligned}$$

For at vise den sidste påstand, antages det nu, at X_1 og X_2 er uafhængige. Dermed er $p(x_1, x_2) = p_1(x_1)p_2(x_2)$, således at

$$\begin{aligned} \int_{\mathbb{R}^2} |x_1 x_2| p(x_1, x_2) dx_1 dx_2 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |x_1| p_1(x_1) |x_2| p_2(x_2) dx_1 \right) dx_2 \\ &= \int_{\mathbb{R}} |x_2| p_2(x_2) \left(\int_{\mathbb{R}} |x_1| p_1(x_1) dx_1 \right) dx_2 \\ &= \left(\int_{\mathbb{R}} |x_1| p_1(x_1) dx_1 \right) \left(\int_{\mathbb{R}} |x_2| p_2(x_2) dx_2 \right). \end{aligned}$$

De to integraler i det sidste udtryk er endelige, da X_1 og X_2 begge har middelværdi, hvorfor vi igen kan bruge Sætning 6.4.1 til at slutte, at den stokastiske variable $X_1 \cdot X_2$ har middelværdi. Efter samme omskrivning af integralet uden numerisk tegn ser vi, at middelværdien kan beregnes ved (6.4.4).

□

For at kunne bevise de samme resultater som i Kapitel 3 om kovarians og korrelation mellem to kontinuerte stokastiske variable, skal vi bruge de følgende to sætninger, der er lette at bevise.

Sætning 6.4.3 *Lad (X_1, X_2) være en to-dimensional kontinuert stokastisk vektor, som opfylder, at $X_1 \leq X_2$, samt at X_1 og X_2 har middelværdi. Da er $E(X_1) \leq E(X_2)$.*

Læg mærke til, at Sætning 6.4.2 og Sætning 6.4.3 medfører Sætning 5.2.4 som et specialtilfælde.

Sætning 6.4.4 *Lad (X_1, X_2) være en kontinuert to-dimensional stokastisk vektor, som opfylder, at X_2 har middelværdi, samt at $|X_1| \leq |X_2|$. Da har også X_1 middelværdi.*

Hvis to kontinuerte stokastiske variable X og Y opfylder, at

$$E(X^2) < \infty \quad \text{og} \quad E(Y^2) < \infty, \quad (6.4.5)$$

kan vi definere deres *kovarians* ved (3.8.1). Kravet er altså, at de to stokastiske variable har varians. Vi så i forrige kapitel, at dette krav medfører, at middelværdien af X og Y også eksisterer. Da

$$(X - E(X))(Y - E(Y)) = XY - XE(Y) - YE(X) + E(X)E(Y)$$

og

$$|XY| \leq \frac{1}{2}(X^2 + Y^2),$$

ser vi af Sætningerne 6.4.4 og 6.4.2, at betingelsen (6.4.5) også sikrer, at den stokastiske variable $(X - E(X))(Y - E(Y))$ har middelværdi, således at vi kan definere kovariansen ved (3.8.1). Det er klart, at vi stadig kan beregne kovariansen ved (3.8.2), samt at regnereglerne (3.8.3) – (3.8.5) gælder for tre kontinuerte stokastiske, forudsat at de alle har varians.

Formlen for variansen af en sum af stokastiske variable (3.8.7) og Sætningerne 3.8.3 og 3.8.8 gælder også for kontinuerte stokastiske variable. Det skal bare altid antages, at alle indgående stokastiske variable har varians, altså opfylder (5.2.8). Beviserne i Afsnit 3.8 bruger slet ikke, at de stokastiske variable i det afsnit er koncentreret på endelige mængder; der bruges kun generelle regneregler for middelværdi, som vi nu har bevist også gælder for kontinuerte stokastiske variable. Der er derfor ingen grund til at gentage beviserne her. Vi kan bare henvise til Afsnit 3.8.

Korrelationen mellem to kontinuerte stokastiske variable X og Y med varians defineres ved (3.8.8). Også for kontinuerte stokastiske variable kan korrelationen fortolkes som kovariansen mellem standardiserede variable, og Sætning 3.8.7 holder også her. Specielt er $|\text{corr}(X, Y)| \leq 1$. Igen henvises der til beviserne i Afsnit 3.8.

6.5 Kontinuerte betingede fordelinger

Lad P være en kontinuert fordeling på \mathbb{R}^n med sandsynlighedstæthed p . Vi så i Sætning 1.4.8, at funktionen $B \mapsto P(B|A)$ for en fast delmængde $A \subseteq \mathbb{R}^n$ med $P(A) > 0$ er et sandsynlighedsmål, den betingede fordeling givet A . Den betingede fordeling er også kontinuert. Dens sandsynlighedstæthed er givet ved

$$p(x_1, \dots, x_n|A) = p(x_1, \dots, x_n)1_A(x_1, \dots, x_n)/P(A). \quad (6.5.1)$$

At (6.5.1) er en sandsynlighedstæthed er klart:

$$\begin{aligned} & \int_{\mathbb{R}^n} p(x_1, \dots, x_n|A) dx_1 \cdots dx_n \\ &= \frac{1}{P(A)} \int_{\mathbb{R}^n} 1_A(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n = \frac{P(A)}{P(A)} = 1. \end{aligned}$$

Endvidere er

$$\begin{aligned} & \int_{\mathbb{R}^n} 1_B(x_1, \dots, x_n) p(x_1, \dots, x_n | A) dx_1 \cdots dx_n \\ &= \frac{1}{P(A)} \int_{\mathbb{R}^n} 1_{A \cap B}(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n = \frac{P(A \cap B)}{P(A)}. \end{aligned}$$

Dermed har vi vist, at (6.5.1) er sandsynlighedstæthed for den betingede fordeling givet A .

Betragt en kontinuert stokastisk variabel X med sandsynlighedstæthed p , og lad A være en delmængde af \mathbb{R} . Da har den betingede fordeling af X givet $X \in A$ sandsynlighedstæthed lig $p(x)1_A(x)/P(X \in A)$

Eksempel 6.5.1 Antag, at X er eksponentialfordelt med parameter λ . Da er $P(X > a) = e^{-\lambda a}$ for $a > 0$, så den betingede fordeling af X givet at $X > a$ har sandsynlighedstætheden

$$q(x) = \lambda e^{-\lambda(x-a)}.$$

Dermed er sandsynlighedstætheden for den betingede fordeling af $X - a$ givet at $X > a$

$$q(x + a) = \lambda e^{-\lambda x},$$

jfr. (5.4.2a).

Eksponentialfordelingen bruges ofte som model for ventetiden på en tilfældigt ankomende hændelse. Vi ser, at denne model har den egenskab, at den betingede fordeling af ventetiden, givet at man allerede har ventet a tidsenheder, er præcis den samme som den ubetingede fordeling. Sådant bør det også være, hvis hændelserne virkelig kommer helt tilfældigt.

□

For en to-dimensional kontinuert stokastisk vektor (X, Y) har man jævnligt brug for at betinge med at $X = x$, men det går jo tilsyneladende ikke, for $P(X = x) = 0!$ Det viser sig imidlertid, at man alligevel kan definere en betinget fordeling af Y givet at $X = x$. Her giver vi kun en heuristisk udledning. Lad altså (X, Y) være en to-dimensional kontinuert stokastisk vektor med sandsynlighedstæthed $p(x, y)$, og betragt et reelt tal x_0 , for hvilket $P(X \in [x_0, x_0 + \epsilon]) > 0$ for alle $\epsilon > 0$. For at lette

notationen sætter vi $K_\epsilon = [x_0, x_0 + \epsilon]$. Da har den betingede fordeling af (X, Y) givet at $X \in K_\epsilon$ sandsynlighedstæthed

$$q(x, y|K_\epsilon) = \frac{p(x, y)1_{K_\epsilon}(x, y)}{P(X \in K_\epsilon)}.$$

Den marginale betingede tæthed for Y givet at $X \in K_\epsilon$ får vi ifølge (6.1.6) ved at integrere m.h.t. x :

$$\begin{aligned} q_2(y|K_\epsilon) &= \int_{\mathbb{R}} q(x, y|K_\epsilon) dx = \frac{\int_{x_0}^{x_0+\epsilon} p(x, y) dx}{P(X \in K_\epsilon)} \\ &\simeq \frac{p(x_0, y)\epsilon}{p_1(x_0)\epsilon} = \frac{p(x_0, y)}{p_1(x_0)}, \end{aligned}$$

hvor p_1 er den marginale tæthed for X , som (igen ifølge (6.1.6)) er givet ved

$$p_1(x) = \int_{\mathbb{R}} p(x, y) dy.$$

Motiveret af disse beregninger, definerer vi for alle x , for hvilke $p_1(x) > 0$, den betingede fordeling af Y givet at $X = x$ som den kontinuerte fordeling med sandsynlighedstæthed

$$q(y) = \frac{p(x, y)}{p_1(x)}. \quad (6.5.2)$$

Læseren opfordres til at overbevise sig om, at (6.5.2) virkelig definerer en sandsynlighedstæthed.

6.6 Sammenfatning

Vi har i dette kapitel studeret kontinuerte fordelinger på \mathbb{R}^n og stokastiske vektorer, hvis fordeling er af denne type, d.v.s. kontinuerte stokastiske vektorer. De fler-dimensionale kontinuerte fordelinger defineres ligesom de endimensionale ved hjælp af en sandsynlighedstæthed, og teorien er et stykke hen ad vejen den samme; blot skal der nu integreres i \mathbb{R}^n .

Der er dog meget nyt. Vi så, at vi kan finde den marginale sandsynlighedstæthed for en stokastiske vektor, der består af en delmængde af de oprindelige koordinater, ved simplethen at integrere den oprindelige

simultane sandsynlighedstæthed over de øvrige koordinater. Vi har også set, at de enkelte koordinater i en kontinuert stokastisk vektor er uafhængige, netop når den simultane sandsynlighedstæthed kan skrives som et produkt af funktioner, der hver kun afhænger af én koordinat.

Videre har vi studeret fordelingen af visse funktioner af flere stokastiske variable. Blandt andet har vi for to stokastiske variable fundet fordelingen af deres sum, produkt og kvotient. Vi har også fundet fordelingen af en lineær transformation af en kontinuert stokastisk vektor. Det er værd at bemærke, at vi har bevist, at summen af n uafhængige normalfordelte stokastiske variable igen er normalfordelt med en middelværdi og en varians, som kan findes ved de sædvanlige regneregler for middelværdi og varians af en sum af stokastiske variable. Vi har nemlig også set, at alle de tidligere viste regneregler for middelværdi, varians, kovarians og korrelation også gælder for kontinuerte stokastiske variable. Endelig har vi kigget lidt på kontinuerte betingede fordelinger.

6.7 Opgaver

- 6.1 I denne opgave vises det, at standard normalfordelingens tæthed faktisk er en sandsynlighedstæthed, altså at dens integral er lig en. Sæt derfor

$$c = \int_0^{\infty} e^{-t^2} dt.$$

Vi skal da vise, at $c = \sqrt{\pi}/2$. Vis først, at

$$\int_0^{\infty} \left(\int_0^{\infty} e^{-t(1+x^2)} dt \right) dx = \int_0^{\infty} \frac{1}{1+x^2} dx = \pi/2.$$

Vink: Stamfunktionen til $1/(1+x^2)$ er \tan^{-1} , d.v.s. den inverse funktion til tangensfunktionen, \arctan (arcus tangens). Vis dernæst ved at skifte integrationsvariabel et par gange, at

$$\int_0^{\infty} \left(\int_0^{\infty} e^{-t(1+x^2)} dx \right) dt = 2c^2,$$

og slut heraf det ønskede resultat.

- 6.2 Vis, at der for en kontinuert stokastisk variabel X på $(0, \infty)$ med fordelingsfunktion F gælder, at

$$E(X) = \int_0^{\infty} (1 - F(x)) dx.$$

- 6.3 Lad den stokastiske vektor (X, Y) være ligefordelt på enhedscirklen, $\{(x, y) | x^2 + y^2 \leq 1\}$. Vis, at X og Y ikke er uafhængige, og find tæthederne for deres marginale fordelinger.

- 6.4 Lad (X, Y) være en kontinuert to-dimensional stokastisk vektor med sandsynlighedstæthed

$$p(x, y) = \begin{cases} 3xy^{-2} & \text{når } x \in (0, 1) \text{ og } y \in (1, 3) \\ 0 & \text{ellers.} \end{cases}$$

Find de marginale sandsynlighedstætheder for X og Y , og vis, at X og Y er uafhængige.

- 6.5 Lad X_1 og X_2 være uafhængige stokastiske variable, som begge er ligefordelte på $(0, 1)$. Vis, at tætheden for $Z = X_1 + X_2$ er

$$p(z) = \begin{cases} z & \text{for } z \in (0, 1] \\ 2 - z & \text{for } z \in (1, 2) \\ 0 & \text{ellers.} \end{cases}$$

- 6.6 Lad X_1 og X_2 være uafhængige stokastiske variable, som begge er eksponentialfordelt med parameter 1. Find tætheden for fordelingen af $X_1 - X_2$.

- 6.7 Lad X_1, \dots, X_n være uafhængige stokastiske variable, der alle er eksponentialfordelte med parameter λ . Vis ved et induktionsargument, at sandsynlighedstætheden for summen $X_1 + \dots + X_n$ er

$$p(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, \quad x > 0.$$

En fordeling af denne type kaldes en *Erlang-fordeling* efter en dansk matematiker. Disse fordelinger er et specialtilfælde af Γ -fordelingerne, som behandles i Kapitel 8.

- 6.8 Lad R og X være uafhængige stokastiske variable, hvor R er ligefordelt på $(0, 1)$, mens X har en vilkårlig tæthed p . Lad F betegne fordelingsfunktionen for X . Vis at tætheden for $S = R + X$ er

$$q(s) = F(s) - F(s - 1).$$

Udregn og tegn tætheden q i tilfældet hvor X er eksponentialfordelt med parameter 1.

- 6.9 Vis Korollar 6.3.5, Korollar 6.3.6 og Korollar 6.3.7.

- 6.10 Lad R_1 og R_2 være uafhængige stokastiske variable, som begge er ligefordelte på $(0, 1)$. Definer $Z = R_1 R_2$.

(a) Hvad er tætheden for Z ?

Fordelingen af Z kan man også nå frem til på følgende måde:

(b) Sæt $X_1 = -\log(R_1)$ og $X_2 = -\log(R_2)$. Hvad er fordelingen af (X_1, X_2) ?

(c) Udled fordelingen af $S = X_1 + X_2$.

(d) Udled fordelingen af $Z = \exp(-S)$.

- 6.11 Lad X og Y være to uafhængige stokastiske variable, som begge er ligefordelte på $(0, 1)$. Vis, at sandsynlighedstætheden for X/Y er

$$q(z) = \begin{cases} \frac{1}{2z^2} & \text{når } z \geq 1 \\ \frac{1}{2} & \text{når } 0 < z < 1 \\ 0 & \text{når } z \leq 0. \end{cases}$$

- 6.12 Lad A være parallelogrammet bestemt af punkterne $(0,0)$, $(1,0)$, $(0,1)$ og $(-1,1)$, og betragt funktionen

$$f(x, y) = \begin{cases} 4xy + 4y^2 & \text{hvis } (x, y) \in A \\ 0 & \text{ellers.} \end{cases}$$

(a) Gør rede for, at f er en sandsynlighedstæthed på A .

- (b) Lad (X, Y) være en kontinuert stokastisk vektor, hvis sandsynlighedstæthed er f . Gør rede for at X og Y ikke er uafhængige.
- (c) Find sandsynlighedstætheden for den marginale fordeling af Y .
- (d) Definer en ny stokastisk vektor (Z_1, Z_2) ved $Z_1 = X + Y$ og $Z_2 = Y$. Find den simultane sandsynlighedstæthed for (Z_1, Z_2) , vis at Z_1 og Z_2 er uafhængige, og find deres marginale fordelinger.
- 6.13 Lad X_1 og X_2 være uafhængige, standard normalfordelte stokastiske variable. Vis at X_1/X_2 er Cauchy fordelt.
- 6.14 Lad X_1 og X_2 være uafhængige stokastiske variable, som begge er eksponentialfordelte med parameter 1.
- (a) Find tætheden for den stokastiske vektor $(X_1 + X_2, X_1)$.
- (b) Vis, at

$$\frac{X_1}{X_1 + X_2}$$

er ligefordelt på $(0, 1)$. Hvad er fordelingen af $X_1/(X_1 + X_2)$, hvis de to eksponentialfordelinger har parameter λ ?

- 6.15 Vis resultatet (6.3.1) ved at kombinere Sætning 6.3.10 med (6.1.5).
- 6.16 Lad X_1 og X_2 være uafhængige stokastiske variable, som begge er eksponentialfordelte. Definer

$$\begin{aligned} M &= \min\{X_1, X_2\} = X_1 \wedge X_2, \\ S &= \max\{X_1, X_2\} = X_1 \vee X_2. \end{aligned}$$

Angiv sandsynlighedstæthederne for fordelingerne af M og S .

- 6.17 Lad X_1, \dots, X_n være uafhængige, identisk fordelte stokastiske variable med en tæthed p , som er strengt positiv på et interval og nul udenfor intervallet. Intervallet kan være hele den reelle akse. Lad F betegne den tilhørende fordelingsfunktion. Lad $X_{(k)}$ være defineret som i Eksempel 6.3.9. Vis, at fordelingen af $X_{(k)}$ har tætheden

$$q(x) = \binom{n}{k} k F(x)^{k-1} (1 - F(x))^{n-k} p(x).$$

Vink: De transformerede variable $R_i = F(X_i)$ er ligefordelte på $(0, 1)$, og fordelingen af $R_{(k)} = F(X_{(k)})$ er derfor den, der blev udledt i Eksempel 6.3.9.

- 6.18 Vis, at to kontinuerte eller diskrete stokastiske variable X og Y er uafhængige hvis og kun hvis $E(g(X)h(Y)) = E(g(X))E(h(Y))$ for ethvert par af begrænsede reelle funktioner g og h .
- 6.19 Lad X_1 og X_2 være uafhængige, identiske fordelte (diskrete eller kontinuerte) stokastiske variable på $(0, \infty)$. Betragt følgende stokastiske variable

$$Y_1 = \frac{X_1}{X_1 + X_2} \quad \text{og} \quad Y_2 = \frac{X_2}{X_1 + X_2}.$$

- (a) Gør rede for, at Y_1 og Y_2 har veldefineret middelværdi.
- (b) Vis, at $E(Y_1) = E(Y_2)$
- (c) Vis endelig, at $E(Y_1) = \frac{1}{2}$.
- 6.20 (a) Lad X være en stokastisk variabel med middelværdi μ og varians σ^2 , og lad f være en to gange differentiabel funktion fra \mathbb{R} ind i \mathbb{R} . Ved at Taylor-udvikle f ses det, at

$$f(X) = f(\mu) + f'(\mu)(X - \mu) + R,$$

hvor R er et stokastisk restled. Benyt dette udtryk til at vise, at

$$\text{Var}(f(X)) = f'(\mu)\sigma^2 + R^*,$$

hvor

$$R^* = \text{Var}(R) + 2f'(\mu)E((X - \mu)R).$$

Hvis R^* er lille, har vi altså approksimationen

$$\text{Var}(f(X)) \simeq f'(\mu)\sigma^2. \quad (*)$$

Vis, at hvis $|f''(x)| \leq M$ for alle x , er

$$|R^*| \leq M^2 E((X - \mu)^4)/4 + |f'(\mu)|M|E((X - \mu)^3)|.$$

Vink: $R = \frac{1}{2}f''(Y)(X - \mu)^2$ for et Y mellem μ og X .

Approximationen (*) og approximationen (**) nedenfor er i fagene fysik og kemi kendt under det noget pompøse navn *fejlophobningsloven* (propagation of error) .

(b) Betragt k uafhængige stokastiske variable X_1, \dots, X_k med middelværdier μ_i og varianser σ_i^2 , $i = 1, \dots, k$. Vis, at man for enhver to gange differentiabel funktion $g : \mathbb{R}^k \mapsto \mathbb{R}$ har følgende approximation

$$\text{Var}(g(X_1, \dots, X_k)) \simeq \sum_{i=1}^k \sigma_i^2 \left(\frac{\partial f}{\partial x_i}(\mu_1, \dots, \mu_k) \right)^2, \quad (**)$$

forudsat at det tilsvarende restled er lille. Man kan naturligvis også vise et resultat for korrelerede stokastiske variable. Dette overlades til særligt interesserede studerende.

- 6.21 Lad X være ligefordelt i intervallet $(-1, 1)$, og definer $Y = X^2$. Vis, at X og Y er ukorrelerede. Er de uafhængige?
- 6.22 Lad X_1 og X_2 være uafhængige stokastiske variable, som begge er eksponentialfordelte med parameter λ .
- (a) Vis, at tætheden for den stokastiske vektor $(X_1, X_1 + X_2)$ er $p(y_1, y_2) = \lambda^2 e^{-\lambda y_2}$, når $y_2 > y_1 > 0$, og ellers er lig nul.
- (b) Vis, at den betingede fordeling af X_1 givet at $X_1 + X_2 = x$, hvor $x > 0$, er ligefordelingen på intervallet $(0, x)$.

Kapitel 7

Grænseresultater for stokastiske variable

I dette kapitel behandles to af sandsynlighedsregningens vigtigste resultater, som begge drejer sig om, hvordan gennemsnittet af n stokastiske variable opfører sig, når n går mod uendelig.

7.1 De store tals lov

Store tals lov siger, at for n ukorrelerede stokastiske variable med middelværdi og varians vil gennemsnittet med stor sandsynlighed ligge nær middelværdien, når n er stor. Dette resultat vises ved hjælp af Chebyshevs ulighed.

Sætning 7.1.1 (*Chebyshevs ulighed*). *Lad X være en diskret eller kontinuert stokastisk variabel med middelværdi μ og varians σ^2 . Da gælder, for ethvert $a > 0$, at*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}. \quad (7.1.1)$$

Bevis: Definer for $a > 0$ mængden

$$A = \{x \in \mathbb{R} \mid |x - \mu| \geq a\}.$$

Da er

$$(x - \mu)^2 1_A(x) \geq a^2 1_A(x)$$

for alle $x \in \mathbb{R}$. Dermed er

$$\begin{aligned}\sigma^2 &= \mathbb{E}((X - \mu)^2) \\ &= \mathbb{E}((X - \mu)^2 1_A(X)) + \mathbb{E}((X - \mu)^2 1_{\mathbb{R} \setminus A}(X)) \\ &\geq \mathbb{E}((X - \mu)^2 1_A(X)) \\ &\geq \mathbb{E}(a^2 1_A(X)) = a^2 P(X \in A) = a^2 P(|X - \mu| \geq a).\end{aligned}$$

Det andet ulighedstegn følger af (5.2.6)

□

Eksempel 7.1.2 Hvis X har middelværdi 0 og varians 1 er $P(|X| \geq 10) \leq 0.01$.

□

Vi kan nu uden større besvær bevise de store tals lov.

Sætning 7.1.3 (*De store tals lov*). Lad X_1, X_2, \dots være en følge af ukorrelerede diskrete eller kontinuerte stokastiske variable, som alle har samme middelværdi og varians. Hvis $\mu = E(X_i)$, gælder for ethvert $\epsilon > 0$ at

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| < \epsilon\right) \rightarrow 1 \quad (7.1.2)$$

for $n \rightarrow \infty$.

Bevis: Gennemsnittet $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ har middelværdi μ og varians

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n)) = \frac{\sigma^2}{n},$$

hvor $\sigma^2 = \text{Var}(X_i)$. Ved at anvende Chebyshevs ulighed (7.1.1) fås

$$\begin{aligned}1 &\geq P(|\bar{X}_n - \mu| < \epsilon) = 1 - P(|\bar{X}_n - \mu| \geq \epsilon) \\ &\geq 1 - \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2} \rightarrow 1,\end{aligned}$$

for $n \rightarrow \infty$.

□

Den eneste grund til at det i Sætningerne 7.1.1 og 7.1.3 kræves, at de stokastiske variable er diskrete eller kontinuerte, er at vi kun har defineret middelværdi og varians for disse to typer af fordelinger. Chebyshevs ulighed og store tals lov gælder for alle typer stokastiske variable, som har middelværdi og varians.

Sætning 7.1.3 er den enkleste version af de store tals lov. Der findes andre varianter, som vi ikke vil bevise her (se dog Opgave 7.7). F. eks. er det ikke nødvendigt at antage, at de stokastiske variable har varians. Det er nok, at $E(|X_i|) < \infty$, hvilket til gengæld klart nok er nødvendigt. I den situation antages uafhængighed i stedet for ukorrelerethed, men de stokastiske variable behøver faktisk ikke være ukorrelerede eller uafhængige; resultatet (7.1.2) gælder, hvis blot afhængigheden er tilstrækkeligt svag.

Vi har nu bevist, at *sandsynlighedsregningens frekvensfortolkning* holder. Lad nemlig Y_1, Y_2, \dots være en følge af uafhængige stokastiske variable med samme fordeling, og lad A være en delmængde af \mathbb{R} . Sæt $p = P(Y_i \in A)$. Da er $E(1_A(Y_i)) = p$ og

$$H_n = \frac{1_A(Y_1) + \dots + 1_A(Y_n)}{n}$$

er den relative hyppighed af udfald i A blandt de n første stokastiske variable. Da $1_A(Y_i)$ er begrænset, har den varians, så ifølge store tals lov gælder

$$P(|H_n - p| < \epsilon) \rightarrow 1$$

for $n \rightarrow \infty$ for ethvert $\epsilon > 0$.

Det skal også lige nævnes, at vi her har stiftet bekendtskab med et af sandsynlighedsregningens vigtigste begreber, *konvergens i sandsynlighed*. Hvis Y_1, Y_2, \dots er en følge af stokastiske variable, som opfylder, at der findes en stokastisk variabel Y , så $P(|Y_n - Y| < \epsilon) \rightarrow 1$ for $n \rightarrow \infty$ for ethvert $\epsilon > 0$, så siger man, at følgen $\{Y_n\}$ konvergerer i sandsynlighed mod Y . Store tals lov siger altså, at gennemsnittet af n uafhængige stokastiske variable med samme middelværdi og varians konvergerer i sandsynlighed mod den fælles middelværdi. Her er Y lig med en konstant, nemlig middelværdien.

7.2 Den centrale grænseværdisætning

Betragt en følge af uafhængige diskrete eller kontinuerte stokastiske variable X_1, X_2, \dots , som alle har samme fordeling med middelværdi μ og varians σ^2 . De store tals lov fortæller os da, at gennemsnittet af de n første stokastiske variable

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

er tæt på μ , når n er stor. Gennemsnittet er jo en stokastisk variabel, så det er naturligt at spørge, hvad dets fordeling er, når n er stor. Variansen er meget lille, så det varierer kun lidt i nærheden af μ . Imidlertid kan man studere gennemsnittets stokastiske variation ved at betragte den standardiserede variabel

$$U_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}. \quad (7.2.1)$$

Da $E(\bar{X}_n) = \mu$ og $\text{Var}(\bar{X}_n) = \sigma^2/n$, har U_n middelværdi nul og varians en. Vi har ved at gange med \sqrt{n}/σ forstørret den stokastiske variation af \bar{X}_n , så vi kan studere den nøjere. Der gælder følgende resultat.

Sætning 7.2.1 (*Den centrale grænseværdisætning*). *Lad X_1, X_2, \dots være en følge af uafhængige, identisk fordelte kontinuerte eller diskrete stokastiske variable med middelværdi μ og varians σ^2 . Da vil fordelingen af U_n , defineret ved (7.2.1), konvergere mod en standard normalfordeling, i den forstand at for alle $u \in \mathbb{R}$ vil*

$$P(U_n \leq u) \rightarrow \Phi(u), \quad (7.2.2)$$

når $n \rightarrow \infty$. Som sædvanlig betegner Φ fordelingsfunktionen for standard normalfordelingen givet ved (5.3.2).

Vi har på dette kursus ikke de nødvendige tekniske hjælpemidler til at bevise den centrale grænseværdisætning. Med de rette hjælpemidler er sætningen faktisk ikke svær at vise.

Der er mange gode grunde til, at normalfordelingen dukker op her. Hvis X_i -erne er normalfordelte, følger det af Sætning 6.3.12, at U_n er standard normalfordelt for alle n . Hvis sætningen skal være sand, må

$P(U_n \leq u)$ derfor nødvendigvis konvergere mod $\Phi(u)$. Bemærk, at vi kan slutte af Sætning 7.2.1, at standard normalfordelingen er den eneste fordeling med varians, som har den egenskab, at U_n har samme fordeling som X_i -erne for alle n . Det kan faktisk bevises, at den er den eneste fordeling overhovedet med denne egenskab.

Hvis vi definerer $S_n = X_1 + \dots + X_n$, kan resultatet (7.2.2) også skrives på formen

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq u\right) \rightarrow \Phi(u), \quad (7.2.3)$$

når $n \rightarrow \infty$. Vi kan altså også opfatte den centrale grænseværdisætning som et udsagn om fordelingen af summen S_n , når n er stor.

Den centrale grænseværdisætning giver en forklaring på, hvorfor det i praksis så ofte viser sig, at observationer kan antages at være normalfordelte. Vi kan f. eks. tænke os, at man skal måle en størrelse, som har værdien ξ , men at en række tilfældige forhold indvirker på målingen, så den ikke bliver helt nøjagtig. Det, der faktisk måles, kan derfor antages at være $Y = \xi + X_1 + \dots + X_n$, hvor X_i -erne er virkningen af de forskellige kilder til målefejl. Hvis X_i -erne opfylder betingelserne i den centrale grænseværdisætning, kan fordelingen af målingen Y tilnærmes med en normalfordeling med middelværdi $\xi + n\mu$ og varians $n\sigma^2$. Ved gode målinger er μ lig nul eller i hvert fald meget lille.

Der findes andre udgaver af den centrale grænseværdisætning med væsentligt svagere betingelser, som gør overstående argument for, at målefejl ofte er normalfordelte, mere overbevisende. De stokastiske variable X_i behøver således ikke at have samme fordeling; ikke engang samme middelværdi og varians. De behøver heller ikke at være uafhængige. Det væsentlige er, at de alle er små i forhold til summen S_n , og at afhængigheden mellem dem ikke er for stor.

Den form for konvergens, som udtrykkes ved (7.2.2) kaldes *konvergens i fordeling*. Et andet eksempel på denne form for konvergens så vi i Sætning 4.1.2. Udforskningen af konvergens i fordeling spiller en central rolle i sandsynlighedsregningen og dens anvendelser i statistik.

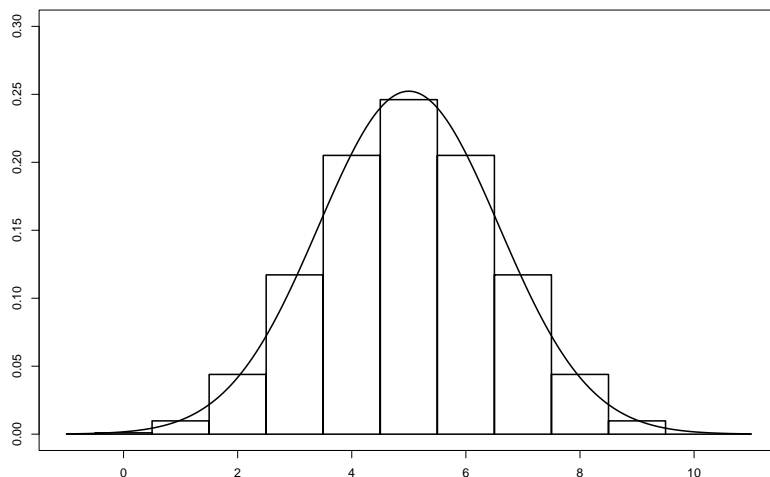
Ikke sjældent ser man, at et interval af "typiske værdier" for en stokastisk variabel angives som middelværdien plus/minus 2 gange spredningen. Hvis vi kan bruge den centrale grænseværdisætning til at argumentere for, at fordelingen af den stokastiske variable kan tilnærmes

godt med en normalfordeling, giver dette god mening. For en normalfordeling gælder nemlig, at 95 procent af sandsynligheden ligger inden for grænserne givet ved middelværdien ± 1.96 gange spredningen.

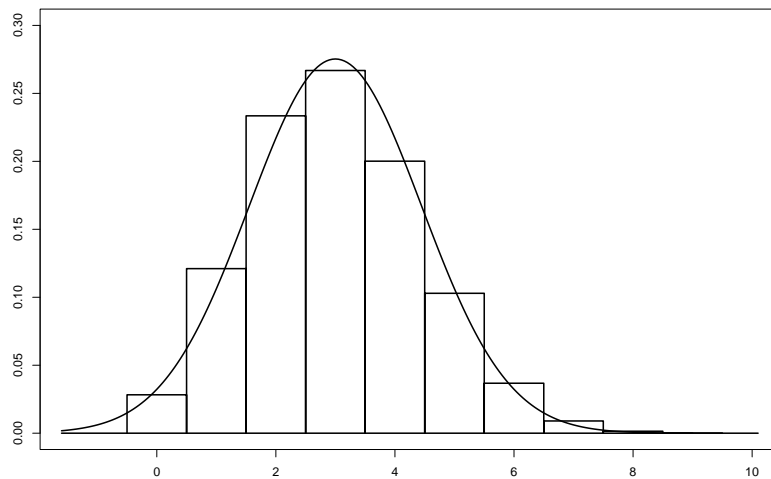
Den centrale grænseværdisætning kan benyttes til at vise, at nogle af de fordelinger, vi har beskæftiget os med, under visse omstændigheder kan tilnærmes med en normalfordeling. Det ser vi nærmere på i de følgende eksempler.

Eksempel 7.2.2 Lad X_1, X_2, \dots være uafhængige stokastiske variable med værdier i $\{0, 1\}$, som alle har samme fordeling givet ved at $P(X_i = 1) = p$ og $P(X_i = 0) = 1 - p$, hvor $p \in (0, 1)$. Sæt $S_n = X_1 + \dots + X_n$. Da $E(X_i) = p$ og $\text{Var}(X_i) = p(1 - p)$ (se Eksemplerne 3.7.2 og 3.7.11), gælder ifølge den centrale grænseværdisætning at

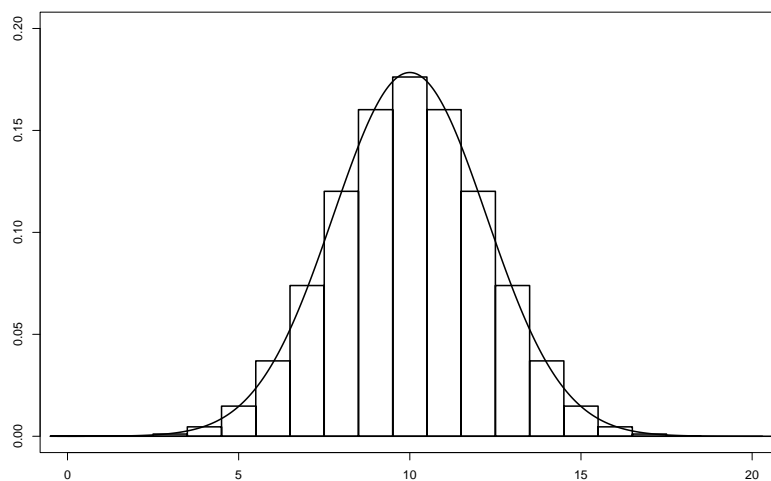
$$P\left(\frac{S_n - np}{\sqrt{np(1 - p)}} \leq u\right) \rightarrow \Phi(u),$$



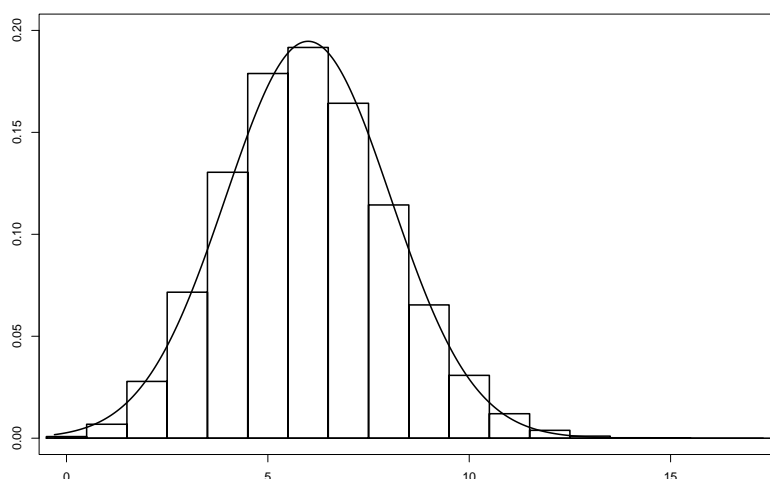
Figur 7.2.1: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 10 og sandsynlighedsparameter 0.5 tegnet som histogram og sandsynlighedstæthed for normalfordelingen med middelværdi 5 og varians 2.5.



Figur 7.2.2: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 10 og sandsynlighedsparameter 0.3 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 3 og varians 2.1.



Figur 7.2.3: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 20 og sandsynlighedsparameter 0.5 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 10 og varians 5.



Figur 7.2.4: Sandsynlighedsfunktionen for binomialfordelingen med antalsparameter 20 og sandsynlighedsparameter 0.3 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 6 og varians 4.2.

når $n \rightarrow \infty$. Da S_n er binomialfordelt med antalsparameter n og sandsynlighedsparameter p , kan vi konkludere, at denne fordeling, når n er stor, ligner en normalfordeling med middelværdi np og varians $np(1-p)$ i den forstand, at

$$P(S_n \leq u) = P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{u - np}{\sqrt{np(1-p)}}\right) \simeq \Phi\left(\frac{u - np}{\sqrt{np(1-p)}}\right).$$

Vi har derfor følgende approximation

$$P(S_n = i) \simeq \Phi\left(\frac{i + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{i - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

for $i \in \{0, \dots, n\}$. Approximationen er god, når blot $n \geq 10$ og $|p - \frac{1}{2}|$ ikke er for stor. Den centrale grænseværdisætning for specialtilfældet binomialfordelingen kaldes de Moivres sætning.

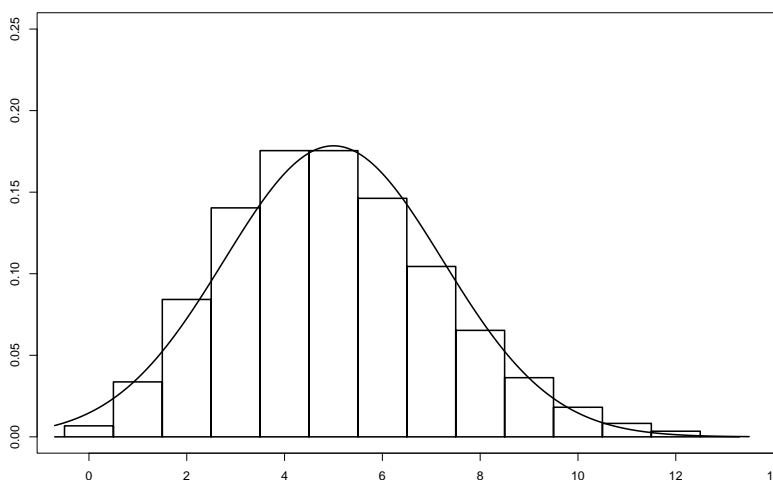
□

Eksempel 7.2.3 Antag, at X_1, X_2, \dots er uafhængige stokastiske variable, som alle er Poisson fordelte med parameter λ , og sæt $S_n = X_1 + \dots + X_n$. Da $E(X_i) = \text{Var}(X_i) = \lambda$ (se Afsnit 4.4), gælder ifølge den centrale grænseværdisætning at

$$P\left(\frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq u\right) \rightarrow \Phi(u),$$

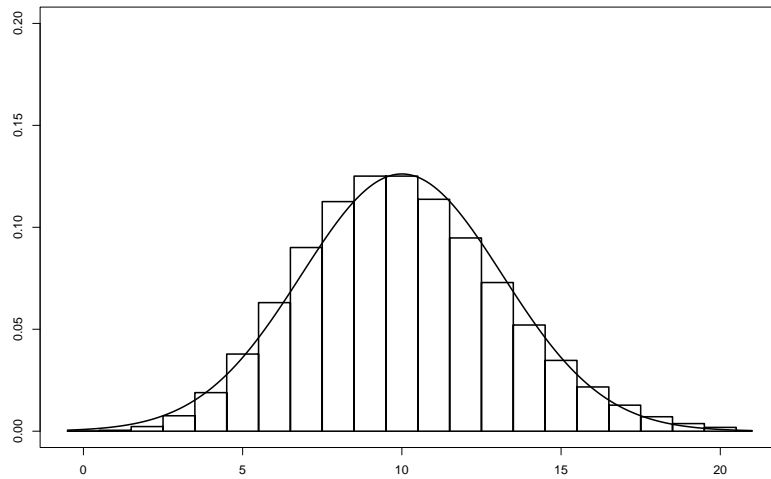
når $n \rightarrow \infty$. Da S_n er Poisson fordelt med parameter $n\lambda$ (se Sætning 4.3.4), kan vi konkludere, at denne fordeling, når n er stor, ligner en normalfordeling med middelværdi $n\lambda$ og varians $n\lambda$ i den forstand, at

$$P(S_n \leq u) \simeq \Phi\left(\frac{u - n\lambda}{\sqrt{n\lambda}}\right).$$

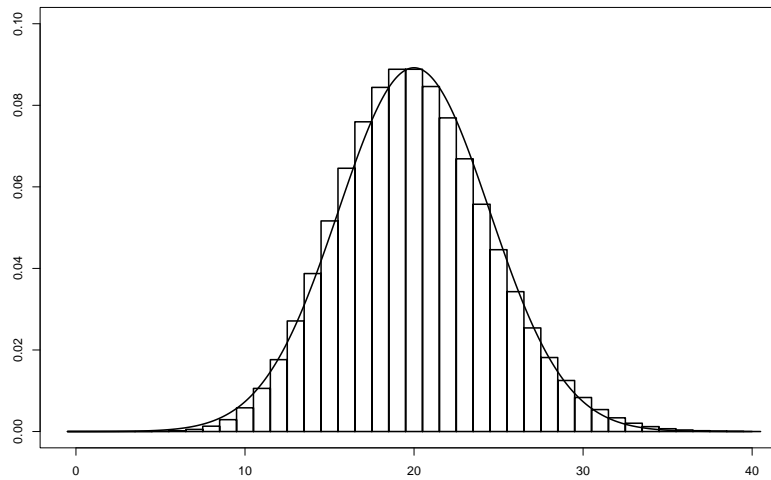


Figur 7.2.5: Sandsynlighedsfunktionen for Poissonfordelingen med parameter 5 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 5 og varians 5.

Vi kan også udtrykke dette resultat ved at sige, at en Poisson fordeling med parameter λ kan approksimeres med en normalfordeling med middelværdi λ og varians λ , når λ er stor. Hvis Y er Poisson fordelt med



Figur 7.2.6: Sandsynlighedsfunktionen for Poissonfordelingen med parameter 10 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 10 og varians 10.



Figur 7.2.7: Sandsynlighedsfunktionen for Poissonfordelingen med parameter 20 tegnet som histogram og sandsynlighedstætheden for normalfordelingen med middelværdi 20 og varians 20.

parameter λ , er altså

$$P(Y = i) \simeq \Phi\left(\frac{i + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{i - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

for $i \in \mathbb{N}$, når λ er stor.

□

7.3 Opgaver

- 7.1 Lad $p \in (0, 1)$ være givet. Find ved hjælp af Chebyshevs ulighed et tal $c > 0$, så der for en stokastisk variabel X gælder, at sandsynlighedsmassen i intervallet $E(X) \pm c\sqrt{\text{Var}(X)}$ er større end eller lig p . Hvad er svaret, når $p = 0.95$?
- 7.2 Betragt en stokastisk variabel $X = \sum_{i=1}^n Y_i$, hvor Y_i erne er uafhængige, identisk fordelte stokastiske variable, som har middelværdi og varians. Ifølge den centrale grænseværdisætning er X tilnærmelsesvist normalfordelt. Brug den approximerende normalfordeling til at give svar på spørgsmålene i Opgave 7.1.
- 7.3 Lad X være en stokastisk variabel, som er Poissonfordelt med parameter λ . Benyt Chebyshevs ulighed til at vise de følgende to uligheder:

$$P\left(X \leq \frac{\lambda}{2}\right) \leq \frac{4}{\lambda} \quad \text{og} \quad P(X \geq 2\lambda) \leq \frac{1}{\lambda}.$$

- 7.4 Lad X være en diskret stokastisk variabel med sandsynlighedsfunktion givet ved $p(1) = p(3) = 1/18$ og $p(2) = 8/9$. Vis, at der findes en værdi af $a > 0$, for hvilken $P(|X - E(X)| \geq a) = \text{Var}(X)/a^2$. Generelt kan Chebyshevs ulighed altså ikke forbedres.
- 7.5 Lad X_1, X_2, \dots være en følge af uafhængige, identisk fordelte stokastiske variable med middelværdi μ og varians σ^2 . Vis, at den empiriske varians s_n^2 for de første n variable (se opgave 3.26) konvergerer i sandsynlighed mod σ^2 , forudsat at $EX_1^4 < \infty$. Vink: Det

bliver lidt enklere, hvis man ved definitionen af s^2 benytter faktoren $\frac{1}{n}$ i stedet for $\frac{1}{n-1}$. I grænsen er dette uden betydning. Udtryk s^2 ved $\frac{1}{n} \sum X_i^2$ og $\bar{X} = \frac{1}{n} \sum X_i$.

- 7.6 En mønt kastes 1000 gange. Giv en approksimation til sandsynligheden for at få mindst 550 plat?
- 7.7 Lad X_1, X_2, \dots være en følge af ukorrelerede diskrete eller kontinuerte stokastiske variable, som alle har samme middelværdi μ . Antag, at der findes et reelt tal $K > 0$, så $\text{Var}(X_i) \leq K$ for alle $i \in \mathbb{N}$. Vis, at der for ethvert $\epsilon > 0$ gælder, at

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| < \epsilon\right) \rightarrow 1$$

for $n \rightarrow \infty$.

- 7.8 Lad Z_1, Z_2, \dots være en følge af uafhængige, identisk fordelte stokastiske variable med middelværdi 0 og varians σ^2 , og definer en følge af stokastiske variable ved $Y_0 = 0$ og

$$Y_i = \theta Y_{i-1} + Z_i, \quad i = 1, 2, \dots,$$

hvor $|\theta| < 1$. Vis, at $\text{Var}(Y_i) = \theta^2 \text{Var}(Y_{i-1}) + \sigma^2$ ($i = 1, 2, \dots$), og slut heraf at $\text{Var}(Y_i) \leq \sigma^2 / (1 - \theta^2)$ for alle i . Vis derefter at de stokastiske variable $X_i = Y_{i-1}(Y_i - \theta Y_{i-1}) = Y_{i-1} Z_i$, $i = 1, 2, \dots$ opfylder at

$$P\left(\left|\frac{X_1 + \dots + X_n}{n}\right| < \epsilon\right) \rightarrow 1$$

for $n \rightarrow \infty$ for ethvert $\epsilon > 0$.

Kapitel 8

Normalfordelingens teori

Dette kapitels første to afsnit behandler nogle fordelinger, som spiller en vigtig rolle i statistisk teori, ikke mindst i teorien for normalfordelte observationer. Dernæst behandles den fler-dimensionale normalfordeling.

8.1 χ^2 -fordelingen og Γ -fordelingen

Definition 8.1.1 Lad U_1, \dots, U_k være uafhængige standard normalfordelte stokastiske variable. Fordelingen af $U_1^2 + \dots + U_k^2$ kaldes da χ^2 -fordelingen med k frihedsgrader.

Sætning 8.1.2 Sandsynlighedstætheden for χ^2 -fordelingen med k frihedsgrader er

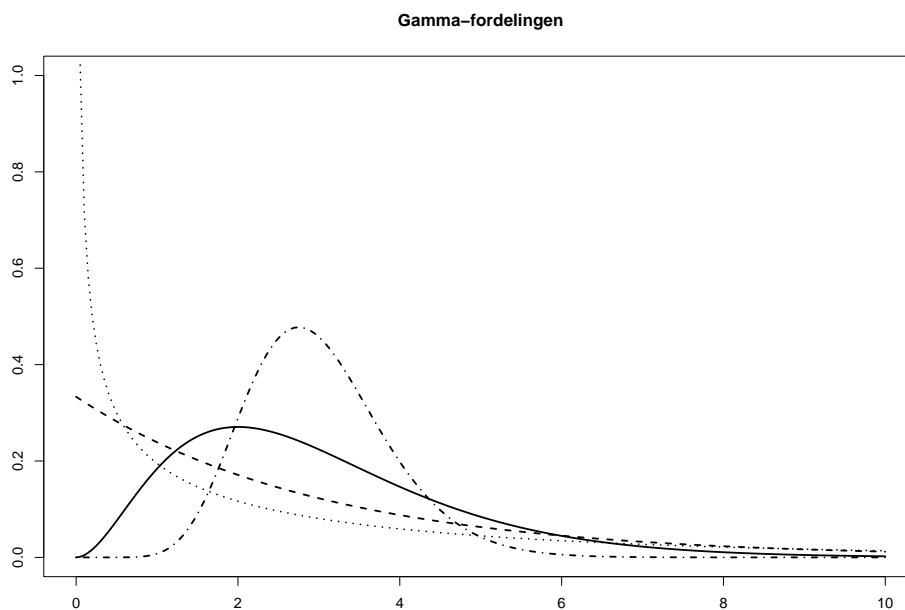
$$p(x) = \frac{x^{\frac{k}{2}-1} e^{-x/2}}{2^{\frac{k}{2}} c_k}, \quad (8.1.1)$$

når $x > 0$, og lig nul når $x \leq 0$. I denne formel er $c_1 = \sqrt{\pi}$, mens $c_k = (\frac{k}{2} - 1)!$ når k er lige og $c_k = (\frac{k}{2} - 1) \cdots \frac{1}{2} \sqrt{\pi}$, hvis $k = 3, 5, 7, \dots$

For $k = 1$ viste vi (8.1.1) i Eksempel 5.4.6, se (5.4.5). Før vi viser (8.1.1) for et generelt k , er det nyttigt at studere den mere generelle klasse af fordelinger med tæthed

$$p(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad (8.1.2)$$

for $x > 0$ og $p(x) = 0$ for $x \leq 0$. Her er $\alpha > 0$ og $\beta > 0$, og $\Gamma(\alpha)$ er en størrelse, som sikrer, at integralet af p over $(0, \infty)$ er lig en. Læseren bør overveje, hvorfor dette integral er endeligt. At $\Gamma(\alpha)$ ikke afhænger af β ses ved at skifte integrationsvariabel som vist nedenfor. Fordelingen med tæthed (8.1.2) kaldes Γ -fordelingen med formparameter α og skalaparameter β . At parameteren β faktisk er en skalaparameter vises i Opgave 8.1: Hvis X er Γ -fordelt med formparameter α og skalaparameter 1, er βX Γ -fordelt med formparameter α og skalaparameter β . For fast α har fordelingen derfor essentielt den samme form. For varierende værdier af α kan formen derimod være helt forskellig. For eksempel går $p(x)$ mod uendelig, β^{-1} eller nul for x gående mod nul, afhængig af om $\alpha < 1$, $\alpha = 1$ eller $\alpha > 1$. Vi ser, at χ^2 -fordelingen med k frihedsgrader er lig Γ -fordelingen med formparameter $k/2$ og skalaparameter 2. Bemærk, at eksponentialfordelingen også er et specialtilfælde svarende til $\alpha = 1$.



Figur 8.1.1: Γ -fordelingens sandsynlighedstæthed for $\alpha = 0.5$ og $\beta = 6$ (prikket kurve), $\alpha = 1$ og $\beta = 3$ (stiplet kurve), $\alpha = 3$ og $\beta = 1$ (fuldt optrukket kurve) og $\alpha = 12$ og $\beta = 0.25$ (prikket og stiplet kurve). Disse fordelinger har alle middelværdi tre.

Lad os først se nærmere på funktionen $\alpha \mapsto \Gamma(\alpha)$, som er defineret for $\alpha > 0$, og som kaldes Γ -funktionen. Da integralet af p skal være lig en, må

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x/\beta} \beta^{-\alpha} dx = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

Af dette udtryk ses det, at $\Gamma(\alpha)$ ikke afhænger af β . Endvidere følger det ved delvis integration, at

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^\infty y^\alpha e^{-y} dy \\ &= \left[y^\alpha (-e^{-y}) \right]_0^\infty - \int_0^\infty \alpha y^{\alpha-1} (-e^{-y}) dy \\ &= \alpha \Gamma(\alpha). \end{aligned}$$

Vi har dermed vist Γ -funktionens funktionalligning

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha). \quad (8.1.3)$$

Da

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1,$$

ses det ved gentagen brug af (8.1.3), at der for et positivt helt tal n gælder, at

$$\Gamma(n) = (n - 1)!$$

Vi har som nævnt vist, at q i (5.4.5) er en sandsynlighedstæthed, og da q er et specialtilfælde af (8.1.2) med $\alpha = \frac{1}{2}$ og $\beta = 2$, følger det, at

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Ved gentagen brug af (8.1.3) ses det derfor, at

$$\Gamma\left(\frac{k}{2}\right) = \left(\frac{k}{2} - 1\right) \cdots \frac{1}{2} \sqrt{\pi},$$

når k er et ulige helt tal større end eller lig 3. Vi har dermed indset, at (8.1.1) er en sandsynlighedstæthed, når c_k er defineret som beskrevet ovenfor.

Vi vil nu vise, at summen af uafhængige Γ -fordelte stokastiske variable med samme skalaparameter igen er Γ -fordelt. Dette resultat kaldes ofte Γ -fordelingens foldningsegenskab.

Sætning 8.1.3 Lad X_1, \dots, X_n være uafhængige stokastiske variable, hvor X_i er Γ -fordelt med formparameter α_i og skalaparameter β , $i = 1, \dots, n$. Da er $X_1 + \dots + X_n$ Γ -fordelt med formparameter $\alpha_1 + \dots + \alpha_n$ og skalaparameter β .

Bevis: Det generelle resultatet vises let ved induktion, når først sætningen er vist for $n = 2$. Vi behøver derfor kun at betragte tilfældet $n = 2$. Ifølge (6.3.2) er tætheden for $X_1 + X_2$

$$\begin{aligned} q(z) &= \int_0^z \frac{x^{\alpha_1-1} e^{-x/\beta} (z-x)^{\alpha_2-1} e^{-(z-x)/\beta}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} dx \\ &= \frac{e^{-z/\beta}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^z x^{\alpha_1-1} (z-x)^{\alpha_2-1} dx \\ &= \frac{z^{\alpha_1+\alpha_2-1} e^{-z/\beta}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^1 y^{\alpha_1-1} (1-y)^{\alpha_2-1} dy, \end{aligned}$$

hvor vi har foretaget substitutionen $y = x/z$. Det ses, at tætheden for $X_1 + X_2$ har samme form som tætheden for en Γ -fordeling med formparameter $\alpha_1 + \alpha_2$ og skalaparameter β , bortset fra at den ser ud til at være normeret anderledes. Sæt

$$c = \frac{\int_0^1 y^{\alpha_1-1} (1-y)^{\alpha_2-1} dy}{\Gamma(\alpha_1) \Gamma(\alpha_2)}.$$

Da vi ved, at q er en sandsynlighedstæthed, er

$$1 = \int_0^\infty q(z) dz = c \int_0^\infty \beta^{-(\alpha_1+\alpha_2)} z^{\alpha_1+\alpha_2-1} e^{-z/\beta} dz = c \Gamma(\alpha_1 + \alpha_2),$$

således at

$$c = \frac{1}{\Gamma(\alpha_1 + \alpha_2)}.$$

Dermed er sætningen vist for $n = 2$. Læseren kan nu selv gennemføre induktionsargumentet. □

Undervejs har vi vist, at

$$\int_0^1 y^{\alpha_1-1} (1-y)^{\alpha_2-1} dy = \frac{\Gamma(\alpha_1) \Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

Denne funktion af (α_1, α_2) kaldes Beta-funktionen. Bemærk, at funktionen

$$f(x) = \frac{y^{\alpha_1-1}(1-y)^{\alpha_2-1}\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, \quad x \in (0, 1),$$

er en sandsynlighedstæthed på $(0, 1)$. Den tilsvarende fordeling kaldes *Beta-fordelingen*. Vi studerede et specialtilfælde i Eksempel 5.1.4, men vil ikke beskæftige os yderligere med Beta-fordelingen på dette kursus, se dog Opgave 8.7.

Vi kan nu vende tilbage til χ^2 -fordelingen og vise Sætning 8.1.2, d.v.s. at χ^2 -fordelingens tæthed er givet ved (8.1.1).

Bevis for Sætning 8.1.2: I Eksempel 5.4.6 viste vi, at U_i^2 har sandsynlighedstæthed (5.4.5), altså at U_i^2 er Γ -fordelt med formparameter $1/2$ og skalaparameter 2. Det følger derfor af Sætning 8.1.3 at $U_1^2 + \dots + U_k^2$ er Γ -fordelt med formparameter $k/2$ og skalaparameter 2, og dermed har sandsynlighedstæthed (8.1.1). □

Det er let at finde middelværdien af χ^2 -fordelingen med k frihedsgrader: Da $E(U_i^2) = \text{Var}(U_i) = 1$, er $E(U_1^2 + \dots + U_k^2) = k$.

Hvis X er χ^2 -fordelt med f frihedsgrader, kaldes fordelingen af $\sigma^2 X$, hvor $\sigma^2 > 0$, for $\sigma^2 \chi^2$ -fordelingen med f frihedsgrader. Af diskussionen tidligere i dette afsnit fremgår det, at $\sigma^2 \chi^2$ -fordelingen med f frihedsgrader er lig Γ -fordelingen med formparameter $f/2$ og skalaparameter $2\sigma^2$. Hvis Y er $\sigma^2 \chi^2$ -fordelt med f frihedsgrader, og $Z = Y/f$, skriver man traditionelt $Z \sim \sigma^2 \chi^2/f$. Fordelingen af Z er en Γ -fordeling med formparameter $f/2$ og skalaparameter $2\sigma^2/f$.

8.2 F-fordelingen og t-fordelingen

Hvis Z_1 og Z_2 er uafhængige stokastiske variable, som er χ^2 -fordelte med henholdsvis f_1 og f_2 frihedsgrader, kaldes fordelingen af den stokastiske variable

$$X = \frac{Z_1/f_1}{Z_2/f_2}$$

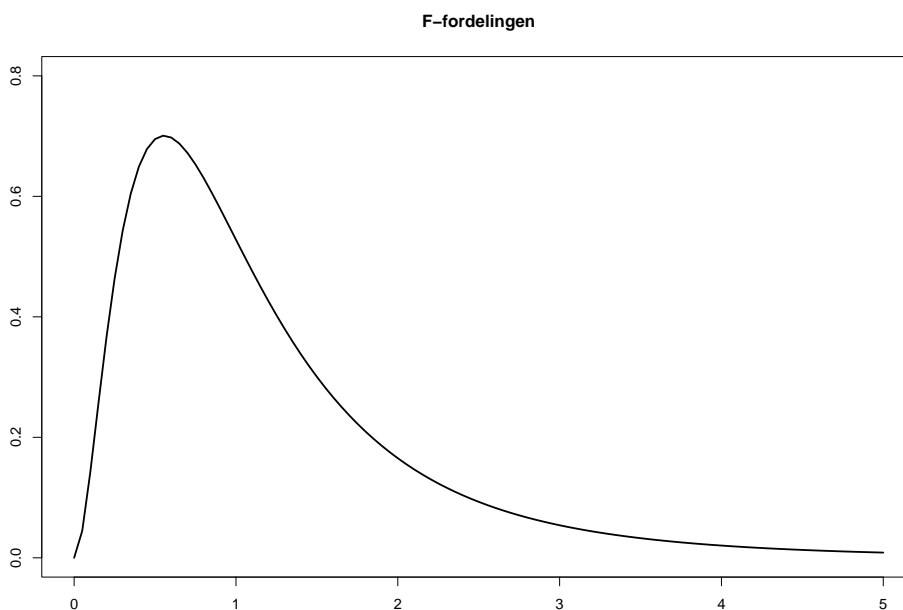
F-fordelingen med (f_1, f_2) frihedsgrader. Denne fordeling har sandsynlighedstæthed

$$p(x) = \frac{f_1^{f_1/2} f_2^{f_2/2} \Gamma((f_1 + f_2)/2) x^{f_1/2-1}}{\Gamma(f_1/2) \Gamma(f_2/2) (f_2 + f_1 x)^{(f_1+f_2)/2}}, \quad x > 0, \quad (8.2.1)$$

hvilket følger af Korollar 6.3.6. Ved at benytte (8.2.1) og sammenligningskriteriet i Sætning D.1.7 kan man indse, at X har middelværdi hvis og kun hvis $f_2 > 2$. Når denne betingelse er opfyldt, kan det vises, at

$$E(X) = \frac{f_2}{f_2 - 2}. \quad (8.2.2)$$

For at bevise (8.2.2) kan man bruge, at middelværdien af $1/Z_2$ er lig $1/(f_2 - 2)$, jfr. opgave 8.4.



Figur 8.2.1: Sandsynlighedstætheden for F-fordelingen med $(6,10)$ frihedsgrader

Bemærk til slut, at hvis U_1, \dots, U_n er uafhængige standard normalfordelte stokastiske variable og $k < n$, så er

$$\frac{(U_1^2 + \dots + U_k^2)/k}{(U_{k+1}^2 + \dots + U_n^2)/(n-k)}$$

F-fordelt med $(k, n-k)$ frihedsgrader.

Hvis U er en standard normalfordelt stokastisk variabel, der er uafhængig af den stokastiske variable Z , som er χ^2 -fordelt med f frihedsgrader, så kaldes fordelingen af

$$T = \frac{U}{\sqrt{Z/f}}$$

t-fordelingen med f frihedsgrader. Denne fordeling har sandsynlighedstætheden

$$p(t) = \frac{\Gamma((f+1)/2)}{\sqrt{\pi f} \Gamma(f/2) (1+t^2/f)^{(f+1)/2}}, \quad t \in \mathbb{R}, \quad (8.2.3)$$

hvilket følger af Korollar 6.3.7. Bemærk, at sandsynlighedstætheden er symmetrisk om nul.

Ved at benytte (8.2.3) og sammenligningskriteriet i Sætning D.1.7 kan det indses, at T har middelværdi hvis og kun hvis $f > 1$. I så fald er

$$E(T) = 0, \quad (8.2.4)$$

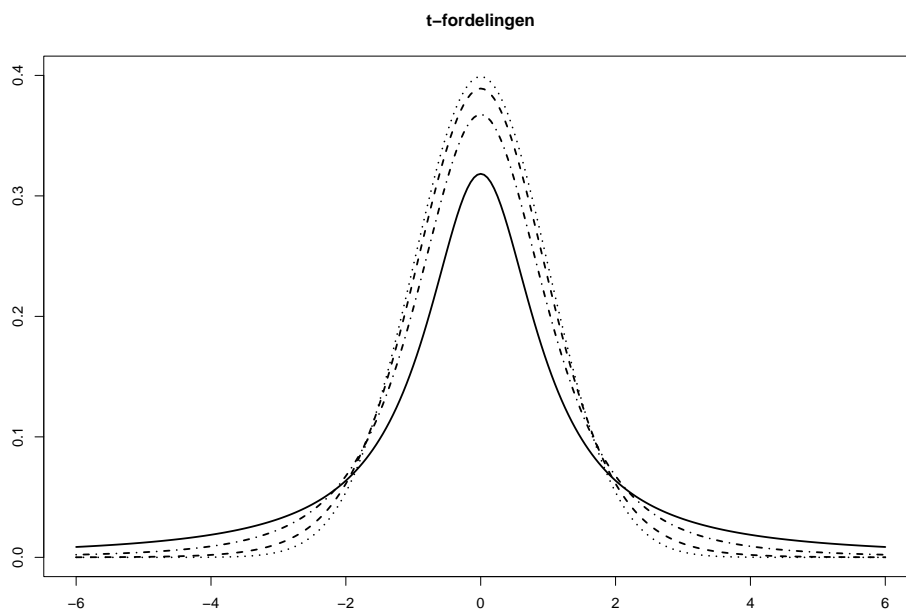
da (8.2.3) er symmetrisk om nul. Det kan på tilsvarende måde vises, at T har varians hvis og kun hvis $f > 2$. I så fald er (se opgave 8.4)

$$\text{Var}(T) = \frac{f}{f-2}. \quad (8.2.5)$$

Bemærk, at T^2 er F-fordelt med $(1, f)$ frihedsgrader. Når $f = 1$ er sandsynlighedstætheden for t-fordelingen

$$p(t) = \frac{1}{\pi(1+t^2)},$$

som vi genkender som sandsynlighedstætheden for Cauchy-fordelingen, se Eksempel 5.4.5. Når derimod f er stor ligner t-fordelingen standard



Figur 8.2.2: Sandsynlighedstætheden for t-fordelingen med 1 (fuldt optrukket kurve), 3 (prikket og stiplede kurve) og 10 frihedsgrader (stiplede kurve). Endvidere er standard normalfordelingens sandsynlighedstæthed indtegnet (prikket kurve).

normalfordelingen. Det har vi ikke redskaberne til at vise på dette kursus, men det er ikke overraskende. Hvis nemlig U_0, \dots, U_n er uafhængige standard normalfordelte stokastiske variable, er

$$T = \frac{U_0}{\sqrt{(U_1^2 + \dots + U_n^2)/n}}$$

t-fordelt med n frihedsgrader, og når n er stor, er $(U_1^2 + \dots + U_n^2)/n$ ifølge store tals lov tæt på $E(U_1^2) = 1$, således at $T \simeq U_0$.

8.3 Den fler-dimensionale normalfordeling

8.3.1 Den fler-dimensionale standard normalfordeling

Hvis U_1, \dots, U_n er uafhængige standard normalfordelte stokastiske variable, kaldes fordelingen af den stokastiske vektor (U_1, \dots, U_n) *den n -dimensionale standard normalfordeling*. Sandsynlighedstætheden for den n -dimensionale standard normalfordeling er ifølge Sætning 6.2.1

$$p(x_1, \dots, x_n) = \varphi(x_1) \cdots \varphi(x_n) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}, \quad (8.3.1)$$

for alle $(x_1, \dots, x_n) \in \mathbb{R}^n$. Som sædvanlig betegner φ sandsynlighedstætheden for den en-dimensionale standard normalfordeling, der er givet ved (5.3.1). Fordelingen med tæthed (8.3.1) kaldes også den normerede fler-dimensionale normalfordeling.

I dette underafsnit vil vi vise nogle resultater, som er vigtige i forbindelse med statistiske normalfordelingsmodeller. Det vil bl.a. vise sig, hvorledes χ^2 -fordelingen og t -fordelingen dukker op. Vi får brug for nogle resultater fra lineær algebra, som vi først vil minde om.

Vi vil altid opfatte elementerne i \mathbb{R}^n som søjlevektorer, d.v.s. som $n \times 1$ matrixer. Koordinaterne af $x \in \mathbb{R}^n$ betegnes med x_1, \dots, x_n . Når vi har brug for at skrive $x \in \mathbb{R}^n$ som en rækkevektor, benytter vi transponeringsoperatoren. Den transponerede af x er en $1 \times n$ matrix, der betegnes med x^t . Også den transponerede af en matrix M betegnes med M^t . Vi benytter det sædvanlige *indre produkt* i \mathbb{R}^n , det såkaldte prik-produkt. For x og $y \in \mathbb{R}^n$ er det givet ved

$$x \cdot y = x_1 y_1 + \dots + x_n y_n = x^t y,$$

hvor det sidste produkt er et matrixprodukt. Den tilsvarende norm er givet ved

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + \dots + x_n^2}.$$

Afstanden mellem to vektorer $x, y \in \mathbb{R}^n$ er således $\|x - y\|$.

Et sæt af vektorer $\{e_1, \dots, e_n\}$ kaldes en *ortonormalbasis* for \mathbb{R}^n , hvis

$$e_i \cdot e_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

En $n \times n$ -matrix M kaldes en *ortonormalmatrix*, hvis $M^t M = I$, hvor I betegner $n \times n$ -identitetsmatricen. Identitetsmatricen er den $n \times n$ matrix, hvis diagonalelementer er lig en, mens alle andre elementer er lig nul. En ortonormalmatrix kaldes somme tider en ortogonalmatrix. Det er let at indse, at M er en ortonormalmatrix hvis og kun hvis dens søjler udgør en ortonormalbasis for \mathbb{R}^n . Identiteten $M^t M = I$ viser, at en ortonormalmatrix er inverterbar (ikke-singulær), og at dens inverse matrix er lig den transponerede, d.v.s. $M^{-1} = M^t$. Da determinanten af M^{-1} er lig $(\det(M))^{-1}$, ser vi, at

$$\det(M)^2 = \det(M^t) \det(M) = (\det(M))^{-1} \det(M) = 1,$$

hvorfor

$$|\det(M)| = 1.$$

Bemærk også, at da $M^{-1} = M^t$, er $MM^t = I$, således at M^t også er en ortonormalmatrix. Vi ser, at M er en ortonormalmatrix hvis og kun hvis dens rækker udgør en ortonormalbasis for \mathbb{R}^n .

Den lineære funktion, som svarer til en ortonormalmatrix M (altså funktionen $x \mapsto Mx$), bevarer det indre produkt. For x og y i \mathbb{R}^n er

$$(Mx) \cdot (My) = (Mx)^t My = x^t M^t My = x^t y = x \cdot y.$$

Specielt bevarer funktionen også normen

$$\|Mx\| = \sqrt{(Mx) \cdot (Mx)} = \sqrt{x \cdot x} = \|x\|.$$

Vi vender nu tilbage til sandsynlighedsregningen og viser en sætning, som vi tidligere har vist i det specialtilfælde, hvor $n = 2$ og $\det(M) = 1$. Bemærk først, at sandsynlighedstætheden for den n -dimensionale standard normalfordeling kan skrives som

$$p(x) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\|x\|^2}, \quad x \in \mathbb{R}^n. \quad (8.3.2)$$

Sætning 8.3.1 *Lad U være en n -dimensional stokastisk vektor, som er standard normalfordelt, og lad M være en $n \times n$ ortonormalmatrix M . Da er den stokastiske vektor $V = MU$ også n -dimensionalt standard normalfordelt.*

Bevis: Da $|\det(M)| = 1 \neq 0$, følger det af Sætning 6.3.11 og af (8.3.2), at sandsynlighedstætheden for V er

$$q(y) = \frac{p(M^{-1}y)}{|\det(M)|} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\|M^t y\|^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\|y\|^2},$$

hvilket ifølge (8.3.2) netop er sandsynlighedstætheden for den n -dimensionale standard normalfordeling. □

Sætning 8.3.1 kan også udtrykkes således:

Korollar 8.3.2 *Lad U_1, \dots, U_n være uafhængige standard normalfordelte stokastiske variable, og lad $\{e_1, \dots, e_n\}$ være en ortonormalbasis for \mathbb{R}^n . Sæt $U = (U_1, \dots, U_n)^t$. Da er de stokastiske variable $V_i = e_i \cdot U$, $i = 1, \dots, n$, uafhængige og standard normalfordelte.*

Bevis: Lad M være ortonormalmatricen, som opfylder, at søjlerne i M^t er vektorerne e_1, \dots, e_n , og anvend Sætning 8.3.1. □

Sætning 8.3.3 *Lad X_1, \dots, X_n være uafhængige stokastiske variable, som alle er normalfordelte med middelværdi μ og varians σ^2 . Definer gennemsnittet \bar{X} og kvadratafvigelsessummen SSD ved*

$$\begin{aligned}\bar{X} &= \frac{1}{n}(X_1 + \dots + X_n), \\ \text{SSD} &= \sum_{i=1}^n (X_i - \bar{X})^2.\end{aligned}$$

Da er \bar{X} og SSD stokastisk uafhængige, \bar{X} er normalfordelt med middelværdi μ og varians σ^2/n og SSD er $\sigma^2\chi^2$ -fordelt med $n - 1$ frihedsgrader. Endelig er størrelsen

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\text{SSD}/(n - 1)}}$$

t-fordelt med $n - 1$ frihedsgrader.

Bevis: De stokastiske variable $U_i = (X_i - \mu)/\sigma$, $i = 1, \dots, n$ er uafhængige og standard normalfordelte. Sæt

$$e_1 = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)^t.$$

Da er $\|e_1\| = 1$. Man kan supplere e_1 med $n - 1$ vektorer e_2, \dots, e_n , så $\{e_1, \dots, e_n\}$ er en ortonormalbasis for \mathbb{R}^n . Ifølge Korollar 8.3.2 er de stokastiske variable $V_i = e_i \cdot U$, hvor $U = (U_1, \dots, U_n)^t$, uafhængige og standard normalfordelte. Sæt $V = (V_1, \dots, V_n)^t$, og lad M være ortonormalmatricen, som opfylder, at søjlene i M^t er vektorerne e_1, \dots, e_n . Da $X_i = \mu + \sigma U_i$, er

$$\begin{aligned} \bar{X} &= \mu + \sigma(U_1 + \dots + U_n)/n \\ &= \mu + \frac{\sigma}{\sqrt{n}} \left(\frac{1}{\sqrt{n}}U_1 + \dots + \frac{1}{\sqrt{n}}U_n \right) \\ &= \mu + \sigma V_1/\sqrt{n}. \end{aligned}$$

Med $\bar{U} = \frac{1}{n}(U_1 + \dots + U_n)$ er

$$\begin{aligned} \text{SSD} &= (X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \\ &= \sigma^2 \left((U_1 - \bar{U})^2 + \dots + (U_n - \bar{U})^2 \right) \\ &= \sigma^2 \left((U_1^2 + \bar{U}^2 - 2U_1\bar{U}) + \dots + (U_n^2 + \bar{U}^2 - 2U_n\bar{U}) \right) \\ &= \sigma^2 (U_1^2 + \dots + U_n^2 - n\bar{U}^2) \\ &= \sigma^2 \left(\|U\|^2 - \left(U_1/\sqrt{n} + \dots + U_n/\sqrt{n} \right)^2 \right) \\ &= \sigma^2 \left(\|M^t V\|^2 - V_1^2 \right) \\ &= \sigma^2 \left(\|V\|^2 - V_1^2 \right) \\ &= \sigma^2 (V_2^2 + \dots + V_n^2). \end{aligned}$$

Heraf ses det, at \bar{X} og SSD er uafhængige og har de påståede fordelinger. Videre er

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\text{SSD}/(n-1)}}$$

$$\begin{aligned}
&= \frac{\sigma\sqrt{n}\bar{U}}{\sigma\sqrt{(V_2^2 + \dots + V_n^2)/(n-1)}} \\
&= \frac{V_1}{\sqrt{(V_2^2 + \dots + V_n^2)/(n-1)}},
\end{aligned}$$

hvoraf det ses, at T er t-fordelt med $n - 1$ frihedsgrader. □

I statistisk sammenhæng er \bar{X} et estimat for μ og $s^2 = \text{SSD}/(n - 1)$ er et estimat for σ^2 . Størrelsen T benyttes til test af hypotesen om, at middelværdien har en bestemt værdi.

8.3.2 Den generelle fler-dimensionale normalfordeling

Lad den stokastiske vektor $U = (U_1, \dots, U_n)$ være n -dimensionalt standard normalfordelt. Den generelle n -dimensionale normalfordeling defineres da som fordelingen af en stokastisk vektor på formen

$$X = \mu + CU,$$

hvor $\mu \in \mathbb{R}^n$, og hvor C er en $n \times n$ matrix. Bemærk, at denne definition er analog til definitionen af en generel normalfordeling på \mathbb{R} , idet positionsparameteren μ blot er blevet til en vektor, og skalaparameteren er blevet erstattet med en matrix C .

Først bemærker vi, at hver koordinat af X er normalfordelt. Idet

$$X_i = \mu_i + c_{i1}U_1 + \dots + c_{in}U_n,$$

følger det af Sætning 6.3.12, at X_i er normalfordelt med middelværdi μ_i og varians $c_{i1}^2 + \dots + c_{in}^2$.

Vi kan nu finde sandsynlighedstætheden q for X , når C er inverterbar (ikke-singulær), ved hjælp af (6.3.13). Lad p være sandsynlighedstætheden for den n -dimensionale standard normalfordeling, som er givet ved (8.3.2). Da er

$$\begin{aligned}
q(x) &= \frac{p(C^{-1}(x - \mu))}{|\det(C)|} \\
&= \frac{1}{\sqrt{(2\pi)^n |\det(C)|}} \exp\left(-\frac{1}{2}\|C^{-1}(x - \mu)\|^2\right) \\
&= \frac{1}{\sqrt{(2\pi)^n |\det(C)|}} \exp\left(-\frac{1}{2}\left(C^{-1}(x - \mu)\right)^t \left(C^{-1}(x - \mu)\right)\right) \\
&= \frac{1}{\sqrt{(2\pi)^n |\det(C)|}} \exp\left(-\frac{1}{2}(x - \mu)^t (C^{-1})^t C^{-1}(x - \mu)\right) \\
&= \frac{1}{\sqrt{(2\pi)^n |\det(C)|}} \exp\left(-\frac{1}{2}(x - \mu)^t (CC^t)^{-1}(x - \mu)\right).
\end{aligned}$$

Man skriver som regel sandsynlighedstætheden ved hjælp af matricen $\Sigma = CC^t$. Da

$$(\det C)^2 = \det(C) \det(C^t) = \det(CC^t) = \det(\Sigma),$$

kan sandsynlighedstætheden for den generelle n -dimensionale normalfordeling skrives som

$$q(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1}(x - \mu)\right), \quad x \in \mathbb{R}^n. \quad (8.3.3)$$

Her svarer μ og Σ til middelværdi og varians i det en-dimensionale tilfælde. Som vi har set, er den i te koordinat af μ lig middelværdien af X_i . Matricen Σ er fordelingsens *kovariansmatrix*, d.v.s. den $n \times n$ -matrix, hvis (i, j) te element er kovariansen mellem X_i og X_j , når $i \neq j$, og hvis i te diagonalelement er $\text{Var}(X_i)$. Vi giver et noget komprimeret bevis for den sidste påstand. Da det (i, j) te element af $n \times n$ -matricen $(X - \mu)(X - \mu)^t$ er $(X_i - \mu_i)(X_j - \mu_j)$, er kovariansmatricen givet ved $E((X - \mu)(X - \mu)^t)$, hvor der tages middelværdi af alle matrixens elementer. Da $X - \mu = CU$, er

$$\begin{aligned}
E\left((X - \mu)(X - \mu)^t\right) &= E\left(CU(CU)^t\right) = E\left(CUU^tC^t\right) \\
&= CE\left(UU^t\right)C^t = CIC^t = CC^t = \Sigma,
\end{aligned}$$

hvor regneregler for middelværdien er blevet anvendt. Vi har også brugt, at kovariansmatricen for U , d.v.s. $E(UU^t)$ er lig identitetsmatricen. Vi går mere i detaljer med beregningen for det to-dimensionale tilfælde i næste delafsnit.

8.3.3 Den to-dimensionale normalfordeling

Lad U_1 og U_2 være uafhængige standard normalfordelte stokastiske variable. Den stokastiske vektor (X_1, X_2) givet ved

$$\begin{aligned} X_1 &= \mu_1 + c_{11}U_1 + c_{12}U_2, \\ X_2 &= \mu_2 + c_{21}U_1 + c_{22}U_2, \end{aligned} \quad (8.3.4)$$

er da to-dimensionalt normalfordelt. Under antagelse af, at matricen

$$C = \begin{Bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{Bmatrix}$$

er ikke-singulær, kan vi opskrive fordelings tæthed på samme måde som i det generelle tilfælde. Først vil vi dog modificere konstruktionen lidt. Vi kan antage, at

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad (8.3.5)$$

hvor V_1 og V_2 er to andre uafhængige standard normalfordelte stokastiske variable, og hvor α og β er vilkårlige tal med $\alpha^2 + \beta^2 = 1$. Da matricen

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$$

er en ortonormalmatrix, følger det af Sætning 8.3.1 (eller af beviset for Sætning 6.3.12), at U_1 og U_2 givet ved (8.3.5) er uafhængige og standard normalfordelte stokastiske variable, som de skal være. Vi kan for eksempel definere V_1 og V_2 ved

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}.$$

Ved at indsætte (8.3.5) i (8.3.4) fås

$$\begin{aligned} X_1 &= \mu_1 + (\alpha c_{11} - \beta c_{12})V_1 + (\beta c_{11} + \alpha c_{12})V_2, \\ X_2 &= \mu_2 + (\alpha c_{21} - \beta c_{22})V_1 + (\beta c_{21} + \alpha c_{22})V_2. \end{aligned}$$

Hermed har vi indset, at den samme to-dimensionale normalfordeling kan fremkomme på adskillige måder ved transformation af en standard normalfordeling. Vi kan lige så godt vælge koefficientmatricen på en måde, som letter de følgende overvejelser. Ved at vælge α og β , så vektoren (α, β) er vinkelret på vektoren (c_{12}, c_{11}) , d.v.s. så $\alpha c_{12} + \beta c_{11} = 0$, opnår vi en fremstilling, hvor den anden koefficient i første linie er 0. Med passende nye betegnelser har vi at

$$\begin{aligned} X_1 &= \mu_1 + aV_1, \\ X_2 &= \mu_2 + bV_1 + cV_2. \end{aligned}$$

Herefter kan vi glemme alt om den mere generelle fremstilling, vi begyndte med, idet enhver todimensional normalfordeling kan konstrueres på denne måde. Bemærk, at V_1 repræsenterer den stokastiske variation, som X_1 og X_2 har til fælles, mens V_2 er den stokastiske variation af X_2 , som intet har med X_1 at gøre.

Det ses umiddelbart, at X_i er normalfordelt med middelværdi μ_i og varians σ_i^2 , $i = 1, 2$, hvor

$$\sigma_1^2 = a^2 \quad \text{og} \quad \sigma_2^2 = b^2 + c^2.$$

Endvidere er

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E((X_1 - \mu_1)(X_2 - \mu_2)) = E(aV_1(bV_1 + cV_2)) \\ &= abE(V_1^2) + ac E(V_1V_2) = ab, \end{aligned}$$

således, at korrelationen mellem X_1 og X_2 , som vi vil betegne med ρ , er givet ved

$$\rho = \frac{ab}{|a|\sqrt{b^2 + c^2}}.$$

For at opskrive sandsynlighedstætheden for den to-dimensionale normalfordeling på formen (8.3.3), skal vi finde matricen

$$\Sigma = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Bemærk, at Σ er kovariansmatricen for (X_1, X_2) , som vist i forrige delafsnit. Da $\det(\Sigma) = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$, er Σ inverterbar, hvis vi forudsætter, at $|\rho| \neq 1$, og

$$\Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} \sigma_1^{-2} & -\rho \sigma_1^{-1} \sigma_2^{-1} \\ -\rho \sigma_1^{-1} \sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix},$$

Ved at indsætte i (8.3.3) fås, at sandsynlighedstætheden for den to-dimensionale normalfordeling er

$$\begin{aligned} q(x, y) &= \frac{\exp\left(-\frac{1}{2}(x - \mu_1, y - \mu_2)\Sigma^{-1}\begin{pmatrix} x - \mu_1 \\ y - \mu_2 \end{pmatrix}\right)}{\sqrt{(2\pi)^2 \det(\Sigma)}} \\ &= \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}. \end{aligned} \quad (8.3.6)$$

Niveaukurverne for denne funktion af to variable er koncentriske, liggedannede ellipser med centrum i (μ_1, μ_2) . Disse ellipsers form og orientering i forhold til akserne bestemmes af parametrene σ_1 , σ_2 og ρ . For $\rho = 0$ fås ellipser med hovedakser, der er parallelle med koordinataksene. Forholdet mellem længden af den vandrette og den lodrette akse er σ_1/σ_2 . For $\rho \neq 0$ fås ellipser, som ligger skævt i forhold til koordinatsystemet. Hældningen af den længste hovedakse har samme fortegn som ρ , og hovedreglen er, at værdier af ρ nær ± 1 svarer til langstrakte ellipser, som ligger tydeligt skævt i forhold til akserne. I tilfældet $\sigma_1 = \sigma_2$ har den længste hovedakse hældning ± 1 (med samme fortegn som ρ). Ellipsens "fladtrykthed" er i dette tilfælde givet ved at forholdet mellem længste og korteste hovedakse er $\sqrt{(1 + |\rho|)/(1 - |\rho|)}$. Det ekstreme tilfælde $\rho = \pm 1$, hvor kovariansmatricen Σ er singular, svarer til, at der er en lineær sammenhæng mellem X_1 og X_2 , således at fordelingen af (X_1, X_2) er koncentreret på en linie.

Korrelationens fortolkning som et mål for graden af samvariation mellem de to variable fremgår tydeligt af ovenstående. Bemærk også, at hvis korrelationen er lig nul, er

$$q(x, y) = \varphi\left((x - \mu_1)^2/\sigma_1^2\right) \varphi\left((y - \mu_2)^2/\sigma_2^2\right),$$

således at X_1 og X_2 er uafhængige. Når den simultane fordeling af (X_1, X_2) er en normalfordeling, er ukorrelerethed altså ensbetydende med uafhængighed. For simultant normalfordelte stokastiske variable, er korrelationen et nyttigt begreb, som er let at fortolke.

En alternativ beskrivelse af, hvordan korrelationen bestemmer samvariation mellem X_1 og X_2 , får man ved at se på, hvorledes den betingede fordeling af X_2 , givet $X_1 = x$, afhænger af x . Ifølge (6.5.2) er sandsynlighedstætheden for den betingede fordeling af X_2 givet at $X_1 = x$

$$\begin{aligned} & \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - 2\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}\exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right)/\sqrt{2\pi\sigma_1^2}} \\ &= \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left((1-\rho^2)\frac{(x-\mu_1)^2}{\sigma_1^2} + \left(\frac{(y-\mu_2)}{\sigma_2} - \rho\frac{(x-\mu_1)}{\sigma_1}\right)^2\right)\right)}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}\exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right)} \\ &= \frac{\exp\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y - (\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1))\right)^2\right)}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}. \end{aligned}$$

Vi ser, at denne betingede fordeling er en normalfordeling med middelværdi $\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$ og varians $\sigma_2^2(1 - \rho^2)$. Bemærk, at denne varians ikke afhænger af x . For $\rho \neq 0$ er variansen mindre end variansen σ_1^2 i den marginale fordeling af X_2 , hvilket intuitivt har den forklaring, at kendskab til X_1 sætter os i stand til at forudsige X_2 mere præcist end det er muligt uden kendskab til X_1 . Middelværdien $\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$ i den betingede fordeling kan fortolkes som det bedst mulige gæt på værdien af X_2 , baseret på en observeret værdi x af X_1 . Denne *betingede middelværdi* (som den kaldes) afhænger lineært af x . Dette er, ligesom den ovenfor bemærkede egenskab at variansen i den betingede fordeling er konstant, noget helt specielt for den to-dimensionale normalfordeling. Hældningen $\rho\frac{\sigma_2}{\sigma_1}$ af den linie, der beskriver X_2 's betingede middelværdi som funktion af x , kaldes *regressionskoefficienten*. Bemærk at den har samme fortegn som ρ , og for $\sigma_1 = \sigma_2$ simpelthen er lig med ρ . Heri ligger en klar bekræftelse af korrelationens relevans som et mål for graden af samvariationen for den todimensionale normalfordeling.

8.4 Sammenfatning

I Kapitel 8 har vi indført og studeret nogle fordelinger, der spiller en vigtig rolle i den statistiske teori, ikke mindst teorien for normalfordelte observationer. Det drejer sig om χ^2 -fordelingen, Γ -fordelingen (som χ^2 -fordelingen er et specialtilfælde af), F -fordelingen og t -fordelingen.

Videre har vi defineret den fler-dimensionale standard normalfordeling og set, at hvis en sådan fordeling transformeres med en ortonormalmatrix opnås en fordeling af samme slags. Dette resultat blev benyttet til at vise nogle vigtige fordelings- og uafhængighedsresultater for størrelser, der optræder i den statistiske teori for normalfordelte observationer. Endelig blev sandsynlighedstætheden for den generelle fler-dimensionale normalfordeling udledt, og den to-dimensionale normalfordeling blev studeret mere detaljeret.

8.5 Opgaver

- 8.1 Lad den stokastiske variable X være Γ -fordelt med formparameter α og skalaparameter β . Vis, at cX er Γ -fordelt med formparameter α og skalaparameter $c\beta$ for alle $c > 0$.
- 8.2 Lad X være Γ -fordelt med formparameter α og skalaparameter β . Vis, at $E(X) = \alpha\beta$ og $\text{Var}(X) = \alpha\beta^2$. Hvad er middelværdien og variansen af en χ^2 -fordelt stokastisk variabel?
- 8.3 Lad X være Γ -fordelt med formparameter α og skalaparameter β . Gør rede for, at

$$P(X \leq x) \simeq \Phi\left(\frac{x - \alpha\beta}{\sqrt{n\alpha\beta^2}}\right),$$

når α er stor. Som altid betegner Φ fordelingsfunktionen for standard normalfordelingen Vink: Den centrale grænseværdisætning.

- 8.4 Lad X være Γ -fordelt med formparameter α og skalaparameter β . Vis, at $1/X$ har middelværdi hvis og kun hvis $\alpha > 1$, samt at i så fald

$$E(1/X) = \frac{1}{(\alpha - 1)\beta}.$$

- 8.5 Vis (8.2.1) og (8.2.2). Vink til (8.2.1): Korollar 6.3.6. Vink til (8.2.2): Opgave 8.4.
- 8.6 Vis (8.2.3), (8.2.4) and (8.2.5). Vink til (8.2.3): Korollar 6.3.7. Vink til (8.2.5): Opgave 8.4.
- 8.7 Lad X_1 og X_2 være uafhængige gammafordelte stokastiske variable, hvor fordelingen af X_i har formparameter α_i , $i = 1, 2$, mens begge fordelinger har samme skalaparameter β .
- (a) Find den simultane sandsynlighedstæthed for $(X_1 + X_2, X_1)$.
- (b) Vis, at sandsynlighedstætheden for

$$Z = \frac{X_1}{X_1 + X_2}$$

er

$$p(z) = \frac{z^{\alpha_1-1}(1-z)^{\alpha_2-1}\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, \quad z \in (0, 1).$$

Fordelingen med denne sandsynlighedstæthed kaldes Beta-fordelingen.

(c) Antag, at $\alpha_1 = f_1/2$ og $\alpha_2 = f_2/2$, hvor f_1 og f_2 er hele tal. Vis, at

$$V = \frac{f_2 Z}{f_1(1-Z)}$$

er F-fordelt med f_1 og f_2 frihedsgrader.

Kapitel 9

Markovkæder med endeligt tilstandsrum

En følge af stokastiske variable $\{X_t\} = \{X_0, X_1, X_2, \dots\}$ kaldes en stokastisk proces. Vi kan nemlig tænke på de stokastiske variable som tilstanden til tidspunkterne $t = 0, 1, 2, \dots$ af et system, der udvikler sig tilfældigt. Det kan f.eks. være daglige målinger af temperaturen på Rådhuspladsen, månedlige observationer af markedsrenten på et tre-årigt lån, målinger hvert minut af vindhastigheden i Kastrup eller det årlige antal stormskader der indberettes til et forsikringsselskab. Mængden af mulige værdier af de stokastiske variable kaldes processens *tilstandsrum*. Den enkleste type stokastisk proces er den, hvor det antages at de stokastiske variable $\{X_t\}$ er uafhængige. En anden relativt enkel type stokastisk proces er Markovkæderne med endeligt tilstandsrum, som er emnet i dette kapitel.

9.1 Definition og eksempler

Om et isoleret dynamisk system bestående af N partikler gælder, at hvis man kender partiklernes positioner og hastigheder til et vist tidspunkt t_0 , så kan systemets tilstand på ethvert senere tidspunkt $t > t_0$ beregnes, og kendskab til systemets opførsel før tid t_0 er overflødig. En Markovkæde er en sandsynlighedsteoretisk model med en tilsvarende egenskab.

Definition 9.1.1 *Lad $\{X_t\} = \{X_0, X_1, X_2, \dots\}$ være en følge af stokastiske variable, der alle antager værdier i den endelige mængde $S =$*

$\{a_1, \dots, a_k\}$. Hvis der for alle $(i, j) \in \{1, \dots, k\}^2$ gælder, at

$$\begin{aligned} P(X_{t+1} = a_j \mid X_t = a_i, X_{t_m} = a_{i_m}, \dots, X_{t_1} = a_{i_1}) \\ = P(X_{t+1} = a_j \mid X_t = a_i) \end{aligned} \quad (9.1.1)$$

for alle $a_{i_1}, \dots, a_{i_m} \in S$, alle $0 \leq t_1 < \dots < t_m < t$ og alle $m \in \{1, \dots, t\}$, så kaldes følgen $\{X_t\}$ en Markovkæde med tilstandsrum S . Hvis yderligere

$$P(X_{t+1} = a_j \mid X_t = a_i) = P(X_1 = a_j \mid X_0 = a_i)$$

for alle $t \in \mathbb{N}$ og alle $(i, j) \in \{1, \dots, k\}^2$ kaldes Markovkæden homogen.

Vi betragter her kun homogene Markovkæder. Derfor udelades ordet homogen i det følgende, så ordet Markovkæde betegner altid en homogen Markovkæde. De betingede sandsynligheder

$$p_{ij} = P(X_1 = a_j \mid X_0 = a_i),$$

kaldes Markovkædens *overgangssandsynligheder*. Ligningen (9.1.1) kaldes *Markovegenskaben*. Den siger, at for en Markovkæde er den betingede sandsynlighed for at $X_{t+1} = a_j$ givet at $X_t = a_i$ lig p_{ij} , uanset hvad vi måtte have af ekstra information om værdierne af X_1, \dots, X_{t-1} .

Eksempel 9.1.2 *Bernoulli-Laplace modellen for diffusion*. Følgende model blev foreslået af Daniel Bernoulli i 1769 som en sandsynlighedsteoretisk analog til strømmen af to usammenlykkelige vædske mellem to beholdere. Modellen blev analyseret af Laplace i 1812.

Antag at vi har $2N$ partikler, hvoraf N er hvide og N er sorte. Partiklerne er fordelt i to beholdere med N i hver beholder. Mellem tidspunkterne $t-1$ og t trækkes der en partikel tilfældigt fra hver beholder og de to partikler bytter plads. Lad X_t være antallet af hvide partikler i den første beholder. Da er $\{X_t\}$ en Markovkæde med tilstandsrum $\{0, 1, \dots, N\}$ og overgangssandsynligheder givet ved

$$\begin{aligned} P(X_t = j - 1 \mid X_{t-1} = j) &= \left(\frac{j}{N}\right)^2 \\ P(X_t = j \mid X_{t-1} = j) &= \frac{2j(N-j)}{N^2} \\ P(X_t = j + 1 \mid X_{t-1} = j) &= \left(\frac{N-j}{N}\right)^2 \end{aligned}$$

$$P(X_t = k | X_{t-1} = j) = 0 \text{ hvis } |j - k| > 1,$$

hvor $j = 0, 1, \dots, N$. Overvej dette. Bemærk, at når $X_t = j$, er der $N - j$ sorte partikler i den første beholder, og henholdsvis $N - j$ og j hvide og sorte partikler i den anden beholder.

□

Eksempel 9.1.3 *En genetisk model.* I en population af N individer vil vi studere to allele gener A og a (A er dominant, mens a er recessivt (vigende)). Ethvert individ har netop to af disse gener og er af en af de følgende tre genotyper: AA , Aa , aa .

Vi følger populationen fra generation til generation. Antallet af individer, N , antages at være det samme i alle generationer. Lad X_t betegne antallet af A -gener i den t 'te generation. Da der er ialt $2N$ gener, er antallet af a -gener lig $2N - X_t$. Processen $\{X_t\}$ er en Markovkæde med tilstandrum $\{0, 1, \dots, 2N\}$. Hvis vi antager at der er tilfældig parring og ingen selektion i populationen, er

$$P(X_t = j | X_{t-1} = i) = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}.$$

Hvis altså $X_{t-1} = i$, er antallet af A -gener i næste generation binomialfordelt med antalsparameter $2N$ og sandsynlighedsparameter $i/(2N)$. Specielt er

$$\begin{aligned} P(X_t = 0 | X_{t-1} = 0) &= 1 \\ P(X_t = 2N | X_{t-1} = 2N) &= 1. \end{aligned}$$

□

Eksempel 9.1.4 *En kømodel.* En kø ved et ekspeditionssted antages at udvikle sig på følgende måde. Køen har N pladser. Imellem tid $t - 1$ og t ankommer et antal kunder Z_t , hvor Z_1, Z_2, \dots er uafhængige stokastiske variable, som alle er Poisson-fordelte med parameter $\lambda > 0$. Hvis køen er fuld, når en kunde ankommer, går kunden igen. Hvis der til tid t er kunder i køen, ekspederes den forreste kunde med sandsynlighed p , mens

der er sandsynlighed $1 - p$ for, at der ikke påbegyndes en ny ekspedition til tid t (for eksempel fordi den ekspedition, der startede til tid $t - 1$, ikke er afsluttet). Hvis der ikke er nogen i køen, sker der selvfølgelig ingen ekspedition. Om første kunde i køen til tid t ekspederes eller ej antages uafhængigt af de stokastiske variable Z_1, Z_2, \dots . Ordene ekspeditionssted og kunde skal ikke tages for bogstavligt. Ekspeditionsstedet kan f.eks. være en internetserver og kunderne kan være ankommande pakker, der står i kø for at blive ekspederet videre af serveren og tabes, hvis køen er fuld.

Lad X_t betegne antallet af kunder i køen til tid t lige før ekspeditionen af den forreste kunde påbegyndes (hvis dette sker). Da er $\{X_t\}$ en Markovkæde (overvej hvorfor). De mulige værdier af X_t er $\{0, 1, \dots, N\}$, som altså er tilstandsrummet. Lad os finde overgangssandsynlighederne. Vi vil først finde $P(X_t = j | X_{t-1} = i)$ for $j \geq i \geq 1$. Overgangen fra i til j kan ske på to måder. Enten bliver den forreste kunde i køen ekspederet til tid $t - 1$, og der ankommer $j - i + 1$ kunder mellem tid $t - 1$ og t , eller den forreste kunde i køen bliver ikke ekspederet til tid $t - 1$, og der ankommer $j - i$ kunder mellem tid $t - 1$ og t . D.v.s. for $N > j \geq i \geq 1$ er

$$\begin{aligned} P(X_t = j | X_{t-1} = i) &= pP(Z_t = j - i + 1) + (1 - p)P(Z_t = j - i) \\ &= p \frac{\lambda^{j-i+1}}{(j-i+1)!} e^{-\lambda} + (1 - p) \frac{\lambda^{j-i}}{(j-i)!} e^{-\lambda} \\ &= \frac{\lambda^{j-i}}{(j-i)!} e^{-\lambda} \left(1 - p + \frac{p\lambda}{j-i+1} \right), \end{aligned}$$

mens

$$\begin{aligned} P(X_t = N | X_{t-1} = i) &= pP(Z_t \geq N - i + 1) + (1 - p)P(Z_t \geq N - i) \\ &= P(Z_t \geq N - i + 1) + (1 - p)P(Z_t = N - i) \\ &= \sum_{j=N-i+1}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} + (1 - p) \frac{\lambda^{N-i}}{(N-i)!} e^{-\lambda}, \end{aligned}$$

for $N \geq i \geq 1$. Læseren bør ved tilsvarende overvejelser overbevise sig om, at de øvrige overgangssandsynligheder er som følger:

$$P(X_t = j | X_{t-1} = 0) = \frac{\lambda^j}{j!} e^{-\lambda} \quad \text{for } j = 0, 1, \dots, N - 1$$

$$\begin{aligned}
 P(X_t = N \mid X_{t-1} = 0) &= \sum_{j=N}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} \\
 P(X_t = i - 1 \mid X_{t-1} = i) &= pe^{-\lambda} \quad \text{for } i = 1, 2, \dots, N \\
 P(X_t = j \mid X_{t-1} = i) &= 0 \quad \text{for } 0 \leq j \leq i - 2.
 \end{aligned}$$

□

Eksempel 9.1.5 *En populationsmodel.* Lad X_t betegne antallet af individer i en population, der udvikler sig på følgende simple vis. Mellem tidspunkterne $t - 1$ og t får hvert individ i populationen uafhængigt af hinanden et enkelt afkom med sandsynlighed $p(X_{t-1})$, hvor

$$p(x) = \begin{cases} \alpha(1 - x/K) & \text{hvis } 0 \leq x \leq K \\ 0 & \text{ellers.} \end{cases}$$

Her er $K \in \mathbb{N}$ den største bæredygtige populationsstørrelse i det miljø, som populationen lever i, mens $\alpha \in [0, 1]$ for at sikre, at $p(x) \in [0, 1]$. Med sandsynlighed $1 - p(X_{t-1})$ får et individ ikke noget afkom mellem tidspunkterne $t - 1$ og t . Der kan naturligvis også ske dødsfald i populationen. For at forenkle modellen antages det, at der mellem tidspunkterne $t - 1$ og t dør netop et individ. Det antages at fødsler sker før dødsfald, så det individ, der dør i et bestemt tidsinterval, også har mulighed for at få et afkom i samme tidsinterval.

Disse antagelser kan formaliseres ved at antage, at

$$X_t = X_{t-1} + F_t - 1,$$

hvor den betingede fordeling af den stokastiske variable F_t givet $X_{t-1} = x$ er en binomialfordeling med antalsparameter x og sandsynlighedsparameter $p(x)$. Processen $\{X_t\}$ er tydeligvis en Markovkæde. Så længe $X_{t-1} \leq K - 1$, er den størst mulige værdi af X_t lig $2X_{t-1} - 1$, mens $X_t = X_{t-1} - 1$ hvis $X_{t-1} \geq K$. Derfor er tilstandsrummet $\{0, 1, \dots, 2K - 3\}$. Vi har følgende overgangssandsynligheder (overvej hvorfor). For $i = 1, 2, \dots, K -$

1 og $i - 1 \leq j \leq 2i - 1$ er

$$P(X_t = j | X_{t-1} = i) = \binom{i}{j-i+1} p(i)^{j-i+1} (1-p(i))^{2i-j-1},$$

mens

$$\begin{aligned} P(X_t = 0 | X_{t-1} = 0) &= 1 \\ P(X_t = i - 1 | X_{t-1} = i) &= 1 \quad \text{for } i = K, K + 1, \dots, 2K - 3. \end{aligned}$$

□

9.2 Nogle regneregler

I dette afsnit gennemgås nogle grundlæggende resultater for Markovkæder med endeligt tilstandsrum. Vi begynder med at finde fordelingen af den stokastiske vektor (X_0, X_1, \dots, X_n) . Som i forrige afsnit betegnes elementerne i tilstandsrummet med $S = \{a_1, \dots, a_k\}$.

Sætning 9.2.1 *Den simultane sandsynlighedsfunktion for den stokastiske vektor (X_0, X_1, \dots, X_n) er*

$$p(x_0, x_1, \dots, x_n) = q(x_0) \prod_{i=1}^k \prod_{j=1}^k p_{ij}^{n_{ij}}, \quad (9.2.1)$$

hvor q er sandsynlighedsfunktionen for X_0 , og hvor n_{ij} er lig antallet af overgange fra a_i til a_j i vektoren (x_0, x_1, \dots, x_n) , det vil sige

$$n_{ij} = \#\{(x_{t-1}, x_t) \mid x_{t-1} = a_i, x_t = a_j, t = 1, \dots, n\}. \quad (9.2.2)$$

Bevis: Ved at bruge Markovegenskaben (9.1.1) fås for ethvert $m \in \mathbb{N}$

$$\begin{aligned} &P(X_m = x_m, X_{m-1} = x_{m-1}, \dots, X_0 = x_0) \\ &= P(X_m = x_m \mid X_{m-1} = x_{m-1}, \dots, X_0 = x_0) \\ &\quad \cdot P(X_{m-1} = x_{m-1}, \dots, X_0 = x_0) \\ &= P(X_m = x_m \mid X_{m-1} = x_{m-1}) P(X_{m-1} = x_{m-1}, \dots, X_0 = x_0). \end{aligned}$$

Bruges dette resultat nu n gange (for $m = n, n - 1, \dots, 1$), ser vi at

$$\begin{aligned} p(x_0, x_1, \dots, x_n) &= P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= P(X_n = x_n | X_{n-1} = x_{n-1}) P(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &\quad \vdots \\ &= P(X_n = x_n | X_{n-1} = x_{n-1}) P(X_{n-1} = x_{n-1} | X_{n-2} = x_{n-2}) \\ &\quad \dots P(X_1 = x_1 | X_0 = x_0) P(X_0 = x_0). \end{aligned}$$

Resultatet (9.2.1) følger nu ved for hvert i og j at samle de faktorer, hvor $x_{t-1} = a_i$ og $x_t = a_j$, og hvor derfor $P(X_t = a_j | X_{t-1} = a_i) = p_{ij}$. \square

Matricen

$$\mathbf{P} = \{p_{ij}\}$$

($k \times k$ -matrix) kaldes Markovkædens *overgangsmatrix*. Bemærk at

$$\sum_{j=1}^k p_{ij} = 1, \quad (9.2.3)$$

da

$$\sum_{j=1}^k p_{ij} = \sum_{j=1}^k P(X_{t+1} = a_j | X_t = a_i) = P(X_{t+1} \in S | X_t = a_i) = 1$$

En $k \times k$ -matrix $\{p_{ij}\}$ er overgangsmatrix for en Markovkæde, hvis alle dens elementer er ikke-negative og (9.2.3) er opfyldt.

For ethvert $m \in \mathbb{N}$ defineres m -trins overgangssandsynlighederne ved

$$p_{ij}^{(m)} = P(X_{t+m} = a_j | X_t = a_i) \quad (9.2.4)$$

og m -trins overgangsmatricen ved

$$\mathbf{P}^{(m)} = \{p_{ij}^{(m)}\}.$$

Specielt er $\mathbf{P}^{(1)} = \mathbf{P}$. Ved hjælp af m -trins overgangssandsynlighederne kan vi finde sandsynlighedsfunktionen for X_m .

Sætning 9.2.2 Sandsynlighedsfunktionen for X_t er givet ved

$$p_t(a_j) = \sum_{i=1}^k p_{ij}^{(t)} q(a_i), \quad \text{for } j = 1, \dots, k, \quad (9.2.5)$$

hvor q er sandsynlighedsfunktionen for X_0 . Hvis p_t betegner rækkevektoren $(p_t(a_1), \dots, p_t(a_k))$ og q rækkevektoren $(q(a_1), \dots, q(a_k))$ kan (9.2.5) også udtrykkes således:

$$p_t = q\mathbf{P}^{(t)}.$$

Bevis: Ifølge (1.4.3) er

$$P(X_t = a_j) = \sum_{i=1}^k P(X_t = a_j \mid X_0 = a_i) P(X_0 = a_i) = \sum_{i=1}^k p_{ij}^{(t)} q(a_i).$$

□

Næste sætning siger, at Markovegenskaben også gælder for m -trins overgange. Det følger af sætningen, at $\{X_{tm}\} = \{X_0, X_m, X_{2m}, X_{3m} \dots\}$ er en Markovkæde.

Sætning 9.2.3 For alle $a_i, a_j, a_{i_1}, \dots, a_{i_n} \in S$ og alle $0 \leq t_1 < \dots < t_n < t$ er

$$P(X_{t+m} = a_j \mid X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) = p_{ij}^{(m)}. \quad (9.2.6)$$

Bevis: For $m = 2$ er

$$\begin{aligned} & P(X_{t+2} = a_j \mid X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) \\ &= P(X_{t+2} = a_j, X_{t+1} \in S \mid X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) \\ &= \sum_{j_1=1}^k P(X_{t+2} = a_j, X_{t+1} = a_{j_1} \mid X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) \\ &= \sum_{j_1=1}^k P(X_{t+2} = a_j \mid X_{t+1} = a_{j_1}, X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) \cdot \\ & \quad P(X_{t+1} = a_{j_1} \mid X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}), \end{aligned}$$

hvor vi til sidst har brugt formelen $P(A \cap B | C) = P(A | B \cap C)P(B | C)$ (eftersig denne formel, som bruges gang på gang i teorien for Markovkæder). Anvender vi nu Markovegenskaben (9.1.1), fås

$$\begin{aligned}
 & P(X_{t+2} = a_j | X_t = a_i, X_{t_n} = a_{i_n}, \dots, X_{t_1} = a_{i_1}) \\
 &= \sum_{j_1=1}^k P(X_{t+2} = a_j | X_{t+1} = a_{j_1})P(X_{t+1} = a_{j_1} | X_t = a_i) \\
 &= \sum_{j_1=1}^k P(X_{t+2} = a_j | X_{t+1} = a_{j_1}, X_t = a_i)P(X_{t+1} = a_{j_1} | X_t = a_i) \\
 &= \sum_{j_1=1}^k P(X_{t+2} = a_j, X_{t+1} = a_{j_1} | X_t = a_i) \\
 &= P(X_{t+2} = a_j | X_t = a_i) = p_{ij}^{(2)}.
 \end{aligned}$$

For generelt m følger sætningen nu ved induktion i m under brug af tilsvarende regnerier. □

Det viser sig, at det er let at finde m -trinsovergangssandsynligheder ved hjælp af matrixmultiplikation.

Sætning 9.2.4 For alle $m, n \in \mathbb{N}$ gælder, at

$$\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)} \cdot \mathbf{P}^{(n)}, \quad (9.2.7)$$

hvor \cdot betegner matrixmultiplikation. For alle $m \in \mathbb{N}$ er

$$\mathbf{P}^{(m)} = \mathbf{P}^m. \quad (9.2.8)$$

Specielt er

$$p_{ij}^{(m)} \geq p_{i\ell_1} p_{\ell_1 \ell_2} \cdots p_{\ell_{m-2} \ell_{m-1}} p_{\ell_{m-1} j} \quad (9.2.9)$$

for alle $\ell_1, \dots, \ell_{m-1} \in \{1, \dots, k\}$.

Bevis:

$$\begin{aligned}
 p_{ij}^{(m+n)} &= P(X_{m+n} = a_j | X_0 = a_i) = P(X_{m+n} = a_j, X_m \in S | X_0 = a_i) \\
 &= \sum_{\ell=1}^k P(X_{m+n} = a_j, X_m = a_\ell | X_0 = a_i)
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{\ell=1}^k P(X_{m+n} = a_j | X_m = a_\ell, X_0 = a_i) P(X_m = a_\ell | X_0 = a_i) \\
&= \sum_{\ell=1}^k p_{\ell j}^{(n)} p_{i \ell}^{(m)},
\end{aligned}$$

hvor vi har brugt Markovegenskaben (9.2.6). Ved at anvende (9.2.7) m gange, fås

$$\mathbf{P}^{(m)} = \mathbf{P} \cdot \mathbf{P}^{(m-1)} = \dots = \underbrace{\mathbf{P} \cdot \dots \cdot \mathbf{P}}_m.$$

Endelig følger (9.2.9) af (9.2.8), som netop siger at $p_{ij}^{(m)}$ er lig summen af alle de mulige værdier af højresiden af (9.2.9), og dermed større end eller lig et enkelt led i denne sum af ikke-negative størrelser. \square

Ligningerne (9.2.7) kaldes *Chapman-Kolmogorov ligningerne*.

Eksempel 9.2.5 Betragt en Markovkæde med to tilstande. Da har overgangsmatricen formen

$$\mathbf{P} = \begin{Bmatrix} 1-p & p \\ \alpha & 1-\alpha \end{Bmatrix}, \quad (9.2.10)$$

hvor $p, \alpha \in [0, 1]$, og

$$\mathbf{P}^2 = \begin{Bmatrix} (1-p)^2 + p\alpha & p(2-p-\alpha) \\ \alpha(2-p-\alpha) & p\alpha + (1-\alpha)^2 \end{Bmatrix}.$$

For eksempel er $p_{12}^{(2)} = p(2-p-\alpha)$. Vi antager, at p og α ikke begge er lig nul. Ved induktion kan det vises, at

$$\mathbf{P}^m = \frac{1}{p+\alpha} \begin{Bmatrix} \alpha & p \\ \alpha & p \end{Bmatrix} + \frac{(1-p-\alpha)^m}{p+\alpha} \begin{Bmatrix} p & -p \\ -\alpha & \alpha \end{Bmatrix}, \quad (9.2.11)$$

for alle $m \in \mathbb{N}$. For eksempel er altså $p_{12}^{(m)} = p(1 - (1-p-\alpha)^m)/(p+\alpha)$. \square

Vi slutter dette afsnit med et resultat, som mere klart viser, at den betingede sandsynlighed for at $X_{t+m} = a_j$ givet at $X_t = a_i$ er lig $p_{ij}^{(m)}$, helt uanset hvad vi måtte vide om værdierne af X_1, \dots, X_{t-1} .

Sætning 9.2.6 For alle delmængder $A_\ell \subseteq S$, $\ell = 1, \dots, n$ og alle $0 \leq t_1 < \dots < t_n < t$ er

$$P(X_{t+m} = a_j \mid X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1) = p_{ij}^{(m)}. \quad (9.2.12)$$

Bevis:

$$\begin{aligned} & P(X_{t+m} = a_j \mid X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1) \\ &= \frac{P(X_{t+m} = a_j, X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)}{P(X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)} \\ &= \frac{\sum_{\ell=1}^n \sum_{x_\ell \in A_\ell} P(X_{t+m} = a_j, X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1)}{P(X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)} \\ &= \frac{p_{ij}^{(m)} \sum_{\ell=1}^n \sum_{x_\ell \in A_\ell} P(X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1)}{P(X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)} \\ &= p_{ij}^{(m)} \frac{P(X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)}{P(X_t = a_i, X_{t_n} \in A_n, \dots, X_{t_1} \in A_1)} = p_{ij}^{(m)}. \end{aligned}$$

Det tredje lighedstegn følger, fordi

$$\begin{aligned} & P(X_{t+m} = a_j, X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1) \\ &= P(X_{t+m} = a_j \mid X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1) \\ &\quad \cdot P(X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1) \\ &= p_{ij}^{(m)} P(X_t = a_i, X_{t_n} = x_n, \dots, X_{t_1} = x_1), \end{aligned}$$

hvor vi har brugt Markovegenskaben (9.2.6). □

9.3 Ergodicitet og stationær fordeling

I dette afsnit behandles det helt centrale begreb en stationær fordeling, som spiller en vigtig rolle i mange anvendelser af teorien for Markovkæder. For eksempel er det den grundlæggende idé i søgemaskinen Google. Vi begynder med et par definitioner.

Man siger, at en tilstand a_i fører til en anden tilstand a_j (og skriver $a_i \rightarrow a_j$), hvis der eksisterer et $m \in \mathbb{N}$, så $p_{ij}^{(m)} > 0$, altså hvis der er

positiv sandsynlighed for at komme til a_j på et eller andet tidspunkt, givet at vi starter i tilstanden a_i . Det følger af (9.2.7), at

Hvis $a_i \rightarrow a_j$ og $a_j \rightarrow a_\ell$, så gælder at $a_i \rightarrow a_\ell$.

En Markovkæde kaldes *irreducibel*, hvis enhver tilstand fører til enhver anden, d.v.s. hvis $a_i \rightarrow a_j$ for alle $i, j = 1, \dots, k$. En tilstand a_i kaldes *absorberende*, hvis $p_{ii} = 1$, fordi dermed $p_{ij} = 0$ for alle $j \neq i$. Det følger af (9.2.9), at $p_{ii}^{(m)} = 1$ for alle $m \in \mathbb{N}$. Dermed er $p_{ij}^{(m)} = 0$ for alle $j \neq i$. Hvis en Markovkæde har en absorberende tilstand, kan den altså ikke være irreducibel.

Definition 9.3.1 *En irreducibel Markovkæde kaldes ergodisk, hvis der eksisterer $p_j \geq 0$, så*

$$\lim_{m \rightarrow \infty} p_{ij}^{(m)} = p_j \quad (9.3.1)$$

for alle $i, j = 1, \dots, k$.

Bemærk, at grænseværdien er uafhængig af i . Da $p_{ij}^{(t)}$ er sandsynlighedsfunktionen for den betingede fordeling af X_t givet at $X_0 = i$, betyder (9.3.1), at fordelingen af X_t hverken afhænger af t eller begyndelsestilstanden i , når t er tilstrækkeligt stor.

Sætning 9.3.2 *Lad $\{X_t\}$ være en ergodisk Markovkæde med grænseværdier p_j , $j = 1, \dots, k$. Da er*

$$\sum_{j=1}^k p_j = 1 \quad (9.3.2)$$

og

$$\lim_{t \rightarrow \infty} P(X_t = a_j) = p_j \quad (9.3.3)$$

for $j = 1, \dots, k$.

Bevis:

$$\sum_{j=1}^k p_j = \sum_{j=1}^k \lim_{m \rightarrow \infty} p_{ij}^{(m)} = \lim_{m \rightarrow \infty} \sum_{j=1}^k p_{ij}^{(m)} = 1.$$

Hvis q betegner sandsynlighedsfunktionen for X_0 , så følger det af (9.2.5), at

$$\lim_{t \rightarrow \infty} P(X_t = a_j) = \sum_{i=1}^k \lim_{t \rightarrow \infty} p_{ij}^{(t)} q(a_i) = \sum_{i=1}^k p_j q(a_i) = p_j \sum_{i=1}^k q(a_i) = p_j.$$

□

Eksempel 9.2.5 (fortsat). For en Markovkæde med to tilstande med overgangsmatrix (9.2.10), hvor p og α ikke begge er lig nul, er m -trins overgangsmatricen lig

$$\mathbf{P}^{(m)} = \frac{1}{p + \alpha} \begin{Bmatrix} \alpha & p \\ \alpha & p \end{Bmatrix} + \frac{(1 - p - \alpha)^m}{p + \alpha} \begin{Bmatrix} p & -p \\ -\alpha & \alpha \end{Bmatrix}.$$

Hvis p og α heller ikke begge er lig en, er $|1 - p - \alpha| < 1$, så

$$\lim_{m \rightarrow \infty} \mathbf{P}^m = \frac{1}{p + \alpha} \begin{Bmatrix} \alpha & p \\ \alpha & p \end{Bmatrix}.$$

Hvis p og α begge er forskellige fra nul, og de ikke begge er lig en, er denne Markovkæde irreducibel og ergodisk med

$$p_1 = \frac{\alpha}{p + \alpha} \quad \text{og} \quad p_2 = \frac{p}{p + \alpha}.$$

Ligning (9.3.2) er klart opfyldt.

□

Definition 9.3.3 *En fordeling på tilstandsrummet S kaldes stationær, hvis dens sandsynlighedsfunktion π opfylder, at*

$$\sum_{i=1}^k \pi(a_i) p_{ij} = \pi(a_j), \quad j = 1, \dots, k. \quad (9.3.4)$$

Ifølge (9.2.5) siger (9.3.4), at hvis X_0 har sandsynlighedsfunktion π , så har også X_1 sandsynlighedsfunktion π . Eller mere generelt, hvis X_t har sandsynlighedsfunktion π , så har også X_{t+1} sandsynlighedsfunktion π . Hvis π betegner rækkevektoren $(\pi(a_1), \dots, \pi(a_k))$, kan (9.3.4) skrives som

$$\pi \mathbf{P} = \pi, \quad (9.3.5)$$

hvoraf det ved multiplikation med \mathbf{P} fra venstre og ved at anvende (9.2.8) ses at

$$\pi \mathbf{P}^{(m)} = \pi \quad (9.3.6)$$

for alle $m \in \mathbb{N}$. Den sidste ligning kan mere detaljeret skrives som følger

$$\sum_{i=1}^k \pi(a_i) p_{ij}^{(m)} = \pi(a_j), \quad j = 1, \dots, k, \quad m \in \mathbb{N}. \quad (9.3.7)$$

Hvis der findes en stationær fordeling, kan man finde den ud fra overgangssandsynlighederne ved at løse ligningssystemet (9.3.4) suppleret med ligningen $\sum_{i=1}^k \pi(a_i) = 1$ med hensyn til $(\pi(a_1), \dots, \pi(a_k))$.

Den næste sætning belyser, hvorfor det er rimeligt at kalde en fordeling, som opfylder (9.3.4), stationær.

Sætning 9.3.4 *Lad π være en stationær fordeling, og antag at fordelingen af X_0 har sandsynlighedsfunktion π . Da er*

$$P(X_t = a_j) = \pi(a_j) \quad \text{for alle } t \in \mathbb{N} \text{ og } j = 1, \dots, k, \quad (9.3.8)$$

og den simultane fordeling af den stokastiske vektor $(X_t, X_{t+1}, \dots, X_{t+n})$ er for alle $t \in \mathbb{N}$ og alle $n \in \mathbb{N}$ lig fordelingen af (X_0, X_1, \dots, X_n) .

Bevis: (9.3.8) følger umiddelbart af (9.2.5) og (9.3.7). Helt på samme måde som beviset for Sætning 9.2.1, ses det at sandsynlighedsfunktionen for $(X_t, X_{t+1}, \dots, X_{t+n})$ er

$$P(X_{t+n} = x_n, X_{t+n-1} = x_{n-1}, \dots, X_t = x_0) = P(X_t = x_0) \prod_{i=1}^k \prod_{j=1}^k p_{ij}^{n_{ij}},$$

hvor n_{ij} er givet ved (9.2.2). Da $P(X_t = x_0) = \pi(x_0)$, er dette ifølge (9.2.1) lig sandsynlighedsfunktionen for (X_0, X_1, \dots, X_n) . □

En stokastisk process $\{X_t\}$, som opfylder at den simultane fordeling af $(X_t, X_{t+1}, \dots, X_{t+n})$ er lig fordelingen af (X_0, X_1, \dots, X_n) for alle $t \in \mathbb{N}$ og alle $n \in \mathbb{N}$, kaldes en *stationær proces*.

Eksempel 9.1.2 (fortsat). Fordelingen givet ved

$$\pi(j) = \frac{\binom{N}{j}^2}{\binom{2N}{N}}, \quad j = 0, \dots, N$$

er stationær fordeling for Bernoulli-Laplace modellen for diffusion. Dette ses af følgende regneri:

$$\begin{aligned} \sum_{i=0}^N \pi(i)p_{ij} &= \pi(j-1)p_{j-1,j} + \pi(j)p_{j,j} + \pi(j+1)p_{j+1,j} \\ &= \left(\frac{N-j+1}{N}\right)^2 \frac{\binom{N}{j-1}^2}{\binom{2N}{N}} + 2\frac{j(N-j)}{N^2} \frac{\binom{N}{j}^2}{\binom{2N}{N}} + \left(\frac{j+1}{N}\right)^2 \frac{\binom{N}{j+1}^2}{\binom{2N}{N}} \\ &= \frac{\binom{N-1}{j-1}^2 + 2\binom{N-1}{j-1}\binom{N-1}{j} + \binom{N-1}{j}^2}{\binom{2N}{N}} \\ &= \frac{\left(\binom{N-1}{j-1} + \binom{N-1}{j}\right)^2}{\binom{2N}{N}} = \frac{\binom{N}{j}^2}{\binom{2N}{N}} = \pi(j). \end{aligned}$$

Til slut brugte vi formel (B.3.2) (Pascals trekant). Bemærk at den stationære fordeling er en hypergeometrisk fordeling med parametre $2N$, N og N .

□

Sætning 9.3.5 *Antag Markovkæden $\{X_t\}$ er ergodisk med grænseværdier p_j , $j = 1, \dots, k$. Da eksisterer der en entydig stationær fordeling, som er givet ved $\pi(a_j) = p_j$, $j = 1, \dots, k$. Endvidere er $\pi(a_j) > 0$ for alle $j = 1, \dots, k$.*

Bevis: Først vises eksistensen af en stationær fordeling. Ifølge (9.2.7) er

$$p_{ij}^{(n+1)} = \sum_{\ell=1}^k p_{i\ell}^{(n)} p_{\ell j}.$$

Heraf følger, at

$$p_j = \lim_{n \rightarrow \infty} p_{ij}^{(n+1)} = \sum_{\ell=1}^k \lim_{n \rightarrow \infty} p_{i\ell}^{(n)} p_{\ell j} = \sum_{\ell=1}^k p_{\ell} p_{\ell j},$$

hvoraf det ses, at fordelingen med sandsynlighedsfunktion $\pi(a_j) = p_j$, $j = 1, \dots, k$, er stationær.

Dernæst skal det vises, at den stationære fordeling er entydig. Antag derfor at r er sandsynlighedsfunktion for en stationær fordeling, og lad $\{X_t^*\}$ være en Markovkæde med overgangssandsynligheder $\{p_{ij}\}$ og med begyndelsesfordeling $P(X_0^* = a_j) = r(a_j)$ for $j = 1, \dots, k$. Ifølge Sætning 9.3.4 er

$$P(X_n^* = a_j) = r(a_j), \quad j = 1, \dots, k,$$

for alle $n \in \mathbb{N}$, så ved at anvende Sætning 9.3.2 ser vi, at

$$r(a_j) = \lim_{n \rightarrow \infty} P(X_n^* = a_j) = p_j.$$

Altså er $r = \pi$.

Til slut vises det, at $p_j > 0$ for $j = 1, \dots, k$. Bemærk, at da $\sum_{j=1}^k p_j = 1$, må der findes et j_0 , så $p_{j_0} > 0$. Lad videre i være vilkårlig. Da en ergodisk Markovkæde er irreducibel, findes der et $m \in \mathbb{N}$, så $p_{j_0 i}^{(m)} > 0$. Dermed er

$$p_i = \sum_{j=1}^k p_{ji}^{(m)} p_j \geq p_{j_0 i}^{(m)} p_{j_0} > 0.$$

□

Eksempel 9.2.5 (fortsat). Lad os checke at grænsefordelingen givet ved

$$\pi(1) = \frac{\alpha}{p + \alpha} \quad \text{og} \quad \pi(2) = \frac{p}{p + \alpha}. \quad (9.3.9)$$

er den stationære fordeling for en Markovkæde med to tilstande. Dette ses af at

$$\pi(1)p_{11} + \pi(2)p_{21} = \frac{\alpha}{p + \alpha}(1 - p) + \frac{p}{p + \alpha}\alpha = \frac{\alpha}{p + \alpha} = \pi(1)$$

$$\pi(1)p_{12} + \pi(2)p_{22} = \frac{\alpha}{p + \alpha}p + \frac{p}{p + \alpha}(1 - \alpha) = \frac{p}{p + \alpha} = \pi(2)$$

□

Eksempel 9.1.3 (fortsat). Den genetiske model er ikke nogen irreducibel Markovkæde, da både 0 og $2N$ er absorberende tilstande. For denne Markovkæde er fordelingen givet ved

$$\pi_\theta(0) = \theta, \quad \pi_\theta(2N) = 1 - \theta, \quad \pi_\theta(j) = 0, \quad j = 1, \dots, 2N - 1,$$

en stationær fordeling for ethvert $\theta \in [0, 1]$. Der er altså uendeligt mange stationære fordelinger. Og ingen af dem opfylder det sidste udsagn i Sætning 9.3.5. At π_θ er stationær ses af at

$$\sum_{i=0}^{2N} \pi_\theta(i)p_{ij} = \theta p_{0j} + (1 - \theta)p_{2N,j}.$$

For $0 < j < 2N$ er

$$\theta p_{0j} + (1 - \theta)p_{2N,j} = 0 = \pi_\theta(j).$$

For $j = 0, 2N$ er

$$\begin{aligned} \theta p_{00} + (1 - \theta)p_{2N,0} &= \theta p_{00} = \theta = \pi_\theta(0) \\ \theta p_{0,2N} + (1 - \theta)p_{2N,2N} &= (1 - \theta)p_{2N,2N} = 1 - \theta = \pi_\theta(2N). \end{aligned}$$

Man kan vise, at fra et vist stokastisk tidspunkt vil Markovkæden være konstant lig enten 0 eller $2N$, således at det ene gen fuldstændigt forsvinder fra populationen. Den betingede sandsynlighed for at ende i tilstanden $2N$ givet at $X_0 = i$ er lig $i/(2N)$. Dermed er den betingede sandsynlighed for at ende i tilstanden 0 lig $1 - i/(2N)$. Bemærk at grænseværdierne af $p_{ij}^{(m)}$ således afhænger af starttilstanden i , modsat (9.3.1).

□

En irreducibel Markovkæde er ergodisk, hvis den opfylder en ekstra betingelse, som er givet i følgende definition.

Definition 9.3.6 *En irreducibel Markovkæde kaldes aperiodisk, hvis der for ethvert $i, j = 1, \dots, k$ findes et m_0 , så*

$$p_{ij}^{(m)} > 0 \quad \text{for alle } m \geq m_0.$$

Bemærk, at det på grund af det sidste resultat i Sætning 9.3.5 er en nødvendig forudsætning for, at en irreducibel Markovkæde er ergodisk, at den er aperiodisk. At det også er en tilstrækkelig betingelse, fremgår af det følgende hovedresultat om ergodicitet af Markovkæder med endeligt tilstandsrum.

Sætning 9.3.7 *En irreducibel og aperiodisk Markovkæde $\{X_t\}$ er ergodisk, og der eksisterer et $c > 0$ og et $\rho \in (0, 1)$, så*

$$|p_{ij}^{(t)} - \pi(a_j)| \leq c\rho^t \quad (9.3.10)$$

og

$$|P(X_t = a_j) - \pi(a_j)| \leq c\rho^t \quad (9.3.11)$$

for alle $i, j = 1, \dots, k$ og alle $t \in \mathbb{N}$, hvor π betegner sandsynlighedsfunktionen for den stationære fordeling.

Bemærk, at Sætningen 9.3.7 ikke kun fortæller os, at $\{X_t\}$ er ergodisk, men også at konvergenen mod $\pi(a_j)$ i (9.3.1) og (9.3.3) går eksponentielt hurtigt. Det betyder, at man i praksis kan benytte approximationen $P(X_t = a_j) \simeq \pi(a_j)$, selv når t ikke er særligt stor.

Før vi beviser Sætning 9.3.7, vender vi lige tilbage til nogle af eksemplerne for at illustrere nogle pointer.

Eksempel 9.2.5 (fortsat). Vi har tidligere ved direkte beregninger set, at m -trins overgangssandsynlighederne konvergerer for en Markovkæde med to tilstande med overgangsmatrix (9.2.10), forudsat at p og α ikke begge er nul, og heller ikke begge er en. Hvis et af disse tal er lig nul, er den tilsvarende tilstand absorberende, og Markovkæden er derfor ikke irreducibel. Hvis $p > 0$ og $\alpha > 0$, og hvis de ikke begge er lig en, ses det af (9.2.11), at Markovkæden er irreducibel og aperiodisk.

Hvis $p = \alpha = 1$, er overgangsmatricen

$$\mathbf{P} = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix},$$

d.v.s. processen springer frem og tilbage mellem de to tilstande. I dette tilfælde er processen irreducibel, men den er ikke aperiodisk. For eksempel er $p_{11}^{(m)} = 0$ for alle ulige m . Da m -trins overgangsmatricen ifølge (9.2.11) er lig

$$\mathbf{P}^{(m)} = \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix} + (-1)^m \begin{Bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{Bmatrix},$$

ser vi, at $\mathbf{P}^{(m)}$ ikke konvergerer i dette tilfælde. Processen er altså ikke ergodisk, hvad der heller ikke var grund til at forvente.

□

Eksempel 9.1.2 (fortsat). Lad os eftervise, at Bernoulli-Laplace modellen for diffusion er irreducibel og aperiodisk. For $N > j > i$ er ifølge (9.2.9)

$$p_{ij}^{(n+j-i)} \geq p_{i,i+1} p_{i+1,i+2} \cdots p_{j-1,j} p_{jj}^n > 0$$

for alle $n \in \mathbb{N}$. Alle andre kombinationer af i og j klares på lignende vis, således at vi kan konkludere, at

$$p_{ij}^{(m)} > 0 \text{ for alle } m > N.$$

Processen er dermed både irreducibel og aperiodisk, således at (9.3.11) sikrer, at når t er stor nok, er

$$P(X_t = j) \simeq \frac{\binom{N}{j}^2}{\binom{2N}{N}}, \quad j = 0, \dots, N$$

Den fysiske fortolkning af dette resultat er, at uanset hvor mange hvide partikler der oprindeligt var i den første beholder, så vil der med tiden (og

ret hurtigt med mindre der startes i en meget ekstrem tilstand) indstille sig en statistisk ligevægt, hvor antallet af hvide partikler i den første beholder er hypergeometrisk fordelt med parametre $2N$, N og N . Nu vil N i praksis være et meget stort tal, så ifølge Sætning 3.3.4 er fordelingen af X_t godt tilnærmet af en binomialfordeling med antalsparameter N og sandsynlighedsparameter $1/2$. Videre kan vi, igen fordi N er meget stor, slutte af den centrale grænseværdisætning, at fordelingen af X_t også approksimeres godt af en normalfordeling med middelværdi $N/2$ og varians $N/4$.

□

Eksempel 9.1.4 (fortsat). Kømodellen er irreducibel og aperiodisk. Da $p_{ij} > 0$ for $j \geq i$, giver (9.2.9) at

$$p_{ij}^{(m)} \geq p_{ii}^{m-1} p_{ij} > 0 \quad \text{for alle } m \in \mathbb{N},$$

når $j \geq i$. Da $p_{j,j-1} > 0$ for $j = 1, \dots, N$, gælder endvidere, at

$$p_{j,j-k}^{(n+k)} \geq p_{jj}^n p_{j,j-1} \cdots p_{j-k+1,j-k} > 0$$

for alle $n \in \mathbb{N}_0$, $k = 1, \dots, j$ og $j = 1, \dots, N$. Vi har dermed vist, at

$$p_{ij}^{(m)} > 0 \quad \text{for alle } m \geq N,$$

således at Markovkæden er irreducibel og aperiodisk. Den er derfor ergodisk, og der eksisterer en entydig stationær fordeling, men et eksplicit udtryk for denne fordeling er særdeles kompliceret. For konkrete værdier af parametrene λ og p kan den stationære fordeling, som vi har set, findes ved at løse et lineært ligningssystem. Ved dimensionering af køsystemer, spiller den stationære fordeling π en vigtig rolle. Man vil naturligvis gerne sikre, at $\pi(N)$ er meget lille, da det er sandsynligheden for at køen er fuld, således at kunderne går uden at handle eller internetpakker går tabt. En kunde vil sætte pris på at $\pi(0)$ er stor, mens forretningen måske synes det er spild af arbejdskraft.

Når N er meget lille, er udtrykket for den stationære fordeling til at håndtere. For $N = 1$ har kømodellen kun to tilstande, hvorfor vi allerede har fundet udtrykket (9.3.9) for den stationære fordeling. Da overgangsmatricen for kømodellen med $N = 1$ er

$$\mathbf{P} = \begin{Bmatrix} e^{-\lambda} & 1 - e^{-\lambda} \\ pe^{-\lambda} & 1 - pe^{-\lambda} \end{Bmatrix},$$

så den stationære fordeling er givet ved

$$\pi(0) = \frac{pe^{-\lambda}}{1 - e^{-\lambda} + pe^{-\lambda}} \quad \text{og} \quad \pi(1) = \frac{1 - e^{-\lambda}}{1 - e^{-\lambda} + pe^{-\lambda}}.$$

□

Vi skal nu i gang med at bevise Sætning 9.3.7. Det gør vi ved hjælp af den såkaldte *koblingsmetode*. Til det formål lader vi $\{X_t\}$ og $\{Y_t\}$ være to uafhængige Markovkæder med samme overgangssandsynligheder $\{p_{ij}\}$, men med forskellige begyndelsesbetingelser. Det antages at $P(X_0 = a_i) = q_1(a_i)$ og $P(Y_0 = a_i) = q_2(a_i)$ for $i = 1, \dots, k$, hvor q_1 og q_2 er sandsynlighedsfunktioner på S . Koblingsmetoden benytter det første tidspunkt, hvor de to Markovkæder er i samme tilstand. Definer derfor den stokastiske variable

$$T_i = \inf\{t \geq 0 \mid X_t = a_i \text{ og } Y_t = a_i\}. \quad (9.3.12)$$

Som sædvanlig sættes $\inf \emptyset = \infty$, så ∞ er som udgangspunkt en mulig værdi af T_i . Vi vil nu bevise to lemmaer, der giver os det rigtige værktøj til at bevise Sætning 9.3.7.

Lemma 9.3.8 *For alle $i = 1, \dots, k$ og alle $t \in \mathbb{N}$ gælder, at*

$$\sum_{j=1}^k |P(X_t = a_j) - P(Y_t = a_j)| \leq 2P(T_i \geq t). \quad (9.3.13)$$

Bevis: Ifølge (1.3.14) er

$$P(X_t = a_j) = \sum_{n=1}^{t-1} P(T_i = n, X_t = a_j) + P(T_i \geq t, X_t = a_j),$$

og for $t > n$ er

$$\begin{aligned} & P(T_i = n, X_t = a_j) \\ &= P((X_r, Y_r) \neq (a_i, a_i), r = 0, \dots, n-1, (X_n, Y_n) = (a_i, a_i), X_t = a_j) \\ &= P(X_t = a_j \mid (X_n, Y_n) = (a_i, a_i), (X_r, Y_r) \neq (a_i, a_i), r = 0, \dots, n-1) \\ &\quad \cdot P((X_r, Y_r) \neq (a_i, a_i), r = 0, \dots, n-1, (X_n, Y_n) = (a_i, a_i)) \\ &= p_{ij}^{(t-n)} P((X_r, Y_r) \neq (a_i, a_i), r = 0, \dots, n-1, (X_n, Y_n) = (a_i, a_i)) \\ &= p_{ij}^{(t-n)} P(T_i = n), \end{aligned}$$

hvor vi har brugt Markovegenskaben (9.2.12) og at $\{X_t\}$ og $\{Y_t\}$ er uafhængige. Dermed er

$$P(X_t = a_j) = \sum_{n=1}^{t-1} p_{ij}^{(t-n)} P(T_i = n) + P(T_i \geq t, X_t = a_j).$$

På tilsvarende måde vises at

$$P(Y_t = a_j) = \sum_{n=1}^{t-1} p_{ij}^{(t-n)} P(T_i = n) + P(T_i \geq t, Y_t = a_j),$$

således at det følger af trekantsuligheden, at

$$|P(X_t = a_j) - P(Y_t = a_j)| \leq P(T_i \geq t, X_t = a_j) + P(T_i \geq t, Y_t = a_j).$$

Nu følger (9.3.13) af (1.3.14) ved summation over j .

□

For at kunne udnytte uligheden (9.3.13), skal vi kunne sige noget om størrelsen af $P(T_i \geq t)$. Følgende lemma er lige, hvad vi skal bruge.

Lemma 9.3.9 *Antag at $\{X_t\}$ er irreducibel og aperiodisk. Da eksisterer der et $c > 0$ og et $\rho \in (0, 1)$, så*

$$P(T_i \geq t) \leq c\rho^t \tag{9.3.14}$$

for alle $t \in \mathbb{N}$.

Bevis: Da $\{X_t\}$ er aperiodisk, findes der et m_0 og et $\delta \in (0, 1)$, så $p_{ij}^{(m_0)} > \delta$ for alle $i, j = 1, \dots, k$. Lad $A_i = S^2 \setminus \{(a_i, a_i)\}$. Da er for alle $t \in \mathbb{N}$

$$\{T_i > nm_0\} \subseteq \{(X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n\},$$

således at

$$P(T_i > nm_0) \leq P((X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n).$$

Med definitionen $z = ((x_1, y_1), \dots, (x_{n-1}, y_{n-1}))$, er nu

$$\begin{aligned} & P((X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n) \\ &= \sum_{z \in A_i^{n-1}} P((X_{nm_0}, Y_{nm_0}) \in A_i, (X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1) \\ &= \sum_{z \in A_i^{n-1}} P((X_{nm_0}, Y_{nm_0}) \in A_i \mid (X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1) \\ &\quad \cdot P((X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1). \end{aligned}$$

og

$$\begin{aligned}
 & P((X_{nm_0}, Y_{nm_0}) \in A_i \mid (X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1) \\
 &= 1 - P((X_{nm_0}, Y_{nm_0}) = (a_i, a_i) \mid (X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1) \\
 &= 1 - P((X_{nm_0} = a_i \mid X_{\ell m_0} = x_\ell, \ell < n) \cdot P(Y_{nm_0} = a_i \mid Y_{\ell m_0} = y_\ell, \ell < n)) \\
 &= 1 - P((X_{nm_0} = a_i \mid X_{(n-1)m_0} = x_{n-1}) \cdot P(Y_{nm_0} = a_i \mid Y_{(n-1)m_0} = y_{n-1})) \\
 &\leq 1 - \delta^2.
 \end{aligned}$$

Læseren bør overbevise sig om rigtigheden af det andet lighedstegn ved hjælp af elementære regneregler for sandsynligheder fra Kapitel 1. Sammenfattende har vi, at

$$\begin{aligned}
 & P((X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n) \\
 &\leq (1 - \delta^2) \sum_{z \in A_i^{n-1}} P((X_{\ell m_0}, Y_{\ell m_0}) = (x_\ell, y_\ell), \ell = 1, \dots, n-1) \\
 &= (1 - \delta^2) P((X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n-1).
 \end{aligned}$$

Ved at gentage disse argumenter $n - 1$ gange mere, fås

$$P(T_i > nm_0) \leq P((X_{\ell m_0}, Y_{\ell m_0}) \in A_i, \ell = 1, \dots, n) \leq (1 - \delta^2)^n.$$

Vælg nu det hele tal n , så $t/m_0 > n \geq t/m_0 - 1$. Da er

$$P(T_i \geq t) \leq P(T_i > nm_0) \leq (1 - \delta^2)^n \leq (1 - \delta^2)^{(t/m_0 - 1)},$$

som er på den ønskede form med $\rho = (1 - \delta^2)^{1/m_0}$.

□

Bemærk, at (9.3.14) medfører at

$$P(T_i = \infty) \leq P(T_i \geq t) \leq c\rho^t$$

for alle $t \in \mathbb{N}$ således at

$$P(T_i = \infty) \leq \lim_{t \rightarrow \infty} c\rho^t = 0.$$

Vi kan nu vise dette afsnits hovedresultat, Sætning 9.3.7 om ergodicitet af Markovkæder.

Bevis for Sætning 9.3.7: Følgen $(p_{11}^{(n)}, p_{12}^{(n)}, \dots, p_{kk}^{(n)})$ i den kompakte mængde $[0, 1]^{k^2}$ har en konvergent delfølge $(p_{11}^{(n_t)}, p_{12}^{(n_t)}, \dots, p_{kk}^{(n_t)})$, hvor $n_t \rightarrow \infty$. D.v.s. der findes c_{ij} , så

$$\lim_{n_t \rightarrow \infty} p_{ij}^{(n_t)} = c_{ij}$$

for alle $i, j = 1, \dots, k$. Vi anvender nu Lemmaerne 9.3.8 og 9.3.9 med begyndelsesfordelingerne for $\{X_t\}$ og $\{Y_t\}$ givet ved $q_1(a_r) = 1$ og $q_2(a_s) = 1$, hvor r og s er vilkårlige tal i $\{1, \dots, k\}$. Ifølge (9.2.5) er

$$P(X_{n_t} = a_j) = p_{rj}^{(n_t)} \quad \text{og} \quad P(Y_{n_t} = a_j) = p_{sj}^{(n_t)},$$

så Lemmaerne 9.3.8 og 9.3.9 giver at

$$|c_{rj} - c_{sj}| = \lim_{n_t \rightarrow \infty} |p_{rj}^{(n_t)} - p_{sj}^{(n_t)}| = 0.$$

Dermed har vi vist at $c_{rj} = c_{sj}$ for alle $r, s = 1, \dots, k$. Grænseværdien $\lim_{n_t \rightarrow \infty} p_{ij}^{(n_t)}$ afhænger altså ikke af i , og vi betegner den med p_j .

Det følger nu på helt samme måde som i beviset for Sætning 9.3.2 at $\sum_{j=1}^k p_j = 1$. Vi kan derfor anvende Lemmaerne 9.3.8 og 9.3.9 med begyndelsesfordelingen for $\{Y_t\}$ givet ved $q_2(a_j) = p_j$, $j = 1, \dots, k$, mens begyndelsesfordelingen for $\{X_t\}$ er givet ved $q_1(a_i) = 1$, hvor i er et vilkårligt tal i $\{1, \dots, k\}$. Da er $P(X_t = a_j) = p_{ij}^{(m)}$, og ifølge Sætning 9.3.4 er $P(Y_t = a_j) = p_j$, så

$$|p_{ij}^{(m)} - p_j| \leq c\rho^t \rightarrow 0$$

for $t \rightarrow \infty$. Dermed har vi vist at $\{X_t\}$ er ergodisk, og ifølge Sætning 9.3.5 er den stationære fordeling givet ved $\pi(a_j) = p_j$, $j = 1, \dots, k$. Lader vi nu begyndelsesfordelingen for $\{X_t\}$ være vilkårlig, mens begyndelsesfordelingen for $\{Y_t\}$ igen er den stationære fordeling, får vi (9.3.11)

□

9.4 Sammenfatning

Vi har i dette kapitel set nogle typiske eksempler på Markovkæder med endeligt tilstandsrum. Dernæst blev nogle vigtige regneregler etableret.

Herunder hvordan man kan finde fordelingen af en observeret Markovkæde ud fra overgangssandsynlighederne og begyndelsesfordelingen, og hvordan man kan finde m -trins overgangssandsynlighederne ud fra 1-trins overgangssandsynlighederne ved hjælp af Chapman-Kolmogorov ligningerne. Endelig blev de vigtige begreber ergodicitet og stationær fordeling behandlet. Vi så at grænsefordelingen for en ergodisk Markovkæde med endeligt tilstandsrum er en entydig stationær fordeling, samt at hvis begyndelsesfordelingen for en Markovkæde er stationær, så er hele processen stationær. Til slut blev der givet betingelser, som sikrer at en Markovkæde med endeligt tilstandsrum er ergodisk, og at konvergensten mod grænsefordelingen går eksponentielt hurtigt.

9.5 Opgaver

9.1 En bærbar PCs trådløse LAN-kort rapporterer hvert sekund radioforbindelsens tilstand. Der er tre muligheder: dårlig, rimelig, god. Hvis forbindelsen er dårlig, er dens tilstand næste sekund enten dårlig eller rimelig, og de to muligheder er da lige sandsynlige. Hvis tilstanden er rimelig eller god, er tilstanden næste sekund med sandsynlighed 0.9 uændret, mens sandsynligheden for at tilstanden næste sekund er dårlig er 0.04.

- (a) Opskriv overgangsmatricen.
- (b) Find den stationære fordeling.
- (c) Er Markovkæden irreducibel og aperiodisk?

9.2 Betragt en Markovkæde med tre tilstande og med overgangsmatrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p \end{pmatrix},$$

- (a) Find 2-trins overgangssandsynlighederne.
- (b) Gør dernæst rede for, at Markovkæden er irreducibel og aperiodisk.
- (c) Find til slut den stationære fordeling for kæden.

- 9.3 Rentens udvikling er vigtig i mange finansielle problemer, for eksempel i livsforsikring. Ofte benyttes en Markovproces som model for rentens udvikling. Her er en simpel model. Antag at renten kun kan have tre niveauer: 2, 4 og 6 procent, og at den er konstant hele året. Overgangssandsynlighederne er som følger. Uanset tilstanden er renten med sandsynlighed 0.6 uændret næste år. Hvis renten et år er 2 procent, er sandsynligheden for at den næste år er 4 procent lig 0.3, mens den med sandsynlighed 0.1 er 6 procent næste år. Hvis renten et år er 4 procent, er der samme sandsynlighed for at den går op og ned næste år. Hvis endelig renten er 6 procent, er der sandsynlighed 0.3 for at den næste år er gået ned til 4 procent, mens der er sandsynlighed 0.1 for at den er hoppet helt ned på 2 procent.
- Opskriv overgangsmatricen.
 - Givet at renten et år er 2 procent, hvad er da sandsynligheden for, at den tre år senere er 6 procent?
 - Gør rede for, at Markovkæden er irreducibel og aperiodisk.
 - Find den stationære fordeling for kæden.
- 9.4 Julie har to paraplyer. Hvis det regner, når hun om morgenen skal på arbejde, tager hun en paraply med, hvis der er en paraply i hendes lejlighed. Hvis det ikke regner, tager hun ikke en paraply med. På samme måde, når hun skal hjem fra arbejde om aftenen. Hvis det regner, tager hun en paraply med, såfremt mindst en af hendes paraplyer ligger på hendes kontor. Hvis det ikke regner, lader hun sin paraply/sine paraplyer ligge på kontoret. Sandsynligheden for at det regner, når hun skal på arbejde, er 0.1. Der er samme sandsynlighed for regn, når hun skal hjem fra arbejde. Dette har stået på i meget lang tid. Hvad er sandsynligheden for, at Julie en morgen må gå på arbejde i regnvejr uden paraply. Vink: Lad X_t være antallet af paraplyer i Julies lejlighed efter hun er kommet hjem fra arbejde den t -te aften.
- 9.5 Markovkæder kan benyttes til at finde sandsynligheder, der er vigtige i forbindelse med livsforsikringskontrakter. Vi betragter en kontrakt, hvor der ud over udbetalingen i forbindelse med død også

er en årlig udbetaling i forbindelse med invalidering. Størrelsen af udbetalingen ved invalidering afhænger af graden af invalidering. En forsikringstager kan være i fire mulige tilstande: rask, lettere invalid, stærkt invalid og død. Vi betegner disse tilstande med henholdsvis 1, 2, 3 og 4. Antag at forsikringstagerens tilstand kan modelleres ved en Markovkæde med overgangsmatrix

$$\begin{pmatrix} 0.98 & 0.01 & 0.005 & 0.005 \\ 0.04 & 0.85 & 0.1 & 0.01 \\ 0 & 0.05 & 0.9 & 0.05 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Her svarer første række/søjle til tilstand 1 og så videre.

(a) Find sandsynligheden for, at en lettere invalideret forsikringstager efter tre år er død.

(b) Er Markovkæden irreducibel?

Den betingede kæde, hvor der betinges med, at forsikringstageren ikke er død har (efter afrunding) overgangsmatricen

$$\begin{pmatrix} 0.985 & 0.01 & 0.005 \\ 0.04 & 0.86 & 0.1 \\ 0 & 0.053 & 0.947 \end{pmatrix}.$$

(c) Overvej, hvordan den sidste overgangsmatrix er fremkommet fra den første.

(d) Gør rede for, at Markovkæden svarende til den sidste overgangsmatrix er irreducibel og aperiodisk, og find den stationære fordeling.

9.6 Følgende er en generalisering af Bernoulli-Laplace modellen for diffusion til en situation, hvor der ikke er lige meget af de to vædsker. Antag at vi har $2N$ partikler, hvoraf $H \geq N$ er hvide og $N - H$ er sorte. Partiklerne er fordelt i to beholdere med N i hver beholder. Mellem tidspunkterne $t-1$ og t trækkes der en partikel tilfældigt fra hver beholder og de to partikler bytter plads. Lad X_t være antallet af hvide partikler i den første beholder.

- (a) Gør rede for at $\{X_t\}$ er en Markovkæde med tilstandsrum $\{0, 1, \dots, N\}$ og find overgangssandsynlighederne.
- (b) Vis dernæst at Markovkæden er ergodisk og find den stationære fordeling.
- (c) Diskuter til slut approximationer til den stationære fordeling.
- 9.7 Lad X_t være en Markovkæde med tilstandsrum $\{1, 2, \dots, k\}$ og overgangssandsynligheder p_{ij} . Vis at det for alle $t \in \mathbb{N}$ gælder, at

$$P(X_t = 1, X_{t-1} = 1, \dots, X_1 = 1 | X_0 = 1) = p_{11}^t.$$

Definer en stokastisk variable T ved, at den er lig $t \in \mathbb{N}_0$, hvis $X_1 = \dots = X_t = 1$ og $X_{t+1} \neq 1$. Det er altså ventetiden på at forlade tilstand 1. Vis at den betingede fordeling af T givet at $X_0 = 1$ er den geometriske fordeling med parameter $1 - p_{11}$, d.v.s. at

$$P(T = t | X_0 = 1) = p_{11}^t(1 - p_{11})$$

for alle $t \in \mathbb{N}_0$.

- 9.8 Lad $\{X_t\} = \{X_0, X_1, X_2, \dots\}$ være en følge af stokastiske variable, der alle antager værdier i den endelige mængde $S = \{a_1, \dots, a_k\}$. Antag at der for alle $(i, j, \ell) \in \{1, \dots, k\}^3$ findes et tal $p_{ij\ell} \in [0, 1]$, så

$$P(X_{t+1} = a_\ell | X_t = a_i, X_{t-1} = a_j, X_{t_m} = a_{i_m}, \dots, X_{t_1} = a_{i_1}) = p_{ij\ell}$$

for alle $a_{i_1}, \dots, a_{i_m} \in S$ og alle $0 \leq t_1 < \dots < t_m < t-1$. Tilstanden af processen $\{X_t\}$ afhænger altså af de to forrige tilstande. Den er således ikke en Markovkæde, men kaldes en anden-ordens Markovkæde. Definer en ny proces ved $Y_t = (X_t, X_{t-1})$. Vis at $\{Y_t\}$ er en Markovkæde med tilstandsrum S^2 og overgangssandsynligheder

$$q_{(i,j)(\ell,n)} = \begin{cases} p_{ij\ell} & \text{hvis } i = n \\ 0 & \text{hvis } i \neq n. \end{cases}$$

Kapitel 10

Nogle stokastiske modeller

10.1 Aktier

Vi betragter en meget enkel model for udviklingen af en akties værdi. Hvert år er der to muligheder, som hver indtræffer med sandsynlighed $1/2$: Enten fordobles aktiens værdi, eller dens værdi halveres. Værdiændringerne i forskellige år antages at være stokastisk uafhængige. Hvis altså V_i ($i = 0, 1, 2, \dots$) betegner værdien af aktien ved slutningen af år nummer i , er de stokastiske variable $U_i = V_i/V_{i-1}$ ($i = 1, 2, \dots$) uafhængige og har alle samme fordeling givet ved $P(U_i = 2) = P(U_i = 1/2) = 1/2$. Vi antager, at værdien af en aktie ved starten af år nummer 1 er et givet ikke-stokastisk tal: $V_0 = v_0 > 0$.

Da $V_i = U_i U_{i-1} \cdots U_1 v_0$, og da $E(U_i) = 2 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{4}$, er

$$E(V_i) = E(U_i)E(U_{i-1}) \cdots E(U_1)v_0 = \left(\frac{5}{4}\right)^i v_0, \quad i = 0, 1, \dots,$$

hvor vi har brugt (3.7.7). Vi ser, at middelværdien af aktiens værdi vokser eksponentielt med tiden.

Antag at vi ved starten af år nummer t har en formue på $x_0 > 0$ kroner (et givet ikke-stokastisk tal). Denne formue investerer vi efter følgende strategi: Ved begyndelsen af hvert år omfordeler vi den formue, vi på det tidspunkt har, således at brøkdelen $\alpha \in [0, 1]$ er investeret i aktier af ovennævnte type, mens vi har resten af formuen i kontanter. Kontanter antages at beholde deres værdi fra år til år, og vi antager, at det er muligt at købe aktier med en værdi svarende til ethvert positivt

reelt kronebeløb. Hvis altså X_i betegner vores formue ved slutningen af år nummer i (= begyndelsen af år nummer $i + 1$), investerer vi i år nummer $i + 1$ αX_i kroner i aktier af ovennævnte type, mens vi beholder $(1 - \alpha)X_i$ kroner i kontanter.

Definer stokastiske variable R_i ved $R_i = X_i/X_{i-1}$, $i = 1, 2, \dots$, hvor $X_0 = x_0$. Da

$$X_{i+1} = \alpha X_i U_i + (1 - \alpha)X_i,$$

er

$$R_i = \alpha U_i + 1 - \alpha,$$

så

$$E(R_i) = 1 + \alpha/4.$$

Dermed er

$$E(X_i) = E(R_i R_{i-1} \cdots R_1 x_0) = (1 + \alpha/4)^i x_0 \quad i = 0, 1, \dots$$

Det er klart, at $E(X_i)$ antager sin største værdi for $\alpha = 1$. Vi maksimerer altså middelværdien af vores formue ved hele tiden at sætte alle pengene i aktier.

Man kan imidlertid også lave en anden overvejelse. Man kunne få den idé, at man ønsker at maksimere middelværdien af logaritmen til formuen i stedet for at maksimere middelværdien af formuen. Økonomer kalder i den forbindelse logaritmen en nyttefunktion. Logaritmen til formuen er

$$\log(X_i) = \log(x_0) + \sum_{j=1}^i \log(R_j)$$

Ifølge (3.7.3) er

$$\begin{aligned} E(\log(R_i)) &= \frac{1}{2} \log(2\alpha + 1 - \alpha) + \frac{1}{2} \log(\alpha/2 + 1 - \alpha) \\ &= \frac{1}{2} \log\left(1 + \frac{1}{2}\alpha(1 - \alpha)\right), \end{aligned}$$

så

$$E(\log(X_i)) = \log(x_0) + i \frac{1}{2} \log\left(1 + \frac{1}{2}\alpha(1 - \alpha)\right).$$

Det er ikke vanskeligt at indse, at $E(\log(X_i))$ er maksimal, når $\alpha = 1/2$.

Der kan være gode grunde til at maksimere $E(\log(X_i))$. Her er en. Sæt $m = E(\log(R_i))$. Ifølge store tals lov (Sætning 7.1.3) har vi for ethvert $\epsilon > 0$, at

$$\begin{aligned} & P\left(x_0 e^{(m-\epsilon)n} \leq X_n \leq x_0 e^{(m+\epsilon)n}\right) \\ &= P\left((m-\epsilon)n \leq \log(X_n) - \log(x_0) \leq (m+\epsilon)n\right) \\ &= P\left(\left|\frac{1}{n} \sum_{j=1}^n \log(R_j) - m\right| \leq \epsilon\right) \rightarrow 1, \end{aligned}$$

for $n \rightarrow \infty$. Vi ser, at vores formue efter n år X_n , når n er stor, med meget stor sandsynlighed ligger nær $x_0 e^{mn}$. I hvert fald for $m > 0$. Det synes derfor fornuftigt på langt sigt at vælge $\alpha = 1/2$.

Strategien med $\alpha = 1/2$ svarer til at sælge aktier, når deres værdi er steget, og købe dem, når deres værdi er faldet. Strategien kaldes *volatility pumping*.

10.2 Poisson processen

Vi observerer en følge af begivenheder, der indtræffer til tidspunkterne $0 < T_1 < T_2 < \dots < T_n < \dots$. Her er T_i altså tidspunktet for den i te begivenhed. Begivenhederne kan være udsendelser af en α -partikel fra et radioaktivt materiale, skadesanmeldelser til et forsikringselskab, telefonopkald og meget andet.

Vi vil forudsætte, at tidsrummene mellem de på hinanden følgende begivenheder $X_n = T_n - T_{n-1}$, $i = 1, 2, \dots$ (vi sætter $T_0 = 0$) er uafhængige stokastiske variable med samme fordeling. Yderligere vil vi antage, at $P(X_n > t) > 0$ for alle $t > 0$, og at begivenhederne ikke kan huske, hvornår den foregående begivenhed indtraf, d.v.s. at

$$P(X_n > t + s | X_n > t) = P(X_n > s)$$

for alle $t, s > 0$. Vi så i Eksempel 2.2.2, at denne antagelse (under en svag ekstra betingelse, som faktisk kan undværes) medfører, at X_n er eksponentialfordelt. Vi kan derfor skrive vores model således:

- 1) $T_n = X_1 + \dots + X_n$ for alle $n \in \mathbb{N}$,
- 2) X_1, X_2, \dots er uafhængige,
- 3) X_n er eksponentialfordelt med parameter λ for alle $n \in \mathbb{N}$.

For ethvert $t > 0$ lader vi $N(t)$ betegne antallet af hændelser i tidsintervallet $[0, t]$. Den stokastiske variable $N(t)$ antager altså værdier i \mathbb{N}_0 . Da $N(t) \geq n$ hvis og kun hvis $T_n \leq t$, ser vi at

$$P(N(t) \geq n) = P(T_n \leq t) = F_n(t),$$

hvor $F_n(t)$ er fordelingsfunktionen for T_n . Da X_i erne er eksponentialfordelte med parameter λ , og da denne fordeling er den samme som en Γ -fordeling med formparameter 1, følger det af Sætning 8.1.3, at T_n er Γ -fordelt med formparameter n og skalaparameter λ^{-1} . Derfor er

$$P(N(t) \geq n) = \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx.$$

For $n > 0$ er altså

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= \frac{\lambda^n}{n!} \int_0^t (nx^{n-1} - \lambda x^n) e^{-\lambda x} dx \\ &= \frac{\lambda^n}{n!} [x^n e^{-\lambda x}]_0^t = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \end{aligned}$$

Endvidere er

$$P(N(t) = 0) = P(X_1 > t) = 1 - P(X_1 \leq t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

Vi ser således, at $N(t)$ er Poisson-fordelt med parameter λt , og har dermed endnu et argument for at benytte Poissonfordelingen som model for tilfældigt ankomende begivenheder.

Vi har her defineret en hel familie af stokastiske variable $\{N(t) \mid t \geq 0\}$. En sådan familie af stokastiske variable kaldes en stokastisk proces.

Den stokastiske proces $\{N(t) \mid t \geq 0\}$ kaldes *Poisson processen med parameter λ* .

Vi slutter med at bevise, at de stokastiske variable $N(s)$ og $N(t) - N(s)$, hvor $t > s$, er uafhængige, samt at $N(t) - N(s)$ er Poisson fordelt med parameter $\lambda(t - s)$. For $n, m \in \mathbb{N}$, hvor $n > m$, sættes $V_{n,m} = T_n - T_m = X_{m+1} + \cdots + X_n$. Da er

$$\begin{aligned} P(N(s) \geq m, N(t) \geq n) &= P(T_m \leq s, T_n \leq t) \\ &= P((T_m, T_m + V_{n,m}) \in [0, s] \times [0, t]). \end{aligned}$$

Da T_m og $V_{n,m}$ er uafhængige og Γ -fordelte, T_m med formparameter m og $V_{n,m}$ med formparameter $n - m$, begge med skalaparameter λ^{-1} , følger det af Sætning 6.3.10, at

$$\begin{aligned} P((T_m, T_m + V_{n,m}) \in [0, s] \times [0, t]) &= \int_0^s \left(\int_0^t \frac{\lambda^m x^{m-1} e^{-\lambda x}}{(m-1)!} \frac{\lambda^{n-m} (y-x)^{n-m-1} e^{-\lambda(y-x)}}{(n-m-1)!} 1_{(0,\infty)}(y-x) dy \right) dx \\ &= \int_0^s \left(\int_x^t \frac{\lambda^n x^{m-1} (y-x)^{n-m-1} e^{-\lambda y}}{(m-1)!(n-m-1)!} dy \right) dx \\ &= \int_0^s \left(\int_0^{t-x} \frac{\lambda^n x^{m-1} z^{n-m-1} e^{-\lambda(x+z)}}{(m-1)!(n-m-1)!} dz \right) dx, \end{aligned}$$

hvor vi har skiftet integrationsvariabel i det inderste integral til $z = y - x$.

Vi kaster os nu ud i et større regneri. For alle $m, k \in \mathbb{N}$, hvor $k \geq 2$ er

$$\begin{aligned} P(N(s) = m, N(t) - N(s) = k) &= P(N(s) = m, N(t) = m + k) \\ &= P(N(s) \geq m, N(t) \geq m + k) - P(N(s) \geq m + 1, N(t) \geq m + k) \\ &\quad - P(N(s) \geq m, N(t) \geq m + k + 1) \\ &\quad + P(N(s) \geq m + 1, N(t) \geq m + k + 1) \end{aligned}$$

$$= \frac{\lambda^{m+k}}{m!k!} \int_0^s \left(\int_0^{t-x} \{mx^{m-1}kz^{k-1} - x^m k(k-1)z^{k-2} - \lambda mx^{m-1}z^k + \lambda x^m k z^{k-1}\} e^{-\lambda(x+z)} dz \right) dx$$

Det inderste integral er lig

$$\begin{aligned} & mx^{m-1} e^{-\lambda x} \int_0^{t-x} \{kz^{k-1} - \lambda z^k\} e^{-\lambda z} dz \\ & - x^m e^{-\lambda x} \int_0^{t-x} \{k(k-1)z^{k-2} - \lambda k z^{k-1}\} e^{-\lambda z} dz \\ & = mx^{m-1} e^{-\lambda x} \int_0^{t-x} (z^k e^{-\lambda z})' dz - x^m e^{-\lambda x} \int_0^{t-x} (kz^{k-1} e^{-\lambda z})' dz \\ & = mx^{m-1} e^{-\lambda x} (t-x)^k e^{-\lambda(t-x)} - x^m e^{-\lambda x} k(t-x)^{k-1} e^{-\lambda(t-x)} \\ & = e^{-\lambda t} (x^m (t-x)^k)'. \end{aligned}$$

Derfor er

$$\begin{aligned} & P(N(s) = m, N(t) - N(s) = k) \\ & = \frac{\lambda^{m+k}}{m!k!} e^{-\lambda t} \int_0^s (x^m (t-x)^k)' dx = \frac{(\lambda s)^m e^{-\lambda s}}{m!} \cdot \frac{(\lambda(t-s))^k e^{-\lambda(t-s)}}{k!}. \end{aligned}$$

Vi mangler at behandle tilfældene $m = 0$ og $k = 0, 1$. Disse tilfælde er lettere og overlades til læseren. Resultatet bliver det samme. Da altså $P(N(s) = m, N(t) - N(s) = k)$ er et produkt af to Poissonsandsynligheder, kan vi konkludere af Sætning 4.3.1, at de stokastiske variable $N(s)$ og $N(t) - N(s)$ er uafhængige, samt at $N(s)$ er Poisson fordelt med parameter λs (det vidste vi allerede), og at $N(t) - N(s)$ er Poisson fordelt med parameter $\lambda(t-s)$.

10.3 Brownsk bevægelse

En partikel bevæger sig i planen efter en Brownsk bevægelse. Vi vil forsøge at opstille en model for dette fænomen, som er kendt fra fysikken.

Til tid 0 befinder partiklen sig i punktet $(0, 0)$. Lad partiklens position en tidsenhed senere være (X_1, X_2) . Det er rimeligt at antage, at X_1 og X_2 er uafhængige, og at de har samme marginale fordeling. Lad os endvidere forudsætte, at fordelingen af (X_1, X_2) er kontinuert med simultan sandsynlighedstæthed f , og lad os betegne den fælles marginale tæthed for X_1 og X_2 med p . Til slut vil vi antage, at ingen retninger i planen er foretrukket af partiklen, hvilket vi kan formulere præcist ved at antage, at tætheden $f(x, y)$ kun afhænger af (x, y) 's afstand til $(0, 0)$, d.v.s. at der findes en funktion g fra $[0, \infty)$ ind i $[0, \infty)$, så

$$f(x, y) = g\left(\sqrt{x^2 + y^2}\right)$$

for alle $(x, y) \in \mathbb{R}^2$. Da $f(x, y) = p(x)p(y)$, er

$$p(x)p(y) = g\left(\sqrt{x^2 + y^2}\right)$$

for alle $(x, y) \in \mathbb{R}^2$. Sætter vi her $y = 0$, får vi, at $g(|x|) = p(0)p(x)$, og vi ser, at p opfylder ligningen

$$p(x)p(y) = p(0)p\left(\sqrt{x^2 + y^2}\right)$$

for alle $(x, y) \in \mathbb{R}^2$. Vi vil nu antage, at $p(x) > 0$ for alle $x \in \mathbb{R}$, og sætte $k(x) = \log(p(x))$. Da er

$$k(x) + k(y) = k(0) + k\left(\sqrt{x^2 + y^2}\right).$$

Sættes $x = (s + t)/\sqrt{2}$ og $y = (s - t)/\sqrt{2}$, fås

$$k((s + t)/\sqrt{2}) + k((s - t)/\sqrt{2}) = k(0) + k\left(\sqrt{s^2 + t^2}\right) = k(s) + k(t).$$

Hvis vi antager, at p er to gange differentiabel, er også k to gange differentiabel, og differentierer vi med hensyn til s får vi

$$\left[k'((s + t)/\sqrt{2}) + k'((s - t)/\sqrt{2})\right] / \sqrt{2} = k'(s). \quad (10.3.1)$$

Differentieres denne ligning med hensyn til t , fås

$$\left[k''((s + t)/\sqrt{2}) - k''((s - t)/\sqrt{2})\right] / 2 = 0$$

for alle $s, t \in \mathbb{R}$. Specielt får vi ved at sætte $t = s = u/\sqrt{2}$, at

$$k''(u) = k''(0) \quad (10.3.2)$$

for alle $u \in \mathbb{R}$. Sætter vi $k''(0) = -c$, følger det af (10.3.2), at

$$k'(u) = a - cu$$

for et eller andet $a \in \mathbb{R}$. Med $s = t = 0$ i (10.3.1) fås

$$\sqrt{2}k'(0) = k'(0),$$

så $k'(0) = 0$. Konstanten a må derfor være nul, således at $k'(u) = -cu$ for alle $u \in \mathbb{R}$. Integrerer vi k' , finder vi, at k må have formen

$$k(x) = b - \frac{c}{2}x^2,$$

hvoraf vi får, at den marginale tæthed for X_i er

$$p(x) = e^b e^{-cx^2/2}$$

for alle $x \in \mathbb{R}$. Denne funktion kan kun være integrabel hvis $c > 0$, og konstanten e^b skal være sådan, at integralet af p over \mathbb{R} er lig en. Sætter vi $c = 1/\sigma^2$, genkender vi normalfordelingen med middelværdi nul og varians σ^2 , så der må gælde, at

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$$

for alle $x \in \mathbb{R}$.

Normalfordelingen er altså en fornuftig model for en koordinat af positionen efter en tidsenhed af en partikel, som udfører Brownsk bevægelse. Lad $X_1(t)$ betegne første koordinat af partiklens position til et vilkårligt tidspunkt t . En dyberegående analyse viser, at det kan antages, at $X_1(t)$ er normalfordelt med middelværdi nul og varians $\sigma^2 t$. Den stokastiske proces $\{X_1(t) | t \geq 0\}$ kaldes *Wienerprocessen* eller *den Brownske bevægelse*. Hvis $X_2(t)$ betegner andenkoordinaten af partiklens position til tidspunktet t , er den stokastiske proces $\{X_2(t) | t \geq 0\}$ ikke overraskende også en Wienerproces. Den to-dimensionale stokastiske proces $\{(X_1(t), X_2(t)) | t \geq 0\}$ kaldes en to-dimensional Wienerproces eller Brownsk bevægelse. Det viser sig, at de to koordinater er stokastisk uafhængige.

Kapitel 11

Efterskrift

Det er på sin plads at slutte med et par bemærkninger om nogle mangler ved disse noters fremstilling af den elementære sandsynlighedsregning.

Nogle læsere har måske tænkt over, om det virkelig er nødvendigt at gennemgå stort set den samme teori to gange: først for diskrete fordelinger, så for kontinuerte fordelinger. Det er faktisk ikke nødvendigt. Hvis man baserer fremstillingen af sandsynlighedsregningen på målteorien, kan man klare disse to tilfælde - plus mange andre - på én gang. Det er imidlertid ikke nogen god måde at starte bekendtskabet med sandsynlighedsregningen på, og det ville iøvrigt være for svært for et første-års kursus. Formålet i disse noter har været at give en vis fortrolighed med nogle af sandsynlighedsregningens grundlæggende begreber samt øvelse i at løse enkle sandsynlighedsteoretiske problemer. Fremstillingen har så måttet forenkles lidt her og der. Det skal nævnes, at vi i definitionen af et sandsynlighedsmål kun har krævet, at (1.3.3) er opfyldt, mens man normalt kræver, at (1.3.14) holder ikke blot for endeligt mange disjunkte mængder, men for tælleligt mange. Vores krav til et sandsynlighedsmål er altså svagere end det sædvanlige. Til gengæld er vi blevet sparet for nogle målteoretiske detaljer, som ville have kunnet overskygge formålet med disse noter. I forbindelse med kontinuerte fordelinger, har vi tilladt os at integrere over vilkårlige delmængder af \mathbb{R} eller \mathbb{R}^n . Der findes imidlertid delmængder af \mathbb{R} og \mathbb{R}^n , som er så aparte, at man ikke kan definere integralet over dem. Hvis man holder sig til en klasse af delmængder af \mathbb{R} eller \mathbb{R}^n , som kaldes Borelmængderne, og som ikke inkluderer disse særegne mængder, er man på den sikre side. Egentlig burde vi derfor alle

steder have skrevet “en Borel delmængde” af \mathbb{R} eller \mathbb{R}^n i stedet for “en vilkårlig delmængde”, men der havde været risiko for, at en sådan formulering, som alligevel ikke kan forklares ordentligt i dette kursus, ville have afledt interessen fra det væsentlige. Hvad mere er, det er uhyre sjældent, at man i praksis støder på en mængde, der ikke er en Borelmængde.

Det skal også lige nævnes, at vi kun har kradset lidt i overfladen af teorien for betingede sandsynligheder. Hvis man forsøger generelt at definere betingede sandsynligheder givet hændelser med sandsynlighed nul ved den grænseovergang, som vi reelt brugte i Afsnit 6.5, kommer man galt af sted. Lad for eksempel X være en kontinuert stokastisk variabel. Det er da fristende at definere den betingede sandsynlighed af hændelsen B givet at $X = x$ ved grænseværdien

$$P(B|X = x) = \lim_{\epsilon \rightarrow 0} \frac{P(B \cap \{|X - x| < \epsilon\})}{P(|X - x| < \epsilon)},$$

men selv når denne grænseværdi eksisterer, er den helt umulig. Da $P(X = x) = 0$, er

$$\begin{aligned} P(\{X \neq x\} \cap \{|X - x| < \epsilon\}) &= P(\{|X - x| < \epsilon\} \setminus \{X = x\}) \\ &= P(|X - x| < \epsilon) - P(X = x) = P(|X - x| < \epsilon), \end{aligned}$$

således at ovenstående “definition” giver $P(X \neq x|X = x) = 1$. Dette er oplagt noget vrøvl. Det kræver en del arbejde, at indføre betingede sandsynligheder generelt.

Alle de problemer, som er nævnt her, vil naturligvis blive behandlet i et mere avanceret kursus i sandsynlighedsteori.

Appendiks A

Om mængder

A.1 Mængder og mængdeoperationer

Dette appendiks handler om mængder og om de elementære operationer, man kan foretage med dem. Det er helt afgørende for tilegnelsen af sandsynlighedsregning, at man behersker de elementære manipulationer med mængder.

En *mængde* er en samling af objekter. Disse objekter kaldes mængdens *elementer*. Eksempler på mængder er de reelle tal \mathbb{R} , intervallet $[0, 1]$, mængden af mulige udfald ved kast med en terning $\{1, 2, 3, 4, 5, 6\}$ og de naturlige tal

$$\mathbb{N} = \{1, 2, 3, \dots\}.$$

Det er almindeligt at angive en mængde ved at opskrive en liste over dens elementer omgivet af krøllede parenteser. Hvis mængden har uendeligt mange elementer, antydes det blot, hvorledes listen forsættes. Mængden, der ingen elementer indeholder, kaldes den *tomme mængde* og betegnes med \emptyset . Hvis antallet af elementer i en mængde er et naturligt tal eller 0, siges mængden at være endelig. Hvis en mængde ikke er endelig, kaldes den uendelig. Mængderne $\{1, 2, 3, 4, 5, 6\}$ og \emptyset er endelige, mens \mathbb{R} , $[0, 1]$ og \mathbb{N} er uendelige mængder.

Hvis x er et element i en mængde A , skrives $x \in A$. Man siger ofte at x tilhører A . Hvis x ikke er et element i A , skrives $x \notin A$. Symbolet \in kan benyttes til at definere mængder. For eksempel består mængden

$$\{\frac{n}{2} \mid n \in \mathbb{N}\}$$

af alle tal på formen $\frac{n}{2}$, hvor n er et positivt helt tal. Et andet eksempel er

$$\{x \in \mathbb{R} \mid \sin(x) > \frac{1}{2}\}$$

som består af de reelle tal x , for hvilke udsagnet $\sin(x) > \frac{1}{2}$ er sandt. Hvis A er en mængde, kan man generelt definere en ny mængde ved

$$\{x \in A \mid x \text{ opfylder en betingelse}\}.$$

Den nye mængde består af de elementer i A , som opfylder betingelsen.

To mængder A og B er identiske, hvis de har de samme elementer. Hvis det er tilfældet, skriver man $A = B$. For eksempel er $\{1, 2, 3, 4\} = \{4, 1, 3, 2\}$. En mængde A siges at være en *delmængde* af en anden mængde B , hvis ethvert element i A også tilhører B . Det skrives på denne måde: $A \subseteq B$ eller $B \supseteq A$. Man siger også at A er indeholdt i B eller at B indeholder A . Tegnet \subseteq kaldes en inklusion. Bemærk, at $A = B$ hvis og kun hvis $A \subseteq B$ og $B \subseteq A$. Man kan ofte vise, at to mængder er identiske ved at bevise disse to modsatrettede inklusioner. Hvis A er en delmængde af B , og B indeholder et element, som ikke tilhører A , kaldes A en ægte delmængde af B , hvilket skrives $A \subset B$ eller $B \supset A$.

I næsten alle sammenhænge findes der en stor mængde E , som er basis for vores mængdeoperationer i den forstand, at vi kun interesserer os for delmængder af E . Mængden E kunne da kaldes vores univers. Hvad vi vælger som vores univers vil typisk variere fra gang til gang afhængigt af sammenhængen. Hvis A og B er to delmængder af E , defineres deres *foreningsmængde*, der skrives $A \cup B$, ved

$$A \cup B = \{x \in E \mid x \in A \text{ eller } x \in B\},$$

hvor “eller” betyder mindst ét af de to udsagn er korrekte, muligvis dem begge. Foreningsmængden består altså af de elementer i E , som tilhører mindst en af mængderne A og B . Bemærk, at $A \cup E = E$, $A \cup \emptyset = A$ og $A \cup A = A$. *Fællesmængden* af A og B , som skrives $A \cap B$, defineres ved

$$A \cap B = \{x \in E \mid x \in A \text{ og } x \in B\}.$$

Det vil sige, at fællesmængden består af de elementer i E , der tilhører både A og B . Bemærk, at $A \cap E = A$, $A \cap \emptyset = \emptyset$ og $A \cap A = A$. Hvis to

mængder A og B ikke har fælles elementer, altså hvis $A \cap B = \emptyset$, siges de to mængder at være *disjunkte*.

For to mængder A og B defineres *mængdedifferensen* mellem A og B , som skrives $A \setminus B$, ved

$$A \setminus B = \{x \in A \mid x \notin B\}.$$

Altså de elementer i A , som ikke tilhører B . Mængdedifferensen $A \setminus B$ omtales ofte som A fraregnet B . En særligt vigtig mængdedifferens er $E \setminus A$, som kaldes *komplementærmængden* til A og betegnes A^c . Bemærk at $A \setminus B = A \cap B^c$, $E^c = \emptyset$, $\emptyset^c = E$, $(A^c)^c = A$ og $A \cup A^c = E$.

De netop omtalte mængdeoperationer kan ikke kombineres efter for-godtbefindende. For tre mængder A , B og C gælder for eksempel i almindelighed at

$$(A \cup B) \cap C \neq A \cup (B \cap C).$$

Dette fremgår af det simple tilfælde $A = \{1, 2\}$, $B = \{3, 4\}$ og $C = \{2, 3\}$, hvor er $(A \cup B) \cap C = C$, mens $A \cup (B \cap C) = \{1, 2, 3\}$. Der gælder følgende regneregler for de indførte mængdeoperationer.

$$A \cup B = B \cup A \tag{A.1.1}$$

$$A \cap B = B \cap A \tag{A.1.2}$$

$$A \cup (B \cup C) = (A \cup B) \cup C \tag{A.1.3}$$

$$A \cap (B \cap C) = (A \cap B) \cap C \tag{A.1.4}$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{A.1.5}$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{A.1.6}$$

$$(A \cap B)^c = A^c \cup B^c \tag{A.1.7}$$

$$(A \cup B)^c = A^c \cap B^c. \tag{A.1.8}$$

Lad os bevise (A.1.5). Vi beviser først, at $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$. Antag derfor, at $x \in A \cup (B \cap C)$. Hvis $x \in A$, så er $x \in A \cup B$ og $x \in A \cup C$, og dermed er $x \in (A \cup B) \cap (A \cup C)$. Hvis $x \notin A$, må jo så $x \in B \cap C$. Det vil sige, at $x \in B$ og $x \in C$, og dermed også $x \in A \cup B$ og $x \in A \cup C$. Også når $x \notin A$, har vi altså, at $x \in (A \cup B) \cap (A \cup C)$. Vi har derfor vist, at $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$. Vi mangler nu at bevise, at $A \cup (B \cap C) \supseteq (A \cup B) \cap (A \cup C)$, før vi kan konkludere, at (A.1.5) holder. For at bevise det antager vi, at $x \in (A \cup B) \cap (A \cup C)$. Hvis $x \in A$, er det klart, at $x \in A \cup (B \cap C)$. Hvis omvendt $x \notin A$, må der gælde, at $x \in B$, eftersom $x \in A \cup B$. Ligeledes må der, da $x \in A \cup C$, gælde, at $x \in C$. Vi kan derfor konkludere, at $x \in B \cap C$, og dermed, at der også når $x \notin A$ gælder, at $x \in A \cup (B \cap C)$. Vi har dermed bevist (A.1.5).

Ved at kombinere regnereglerne (A.1.1) – (A.1.8) kan man manipulere med mange mængder. Hvis for eksempel A, B_1, \dots, B_k er $k+1$ mængder, er

$$A \cap (B_1 \cup B_2 \cup \dots \cup B_k) = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_k). \quad (\text{A.1.9})$$

For to mængder A og B definerer man *produktmængden* $A \times B$ ved

$$A \times B = \{(x, y) \mid x \in A \text{ og } y \in B\},$$

det vil sige mængden af par (x, y) , hvor x tilhører A , og y tilhører B . Man kan også definere produktmængden af tre eller flere mængder. For eksempel er produktmængden af tre mængder A, B og C givet ved

$$A \times B \times C = \{(x, y, z) \mid x \in A \text{ og } y \in B \text{ og } z \in C\}.$$

Man kan godt lave produktmængden $A \times A$ af en mængde A med sig selv. Den betegnes ofte med A^2 . Tilsvarende betegnes produktmængden $A \times A \times \dots \times A$, hvor A forekommer k gange med A^k .

Lad t være en afbildning fra en mængde A ind i en mængde B , lad C være en delmængde af A , og lad D være en delmængde af B . Da defineres *billedmængden* af C ved

$$t(C) = \{t(x) \mid x \in C\}$$

og *originalmængden* til D defineres ved

$$t^{-1}(D) = \{x \in A \mid t(x) \in D\}.$$

Originalmængden består altså af de elementer i A , som afbildes ind i D , når man anvender t på dem. Originalmængden kaldes også somme tider Urbilledet af D ved t . På trods af notationen kan originalmængden defineres for enhver afbildning t , uanset om t har en invers afbildning eller ej. I tilfælde hvor t har en invers afbildning t^{-1} , er originalmængden lig billedmængden af D ved afbildningen t^{-1} .

Eksempel A.1.1 Lad $A = B = \mathbb{R}$, og betragt funktionen $t(x) = x^2$. Da er $t(\mathbb{R}) = [0, \infty)$, $t^{-1}((-\infty, 4]) = [-2, 2]$, $t^{-1}([0, 4]) = [-2, 2]$ og $t^{-1}([1, 4]) = [-2, -1] \cup [1, 2]$.

□

Originalmængdedannelse passer (i modsætning til billedmængdedannelse) fint sammen med mængdeoperationerne. Der gælder således følgende mængdeidentiteter, hvor C_1 og C_2 er delmængder af B :

$$t^{-1}(B) = A \quad (\text{A.1.10})$$

$$t^{-1}(C_1 \cap C_2) = t^{-1}(C_1) \cap t^{-1}(C_2) \quad (\text{A.1.11})$$

$$t^{-1}(C_1 \cup C_2) = t^{-1}(C_1) \cup t^{-1}(C_2) \quad (\text{A.1.12})$$

$$t^{-1}(C_1 \setminus C_2) = t^{-1}(C_1) \setminus t^{-1}(C_2) \quad (\text{A.1.13})$$

$$t^{-1}(C_1^c) = t^{-1}(C_1)^c. \quad (\text{A.1.14})$$

Lad os bevise (A.1.11). Antag, at $x \in t^{-1}(C_1 \cap C_2)$. Ifølge definitionen af originalmængden, betyder det, at $t(x) \in C_1 \cap C_2$. Det følger derfor af definitionen af fællesmængden, at $t(x) \in C_1$ og $t(x) \in C_2$, hvilket per definition af originalmængder er det samme som $x \in t^{-1}(C_1)$ og $x \in t^{-1}(C_2)$. Definitionen af fællesmængden giver til slut, at $x \in t^{-1}(C_1) \cap t^{-1}(C_2)$. Vi har dermed vist, at $t^{-1}(C_1 \cap C_2) \subseteq t^{-1}(C_1) \cap t^{-1}(C_2)$. Det overlades til læseren at bevise, at $t^{-1}(C_1 \cap C_2) \supseteq t^{-1}(C_1) \cap t^{-1}(C_2)$, således at det kan sluttes, at (A.1.11) holder. Læseren opfordres til også at bevise de øvrige identiteter.

Et nyttigt hjælpemiddel i forbindelse med mængder er de såkaldte indikatorfunktioner. *Indikatorfunktionen* for en delmængde A af vores univers E er en funktion fra E ind i \mathbb{R} defineret ved

$$1_A(x) = \begin{cases} 1 & \text{hvis } x \in A \\ 0 & \text{hvis } x \notin A. \end{cases} \quad (\text{A.1.15})$$

Indikatorfunktionen fortæller altså, hvor mængden ligger. For to mængder A og B gælder følgende regneregler:

$$1_{A \cap B}(x) = 1_A(x)1_B(x) \quad (\text{A.1.16})$$

$$1_{A \setminus B}(x) = 1_A(x) - 1_{A \cap B}(x) \quad (\text{A.1.17})$$

$$1_{A^c}(x) = 1 - 1_A(x) \quad (\text{A.1.18})$$

$$1_{A \cup B}(x) = 1_A(x) + 1_B(x) - 1_{A \cap B}(x). \quad (\text{A.1.19})$$

Hvis A og B er disjunkte, er

$$1_{A \cup B}(x) = 1_A(x) + 1_B(x).$$

A.2 Opgaver

A.1 Vis regnereglerne (A.1.6), (A.1.7) og (A.1.8). Vis dernæst, at disse tre regneregler medfører (A.1.5).

A.2 Lad A , B og C være mængder. Som nævnt i teksten gælder det i almindelighed, at $(A \cup B) \cap C \neq A \cup (B \cap C)$. Find en betingelse på A og C , som sikrer, at $(A \cup B) \cap C = A \cup (B \cap C)$.

A.3 Vis, at

$$\begin{aligned} (A \setminus B) \cap C &= (A \cap C) \setminus (B \cap C) \\ (A \cap B) \setminus C &= (A \setminus C) \cap (B \setminus C). \end{aligned}$$

A.4 Vis (A.1.9)

A.5 Vis (A.1.10) – (A.1.14)

A.6 Vis (A.1.16) – (A.1.19)

A.7 Vis (A.1.6) og (A.1.7) ved hjælp af regnereglerne for indikatorfunktioner.

Appendiks B

Om kombinatorik

I hvor mange forskellige rækkefølger kan de 52 kort i et almindeligt kortspil lægges op, således at der aldrig ligger to af samme kulør (hermed menes klør, ruder, hjerter eller spar) ved siden af hinanden? Hvor mange synligt forskellige placeringer af 7 røde og 20 hvide kugler i 5 kasser er mulige? Det er eksempler på spørgsmål, hvis løsning hører ind under det felt der kaldes *kombinatorik*. Generelt kan man sige, at kombinatorikken (i hvert fald i den snævre forstand, vi taler om den her) handler om optælling af kombinationsmuligheder. I denne paragraf skal vi diskutere løsningen af nogle meget simple kombinatoriske problemer, og indføre nogle betegnelser i relation hertil.

B.1 Fakultetsfunktionen

For $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ defineres

$$n! = 1 \cdot 2 \cdot 3 \cdots n \tag{B.1.1}$$

(udtales “ n fakultet”). Man ser at $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$. Definitionsmæssigt sættes $0! = 1$ (idet et tomt produkt altid sættes til 1, ligesom en tom sum sættes til 0). I kombinatorikken er $n!$ antallet af rækkefølger, hvori n forskellige objekter kan opskrives. For eksempel kan tallene fra 1 til 5 ordnes på 120 forskellige måder, fordi vi for første plads har 5 valgmuligheder, for anden plads derefter 4, o.s.v.

B.2 Nedadstigende faktoriel

For $0 \leq k \leq n$ defineres

$$n^{(k)} = n \cdot (n-1) \cdots (n-k+1). \quad (\text{B.2.1})$$

En størrelse af denne slags kaldes et *nedadstigende faktoriel*. Tallet $n^{(k)}$, ofte kaldet “ n i k rund”, er altså produkt af de k på hinanden følgende heltal fra n og nedefter. Definitionsmæssigt sættes $n^{(0)} = 1$ (også for $n = 0$). Kombinatorisk kan denne størrelse fortolkes som antallet af ordnede sæt, bestående af k indbyrdes forskellige elementer fra en mængde med n elementer. Antallet af ordnede sæt, bestående af 25 forskellige tal fra $\{1, \dots, 365\}$, er for eksempel $365^{(25)}$, jvf. Eksempel 1.2.2. Bemærk at der gælder

$$n^{(n)} = n!$$

og

$$n^{(k)} = \frac{n!}{(n-k)!}.$$

B.3 Binomialkoefficienter

For $0 \leq k \leq n$ defineres

$$\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!}. \quad (\text{B.3.1})$$

Denne størrelse kaldes “ n over k ”. Tal af denne art er hele tal og kaldes *binomialkoefficienter*. Det er måske ikke oplagt, at der er tale om hele tal, men det følger af deres kombinatoriske fortolkning: Binomialkoefficienten $\binom{n}{k}$ er antallet af delmængder med netop k elementer af en mængde E med n elementer (f.eks. $E = \{1, \dots, n\}$). For at indse dette kan vi bemærke, at antallet af ordnede sæt, bestående af k forskellige elementer fra en mængde E med n elementer, jo er $n^{(k)}$. Ethvert sådant sæt giver på naturlig måde anledning til en delmængde af E med k elementer, bestående af de elementer, der forekommer i sættet. Men ved optællingen af de ordnede k -sæt får vi hver k -delmængde talt med $k!$ gange, nemlig én gang for hver mulig ordning af dens elementer. Antallet af k -delmængder af E bliver således $n^{(k)}/k! = \binom{n}{k}$.

som for $n = 2, 3$ og 4 antager de (mere eller mindre) velkendte former

$$\begin{aligned}(x + y)^2 &= x^2 + 2xy + y^2, \\(x + y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3, \\(x + y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.\end{aligned}$$

Binomialformlen kan vises ved induktion, men følger i øvrigt også af binomialkoefficienternes kombinatoriske fortolkning. Når man ganger ud i produktet

$$(x + y)^n = (x + y)(x + y) \dots (x + y)$$

får man jo i første omgang 2^n led af formen $x^k y^{n-k}$, og hvert af disse svarer naturligt til en delmængde af $\{1, \dots, n\}$, bestående af numrene på de k faktorer hvorfra “ x -bidragene” stammer. Når man herefter samler led af samme grad, bliver der altså netop $\binom{n}{k}$ led af typen $x^k y^{n-k}$.

B.4 Polynomialkoefficienter

For $n \in \mathbb{N}$ og $(x_1, \dots, x_k) \in D_k(n)$ ($k \in \mathbb{N}$), hvor

$$\begin{aligned}D_k(n) &= \{(x_1, \dots, x_k) \mid x_i \in \{0, \dots, n\}, i = 1, \dots, k, x_1 + \dots + x_k = n\},\end{aligned}\tag{B.4.1}$$

defineres polynomialkoefficienten ved

$$\binom{n}{x_1, \dots, x_k} = \frac{n!}{x_1! \dots x_k!}.\tag{B.4.2}$$

Denne størrelse kaldes også en *multinomialkoefficient*. Også den har en kombinatorisk fortolkning: Den er lig antallet af måder, hvorpå man af en mængde M med n elementer kan vælge k disjunkte delmængder M_1, \dots, M_k , således at $\#M_i = x_i$, $i = 1, \dots, k$ og $M = M_1 \cup \dots \cup M_k$. Dette kan indses på følgende måde. Først vælges de x_1 elementer i M_1 . Det kan gøres på $\binom{n}{x_1}$ måder. Dernæst vælges de x_2 elementer i M_2 blandt de resterende $n - x_1$ elementer i M . Dette kan gøres på $\binom{n-x_1}{x_2}$ måder. Således fortsætter man udvælgelsen af elementerne i mængderne M_i . Når elementerne i M_{k-1} skal vælges, er der $n - (x_1 + \dots + x_{k-2})$ elementer

tilbage at vælge iblandt, hvorfor der er $\binom{n-(x_1+\dots+x_{k-2})}{x_{k-1}}$ måder at vælge elementerne i M_{k-1} på. Der er nu $n - (x_1 + \dots + x_{k-1}) = x_k$ elementer tilbage, som altså udgør M_k . Det ønskede antal er produktet af de fundne antal muligheder for de enkelte delmængder, altså

$$\begin{aligned} & \binom{n}{x_1} \binom{n-x_1}{x_2} \binom{n-x_1-x_2}{x_3} \dots \binom{n-(x_1+\dots+x_{k-2})}{x_{k-1}} \\ &= \frac{n!}{x_1!(n-x_1)!} \cdot \frac{(n-x_1)!}{x_2!(n-x_1-x_2)!} \cdot \frac{(n-x_1-x_2)!}{x_3!(n-x_1-x_2-x_3)!} \\ & \quad \dots \frac{(n-(x_1+\dots+x_{k-2}))!}{x_{k-1}!(n-(x_1+\dots+x_{k-1}))!} \\ &= \frac{n!}{x_1! \dots x_k!}. \end{aligned}$$

B.5 Opgaver

B.1 Vis formel (B.3.2). Overvej den kombinatoriske fortolkning af Pascals trekant.

B.2 Vis binomialformlen (B.3.3) for $x > 0$ og $y > 0$ ved at betragte sandsynlighedsfunktionen for binomialfordelingen med antalsparameter n og sandsynlighedsparameter $x/(x+y)$ og udnytte at summen af sandsynlighederne for alle de mulige udfald er lig en.

B.3 Vis at

$$\binom{n}{x_1, \dots, x_k} = \binom{n}{x_k} \binom{n-x_k}{x_1, \dots, x_{k-1}}.$$

Forklar denne relation kombinatorisk. Skitser et bevis for polynomialkoefficientens kombinatoriske fortolkning ved induktion efter k .

Appendiks C

Om uendelige rækker

Lad $\{x_1, x_2, x_3, \dots\}$ være en følge af reelle tal, og definer

$$s_n = \sum_{i=1}^n x_i.$$

Man siger, at den uendelige række $\sum_{i=1}^{\infty} x_i$ er *konvergent med sum s* , hvis $\lim_{n \rightarrow \infty} s_n = s$. I så fald skriver man

$$\sum_{i=1}^{\infty} x_i = s.$$

Nu følger en liste over resultater vedrørende uendelige rækker, som ofte benyttes i sandsynlighedsregning. Der gives ingen beviser her.

1) Hvis rækken $\sum_{i=1}^{\infty} x_i$ er konvergent, må der gælde, at $x_n \rightarrow 0$.

2) Antag $x_i \geq 0$, og at rækken $\sum_{i=1}^{\infty} x_i$ er konvergent. Hvis $|y_i| \leq x_i$ for alle $i \in \mathbb{N}$, er også rækken $\sum_{i=1}^{\infty} y_i$ konvergent.

3) Hvis rækkerne $\sum_{i=1}^{\infty} x_i$ og $\sum_{i=1}^{\infty} y_i$ er konvergente, og a og b er reelle tal, er

$$\sum_{i=1}^{\infty} (ax_i + by_i) = a \sum_{i=1}^{\infty} x_i + b \sum_{i=1}^{\infty} y_i.$$

4) Lad $\{x_{ij} | i \in \mathbb{N}, j \in \mathbb{N}\}$ være en reel dobbelt talfølge, for hvilken $x_{ij} \geq 0$ for alle i og j . Da er

$$\sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} x_{ij} \right) = \sum_{j=1}^{\infty} \left(\sum_{i=1}^{\infty} x_{ij} \right),$$

i den forstand, at hvis en af siderne eksisterer, så eksisterer den anden også, og de er lige store. Mere præcist gælder, at hvis rækken $\sum_{j=1}^{\infty} x_{ij} = s_i$ er konvergent for alle $i \in \mathbb{N}$, og hvis rækken $\sum_{i=1}^{\infty} s_i = s$ er konvergent, da er rækken $\sum_{i=1}^{\infty} x_{ij} = t_j$ konvergent for ethvert $j \in \mathbb{N}$, og rækken $\sum_{j=1}^{\infty} t_j$ er konvergent med sum s . Da summationsrækkefølgen er uden betydning, skriver man ofte

$$\sum_{i,j} x_{ij} = s.$$

Når dobbeltrækken er konvergent med sum s , gælder det yderligere for enhver bijektion $n \mapsto (i(n), j(n))$ af \mathbb{N} på \mathbb{N}^2 , at

$$\sum_{n=1}^{\infty} x_{i(n)j(n)} = s.$$

5) Lad $\{x_{ij} | i \in \mathbb{N}, j \in \mathbb{N}\}$ være en reel dobbelt talfølge, for hvilken $\sum_{i,j} |x_{ij}|$ er konvergent. Da er

$$\sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} x_{ij} \right) = \sum_{j=1}^{\infty} \left(\sum_{i=1}^{\infty} x_{ij} \right),$$

i den forstand, at hvis en af siderne eksisterer, så eksisterer den anden også, og de er lige store. Yderligere gælder for enhver bijektion $n \mapsto (i(n), j(n))$ af \mathbb{N} på \mathbb{N}^2 , at

$$\sum_{n=1}^{\infty} x_{i(n)j(n)} = s.$$

6) Lad $\{x_{ij} | i \in \mathbb{N}, j = 1, \dots, i\}$ være en såkaldt trekantsfølge af reelle tal. Hvis enten $x_{ij} \geq 0$ for alle i og j , eller hvis $\sum_{i=1}^{\infty} (\sum_{j=1}^i |x_{ij}|)$ er konvergent, så er

$$\sum_{i=1}^{\infty} \left(\sum_{j=1}^i x_{ij} \right) = \sum_{j=1}^{\infty} \left(\sum_{i=j}^{\infty} x_{ij} \right).$$

Eksempel C.1 For ethvert reelt tal $a \neq 1$ er

$$s_n = \sum_{i=0}^n a^i = \frac{1 - a^{n+1}}{1 - a}, \quad (\text{C.0.1})$$

hvilket følger af at $s_n - as_n = 1 - a^{n+1}$. Hvis $|a| < 1$, er rækken $\sum_{i=0}^{\infty} a^i$ derfor konvergent og

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1 - a}.$$

Hvad er $\sum_{i=1}^{\infty} a^i$?

□

Eksempel C.2 Det kan vises, at der for ethvert reelt tal x gælder, at

$$\sum_{i=0}^{\infty} \frac{x^i}{i!} = \exp(x).$$

□

Eksempel C.3 Da det kan indses, at

$$\sum_{i=1}^n \frac{1}{i} \geq \log(n + 1),$$

er rækken $\sum_{i=1}^{\infty} \frac{1}{i}$ ikke konvergent.

□

Appendiks D

Om integraler

D.1 Funktioner af én variabel

1) Det antages, at læseren er fortrolig med integralet af en funktion f , som er kontinuert på et endeligt interval $[a, b]$. Der mindes om, at integralet er grænseværdien af en følge af middelsummer svarende til en inddeling af intervallet $[a, b]$, som for eksempel

$$a < a + \frac{b-a}{n} < a + 2\frac{b-a}{n} < \dots < a + (n-1)\frac{b-a}{n} < b,$$

d.v.s.

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \frac{b-a}{n} \sum_{i=1}^n f\left(a + \left(i - \frac{1}{2}\right) \cdot (b-a)/n\right).$$

2) Antag nu, at f er en begrænset funktion på det endelige interval $[a, b]$, for hvilken der findes reelle tal $a_0 = a < a_1 < \dots < a_{n-1} < a_n = b$, så f er kontinuert på ethvert af intervallerne (a_{i-1}, a_i) , $i = 1, \dots, n$. Vi kalder sådan en funktion en begrænset stykvis kontinuert funktion. Integralet af f kan defineres ved

$$\int_a^b f(x)dx = \sum_{i=1}^n \int_{a_{i-1}}^{a_i} f(x)dx$$

3) Vi har brug for at integrere funktioner, som er defineret på hele den reelle akse. Lad f være en funktion fra \mathbb{R} ind i $[0, \infty)$, som er begrænset

og stykvis kontinuert på ethvert af intervallerne $[-n, n]$, $n = 1, 2, \dots$. Hvis den voksende følge

$$\int_{-n}^n f(x) dx$$

er konvergent for $n \rightarrow \infty$, siger vi, at f er integrabel (på \mathbb{R}) med integralet

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{n \rightarrow \infty} \int_{-n}^n f(x) dx$$

Hvis følgen $\int_{-n}^n f(x) dx$ går mod uendelig, skriver man $\int_{-\infty}^{\infty} f(x) dx = \infty$. I denne situation siger man, at f ikke er integrabel.

Hvis en ikke-negativ funktion f kun er defineret på $[0, \infty)$ og er begrænset og stykvis lineær på ethvert af intervallerne $[0, n]$, $n = 1, 2, \dots$, betragter man i stedet følgen

$$\int_0^n f(x) dx.$$

Hvis denne er konvergent, siger vi, at f er integrabel (på $[0, \infty)$) med integralet

$$\int_0^{\infty} f(x) dx = \lim_{n \rightarrow \infty} \int_0^n f(x) dx.$$

Hvis f er defineret på intervaller af formen $(-\infty, a]$ eller $[a, \infty)$ kan man forsøge sig med en tilsvarende fremgangsmåde.

Eksempel D.1.1 Funktion $f(x) = e^{-x}$ er kontinuert og afbilder $[0, \infty)$ ind i $[0, 1]$. Da

$$\int_0^n e^{-x} dx = 1 - e^{-n} \rightarrow 1,$$

når $n \rightarrow \infty$, er f integrabel på $[0, \infty)$, og

$$\int_0^{\infty} e^{-x} dx = 1.$$

□

Eksempel D.1.2 Funktion $f(x) = 1/(1+x)$ er kontinuert og afbilder $[0, \infty)$ ind i $[0, 1]$. Da

$$\int_0^n 1/(1+x) dx = \log(1+n),$$

er f ikke integrabel på $[0, \infty)$. □

4) Også funktioner, som er defineret på et begrænset interval, kan give problemer. Funktionen $f(x) = 1/\sqrt{x}$ defineret på $(0, 1]$ er et eksempel. Den er ikke begrænset. Lad mere generelt f være en ikke-negativ funktion defineret på intervallet $(a, b]$, som er kontinuert på $(a, b]$, men for hvilken $f(x) \rightarrow \infty$, når $x \rightarrow a$. Da eksisterer integralerne

$$\int_{a+\frac{1}{n}}^b f(x)dx.$$

Hvis de konvergerer for $n \rightarrow \infty$, siger vi, at f er integrabel på $(a, b]$ med integral

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \int_{a+\frac{1}{n}}^b f(x)dx.$$

I modsat fald skriver vi $\int_a^b f(x)dx = \infty$ og siger, at f ikke er integrabel.

Eksempel D.1.3 Betragt funktionen $f(x) = 1/\sqrt{x}$ for $x \in (0, 1]$. Da

$$\int_{\frac{1}{n}}^1 \frac{1}{\sqrt{x}} dx = 2 - \frac{2}{\sqrt{n}} \rightarrow 2$$

for $n \rightarrow \infty$, er f altså integrabel på $(0, 1]$ med integralet 2. □

Eksempel D.1.4 Betragt mere generelt funktionen $f(x) = x^{-\alpha}$ for $x \in (0, 1]$, hvor α er et positivt reelt tal. Da

$$\int_{\frac{1}{n}}^1 x^{-\alpha} dx = \frac{1 - n^{\alpha-1}}{1 - \alpha}$$

for $\alpha \neq 1$, og da

$$\int_{\frac{1}{n}}^1 x^{-1} dx = \log(n),$$

ser vi, at f er integrabel med integral $1/(1 - \alpha)$, når $0 < \alpha < 1$, mens f ikke er integrabel, når $\alpha \geq 1$. □

For endelige intervaller, hvor der er problemer i den anden ende eller i begge ender, kan en tilsvarende metode anvendes.

5) Lad os nu betragte funktioner, som også kan antage negative værdier. Her skal man passe lidt på, som følgende eksempel viser.

Eksempel D.1.5 Funktionen $f(x) = x/(1 + x^2)$ opfylder, at

$$\int_{-n}^n x/(1 + x^2) dx = 0$$

for alle n , da f er en ulige funktion. Man kunne derfor fristes til at mene, at f er integrabel på \mathbb{R} . På den anden side, er

$$\int_0^n x/(1 + x^2) dx = \frac{1}{2} \log(1 + n^2) \rightarrow \infty$$

for $n \rightarrow \infty$, så f er ikke integrabel på $[0, \infty)$. På lignende måde indses det, at f heller ikke er integrabel på $(-\infty, 0]$. Det er derfor urimeligt at sige, at f er integrabel på \mathbb{R} . □

For at undgå urimeligheder, som at kalde funktionen i Eksempel D.1.5 integrabel, siger vi, at en reel funktion f er integrabel, hvis den ikke-negative funktion $|f|$ er integrabel. Hvis $|f|$ er integrabel, kan vi uden problemer definere integralet af f , som vi ovenfor gjorde for ikke-negative funktioner.

Eksempel D.1.6 Funktionen $f(x) = x(1 + x^2)^{-2}$ opfylder, at

$$\int_0^n x(1 + x^2)^{-2} dx = \frac{1}{2}(1 - (1 + n^2)^{-1}) \rightarrow \frac{1}{2}$$

for $n \rightarrow \infty$. Læseren kan overbevise sig om, at dette medfører, at $|f|$, som er symmetrisk om nul, er integrabel på \mathbb{R} . Integralet af den ulige funktion f over hele \mathbb{R} er nul. □

Følgende sammenligningskriterium er ikke vanskeligt at bevise ud fra definitionen på integrabilitet.

Sætning D.1.7 *Lad f og g være to funktioner fra intervallet I ind i \mathbb{R} . Hvis g er integrabel på I , og hvis $|f(x)| \leq |g(x)|$ for alle $x \in I$, så er f også integrabel på I .*

Eksempel D.1.6 (fortsat) Vi kan ved hjælp af Sætning D.1.7 give et enklere argument for, at funktionen $f(x) = x(1+x^2)^{-2}$ er integrabel på \mathbb{R} . Den opfylder nemlig, at $|f(x)| \leq g(x)$ for alle $x \in \mathbb{R}$, hvor

$$g(x) = \begin{cases} 1 & \text{for } |x| \leq 1 \\ |x|^{-3} & \text{for } |x| > 1, \end{cases}$$

og det er ikke vanskeligt at se, at g er integrabel på \mathbb{R} . □

D.2 Funktioner af flere variable

Vi har også brug for at integrere funktioner af flere variable. For at lette notationen betragter vi i dette appendiks hovedsagelig funktioner af to variable.

Vi vil ikke bevise noget her, kun forsøge at motivere en definition af integration i \mathbb{R}^2 . Betragt en kontinuert funktion $f : [a_1, a_2] \times [b_1, b_2] \mapsto \mathbb{R}$. Det er naturligt at definere integralet af f over $A = [a_1, a_2] \times [b_1, b_2]$ som en grænseværdi:

$$\int_A f(x, y) dx dy = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sum_{j=1}^n f(x_i, y_j) \frac{(a_2 - a_1)(b_2 - b_1)}{n^2}. \quad (\text{D.2.1})$$

hvor $x_i = a_1 + \left(i - \frac{1}{2}\right) \cdot (a_2 - a_1)/n$ og $y_j = b_1 + \left(j - \frac{1}{2}\right) \cdot (b_2 - b_1)/n$. Vi har inddelt A i n^2 små rektangler, hver med arealet $(a_2 - a_1)(b_2 - b_1)n^{-2}$. Punkterne (x_i, y_j) er midtpunkterne i disse rektangler.

Hvordan kan man beregne dette integral? For at belyse det, holder vi først i fast i (D.2.1). Vi kan da opfatte

$$\sum_{j=1}^n f(x_i, y_j) \frac{(b_2 - b_1)}{n}$$

som en middelsum, der approksimerer integralet

$$\int_{b_1}^{b_2} f(x_i, y) dy.$$

Definer funktionen

$$f_1(x) = \int_{b_1}^{b_2} f(x, y) dy, \quad x \in [a_1, a_2].$$

Her holdes x altså fast, mens der integreres med hensyn til y . Vi har da, at

$$\begin{aligned} \int_A f(x, y) dx dy &\simeq \sum_{i=1}^n \left(\sum_{j=1}^n f(x_i, y_j) \frac{(b_2 - b_1)}{n} \right) \frac{(a_2 - a_1)}{n} \\ &\simeq \sum_{i=1}^n f_1(x_i) \frac{(a_2 - a_1)}{n} \\ &\simeq \int_{a_1}^{a_2} f_1(x) dx, \end{aligned}$$

idet summen i næstsidste linie er en middelsum, som approksimerer integralet i sidste linie. Vi har dermed givet et intuitivt argument for, at integralet af f over A kan beregnes ved formlen

$$\int_A f(x, y) dx dy = \int_{a_1}^{a_2} \left(\int_{b_1}^{b_2} f(x, y) dy \right) dx, \quad (\text{D.2.2})$$

altså som et såkaldt dobbeltintegral. Først holdes x fast, og der integreres med hensyn til y . Derved opnås en funktion af x , som dernæst integreres over $[a_1, a_2]$. Det er sjældent, man ulejlig sig med at skrive parentesen omkring det indre integral.

Det er klart, at vi lige så godt kunne have defineret integralet af f over A ved

$$\int_A f(x, y) dx dy = \int_{b_1}^{b_2} \left(\int_{a_1}^{a_2} f(x, y) dx \right) dy. \quad (\text{D.2.3})$$

Det kan vises, at den rækkefølge, i hvilken vi vælger at integrere m.h.t. x og y er ligegyldig. Resultatet bliver det samme.

Eksempel D.2.1 Funktionen $f(x, y) = x \wedge y$ er kontinuert på $A = [0, 1] \times [0, 1]$. Da

$$f_1(x) = \int_0^1 (x \wedge y) dy = \int_0^x y dy + \int_x^1 x dy = x^2/2 + x(1 - x) = x - x^2/2,$$

er

$$\int_A (x \wedge y) dx dy = \int_0^1 (x - x^2/2) dx = \frac{1}{3}.$$

□

Vi kan definere integralet af f over A ved (D.2.2) for enhver funktion f , for hvilken de to integrationer kan udføres. Hvis for eksempel funktionen f opfylder, at $y \mapsto f(x, y)$ er en begrænset stykvis kontinuert funktion for enhver fastholdt værdi af x , og hvis $x \mapsto \int_{b_1}^{b_2} f(x, y) dy$ er af samme type, kan vi definere integralet af f over A ved (D.2.2).

Eksempel D.2.2 Betragt funktionen f defineret på $A = [0, 1] \times [0, 1]$ ved $f(x, y) = 1_B(x, y)$, hvor $B = \{(x, y) \in A \mid x \geq y\}$. Da

$$f_1(x) = \int_0^1 1_B(x, y) dy = \int_0^x 1 dy + \int_x^1 0 dy = x,$$

er

$$\int_A 1_B(x, y) dx dy = \int_0^1 x dx = \frac{1}{2}.$$

Bemærk, at integralet er lig arealet af B .

□

Hvis vi vil integrere en funktion over hele \mathbb{R}^2 , må vi ligesom i Afsnit D.1 passe lidt på. Igen antager vi først, at $f(x, y) \geq 0$ for alle $(x, y) \in \mathbb{R}^2$. Vi definerer $A_n = [-n, n] \times [-n, n]$ og siger, at f er integrabel (på \mathbb{R}^2), hvis den voksende følge

$$\int_{A_n} f(x, y) dx dy$$

konvergerer for $n \rightarrow \infty$. I så fald defineres

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \lim_{n \rightarrow \infty} \int_{A_n} f(x, y) dx dy. \quad (\text{D.2.4})$$

Hvis følgen $\int_{A_n} f(x, y) dx dy$ går mod uendelig, siger man, at f ikke er integrabel. Hvis f kan antage både negative og positive værdier, må man først undersøge, om $|f|$ er integrabel. Hvis den er det, siger man, at f er integrabel, og definere integralet af f over \mathbb{R}^2 ved (D.2.4).

Hvis $f \geq 0$, kan man beregne integralet af f over \mathbb{R}^2 på to måder ved hjælp af dobbeltintegraler

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dx \right) dy.$$

Hvis et af dobbeltintegralerne giver et endeligt tal, er f integrabel, og integralet er lig begge dobbeltintegraler. Hvis f kan antage både negative og positive værdier, kan integralet af f over \mathbb{R}^2 også beregnes ved de to dobbeltintegraler, *hvis* man vel at mærke ved, at f er integrabel. Man må altså først undersøge om $|f|$ er integrabel, f. eks. ved at beregne et dobbeltintegral.

Hvis B er en begrænset delmængde af \mathbb{R}^2 , som er pæn nok til, at den kan tilskrives et areal, gælder det generelt at

$$\int_{\mathbb{R}^2} 1_B(x, y) dx dy = |B|, \quad (\text{D.2.5})$$

hvor $|B|$ betegner arealet af B , se Eksempel D.2.2. Ved igen at se på (D.2.1), indser man, at dette ikke er så mærkeligt. Summen, der approksimerer integralet, er jo netop en sum af arealer af små rektangler, som omtrent dækker B .

Alt hvad der her er sagt om funktioner af to variable, kan gentages ord til andet om funktioner af flere variable. For $f : \mathbb{R}^n \mapsto [0, \infty)$ er altså

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_{-\infty}^{\infty} \cdots \left(\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \right) dx_{n-1} \right) \cdots dx_1.$$

Integrationerne kan også udføres i en anden rækkefølge. Hvis f kan antage både negative og positive værdier, kan integralet af f over \mathbb{R}^n udregnes på samme måde, hvis $|f|$ er integrabel over \mathbb{R}^n .

For en delmængde B af \mathbb{R}^3 er

$$\int_{\mathbb{R}^3} 1_B(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

lig B s volumen (rumfang). Generelt kan vi definere *volumenet af en delmængde B af \mathbb{R}^n* som

$$|B| = \int_{\mathbb{R}^n} 1_B(x_1, \dots, x_n) dx_1 \cdots dx_n. \quad (\text{D.2.6})$$

Hvis indikatorfunktionen 1_B ikke er integrabel, siges B at have uendeligt volumen.

Hvis en funktion f kun er defineret på en delmængde B af \mathbb{R}^n , eller hvis vi kun ønsker at integrere f over mængden B , kan vi definere integralet af f over B ved

$$\int_B f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_{\mathbb{R}^n} 1_B(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

hvor 1_B er indikatorfunktionen for B . Altså ved at sætte funktionen lig nul udenfor B .

Appendiks E

Tabeller

I dette appendiks gives nogle tabeller vedrørende normalfordelingen, χ^2 -fordelingen, t-fordelingen og F-fordelingen, som er nyttige i forbindelse med test af statistiske hypoteser. Endvidere angives det græske alfabet.

E.1 Standard normalfordelingen

Tabellen giver samhørende værdier u og p , så der for en standard normalfordelt stokastisk variabel U gælder, at $P(U > u) = p$.

p	u	p	u	p	u	p	u	p	u
0.500	0.000	0.050	1.645	0.0050	2.576	0.000500	3.291	0.000050	3.891
0.490	0.025	0.049	1.655	0.0049	2.583	0.000490	3.296	0.000049	3.895
0.480	0.050	0.048	1.665	0.0048	2.590	0.000480	3.302	0.000048	3.900
0.470	0.075	0.047	1.675	0.0047	2.597	0.000470	3.308	0.000047	3.906
0.460	0.100	0.046	1.685	0.0046	2.605	0.000460	3.314	0.000046	3.911
0.450	0.126	0.045	1.695	0.0045	2.612	0.000450	3.320	0.000045	3.916
0.440	0.151	0.044	1.706	0.0044	2.620	0.000440	3.326	0.000044	3.921
0.430	0.176	0.043	1.717	0.0043	2.628	0.000430	3.333	0.000043	3.927
0.420	0.202	0.042	1.728	0.0042	2.636	0.000420	3.339	0.000042	3.933
0.410	0.228	0.041	1.739	0.0041	2.644	0.000410	3.346	0.000041	3.938
0.400	0.253	0.040	1.751	0.0040	2.652	0.000400	3.353	0.000040	3.944
0.390	0.279	0.039	1.762	0.0039	2.661	0.000390	3.360	0.000039	3.950
0.380	0.305	0.038	1.774	0.0038	2.669	0.000380	3.367	0.000038	3.957
0.370	0.332	0.037	1.787	0.0037	2.678	0.000370	3.374	0.000037	3.963
0.360	0.358	0.036	1.799	0.0036	2.687	0.000360	3.382	0.000036	3.970
0.350	0.385	0.035	1.812	0.0035	2.697	0.000350	3.390	0.000035	3.976
0.340	0.412	0.034	1.825	0.0034	2.706	0.000340	3.398	0.000034	3.983
0.330	0.440	0.033	1.838	0.0033	2.716	0.000330	3.406	0.000033	3.990
0.320	0.468	0.032	1.852	0.0032	2.727	0.000320	3.414	0.000032	3.998
0.310	0.496	0.031	1.866	0.0031	2.737	0.000310	3.423	0.000031	4.005
0.300	0.524	0.030	1.881	0.0030	2.748	0.000300	3.432	0.000030	4.013
0.290	0.553	0.029	1.896	0.0029	2.759	0.000290	3.441	0.000029	4.021
0.280	0.583	0.028	1.911	0.0028	2.770	0.000280	3.450	0.000028	4.029
0.270	0.613	0.027	1.927	0.0027	2.782	0.000270	3.460	0.000027	4.038
0.260	0.643	0.026	1.943	0.0026	2.794	0.000260	3.470	0.000026	4.046
0.250	0.674	0.025	1.960	0.0025	2.807	0.000250	3.481	0.000025	4.056
0.240	0.706	0.024	1.977	0.0024	2.820	0.000240	3.492	0.000024	4.065
0.230	0.739	0.023	1.995	0.0023	2.834	0.000230	3.503	0.000023	4.075
0.220	0.772	0.022	2.014	0.0022	2.848	0.000220	3.515	0.000022	4.085
0.210	0.806	0.021	2.034	0.0021	2.863	0.000210	3.527	0.000021	4.096
0.200	0.842	0.020	2.054	0.0020	2.878	0.000200	3.540	0.000020	4.107
0.190	0.878	0.019	2.075	0.0019	2.894	0.000190	3.554	0.000019	4.119
0.180	0.915	0.018	2.097	0.0018	2.911	0.000180	3.568	0.000018	4.132
0.170	0.954	0.017	2.120	0.0017	2.929	0.000170	3.583	0.000017	4.145
0.160	0.994	0.016	2.144	0.0016	2.948	0.000160	3.599	0.000016	4.159
0.150	1.036	0.015	2.170	0.0015	2.968	0.000150	3.615	0.000015	4.173
0.140	1.080	0.014	2.197	0.0014	2.989	0.000140	3.633	0.000014	4.189
0.130	1.126	0.013	2.226	0.0013	3.011	0.000130	3.652	0.000013	4.206
0.120	1.175	0.012	2.257	0.0012	3.036	0.000120	3.673	0.000012	4.224
0.110	1.227	0.011	2.290	0.0011	3.062	0.000110	3.695	0.000011	4.244
0.100	1.282	0.010	2.326	0.0010	3.090	0.000100	3.719	0.000010	4.265
0.090	1.341	0.009	2.366	0.0009	3.121	0.000090	3.746	0.000009	4.288
0.080	1.405	0.008	2.409	0.0008	3.156	0.000080	3.775	0.000008	4.314
0.070	1.476	0.007	2.457	0.0007	3.195	0.000070	3.808	0.000007	4.344
0.060	1.555	0.006	2.512	0.0006	3.239	0.000060	3.846	0.000006	4.378

E.2 χ^2 -fordelingen

For hvert p og f angives tallet x , som opfylder, at $P(X > x) = p$, hvis X er χ^2 -fordelt med f frihedsgrader.

f	$p = 0.05$	$p = 0.01$	$p = 0.005$	$p = 0.001$	$p = 0.0001$
1	3.841	6.635	7.879	10.828	15.137
2	5.991	9.210	10.597	13.816	18.421
3	7.815	11.345	12.838	16.266	21.108
4	9.488	13.277	14.860	18.467	23.513
5	11.070	15.086	16.750	20.515	25.745
6	12.592	16.812	18.548	22.458	27.856
7	14.067	18.475	20.278	24.322	29.878
8	15.507	20.090	21.955	26.124	31.828
9	16.919	21.666	23.589	27.877	33.720
10	18.307	23.209	25.188	29.588	35.564
11	19.675	24.725	26.757	31.264	37.367
12	21.026	26.217	28.300	32.909	39.134
13	22.362	27.688	29.819	34.528	40.871
14	23.685	29.141	31.319	36.123	42.579
15	24.996	30.578	32.801	37.697	44.263
16	26.296	32.000	34.267	39.252	45.925
17	27.587	33.409	35.718	40.790	47.566
18	28.869	34.805	37.156	42.312	49.189
19	30.144	36.191	38.582	43.820	50.795
20	31.410	37.566	39.997	45.315	52.386
21	32.671	38.932	41.401	46.797	53.962
22	33.924	40.289	42.796	48.268	55.525
23	35.172	41.638	44.181	49.728	57.075
24	36.415	42.980	45.559	51.179	58.613
25	37.652	44.314	46.928	52.620	60.140
26	38.885	45.642	48.290	54.052	61.657
27	40.113	46.963	49.645	55.476	63.164
28	41.337	48.278	50.993	56.892	64.662
29	42.557	49.588	52.336	58.301	66.152
30	43.773	50.892	53.672	59.703	67.633
31	44.985	52.191	55.003	61.098	69.106
32	46.194	53.486	56.328	62.487	70.571
33	47.400	54.776	57.648	63.870	72.030
34	48.602	56.061	58.964	65.247	73.481
35	49.802	57.342	60.275	66.619	74.926
36	50.998	58.619	61.581	67.985	76.365
37	52.192	59.893	62.883	69.346	77.798
38	53.384	61.162	64.181	70.703	79.225
39	54.572	62.428	65.476	72.055	80.646
40	55.758	63.691	66.766	73.402	82.062
41	56.942	64.950	68.053	74.745	83.473
42	58.124	66.206	69.336	76.084	84.879
43	59.304	67.459	70.616	77.419	86.281
44	60.481	68.710	71.893	78.750	87.677
45	61.656	69.957	73.166	80.077	89.070

E.3 t-fordelingen

For hvert p og f angives tallet x , som opfylder, at $P(X > x) = p$, hvis X er t-fordelt med f frihedsgrader.

f	$p = 0.05$	$p = 0.025$	$p = 0.01$	$p = 0.001$	$p = 0.0001$
1	6.314	12.706	31.821	318.309	3183.099
2	2.920	4.303	6.965	22.327	70.700
3	2.353	3.182	4.541	10.215	22.204
4	2.132	2.776	3.747	7.173	13.034
5	2.015	2.571	3.365	5.893	9.678
6	1.943	2.447	3.143	5.208	8.025
7	1.895	2.365	2.998	4.785	7.063
8	1.860	2.306	2.896	4.501	6.442
9	1.833	2.262	2.821	4.297	6.010
10	1.812	2.228	2.764	4.144	5.694
11	1.796	2.201	2.718	4.025	5.453
12	1.782	2.179	2.681	3.930	5.263
13	1.771	2.160	2.650	3.852	5.111
14	1.761	2.145	2.624	3.787	4.985
15	1.753	2.131	2.602	3.733	4.880
16	1.746	2.120	2.583	3.686	4.791
17	1.740	2.110	2.567	3.646	4.714
18	1.734	2.101	2.552	3.610	4.648
19	1.729	2.093	2.539	3.579	4.590
20	1.725	2.086	2.528	3.552	4.539
22	1.717	2.074	2.508	3.505	4.452
24	1.711	2.064	2.492	3.467	4.382
26	1.706	2.056	2.479	3.435	4.324
28	1.701	2.048	2.467	3.408	4.275
30	1.697	2.042	2.457	3.385	4.234
32	1.694	2.037	2.449	3.365	4.198
34	1.691	2.032	2.441	3.348	4.167
36	1.688	2.028	2.434	3.333	4.140
38	1.686	2.024	2.429	3.319	4.116
40	1.684	2.021	2.423	3.307	4.094
45	1.679	2.014	2.412	3.281	4.049
50	1.676	2.009	2.403	3.261	4.014
55	1.673	2.004	2.396	3.245	3.986
60	1.671	2.000	2.390	3.232	3.962
70	1.667	1.994	2.381	3.211	3.926
80	1.664	1.990	2.374	3.195	3.899
90	1.662	1.987	2.368	3.183	3.878
100	1.660	1.984	2.364	3.174	3.862
125	1.657	1.979	2.357	3.157	3.832
150	1.655	1.976	2.351	3.145	3.813
175	1.654	1.974	2.348	3.137	3.799
200	1.653	1.972	2.345	3.131	3.789
300	1.650	1.968	2.339	3.118	3.765
400	1.649	1.966	2.336	3.111	3.754
500	1.648	1.965	2.334	3.107	3.747
∞	1.645	1.960	2.326	3.090	3.719

E.4 F-fordelingen

95 procent fraktiler.

	f_1														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46
f_2 15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04	2.01	1.99	1.96
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80	1.78
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74	1.72
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.77	1.74	1.71	1.69

	f_1														
	16	18	20	25	30	35	40	50	60	70	80	90	100	200	500
1	246	247	248	249	250	251	251	252	252	252	253	253	253	254	254
2	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	8.69	8.67	8.66	8.63	8.62	8.60	8.59	8.58	8.57	8.57	8.56	8.56	8.55	8.54	8.53
4	5.84	5.82	5.80	5.77	5.75	5.73	5.72	5.70	5.69	5.68	5.67	5.67	5.66	5.65	5.64
5	4.60	4.58	4.56	4.52	4.50	4.48	4.46	4.44	4.43	4.42	4.41	4.41	4.41	4.39	4.37
6	3.92	3.90	3.87	3.83	3.81	3.79	3.77	3.75	3.74	3.73	3.72	3.72	3.71	3.69	3.68
7	3.49	3.47	3.44	3.40	3.38	3.36	3.34	3.32	3.30	3.29	3.29	3.28	3.27	3.25	3.24
f_2 8	3.20	3.17	3.15	3.11	3.08	3.06	3.04	3.02	3.01	2.99	2.99	2.98	2.97	2.95	2.94
9	2.99	2.96	2.94	2.89	2.86	2.84	2.83	2.80	2.79	2.78	2.77	2.76	2.76	2.73	2.72
10	2.83	2.80	2.77	2.73	2.70	2.68	2.66	2.64	2.62	2.61	2.60	2.59	2.59	2.56	2.55
11	2.70	2.67	2.65	2.60	2.57	2.55	2.53	2.51	2.49	2.48	2.47	2.46	2.46	2.43	2.42
12	2.60	2.57	2.54	2.50	2.47	2.44	2.43	2.40	2.38	2.37	2.36	2.36	2.35	2.32	2.31
13	2.51	2.48	2.46	2.41	2.38	2.36	2.34	2.31	2.30	2.28	2.27	2.27	2.26	2.23	2.22
14	2.44	2.41	2.39	2.34	2.31	2.28	2.27	2.24	2.22	2.21	2.20	2.19	2.19	2.16	2.14
15	2.38	2.35	2.33	2.28	2.25	2.22	2.20	2.18	2.16	2.15	2.14	2.13	2.12	2.10	2.08

99 procent fraktiler.

		f_1														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6126	6143	6157
2		98.5	99.0	99.2	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3		34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.4	27.2	27.1	27.1	27.0	26.9	26.9
4		21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.6	14.5	14.4	14.3	14.3	14.2
5		16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82	9.77	9.72
6		13.8	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56
7		12.3	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31
8		11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52
9		10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96
10		10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56
11		9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25
12		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01
13		9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82
14		8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66
f_2 15		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52
16		8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41
18		8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23
20		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
25		7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85
30		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
35		7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64	2.60
40		7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
50		7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.51	2.46	2.42
60		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35
70		7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31
80		6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27
90		6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24
100		6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27	2.22
200		6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17	2.13
500		6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.28	2.22	2.17	2.12	2.07

		f_1														
		16	18	20	25	30	35	40	50	60	70	80	90	100	200	500
1		6170	6192	6209	6240	6261	6276	6287	6303	6313	6321	6326	6331	6334	6350	6360
2		99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3		26.8	26.8	26.7	26.6	26.5	26.5	26.4	26.4	26.3	26.3	26.3	26.3	26.2	26.2	26.2
4		14.2	14.1	14.0	13.9	13.8	13.8	13.8	13.7	13.7	13.6	13.6	13.6	13.6	13.5	13.5
5		9.68	9.61	9.55	9.45	9.38	9.33	9.29	9.24	9.20	9.18	9.16	9.14	9.13	9.08	9.04
6		7.52	7.45	7.40	7.30	7.23	7.18	7.14	7.09	7.06	7.03	7.01	7.00	6.99	6.93	6.90
7		6.28	6.21	6.16	6.06	5.99	5.94	5.91	5.86	5.82	5.80	5.78	5.77	5.75	5.70	5.67
f_2 8		5.48	5.41	5.36	5.26	5.20	5.15	5.12	5.07	5.03	5.01	4.99	4.97	4.96	4.91	4.88
9		4.92	4.86	4.81	4.71	4.65	4.60	4.57	4.52	4.48	4.46	4.44	4.43	4.41	4.36	4.33
10		4.52	4.46	4.41	4.31	4.25	4.20	4.17	4.12	4.08	4.06	4.04	4.03	4.01	3.96	3.93
11		4.21	4.15	4.10	4.01	3.94	3.89	3.86	3.81	3.78	3.75	3.73	3.72	3.71	3.66	3.62
12		3.97	3.91	3.86	3.76	3.70	3.65	3.62	3.57	3.54	3.51	3.49	3.48	3.47	3.41	3.38
13		3.78	3.72	3.66	3.57	3.51	3.46	3.43	3.38	3.34	3.32	3.30	3.28	3.27	3.22	3.19
14		3.62	3.56	3.51	3.41	3.35	3.30	3.27	3.22	3.18	3.16	3.14	3.12	3.11	3.06	3.03
15		3.49	3.42	3.37	3.28	3.21	3.17	3.13	3.08	3.05	3.02	3.00	2.99	2.98	2.92	2.89

99.9 procent fraktiler.

Fraktilerne svarende til $f_2 = 1$ er udeladt, da de ligger mellem 405284 og 635983.

	f_1														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3	167	149	141	137	135	133	132	131	130	129	129	128	128	128	127
4	74.1	61.3	56.2	53.4	51.7	50.5	49.7	49.0	48.5	48.1	47.7	47.4	47.2	47.0	46.8
5	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.7	27.2	26.9	26.7	26.4	26.2	26.1	25.9
6	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7	18.4	18.9	18.0	17.8	17.7	17.6
7	29.3	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.3	14.1	13.9	13.7	13.6	13.4	13.3
8	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.1	11.8	11.5	11.4	11.2	11.1	10.9	10.8
9	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1	9.89	9.72	9.57	9.44	9.33	9.24
10	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.59	8.45	8.32	8.22	8.13
11	19.7	13.8	11.6	10.4	9.58	9.05	8.66	8.35	8.12	7.92	7.76	7.63	7.51	7.41	7.32
12	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.14	7.00	6.89	6.79	6.71
13	17.8	12.3	10.2	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.65	6.52	6.41	6.31	6.23
14	17.1	11.8	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40	6.26	6.13	6.02	5.93	5.85
f_2 15	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.94	5.81	5.71	5.62	5.54
16	16.1	11.0	9.01	7.94	7.27	6.80	6.46	6.19	5.98	5.81	5.67	5.55	5.44	5.35	5.27
18	15.4	10.4	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.25	5.13	5.03	4.94	4.87
20	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.72	4.64	4.56
25	13.9	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71	4.56	4.42	4.31	4.22	4.13	4.06
30	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.91	3.82	3.75
35	12.9	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.90	3.79	3.70	3.62	3.55
40	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.55	3.47	3.40
50	12.2	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	3.55	3.44	3.35	3.27	3.20
60	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.42	3.32	3.23	3.15	3.08
70	11.8	7.64	6.06	5.20	4.66	4.28	3.99	3.77	3.60	3.45	3.33	3.23	3.14	3.06	2.99
80	11.7	7.54	5.97	5.12	4.58	4.20	3.92	3.70	3.53	3.39	3.27	3.16	3.07	3.00	2.93
90	11.6	7.47	5.91	5.06	4.53	4.15	3.87	3.65	3.48	3.34	3.22	3.11	3.02	2.95	2.88
100	11.5	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44	3.30	3.18	3.07	2.99	2.91	2.84
200	11.2	7.15	5.63	4.81	4.29	3.92	3.65	3.43	3.26	3.12	3.00	2.90	2.82	2.74	2.67
500	11.0	7.00	5.51	4.69	4.18	3.81	3.54	3.33	3.16	3.02	2.91	2.81	2.72	2.64	2.58

	f_1														
	16	18	20	25	30	35	40	50	60	70	80	90	100	200	500
2	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3	127	127	126	126	125	125	125	125	124	124	124	124	124	124	124
4	46.6	46.3	46.1	45.7	45.4	45.2	45.1	44.9	44.8	44.7	44.6	44.5	44.5	44.3	44.1
5	25.8	25.6	25.4	25.1	24.9	24.7	24.6	24.4	24.3	24.3	24.2	24.2	24.1	24.0	23.9
6	17.5	17.3	17.1	16.9	16.7	16.5	16.4	16.3	16.2	16.2	16.1	16.1	16.0	15.9	15.8
7	13.2	13.1	12.9	12.7	12.5	12.4	12.3	12.2	12.1	12.1	12.0	12.0	12.0	11.8	11.8
f_2 8	10.8	10.6	10.5	10.3	10.1	10.0	9.92	9.80	9.73	9.67	9.63	9.60	9.57	9.45	9.38
9	9.15	9.01	8.90	8.69	8.55	8.45	8.37	8.26	8.19	8.13	8.09	8.06	8.04	7.93	7.86
10	8.05	7.91	7.80	7.60	7.47	7.37	7.30	7.19	7.12	7.07	7.03	7.00	6.98	6.87	6.81
11	7.24	7.11	7.01	6.81	6.68	6.59	6.52	6.42	6.35	6.30	6.26	6.23	6.21	6.10	6.04
12	6.63	6.51	6.40	6.22	6.09	6.00	5.93	5.83	5.76	5.71	5.68	5.65	5.63	5.52	5.46
13	6.16	6.03	5.93	5.75	5.63	5.54	5.47	5.37	5.30	5.26	5.22	5.19	5.17	5.07	5.01
14	5.78	5.66	5.56	5.38	5.25	5.17	5.10	5.00	4.94	4.89	4.86	4.83	4.81	4.71	4.65
15	5.46	5.35	5.25	5.07	4.95	4.86	4.80	4.70	4.64	4.59	4.56	4.53	4.51	4.41	4.35

E.5 Det græske alfabet

Stort bogstav	Lille bogstav	Navn
A	α	alfa
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ϵ, ε	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ, ϑ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	my
N	ν	ny
Ξ	ξ	xi
O	\omicron	omikron
Π	π	pi
P	ρ, ϱ	ro
Σ	σ, ς	sigma
T	τ	tau
Y	υ	ypsilon
Φ	ϕ, φ	fi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Indeks

- absorberende tilstand, 252
- aktier, 269–271
- antal besøg til en web-site, 114–115
- antal ordnede sæt, 287
- antallet af delmængder, 287
- aperiodisk, 257
- arcus-sinus fordelingen, 174, 177

- Bayes' formel, 29
- begrænset stokastisk variabel, 131
- Bernoulli-Laplace diffusionsmodel, 243–244, 254–255, 258–259, 267
- Beta-fordelingen, 19, 149, 152, 158, 225, 240
- Beta-funktionen, 225
- betinget fordeling, 29–30, 139–140, 200–202
- betinget middelværdi, 238
- betinget sandsynlighed, 24
- betinget uafhængighed, 35
 - af diskrete stokastiske variable, 85–86, 129–130
- billedmængde, 282
- binomialfordelingen, 68–72, 74–75, 78, 88, 92, 94, 101, 114, 115, 214, 243, 245, 259
- binomialformlen, 288
- binomialkoefficient, 287–289
- Brownsk bevægelse, 9–10, 274–276
- Cauchy-fordelingen, 169–170, 176, 227
- centrale grænseværdisætning, 212
- Chapman-Kolmogorov ligningerne, 250
- Chebyshevs ulighed, 209
- χ^2 -fordelingen, 171, 221, 231

- de Moivres sætning, 216
- de store tals lov, 210
- den centrale grænseværdisætning, 212
- den geometriske fordeling, 119, 123–124, 131–132, 137–138, 140, 142, 172, 267
- den hypergeometriske fordeling, 72–75, 92, 98, 255, 259
- den logaritmiske normalfordeling, 176
- den n -dimensionale normalfordeling, 229, 233–235
- den negative binomialfordeling, 142
- den rektangulære fordeling, 148, 151, 154, 157, 161, 180
- den to-dimensionale normalfordeling, 236–238
- den tosidede eksponentialfordeling, 174
- den udartede fordeling, 95
- den uniforme fordeling, 148, 151, 154, 157, 161, 180
- disjunkte mængder, 281

- diskret fordeling, 118
 diskret stokastisk variabel, 120
 dobbeltintegral, 299
 dødelighedstavle, 25
 eksponentialfordelingen, 52–54, 146, 149, 154, 157, 160, 169, 172, 190, 201, 222, 271
 empirisk
 fordeling, 102
 korrelation, 103
 kovarians, 103
 middelværdi, 103
 varians, 103, 111
 ergodisk, 252
 Erlang-fordelingen, 204
 et tilfældigt tal mellem nul og en, 16–17, 51, 56, 169
 fakultetfunktionen, 286
 fejlophobningsloven, 208
 F-fordelingen, 190, 226
 fler-dimensionalt volumen, 301
 foldning
 af binomialfordelinger, 71
 af diskrete fordelinger, 84–85, 128
 af eksponentialfordelinger, 188, 204
 af Γ -fordelinger, 224
 af geometriske fordelinger, 142
 af kontinuerte fordelinger, 188
 af normalfordelinger, 195
 af Poissonfordelinger, 129
 af sandsynlighedsfunktioner, 85
 af sandsynlighedstætheder, 188
 fordeling
 af stokastisk variabel, 46
 af stokastisk vektor, 57
 diskret, 118
 kontinuert, 148, 180
 marginal, 57
 simultan, 57
 fordeling af blandet type, 155
 fordelingsfunktion
 for en fordeling på \mathbb{N}_0 , 122
 for en kontinuert fordeling, 151
 for en stokastisk variabel, 49–52
 for en stokastisk vektor, 58
 marginal, 58
 foreningsmængde, 280
 fraktil, 176
 frihedsgrader, 221, 226, 227
 fællesmængde, 280
 fødselsdage i en klasse, 13, 21
 Γ -fordelingen, 222, 272
 Γ -funktionen, 223
 Gauss-fordelingen, 162
 genetisk model, 243, 256–257
 geometriske fordeling, 119, 123–124, 131–132, 137–138, 140, 142, 172, 267
 hypergeometriske fordeling, 72–75, 92, 98, 255, 259
 hændelse, 12, 14, 18
 indikatorfunktionen, 284
 internetserver, 244, 260
 irreducibel, 252
 kast med en terning, 49–51, 79, 81
 kast med to terninger, 13, 15, 23, 24, 31, 47, 56–58, 89
 koblingsmetoden, 260

- komplementærmængde, 281
koncentreret, 65
kontinuert differentiabel, 153
kontinuert fordeling, 148, 180
kontinuert stokastisk variabel, 148
kontinuert stokastisk vektor, 180
konvergens i fordeling, 213
konvergens i sandsynlighed, 211
korrelation, 99, 138–139, 200, 237–238
kovarians, 95, 138–139, 199
kovariansmatrix, 112, 234
kugler i to kasser, 27, 28, 35, 85
kvadratet af en kontinuert stokastisk variabel, 171
kvotienten mellem to kontinuerte stokastiske variable, 189
kømodel, 243–245, 259–260
Laplace-fordelingen, 174
ligefordelingen
 på en begrænset mængde, 180
 på en endelig mængde, 14, 51, 63, 67
 på et begrænset interval, 148, 151, 154, 157, 161
livsforsikring, 266
logaritmiske normalfordeling, 176
marginal fordeling, 57
 for diskret stokastisk variabel, 124
 for kontinuert stokastisk variabel, 182
 for stokastisk variabel på en endelig mængde, 75
marginal fordelingsfunktion, 58
marginal kontinuert fordeling, 182
marginal sandsynlighedstæthed, 182
Markovegenskaben, 242, 248, 251
Markovkæde, 241–242
 anden-ordens, 268
 aperiodisk, 257
 ergodisk, 252
 irreducibel, 252
 med to tilstande, 250, 253, 256, 258
Markovs ulighed, 110
median, 176
Mendel, 79, 82
middelværdi
 af binomialfordelingen, 92
 af den geometriske fordeling, 131–132
 af den hypergeometriske fordeling, 92
 af eksponentialfordelingen, 157
 af en diskret stokastisk variabel, 87, 130
 af en kontinuert stokastisk variabel, 156
 af en sum af stokastiske variable, 91, 135, 198
 af en transformeret diskret stokastisk variabel, 89, 132
 af en transformeret kontinuert stokastisk variabel, 158
 af en transformeret kontinuert stokastisk vektor, 197
 af et produkt af uafhængige stokastiske variable, 91, 135, 198
af F-fordelingen, 226
af Γ -fordelingen, 239

- af ligefordelingen, 157
 af normalfordelingen, 163
 af Poissonfordelingen, 131
 af t-fordelingen, 227
 multinomialfordelingen, 78–82, 84, 102
 multinomialkoefficient, 79, 289–290
 mængdedifferens, 281
 møntkast, 33, 34, 45, 47, 68

 n-dimensionale normalfordeling, 229, 233–235
 nedadstigende faktoriel, 287
 negative binomialfordeling, 142
 normalfordelingen, 162–165, 195, 212, 259, 276
 den fler-dimensionale, 229, 233–235
 den to-dimensionale, 236–238

 originalmængde, 282
 ortonormal transformation af en normalfordeling, 230
 overgangsmatrix, 247
 m-trins, 247, 249
 overgangssandsynligheder, 242
 m-trins, 247

 Pareto-fordelingen, 157, 161
 Pascals trekant, 9, 288
 planintegral, 298
 Poisson processen, 273
 Poissonfordelingen, 115, 118, 121, 129, 131, 137, 146, 217, 243, 272
 polynomialfordelingen, 78–82, 84, 102
 polynomialkoefficient, 79, 289–290

 populationsmodel, 245–246
 positionsparameter, 168
 produktet af to kontinuerte stokastiske variable, 189
 produktmængde, 282
 punktsandsynlighed, 12, 120

 regressionskoefficient, 238
 rektangulære fordeling, 148, 151, 154, 157, 161, 180
 rencontreproblemet, 110
 rentemodell, 265

 sandsynlighedsfelt
 endeligt, 12
 generelt, 18
 sandsynlighedsfunktion, 12, 66–67, 118
 sandsynlighedsmål, 18
 sandsynlighedsregningens frekvensfortolkning, 9, 11, 102, 211
 sandsynlighedstæthed, 20, 147, 179
 simulation, 169
 simultan fordeling, 57
 simultan fordelingsfunktion, 58
 skadesforsikring, 37, 271
 skalaparameter, 168
 spredning, 93, 137, 160
 standardafvigelse, 93, 137, 160
 stationær fordeling, 253
 stationær proces, 254
 stikprøveudtagning
 med tilbagelægning, 72
 uden tilbagelægning, 72–75
 stokastisk proces, 241, 273
 stokastisk uafhængighed

- af diskrete stokastiske variable, 82, 127
- af kontinuerte stokastiske variable, 183, 186
- af n hændelser, 34, 37
- af stokastiske variable, 58
- af to hændelser, 31
- store tals lov, 9, 210
- sum af
 - binomialfordelinger, 71
 - eksponentialfordelinger, 188, 204
 - Γ -fordelinger, 224
 - geometriske fordelinger, 142
 - normalfordelinger, 195
 - Poissonfordelinger, 129
 - to diskrete stokastiske variable, 77, 126
 - to kontinuerte stokastiske variable, 187
 - to uafhængige kontinuerte stokastiske variable, 188
 - to uafhængige diskrete stokastiske variable, 84–85, 128
- terningkast, 13, 15, 23, 24, 31, 47, 49–51, 56–58, 79, 81, 89
- t-fordelingen, 227, 231
- tilstandsrum, 241, 242
- to-dimensionale normalfordeling, 236–238
- tosidede eksponentialfordeling, 174
- transformation
 - af generelle stokastiske variable, 49
- transformationssætningen
 - for diskrete fordelinger, 125
 - for en-dimensionale kontinuerte fordelinger, 166
 - for fordelinger, 49
 - for fordelinger på endelige mængder, 76
 - for fordelingsfunktioner, 54–55
 - for n -dimensionale kontinuerte fordelinger, 194
 - for to-dimensionale kontinuerte fordelinger, 192
- tællelig mængde, 118
- tæthedsfunktion, 147, 179
- uafhængighed
 - af diskrete stokastiske variable, 82, 127
 - af kontinuerte stokastiske variable, 183, 186
 - af n hændelser, 34, 37
 - af stokastiske variable, 58
 - af to hændelser, 31
 - betinget, 35
- udartede fordeling, 95
- udartet stokastisk variabel, 95
- udfaldsrum, 12, 18
- udtrækning af kort, 31, 33, 70, 73
- u-fordelingen, 162
- ukorrelerede stokastiske variable, 100
- uniforme fordeling, 148, 151, 154, 157, 161, 180
- varians
 - af binomialfordelingen, 101
 - af den geometriske fordeling, 137–138
 - af den hypergeometriske fordeling, 98

- af eksponentialfordelingen, 160
- af en diskret stokastisk variabel, 93, 136
- af en kontinuert stokastisk variabel, 160
- af en sum af stokastiske variable, 98
- af en sum af ukorrelerede stokastiske variable, 101
- af Γ -fordelingen, 239
- af ligefordelingen, 161
- af normalfordelingen, 163
- af Poissonfordeligen, 137
- af t-fordelingen, 227
- volatility pumping, 271
- volumen, 301

- Wienerprocessen, 276