Innhold

Forord v
Figurer xi
Tabeller xii
Eksempel xiv
Definisjoner xv
Symboler xvi

I Innledende manøvre **1**

1 Statistikk og sannsynlighet? **3**
1.1 Statistikk og data over alt og alltid 3
1.2 Oppgaver 12

II Sannsynlighetsteori **15**

2 Eksperiment og Kolmogorovs aksiomer **17**
2.1 Utfallsrom, hendelser og eksperiment 17
2.2 Sannsynlighet og eksperiment 22
2.3 Oppgaver 24

3 Sannsynlighetstetthet **27**
3.1 Diskrete fordelinger 28
3.2 Kontinuerlige fordelinger for $\mathbb{R}$ 31
3.3 Oppgaver 33

4 Betinget sannsynlighet og uavhengighet **35**
4.1 Betinget sannsynlighet 35
4.2 Uavhengighet 39
4.3 Oppgaver 40

5 Kombinatorisk sannsynlighet **43**
5.1 Multiplikasjonsregelen og permutasjoner 43
5.2 Binomialkoeffisienter og kombinatoriske sannsynligheter 46

5.3 Oppgaver . . . 49

6 Observatorer og tilfeldige variabler **51**
6.1 Observatorer og fordelingen til en observator . . . 51
6.2 Tilfeldige variabler . . . 53
6.3 Tilfeldige vektorer og uavhengighet . . . 56
6.4 Oppgaver . . . 59

7 Forventningsverdi **61**
7.1 Definisjon av forventningsverdi og substitusjonsformelen . . . 61
7.2 Forventningsverdier av summer og produkt . . . 64
7.3 Geometrisk tolkning av forventningsverdi og varians . . . 65
7.4 Beregning av forventningsverdi og varians . . . 67
7.5 Chebyshevs ulikhet . . . 69
7.6 Oppgaver . . . 70

8 Beregning av tettheter og betingede tettheter **73**
8.1 Funksjoner av observatorer . . . 73
8.2 Beregning av tettheter . . . 74
8.3 Betingede tettheter . . . 78
8.4 Oppgaver . . . 80

9 De små talls lov: Poissonfordelingen **83**
9.1 Poissonfordelingen og Poisson-punktprosesser . . . 83
9.2 Fordelinger knyttet til en Bernoulli-forsøksrekke . . . 86
9.3 Fra binomisk fordeling til poissonfordelingen . . . 88
9.4 Fra poissonfordelingen til gammafordelingen . . . 90
9.5 Oppgaver . . . 92

10 Normalfordelingen: Loven for målefeil **95**
10.1 Egenskaper til normalfordelingen . . . 96
10.2 Bevis av det viktigste grenseteoremet i statistikken . . . 97
10.3 Khi-kvadrat-, Fisher- og Student-t-fordelingene . . . 98
10.4 Oppgaver . . . 99

III Statistikk **101**

11 Estimering **103**
11.1 Beslutningsteori . . . 105
11.2 Statistiske modeller kontra sannsynlighets modeller . . . 111
11.3 Punktestimering . . . 112
11.4 Nøyaktigheten til en punktestimator . . . 117
11.5 Konsistente estimatorer . . . 119

11.6 Tilstrekkelige observatorer . . . 119
11.7 Metoder for å finne estimatorer . . . 122
11.8 Momentestimatoren . . . 122
11.9 Rimelighetsestimatoren . . . 124
11.10 Bayes-estimatoren . . . 127
11.11 Intervallestimering og konfidensintervall . . . 128
11.12 Metoder for å finne konfidensintervall . . . 130
11.13 Konfidensintervall: Flere eksempler . . . 135
11.14 Oppgaver . . . 140

12 Hypotesetesting **149**
12.1 Terningen på tiltalebenken . . . 149
12.2 Hypotese, forkastningsregel og signifikansnivå . . . 152
12.3 Utledning av en test med et gitt nivå . . . 153
12.4 Styrkefunksjonen og rimelighetsmetoden . . . 154
12.5 Rimelighetstesten . . . 158
12.6 Oppgaver . . . 166

IV Tillegg **171**

A Matematikk **173**
A.1 Litt analyse . . . 173
A.2 Funksjoner og mengder . . . 174

B Statistikk på datamaskin **177**
B.1 Sannsynligheter fra fordelinger . . . 177
B.2 Histogram og fordeling . . . 177

C Faginformasjon 1997 **179**

D Eksamen desember 1997 **181**

E Eksamen juni 1997 **183**

Bibliografi **186**

Register **188**

Figurer

1.1 Histogram og fordeling for alder for mor og far. 5
1.2 Resultatet av tre pilkast. 10

3.1 Atomene til en diskret fordeling og en hendelse. 29
3.2 Den geometriske rekken $1/2 + 1/4 + 1/8 + \cdots = 1$. 30
3.3 Addisjonregelen for en kontinuerlig tetthet tilsvarer addisjon av areal under grafen. 31
3.4 Tettheten til en uniform fordeling. 32
3.5 Tettheten til en normalfordeling. 32
3.6 Tettheten til en eksponensialfordeling. 33

4.1 Terningkastet resulterte i at Ω' inntraff. 36
4.2 Betingede sannsynligheter ved trekking av 4 brikker fra en urne. 37
4.3 En partisjon av Ω. 38

5.1 Tre-diagram for et spesielt Mynt + Terning eksperiment. 44
5.2 Tre-diagram for multiplikasjonsregelen. 45
5.3 Matrise for multiplikasjonsregelen. 45
5.4 Pascals trekant. 47
5.5 Binomialkoeffisientene i Pascals trekant. 47

6.1 Tre typer fordelingsfunksjoner. 55
6.2 Produktet av to intervall er et rektangel. 57
6.3 En y-enkel mengde. 57

7.1 Tyngdepunkt til en fordeling. 66
7.2 Variansen = Treghetsmoment om tyngdepunktet til en fordeling. 66

8.1 Funksjon av en observator. 74
8.2 Integrasjonsområdet ved konvolusjon. 76
8.3 Et trinomisk utfall. 77
8.4 Geometrisk tolkning av betinget tetthet. 78

9.1 Skjønnhetsfeil tilfeldig fordelt på tallerkener. Hver feil er gitt ved et kryss. 85
9.2 De små talls lov. Når $n \to \infty$ er antallet punkter poissonfordelt. Samlingen $T_1, T_2, \ldots$ av punkter er da en Poisson-punktprosess T. 89

11.1 Azorene er en øygruppe langt ute i Atlanterhavet. 106
11.2 Tettheten til observatoren W ved $n = 4$ og $\theta = 1$ 113
11.3 Fordelingen til estimatoren W har bedre posisjon enn fordelingen til V i forhold til $\tau(\theta)$. 117
11.4 Fordelingen til estimatoren W har mindre spredning enn fordelingen til V. 118
11.5 Rimelighetsestimat 124
11.6 Fortegnsskjema 125
11.7 Mengdeestimat for $\mu = (\mu_1, \mu_2)$ 129
11.8 $P(a \leq X \leq b) = 1 - \alpha$ når $P(a > X) = P(X > b) = \alpha/2$. 132
11.9 Kvantilen $u_{\frac{\alpha}{2}}$ til standard normalfordelingen. 133
11.10 Kvantilen $t_{\frac{\alpha}{2},n-1}$ til t$(n-1)$-fordelingen. 134
11.11 Kvantiler i khi-kvadrat fordelingen. 138
11.12 Intervallestimat for p. 139

12.1 Kritisk område $w \geq w^*$ for en test med signifikanskoeffisient α. 154
12.2 Klassifisering av feil i en hypotesetest. 156
12.3 Styrkefunksjonene viser at test A er uniformt bedre enn test B. 158
12.4 Nullhypotesen er ikke enkel. 161
12.5 $P(|T| \geq t_{\alpha/2,n-1}) = \alpha/2 + \alpha/2 = \alpha$ 163
12.6 Nullhypotesen er enkel. 163
12.7 Nullhypotesen er gitt ved en linje i parameterrommet. 165
12.8 Kritisk område i en χ^2-test. 166

A.1 Venn-diagram. 176

Tabeller

1.1 Antall fødsler i 2018 gitt alder til mor og far. 4
1.2 Måling av en lengde. 8
1.3 Antall prøyssiske soldater sparket i hjel av hester. 8
1.4 Data fra pilkasteksperiment. 9
1.5 Geometri i pilkasteksperiment. 9

2.1 Levetid til V805 radar senderør. 20

4.1 Værmelding i forhold til tilsvarende observert vær. 40

9.1 Fordelinger knyttet til poissonfordelingen. 92

11.1 Områdesannsynligheter i en leteaksjon. 106

Eksempel

1.1 Fødselsdata . . . 4
1.2 Lengdemåling . . . 7
1.3 Små talls lov . . . 8
1.4 Pilkast . . . 9
2.5 Et myntkast . . . 18
2.6 Tre myntkast i rekkefølge . . . 19
2.7 Gjentatt myntkast inntil Kron . . . 19
2.8 Levetid til klimaannlegg i fly . . . 19
2.9 Levetiden til V805 radar senderør . . . 20
2.10 Gjentatt myntkast . . . 21
2.11 Eksempel 3.2 fortsatt . . . 21
2.12 Terningkast . . . 22
2.13 To myntkast . . . 22
2.14 Fotballkamp . . . 23
2.15 Eksempel 3.2 fortsatt . . . 24
3.16 To myntkast . . . 28
3.17 Pilkast . . . 29
3.18 Myntkast inntil resultatet er krone . . . 30
3.19 Fri veilengde . . . 33
4.20 Terningkast med informasjon . . . 35
4.21 Kortstokk . . . 36
4.22 Levealder . . . 36
4.23 Trekking uten tilbakelegging . . . 37
4.24 Valg . . . 38
4.25 Varsellampe . . . 39
4.26 Uavhengige myntkast . . . 39
4.27 Uavhengige lyskryss . . . 40
5.28 Mynt + Terning . . . 43
5.29 Passord . . . 44
5.30 Fikset mynt . . . 48
6.31 Antall kron i myntkast med n mynter . . . 51
7.32 Middelverdi . . . 65
7.33 Forventningsverdi og varians til en Bernoulli variabel . . . 67
7.34 Forventningsverdi og varians beregnet ved variabelskifte . . . 67

7.35 Forventningsverdi beregnet ved normaliseringen . . . 67
7.36 Derivasjon av den momentgenererende funksjon . . . 68
8.37 Tettheten til $aX + b$ . . . 75
8.38 Tettheten til $X + Y$ og konvolusjon. . . . 75
8.39 Ordningsobservatoren . . . 76
8.40 Betinget tetthet . . . 79
9.41 Antall skjønnhetsfeil . . . 84
9.42 Telling av biler . . . 84
11.43 Atomubåten Scorpion . . . 105
11.44 Kjøpmannen Bør . . . 109
11.45 Estimat for θ i en uniform fordeling . . . 112
11.46 Estimering av forventning og varians til en Bernoulli fordeling . . . 116
11.47 Mer nøyaktig estimator . . . 118
11.48 Tilstrekkelig observator for p . . . 120
11.49 Tilstrekkelig observator for μ og for (μ, σ^2) . . . 120
11.50 Diskvalifisering av en estimator og datareduksjon . . . 121
11.51 Momentmetoden . . . 123
11.52 Rimelighetsestimator . . . 125
11.53 Rimelighetsestimator for μ, σ^2 . . . 126
11.54 Beta prior . . . 127
11.55 Mengdeestimator . . . 128
11.56 Konfidensområde fra Tsjebysjevs ulikhet . . . 131
11.57 Normalfordeling med kjent varians . . . 133
11.58 Intervallestimator for μ når σ^2 er ukjent . . . 134
11.59 Intervall-invertering . . . 136
11.60 Pivot variabel . . . 137
11.61 Tilnærmede konfidensintervall . . . 139
12.62 Hypotese i terningkast . . . 150
12.63 Rimelighetstest for uniform fordeling . . . 159
12.64 Test for forventning med ukjent varians . . . 160
12.65 Test for forventning med kjent varians . . . 162
12.66 Test for varians . . . 164

Definisjoner

1.1 *Statistikk* . . . 3
1.2 *Parameter* . . . 7
2.3 *Utfallsrom* . . . 18
2.4 *Eksperiment* . . . 21
2.5 *Sannsynlighet* . . . 22
2.6 *Sannsynlighetsmodell* . . . 23
3.7 *Sannsynlighetstetthet* . . . 27
4.8 *Betinget sannsynlighet* . . . 36
4.9 *Uavhengighet* . . . 39
6.10 *Observator* . . . 52
6.11 *Tilfeldig variabel* . . . 53
7.12 *Forventningsverdien* . . . 62
7.13 *Variansen* . . . 64
11.14 *Risiko* . . . 107
11.15 *En statistisk modell* . . . 111

Symboler

μ	forventningsverdi, middelverdi, tyngdepunkt	4
σ	standardavvik	4
X	stokastisk variabel	6
$(X < x)$	hendelsen at $X < x$	6
$\hat{p}$	empirisk sannsynlighet	6
s	antall suksess	6
n	antall forsøk	6
$\overline{x}$	middel	7
s	standardavvik	7
$\mathrm{P}(A)$	sannsynlighet for A	22
nA	hyppighet til A	23
$\mathrm{P}(A \mid B)$	betinget sannsynlighet	36
E_T	Forventningsverdi	63
$\mathcal{D}(f)$	definisjonsmengde til f	174
$\mathcal{R}(f)$	rekkevidden til f	174
f^{-1}	invers til f	174
$\emptyset$	den tomme mengden	174
$\in$	er element i	175
$\notin$	er ikke element i	175
$\subset$	er en delmengde av	175
$\Rightarrow$	medfører at	175
$\exists$	det eksisterer	175
$\forall$	for alle	175
A^c	komplement til A	175
$\setminus$	differans	175
$\uplus$	disjunkt union	175

Del I

Innledende manøvre

Kapittel 1

Statistikk og sannsynlighet?

Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of Man.

F.Galton (1908)

1.1 Statistikk og data over alt og alltid

Hva er statistikk? Ordet *statistikk* brukes med mange ulike betydninger. Det brukes for faget statistikk. Det brukes også om et av de mest sentrale begrep i statistikkfaget.

Definisjon 1.1 (*Statistikk*) En statistikk er en funksjon av data.

Observasjoner gir oss data. Mengden data som er tilgjengelig har eksplodert med utviklingen av datamaskinen. Ordet statistikk brukes også i betydningen *statistikkfaget*. Statistikk er en vitenskap. Følgende spørsmål - og deres samspill - står sentralt.

- Hvordan lære fra data?
- Hva er gode data?
- Hvilke data skal observeres?

Som menneske vet du at det er usikkerhet knyttet til alle observasjoner. Dermed er det usikkerhet knyttet til alle data. Utgangspunktet i teoretisk statistikk er å modellere usikkerheten ved å assosiere sannsynligheter med data. Fordi en statistikk er en funksjon av data, så er statistikken selv også data med sine egne assosierte sannsynligheter. I teoretisk statistikk vil egenskapene til denne sannsynlighetsfordelingen avgjøre om det er en god eller dårlig statistikk. Dette er bokens tema, og vil bli forklart i mye større detalj etterhvert. I første omgang vil vi se på noen eksempel.

Tabell 1.1: Antall fødsler i 2018 gitt alder til mor og far.

Alder (år)	≤ 17	18-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	≥ 60	Uoppgitt
# Mor	56	366	5456	17748	20106	9336	1848	154 (45+)	0	0	0	0
# Far	21	96	2536	12189	19025	12701	5163	1809	565	185	82	698

Eksempel 1.1 (Fødselsdata) Medisinsk fødselsregister inneholder opplysninger om fødsler i Norge. Staten Norge, og andre stater, registrerer og har registrert slike og andre data over lang tid. Dette gir den historiske betydningen av statistikkfaget tilbake til Aristoteles (384-322 f.Kr.). Det er studiet av staten. Dette gir et grunnlag for politiske beslutninger.

Tabell 1.1 gir antall fødsler i 2018 i ulike alderskategorier for kvinner og menn. Tallene i tabellen er en statistikk. Denne statistikken er en funksjon av de underliggende data for fødsler som er registrert av Medisinsk fødselsregister. Dette er da et eksempel på at en statistikk er en funksjon av data. Statistikken selv er også data.

Figur 1.1 viser *normaliserte histogram* tilsvarende Tabell 1.1. De er laget ved opptellingen i intervallene, men høyden er normalisert slik at totalarelaet under kurven er lik 1. I Tillegg B.2 er det vist hvordan grafene er laget ved hjelp av en *datamaskin*. En datamaskin er uunværlig ved beregning av statistikk fra data. Boken her handler i liten grad om hvordan dette gjøres i praksis på en datamaskin. Tillegg B gir allikevel noe informasjon om dette ved eksempel på beregning av ulike statistikker på en datamaskin. Bruk av datamaskin er - ikke uventet - et av de sentrale tema i moderne statistikk.

Figur 1.1 inneholder også grafen til to kontinuerlige funksjoner som tilnærmer histogrammene. Funksjonene er også normalisert slik at arealet under grafen er lik 1. Begge funksjoner er på formen

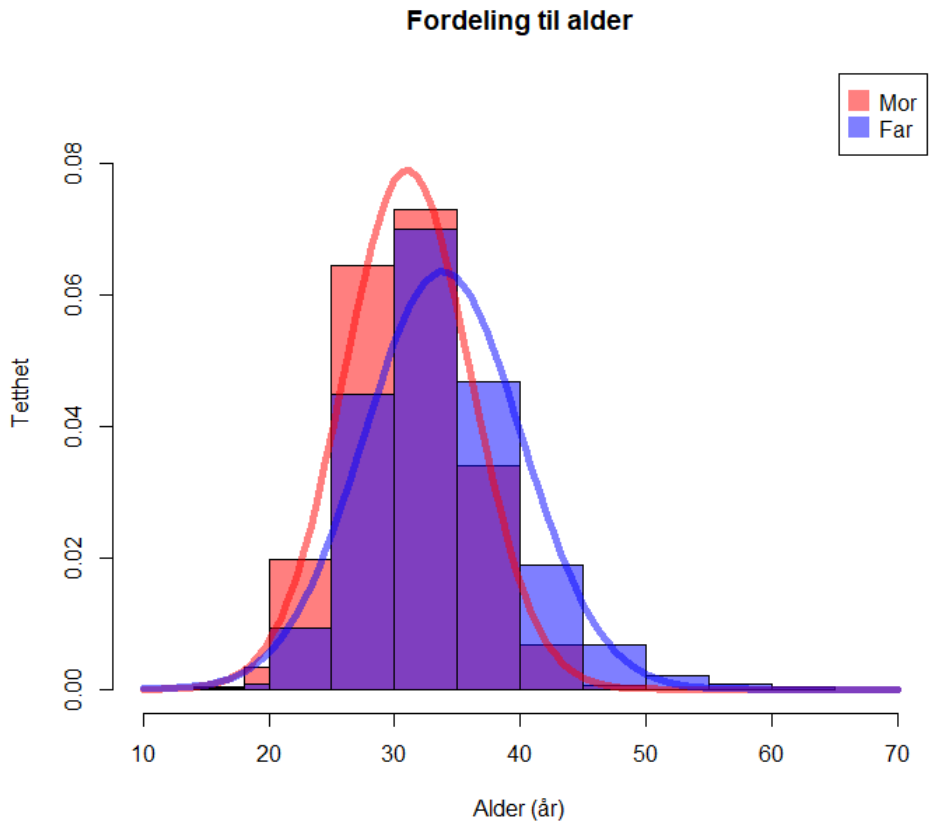
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1.1}$$

Tallene μ og σ gir posisjon og spredning til grafen. Funksjonen i ligning (1.1) definerer *normalfordelingen* med *forventningsverdi* μ og *standardavvik* σ.

Histogrammene og grafene er funksjoner av de innsamlede data. Dette viser at en statistikk ikke behøver å være en tabell med tall. En *statistikk* kan med andre ord også være en *graf*, et *kakediagram*, eller andre geometriske objekter.

Hva kan vi lære av disse statistikkene? Er de gode statistikker? Gode på hvilken måte? Svaret på disse spørsmålene avhenger av hva som ønskes undersøkt. *Grafene er statistikker* som er gode her ved at det kan være lettere å se sammenhenger geometrisk. De inneholder imidlertid mindre informasjon enn det som er gitt i Tabell 1.1. Grafene kan ikke brukes til å gjennvinne tallene i tabellen. På samme måte inneholder tabellen mindre informasjon enn det som er gitt i de underliggende data registrert i Medisinsk fødselsregister. En statistikk inneholder aldri mer informasjon enn det som finnes i de opprinnelige data.

En statistikk kan redusere informasjon. Det er nettopp denne reduksjonen som kan gjøre en statistikk god. Grafene gir oss raskt et bilde av fordelingen til alder til mor og far. Dette bildet er ikke så lett å se direkte ved å stirre på alle tallene i Medisinsk fødselsregister. Der er bildet skjult i mye annen informasjon.



Figur 1.1: Histogram og fordeling for alder for mor og far.

Hvor er så usikkerheten blitt av? Tallene gir at det ble registrert $n = 55070$ fødsler i Norge i 2018. Av disse var det $s = 56 + 366 + 5456 = 5878$ som hadde en mor yngre enn 25 år. Disse registrerte tallene er det ingen usikkerhet i. Dersom en i stedet ser på antall faktiske fødsler, så er det et usikkert antall.

Helene ble født i 2001. Tallene gir

$$\hat{p} = \frac{s}{n} = \frac{5878}{55070} \approx 11\% \tag{1.2}$$

som en approksimasjon til sannsynligheten p for at Helene blir mor før hun er 25 år. Hva betyr dette? Her er det mange og betydelige usikkerheter. En begrunnelse for påstandene kan gis ved å lage en statistisk modell for alder X til Helene ved hennes første fødsel. Parameteren vi vil si noe om er sannsynligheten

$$p = \mathrm{P}(X < 25 \text{ år}) \tag{1.3}$$

for at alderen $X < 25$ år. Symbolet X brukes her for en reell størrelse som det er assosiert usikkerhet med. Slike størrelser kalles *stokastiske variabler*.

Anta at fødselstallene i 2018 er representative for Helene. Hver av de $n = 55070$ observasjonene sees som en tilfeldig trekning av et tall x. Resultatet er $x_1, x_2, \ldots, x_n$ tall som er registrert i Medisinsk fødselsregister. Hendelsen $x_i < 25$ år inntreffer i $\hat{p} = s/n = 5878/55070 \approx 11\%$ av tilfellene. Den observerte *relative hyppigheten* til hendelsen $(X < x)$ er m.a.o. 11% når $x = 25$ år. Den relative hyppigheten definerer den *empiriske sannsynligheten* $\hat{p}$. Ligning (1.2) gir at $\hat{p}$ er gitt ved antall suksess s delt på antall forsøk n.

Sannsynligheten p er, dersom modellen er rett, lik grensen for den tilsvarende relative hyppigheten når antall observasjoner går mot uendelig. Dette er på samme måte som at relativt antall Mynt i en serie myntkast kan gå mot 50% når antall myntkast går mot uendelig. I vårt tilfelle er antall observasjoner $n = 55070$, og vi forventer dermed at $\hat{p}$ er en god approksimasjon til p.

Husk at denne konklusjonen er betinget gitt at fødselstallene i 2018 er representative for Helene. Annen informasjon vil kunne endre konklusjonen. Alle sannsynligheter er betingede sannsynligheter. I tillegg har vi kun funnet en approksimasjon $\hat{p}$ til p. Hvor god er denne approksimasjonen? Metodene i teoretisk statistikk er oppfunnet nettopp for å kunne gi et mer presist svar på dette og lignende spørsmål.

Histogrammet for Mor i Figur 1.1 er laget slik at arealet under kurven opp til alder 25 år er 11%. Arealet under den kontinuerlige kurven opp til alder 25 år er da også tilnærmet lik 11% fordi kurven approksimerer histogrammet. En *statistisk modell* for alderen X er gitt ved å anta at sannsynlighetene assosiert med X er gitt av normalfordelingen i ligning (1.1). Sannsynligheten p for at alderen $X < x = 25$ år er da definert ved

$$p = \mathrm{P}(X < x) = \int_{-\infty}^{x} f(x)\, dx \tag{1.4}$$

Sannsynligheten p er her en *parameter* fordi den er en funksjon av modellparametrene μ og σ som bestemmer f. Hvordan er så den kontinuerlige kurven for Mor i Figur 1.1 laget?

Problemet er at modellparametrene μ og σ er ukjente. Dette er på samme måte som at sannsynligheten p også er ukjent.

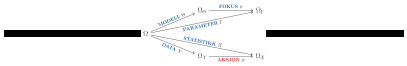
Dersom utvalget gitt av tallene $x_1, x_2, \ldots, x_n$ registrert i Medisinsk fødselsregister var kjente for oss, så kunne vi bruke utvalgets *middel* $\overline{x}$ og *standardavvik* s som estimat for μ og σ. Utvalgets middel $\overline{x}$ og *varians* s^2 er definert ved ligningene

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + \cdots + x_n}{n} \tag{1.5}$$

$$s^2 = \frac{\sum_i (x_i - \overline{x})^2}{n-1} = \frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1} \tag{1.6}$$

Tillegg B.2 gir en mulig alternativ metode basert på data som i Tabell 1.1. Resultatet er estimatene $\tilde{\mu} = 31.01$ og $\tilde{\sigma} = 5.05$. Disse kan brukes til å beregne et estimat $\tilde{p}$ ved integralet i ligning (1.4). Dette integralet kan ikke løses analytisk, men må beregnes ved numerisk integrasjon. Dette er gjort ved metoden i Tillegg B.1. Resultatet er estimatet $\tilde{p} = 12\%$.

Diskusjonen av dette eksempelet kan gjøres mye lenger. Det er mange åpne spørsmål. Er estimatet $\tilde{p}$ basert på en normalfordeling bedre enn estimatet $\hat{p}$ beregnet mer direkte? Hva menes med bedre? Disse, og flere spørsmål, er tema for videre diskusjon i boken. Noen av disse kan du selv arbeide med i oppgavene på side 12-.



Definisjon 1.1 gir en av betydningene til begrepet *statistikk*. En statistikk er en funksjon av data. Funksjonen og verdien til funksjonen er to forskjellige ting. Begrepet *statistikk* brukes om begge tolkningene. Denne konvensjonen skaper ofte forvirring for studenter og brukere av statistikk, men den er hensiktsmessig. Den samme konvensjonen brukes om følgende like sentrale begrep.

Definisjon 1.2 (*Parameter*) En parameter er en funksjon av modellen.

Det er gitt eksempel både eksplisitt og implisitt på både parametre og statistikker i Eksempel 1.1. Det overlates til deg å lage flere eksempel ut fra dette. Mange flere eksempel vil bli gitt senere.

Eksempel 1.2 (Lengdemåling) Tabell 1.2 gir resultatet av å måle en lengde μ med et mikrometer, med et skyvelær (Taraldsen, 2006) og med øyemål. Tallene i tabellen er data og derfor også en statistikk. Inspeksjon av tallene viser at resultatene fra skyvelæret tilfeldigvis også er lik avrundede verdier fra målingene med mikrometer.

En mulig modell for målingene med skyvelæret er at de er lik avrundede verdier av en underliggende verdi x. Resultatet $y = 7.5$mm i måling 1 med skyvelæret tilsvarer da en underliggende verdi x slik at $7.450\text{mm} \leq x < 7.55\text{mm}$. Litt ettertanke gir at avrundingen tilsvarende skyvelæret er avrundet på samme måte som data i Tabell 1.1. Dette gjelder også resultatene fra mikrometeret, men der er det avrunding innenfor et mindre intervall.

Tabell 1.2: Måling av en lengde.

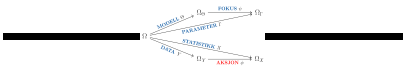
	Lengde (mm)		
Måling	Mikrometer	Skyvelær	Øyemål
1	7.489	7.5	10
2	7.503	7.5	10
3	7.433	7.4	10
4	7.549	7.5	10
5	7.526	7.5	10
6	7.396	7.4	10
7	7.543	7.5	10
8	7.509	7.5	10
9	7.504	7.5	10
10	7.383	7.4	10

Tabell 1.3: Antall prøyssiske soldater sparket i hjel av hester.

Soldater	År	Forventet antall år
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
5-	0	0.1

Målingene med øyemål tilsvarer avrunding innenfor et større intervall. Disse siste målingene var gjort slik at personen kun kunne gi svar i hele 1cm'ere.

Vi skal analysere tallene i Tabell 1.2 senere. Det er mange naturlige spørsmål. Hvordan estimeres lengden μ? Hva er nøyaktigheten? Usikkerheten? Hva er effekten av avrundingen? Alle data lagret på en datamaskin er med endelig nøyaktighet, og i mange tilfeller er data avrundet til kun noen få siffer. Når kan effekten av avrundingen neglisjeres? Hva hvis ikke? Kan effekten av avrunding fjernes ved et alternativt eksperiment?



Eksempel 1.3 (Små talls lov) En hest kan skade et menneske, men det skjer sjelden. von Bortkiewicz (1898, p.25) studerte antall prøyssiske soldater sparket i hjel av hester. Tabell 1.3 er resultatet av å telle antall dødsfall per år for 10 kavalerikorps i 20 år. Rad 1 viser at i 109 av disse 200 kavaleriårene så var det ingen dødsfall, og forventet antall år uten dødsfall er 108.7 år. Hva betyr dette siste? Hva er usikkerheten her?

von Bortkiewicz (1898) brukte tallene i Tabell 1.3 sammen med flere andre eksempel når han formulerte det som nå kalles *Små talls lov*. Han studerte fenomen med relativt liten hyppighet, men modellen han utviklet er ikke begrenset til dette. En mulig modell tilsvarende Tabell 1.3 er gitt ved å anta at for hver soldat, så er det hvert sekund en viss

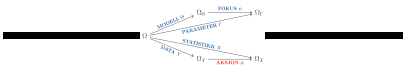
Tabell 1.4: Data fra pilkasteksperiment.

		Koordinater (mm, 1°)			
Kast	Poeng	x	y	r	θ
1	100	-11	-11	16	225°
2	0	-	-	-	-
3	75	32	-17	36	332°
4	75	-25	-61	66	248°
5	75	41	22	47	28°
6	75	16	-52	54	287°
7	100	9	17	19	62°
8	75	-26	-45	52	240°
9	75	26	29	39	48°

Tabell 1.5: Geometri i pilkasteksperiment.

		Radius (mm)	
Område	Poeng	Indre	Ytre
1	200	0	6
2	100	6	29
3	75	29	74
4	50	74	117
5	25	117	169
6	10	169	225
7	0	225	∞

sannsynlighet for å bli sparket i hjel. Et mer positivt eksempel er gitt ved fødselstallene i Tabell 1.1. En mulig modell der er gitt ved at det hvert sekund er en viss sannsynlighet for at en kvinne føder et barn. Her er det igjen mange åpne spørsmål, og vi skal se mer på dette senere.

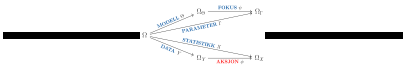


Eksempel 1.4 (Pilkast) Tabell 1.4 gir resultatet av 9 pilkast på en blink. De tre første resultatene er som i Figur 1.2. Kast nummer to feilet ved at pilen datt ned på gulvet, og dermed ble det ingen koordinater for kastet. Tabell 1.5 gir geometrien til blinkens ringer.

Anta at du blir tilbydt et veddemål. Du får 100 kr dersom pilkasteren ikke treffer i midten. Dersom pilkasteren treffer må du betale 200 kr til pilkasteren. Vil du være med på dette veddemålet? Dersom pilkasteren får 3 kast? Annet beløp? Se flere kast først? Hvor mange?

Her er det mange naturlige spørsmål. Slike problemstillinger for spill var en av kildene til utviklingen av en teori for sannsynligheter. Generell beslutningsteori er en av pilarene for teoretisk statistikk. Vi skal bruke dette konkrete problemet og andre spill til illustrasjon av sentrale generelle begrep.

Figur 1.2: Resultatet av tre pilkast.


MODELL γ
Ω_Θ
FOKUS η
Ω_Γ
PARAMETER ℓ
Ω
STATISTIKK χ
DATA χ
Ω_Γ
AKSJON σ
Ω_Σ

1.2 Oppgaver

De neste oppgavene refererer til data gitt i Eksempel 1.1. Oppgaver merket med (**D**) anbefales løst med statistisk programvare på datamaskin. Eksempel på dette er gitt i Tillegg B.

1.1. Helene har ikke fått barn i 2020. Bruk denne opplysningen og Tabell 1.1 til å finne et forbedrede estimat for sannsynligheten for at hun får barn før hun er 25 år. Diskuter resultatet.

1.2. Diskuter antagelsen om at fødselstallene i 2018 er representative for Helene. Er 2018 relevant? Hva med tvillinger? Får alle barn? Bosted? Andre faktorer? Hva hvis det hadde vært oppgitt fødselstall for 1967 i stedet? (Hvis du er nysgjerrig kan du finne disse i statistikkbanken til Medisinsk fødselsregister.)

1.3. Halvor er født i 2002. Bruk dette og Tabell 1.1 til å finne en approksimasjon $\hat{p}_f$ til sannsynligheten p_f for at Halvor blir far før han er 25 år.

1.4 (**D**). Tallene i Tabell 1.1 for Far kan brukes til å begrunne estimatene $\tilde{\mu}_f = 33.84$ og $\tilde{\sigma}_f = 6.28$. Bruk dette, og en antagelse om normalfordeling, til å beregne et estimat $\tilde{p}_f$ til sannsynligheten p_f for at Halvor blir far før han er 25 år.

1.5 (**D**). Kan du estimere sannsynligheten for at Halvor blir far etter at han er 50 år? Hva med før han er 25 år, men etter at han er 15 år? Hva med Helene? Kommenter svarene og antagelsene.

1.6. Kan du finne og beskrive en metode til å estimere forventningsverdi μ_f og standardavvik σ_f ut fra tallene i Tabell 1.1?

1.7. La N være antall registrerte fødsler i 2018 og la M være antall faktiske fødsler. Er det en sammenheng mellom M og N? Hvorfor er det en usikkerhet knyttet til M? Kan du tenke deg et eksperiment som sier noe om sannsynligheten for at en faktisk fødsel blir registrert? Har alder til mor eller far noen betydning for denne sannsynligheten? Kan du tenke deg et eksperiment for å måle dette?

1.8. Gi 3 ulike eksempel på en parameter knyttet til en modell for Helene.

1.9. Gi 3 ulike eksempel på statistikker knyttet til Tabell 1.1 ved å definere funksjonene.

De neste oppgavene refererer til data gitt i Eksempel 1.2. Oppgaver merket med (**D**) anbefales løst med statistisk programvare på datamaskin. Eksempel på dette er gitt i Tillegg B.

1.10 (**D**). Regn ut empirisk middel og empirisk standardavvik til mikrometermålingene i Tabell 1.2. Gi en intuitiv begrunnelse for at empirisk middel er et estimat for lengden μ. Hva med empirisk standardavvik? Hint: Anta at målingene er normalfordelt med forventning μ og standardavvik σ.

1.11 (**D**). Gjør som i Oppgave 1.10, men se på målingene med skyvelær og med øyemål. Diskuter resultatene.

1.12. Gi 3 ulike eksempel på en parameter knyttet til en modell for mikrometermålingen.

1.13. Gi 3 ulike eksempel på statistikker knyttet til mikrometermålingen ved å definere funksjonene.

1.14. Forklar hvordan dataene i Tabell 1.1 kan representeres som avrundede data av samme type som i Tabell 1.2.

De neste oppgavene refererer til data gitt i Eksempel 1.3. Oppgaver merket med (**D**) anbefales løst med statistisk programvare på datamaskin. Eksempel på dette er gitt i Tillegg B.

1.15. Bruk Tabell 1.3 til å estimere sannsynligheten for at ingen soldater dør av hestespark i et regiment i løpet av et år.

1.16. Estimer sannsynligheten for at det dør 3 soldater, eller færre, i et regiment i løpet av et år.

1.17. I løpet av et år i et regiment blir x soldater sparket i hjel av en hest. La $x_1, \ldots, x_N$ tilsvare observasjonene for de $N = 200$ regimentårene gitt i Tabell 1.3. De 200 observasjonene er ikke kjent, men det er allikevel mulig å beregne empirisk middel $\overline{x}$. Forklar dette og regn ut empirisk middel $\overline{x}$.

1.18. Gi 3 ulike eksempel på en parameter knyttet til en modell for dødsfallene.

1.19. Gi 3 ulike eksempel på statistikker knyttet til observasjonene av dødsfall ved å definere funksjonene.

De neste oppgavene omhandler elementær teori og notasjon for funksjoner og mengder.

1.20. Definer funksjonene exp og ln. Er det flere mulige definisjoner?

1.21. Kan du gi to rimelige definisjoner av $\ln : (0, \infty) \to \mathbb{R}$ hvor resultatet er to ulike funksjoner? Hint: Hva er domenet til ln ?

1.22. La A være en 2×2 matrise. Definer en funksjon som tilsvarer $A \mapsto A^{-1}$. Finnes det flere naturlige definisjoner? Hint: Generalisert invers for løsning av lineært ligningssystem.

1.23. La A og B være to mengder i planet som hver inneholder høyst 10 punkter. Definer en funksjon f som tilsvarer $(A, B) \mapsto A \cap B$. Hva er naturlig domene, kodomene og definisjonsmengde. Hva blir verdimengden?

1.24. La $\Omega = \mathbb{R}^2 = \{(x, y) \,|\, x, y \in \mathbb{R}\}$, dvs Ω er mengden av punkter i planet. Illustrer delmengdene $A = \{(x, y) \,|\, x = 5\}$, $B = \{(x, y) \,|\, x \geq 5\}$, $C = \{(x, y) \,|\, x \geq y^2\}$, $D = C \setminus B$, $E = \{(x, y) \,|\, x \geq y\}$, $F = C \cap E$, $G = F \cup A$, $H = \mathbb{N}^2 = \{(x, y) \,|\, x, y \in \mathbb{N}\}$. Er H tellbar?

1.25. La $\mathbb{Q}_+$ være mengden av alle positive rasjonale tall. Er $\mathbb{Q}_+$ tellbar? Er mengden av rasjonale tall tellbar? Svarene må begrunnes!

1.26. Kontroller om følgende regneregler gjelder: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $(A \cup B) \cap (C \cup D) = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D)$, $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$.

Illustrer konklusjonene med Venn-diagram.

Kan regnereglene generaliseres til union/snitt av vilkårlige familier av mengder?

1.27. La $A_n = [0, 1/n]$. Finn $B = \cup_{n=1}^{10} A_n$, $C = \cap_{n=1}^{10} A_n$, $D = \cap_{n=1}^{\infty} A_n$, $E = \cup_{n=1}^{\infty} A_n$.

1.28. La A, B være delmengder av Ω. Vis at $\Omega = A \uplus A^c$.

Vis at $\Omega = (A \cup B)^c \uplus (A \cap B) \uplus (A \setminus B) \uplus (B \setminus A)$. Illustrer dette i et Venn-diagram.

Kan du generalisere dette til tilfellet hvor vi tar utgangspunkt i tre delmengder A, B, C?

Hva med tilfellet med n delmengder, eller et tellbart antall delmengder?

Del II

Sannsynlighetsteori

Kapittel 2

Eksperiment og Kolmogorovs aksiomer

The situations we are going to model can all be thought of as random experiments. Viewed naively, an experiment is an action which consists in observing or preparing a set of circumstances and then observing the outcomes of this situation. We add to this notion the requirement that to be called an experiment such an action must be repeatable, at least conceptually. The adjective "random" is used only to indicate that we do not, in addition, require that every repetition yield the same outcome, although we do not exclude this case. What we expect and observe in practice when we repeat a random experiment many times is that the frequency of each of the possible outcomes will tend to stabilize. This "long term frequency" is to many statisticians, including the authors, the operational interpretation of the mathematical concept of probability. In this sense, almost any kind of activity involving uncertainty, from horse races to genetic experiments, falls under the vague heading, "random experiment".

P. J. Bickel og Kjell A. Doksum (1977)

2.1 Utfallsrom, hendelser og eksperiment

I dagligtale er hendelser eller begivenheter velkjente begrep. Hva er egentlig en hendelse? Dette er ikke så opplagt som du kanskje tror. Aksiomer for geometrien ble formulert for mer enn 2200 år siden av Euklid. Det tok så mer enn 2000 år før en tilsvarende presis teori for hendelser ble formulert. Kolmogorov (1933) formulerte aksiomer for sannsynlighetsteorien og brukte mengder som en matematisk modell for hendelser. Teori og notasjon for mengder antas kjent her. Tillegg A.2 kan konsulteres etter behov.

La Ω være en mengde og la $\mathcal{E}$ være en mengde delmengder av Ω. Dette betyr at

$$(A \in \mathcal{E}) \Rightarrow (A \subset \Omega) \tag{2.1}$$

En mengde av mengder kalles ofte en *familie av mengder*. Familien $\mathcal{E}$ er en *hendelsesfamilie* hvis

$$[E \in \mathcal{E}] \Rightarrow [E^c \in \mathcal{E}], \quad [A_i \in \mathcal{E}] \Rightarrow [\bigcup_i A_i \in \mathcal{E}] \tag{2.2}$$

Det siste skal gjelde for enhver tellbar familie $\{A_i \mid i \in I\}$ av hendelser. En *hendelse* er et medlem av hendelsesfamilien. Et *utfall* ω er et element i utfallsrommet. Dersom vi i et eksperiment får et utfall $\omega \in A$, så har hendelsen A *inntruffet*. Hendelsen Ω inntreffer alltid og hendelsen $\emptyset$ inntreffer aldri.

Definisjon 2.1 (*Utfallsrom*) En ikke-tom mengde utstyrt med hendelser.

Det er kanskje en bekymring for leseren at overstående er meget generelt. Vi ser på noen konkrete eksempler for å belyse begrepene og for å vise anvendelsen i forbindelse med modellering av eksperiment. I eksemplene fremgår det spesielt at en hendelse alltid er en mengde, men en mengde behøver ikke å være en hendelse.

Eksempel 2.1 (Et myntkast) Du kaster en mynt og får resultatet M (Mynt). Dette er et eksperiment med utfallsrom

$$\Omega = \{M, K\} \tag{2.3}$$

Resulatet av et enkelt myntkast er $\omega = M$ (Mynt) eller $\omega = K$ (Kron). Disse utfallene er inneholdt i Ω som dermed er et mulig utfallsrom for eksperimentet.

En naturlig hendelsesfamilien er

$$\mathcal{E} = \{\{M\}, \{K\}, \emptyset, \Omega\} \tag{2.4}$$

Den er naturlig fordi den er den minste hendelsesfamilien som inneholder den enkle hendelsen $A = \{M\}$. En mengde er *enkel* dersom den inneholder kun et element.

Resultatet av eksperimentet var her utfallet $\omega = M$. Hendelsen $A = \{M\}$ har da inntruffet fordi $\omega \in A$. Den motsatte hendelsen $A^c = \{K\}$ har ikke inntruffet.

Et annet ekvivalent eksperiment kan defineres ved å la utfallsrommet være

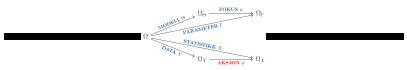
$$\Omega = \{0, 1\} \tag{2.5}$$

Her er $\omega = 0$ dersom resultatet er Mynt og $\omega = 1$ dersom resultatet er Kron. En naturlig hendelsesfamilie $\mathcal{E}$ er gitt ved familien av alle delmengder i Ω.

De to foregående utfallsrom er så små som mulig. Et tredje eksperiment kan defineres ved å la utfallsrommet være

$$\Omega = \mathbb{R} \tag{2.6}$$

Her er igjen $\omega = 0$ dersom resultatet er Mynt og $\omega = 1$ dersom resultatet er Kron. Et *utfallsrom for et eksperiment* er et utfallsrom som inneholder alle mulige utfall av eksperimentet. Utfallsrommet $\Omega = \mathbb{R}$ for dette eksperimentet skal vi bruke mye videre i boken. Hendelsesfamilien $\mathcal{E}$ er den minste [1] som inneholder alle intervall.



[1] Se Oppgave 2.5

Eksempel 2.2 (Tre myntkast i rekkefølge) Du kaster en mynt og får resultatet MMK (Mynt,Mynt,Kron). Dette er et eksperiment med minste utfallsrom

$$\Omega = \{MMM, MMK, MKM, MKK, KMM, KMK, KKM, KKK\} \tag{2.7}$$

Det består av 8 enkeltutfall. Vi lar hendelsesfamilien $\mathcal{E}$ være familien av alle delmengder. La $E = \{MKK, KMK, KKM, KKK\}$. Utfallet av eksperimentet ble $\omega = MMK$. Da inntraff ikke hendelsen E, dvs vi fikk $\omega \in E^c$. Å si at hendelsen E inntraff er det samme som å si at vi fikk flere Kron enn Mynt, og det fikk vi ikke.

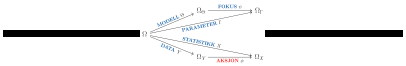
Et alternativt utfallsrom i dette eksemplet er

$$\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\} \tag{2.8}$$

ved å identifisere 0 med Mynt og 1 med Kron. Fordelen med dette valget er at alle de mulige utfallene kommer fram ved å telle binært. Et tredje alternativt utfallsrom er

$$\Omega = \{0, 1\}^3 = \{0, 1\} \times \{0, 1\} \times \{0, 1\} \tag{2.9}$$

Et utfall $\omega = (\omega_1, \omega_2, \omega_3)$ betyr at myntkast nummer i ga resusltat ω_i. I kast i er da $\omega_i = 0$ dersom resultatet er Mynt og $\omega_i = 1$ dersom resultatet er Kron.



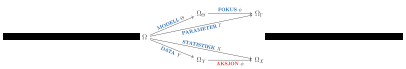
Eksempel 2.3 (Gjentatt myntkast inntil Kron) Du kaster en mynt og får resultatet MMK (Mynt,Mynt,Kron). Dette kan sees som utfallet i et eksperiment hvor du kaster en mynt inntil du får Kron. Utfallsrommet er

$$\Omega = \{K, MK, MMK, MMMK, \ldots\} \tag{2.10}$$

Et alternativt utfallsrom er

$$\Omega = \mathbb{N} \tag{2.11}$$

Resultatet i eksperimentet er da $\omega = 3$. Begge utfallsrom kan være utstyrt med hendelsesfamilien $\mathcal{E}$ lik familien av alle delmengder.



Eksempel 2.4 (Levetid til klimaannlegg i fly) Proschan (1963) studerte levetidsdata for komponenter i klimaanlegg til en flåte med jetfly. Levetidene målt i hele timer for fly nummer 7914 ble $50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14$ (Taraldsen, 2011, s.977). Utfallsrommet for første observasjon $\tilde{\omega} = \omega_1 = 50$ timer er

$$\tilde{\Omega} = \mathbb{R} \tag{2.12}$$

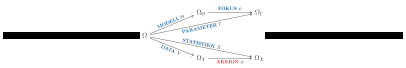
Tabell 2.1: Levetid til V805 radar senderør.

Tid (timer)	Antall	Forventet antall
0-40	166	180.5
40-80	151	145.0
80-120	132	115.4
120-160	98	92.7
160-200	73	73.8
200-240	45	59.3
240-280	53	47.2
280-320	40	37.9
320-360	23	30.2
360-400	26	24.3
400-440	24	19.3
440-480	9	16.0
480-520	9	11.6
520-560	8	9.9
560-600	9	8.0
600-700	17	13.5
700-800	8	8.2
800-	12	10.0

Hendelsesfamilien $\tilde{\mathcal{E}}$ er den minste som inneholder alle hendelser på formen $\tilde{H} = (t_1, \infty)$. Alle levetidene samlet er gitt ved enkeltutfallet $\omega = (\omega_1, \ldots, \omega_n)$ hvor $n = 24$. Utfallsrommet er da

$$\Omega = \mathbb{R}^n \tag{2.13}$$

Hendelsesfamilien $\mathcal{E}$ er den minste som inneholder alle hendelser på formen $H = (t_1, \infty) \times \cdot \times (t_n, \infty)$.



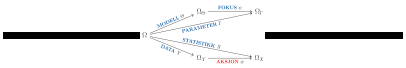
Eksempel 2.5 (Levetiden til V805 radar senderør) V805 senderrøret var en vanlig komponenent i mange flyradarsystemer. Davis (1952, Tabell C på s.133) studerte levetiden. Tabell 2.1 er resultatet av $n = 903$ levetider $\omega_1, \ldots, \omega_n$. Utfallsrommet for eksperimentet tilsvarende den første observasjonen $\tilde{\omega} = \omega_1$ er

$$\tilde{\Omega} = \mathbb{R} \tag{2.14}$$

Utfallsrommet for eksperimentet tilsvarende de $n = 903$ observasjonene er

$$\Omega = \mathbb{R}^n \tag{2.15}$$

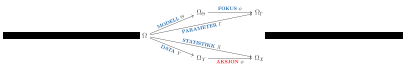
Her er det da observert et utfall $\omega = (\omega_1, \ldots, \omega_n)$. Hendelsesfamilien $\mathcal{E}$ er den minste som inneholder alle hendelser på formen $B = (t_1, \infty) \times \cdot \times (t_n, \infty)$.



Eksempel 2.6 (Gjentatt myntkast) Du kaster en mynt mange ganger og får resultatet $MMKMK\cdots$ (Mynt,Mynt,Kron,Mynt,Kron, ...). Denne uendelige serien kan identifiseres med utfallet $\omega = (0, 0, 1, 0, 1, \ldots)$. Et minste utfallsrom er

$$\Omega = \{0, 1\}^{\mathbb{N}} \tag{2.16}$$

Hendelsesfamilien $\mathcal{E}$ er den minste som inneholder alle hendelser på formen $B_n = \{\omega \,|\, \omega_1 \in I_1, \ldots, \omega_n \in I_n\}$ hvor n er vilkårlig og $I_j \in \{0, 1\}$.



De foregående eksemplene viser at utfallsrommet Ω kan være en endelig mengde, en tellbar mengde, eller en mengde som ikke er tellbar. Ordet eksperiment stammer fra det latinske *experiri* som betyr *prøve*. I det foregående og i det som følger vil ordet eksperiment bli brukt i en mer generell betydning.

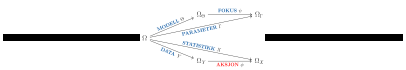
Definisjon 2.2 (*Eksperiment*) Et eksperiment er en prosedyre som kan tenkes gjentatt under identiske idealiserte forhold. De ulike mulige utfall av eksperimentet skal kunne identifiseres med ulike element i et utfallsrom.

Termen *eksperiment* definert over er ikke et matematisk begrep. Et myntkast er et eksperiment, men det er ikke et matematisk begrep. Et eksperiment er et fenomen som ligger utenfor matematikkens verden. Identifiseringen av utfallsrommet for et eksperiment er første skritt i den matematiske beskrivelsen av eksperimentet. Dette er det som er illustrert i alle de foregående eksperimentene. Et gitt eksperiment kan modelleres med mange mulige utfallsrom. Utfallsrommet er ikke bare er en mengde som *inneholder* alle mulige eksperimentelle utfall, men det er en mengde utstyrt med *hendelser*.

Eksempel 2.7 (Eksempel 3.2 fortsatt) Et enkelt pilkast er et eksperiment: Det kan tenkes gjentatt vilkårlig mange ganger med gitte idealiserte betingelser og det har et veldefinert utfall. Tabell 1.4 gir resultatet av å gjenta pilkasteksperimentet 9 ganger.

For å være kvalifisert til å kunne kalles et eksperiment, så må et tilhørende utfallsrom defineres. Et eksperiment er gitt ved å la utfallsrommet være

Et pilkast er et eksperiment: Det kan tenkes gjentatt vilkårlig mange ganger med gitte idealiserte betingelser og det har et veldefinert utfall. En matematisk modell er gitt ved et utfallsrom som *inneholder* alle mulige utfall. Et utfallsrom er en mengde utstyrt med hendelser E. I dette eksperimentet er det knyttet usikkerhet til utfallet. En matematisk modell er da gitt ved å tilordne en sannsynlighet $\mathrm{P}(E)$ til hver hendelse E.



2.2 Sannsynlighet og eksperiment

La Ω være et utfallsrom med en hendelsesfamilie $\mathcal{E}$.

Definisjon 2.3 (*Sannsynlighet*) En funksjon $\mathrm{P} : \mathcal{E} \to \mathbb{R}$ som oppfyller

(1) $\mathrm{P}(A_1 \cup A_2 \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots$ når $A_1, A_2, \ldots$ er disjunkte.

(2) $\mathrm{P}(A) \geq 0$.

(3) $\mathrm{P}(\Omega) = 1$.

En sannsynlighet P kalles også en *sannsynlighetsfordeling*, eller bare en *fordeling*, og også et *sannsynlighetsmål*. Tallet $\mathrm{P}(A)$ er sannsynligheten til hendelsen A. Et *sannsynlighetsrom* er et utfallsrom Ω utstyrt med en hendelsesfamilie $\mathcal{E}$ og en sannsynlighetsfordeling P.

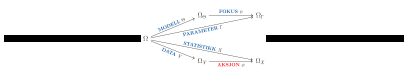
En sannsynlighetsfordeling kan også defineres som en funksjon $\mathrm{P} : \mathcal{E} \to [0, \infty]$ som oppfyller kravene (1) og (3). Dersom normaliseringskravet $P(\Omega) = 1$ fjernes, så sies P å være et mål. En fordeling er dermed et normalisert mål. Det første kravet kalles *tellbar additivitet*. Dersom denne addisjonsregelen bare kreves for endelige familier, så sies P å være en endelig additiv fordeling. I det følgende vil vi bevise noen av de viktigste regnereglene for fordelinger. Det vil fremgå av bevisene at mange av regnereglene også er gyldige for endelig additive mål fordi det kun er den endelige addisjonsregelen som benyttes.

En sannsynlighetsfordeling benyttes for å beskrive graden av usikkerhet i eksperiment. Vi ser på noen eksempler for å belyse dette.

Eksempel 2.8 (Terningkast) Utfallsrommet for et terningkast er

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Ved symmetri er det rimelig å anta at hvert enkelt utfall er like sannsynlig, dvs $P\{\omega\} = p$. Sannsynligheten til en generell hendelse E er da gitt ved addisjonsregelen, dvs $PE = \sum_{\omega \in E} P\{\omega\}$. Normaliseringen gir $1 = P\Omega = \sum_{\omega=1}^{6} P\{\omega\} = 6p$ gir $p = 1/6$, og det følger at P er en fordeling, Når fordelingen er gitt kan f eks sannsynligheten for at terningkastet gir 5 eller 6 øyne beregnes ved $P\{5, 6\} = \sum_{\omega \in \{5,6\}} p = 2p = 1/3$. Videre finnes $P(\text{odde antall øyne}) = P\{1, 3, 5\} = 1/2$.



Eksempel 2.9 (To myntkast) Utfallsrommet er

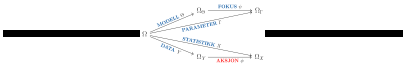
$$\Omega = \{MM, MK, KM, KK\}.$$

Det er rimelig å anta at enkeltutfallene er like sannsynlige. Dermed kan en f eks beregne $P(\text{To kronesider opp}) = P\{KK\} = 1/4$, og $P(\text{Minst en myntside opp}) = 1 - P(\text{Ingen myntsider opp}) = 1 - P\{KK\} = 3/4$.

Dersom myntene er identiske og eksperimentet er å kaste de samtidig, så er utfallsrommet

$$\Omega = \{0, 1, 2\},$$

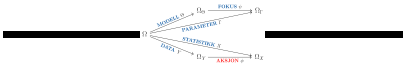
hvor enkeltutfallene teller antall myntsider opp. I dette tilfellet er en rimelig fordeling gitt ved $P\{0\} = P\{2\} = 1/4$ og $P\{1\} = 1/2$, som kan begrunnes ved likheten med eksperimentet hvor myntkastene ble utført i rekkefølge.



Eksempel 2.10 (Fotballkamp) La eksperimentet være gitt ved Norges neste hjemmelandskamp i fotball med utfallsrom

$$\Omega = \{0-0, 0-1, 1-0, 1-1, 0-3, \ldots\}.$$

En mulig vurdering er gitt ved $P(\text{Norge vinner}) = P\{1-0, 2-0, 2-1, 3-1, \ldots\} = 80\%$.



I alle eksemplene over er eksperimentene karakterisert ved et utfallsrom, og utfallet av eksperimentet er usikkert. Slike eksperiment kan kalles *sjanseeksperiment* for å understreke at utfallet er usikkert. Termen *statistisk eksperiment* er også i bruk. Dersom resultatet av et eksperiment er ω, så sier vi at hendelsen E har inntruffet dersom $\omega \in E$. En mulig modell for et sjanseeksperiment er gitt ved at utfallsrommet er et sannsynlighetsrom. Fordelingen P gir sannsynligheten $\mathrm{P}(E)$ for at hendelsen E inntreffer. Sannsynligheten tolkes i dette mest generelle tilfellet som en vurdering av rimeligheten for at hendelsen E inntreffer.

Definisjon 2.4 (*Sannsynlighetsmodell*) En sannsynlighetsmodell for et eksperiment er gitt ved at utfallsrommet er et sannsynlighetsrom.

Bortsett fra fotballkamp eksemplet, så kan alle overnevnte eksperiment gjentas et vilkårlig antall ganger. Alle kan tenkes gjentatt i et tankeeksperiment. Et sjanseeksperiment kan gjentas et vilkårlig antall ganger under idealiserte identiske betingelser. Termen *modell* brukes i stedet for sannsynlighetsmodell når betydningen er klar fra konteksten. Likeledes brukes termen *eksperiment* i stedet for sjanseeksperiment når betydningen er klar fra konteksten. I alle tilfellene er det ønskelig å formulere en matematisk modell for et eksperiment.

Hyppigheten nA til A er antall ganger hendelsen A inntreffer når et statistisk eksperiment er utført n ganger. Den *relative hyppigheten* til A er da definert ved

$$\mathrm{P}_n(A) = \frac{nA}{n} \tag{2.17}$$

Det følger at P_n er en fordeling. Dette er den *empiriske fordelingen*. Et eksperiment er et *statistisk lovmessig eksperiment* dersom grenseverdien

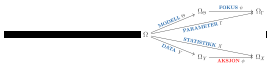
$$\mathrm{P}(E) = \lim_{n\to\infty} \mathrm{P}_n(E) \tag{2.18}$$

eksisterer for alle hendelser E. Det følger at dette definerer en fordeling P. Denne unike fordelingen er da den *sanne fordelingen* til eksperimentet. Vi sier i så fall at hendelsen E inntreffer i $\mathrm{P}(E) \cdot 100\%$ av eksperimentene når eksperimentet *gjentas i det lange løp*.

I praksis vil et eksperiment kun gjentas et endelig antall ganger, men vi vil analysere eksperimentet ved å tenke oss at eksperimentet kan gjentas et vilkårlig antall ganger. Observasjonen av en statistisk lovmessighet vil være gitt ved at den relative frekvensen til en hendelse synes å nå en grenseverdi når eksperimentet gjentas mange ganger. Mer generelt tolkes $\mathrm{P}(E)$ som en subjektiv vurdering av rimeligheten til hendelsen E. Tolkningen av dette kan fortsatt tolkes ved at en gitt sannsynlighet er lik grensen av en relativ hyppighet i et tilsvarende tankeeksperiment.

Eksempel 2.11 (Eksempel 3.2 fortsatt) Tabell 1.4 gir resultatet av 9 pilkast på blink.

Et pilkast er et eksperiment: Det kan tenkes gjentatt vilkårlig mange ganger med gitte idealiserte betingelser og det har et veldefinert utfall. En matematisk modell er gitt ved et utfallsrom som *inneholder* alle mulige utfall. Et utfallsrom er en mengde utstyrt med hendelser E. I dette eksperimentet er det knyttet usikkerhet til utfallet. En matematisk modell er da gitt ved å tilordne en sannsynlighet $\mathrm{P}(E)$ til hver hendelse E.



2.3 Oppgaver

Oppgaver merket med (**T**) er teorioppgaver som kan oppfattes som vanskelige.

2.1. Finn en hendelse B som ikke har inntruffet i Eksempel 2.1. Finnes det flere?

2.2. Vis at den tomme mengden $\emptyset$ og utfallsrommet Ω er hendelser.

2.3. Vis at $\mathcal{E}$ i ligning (2.4) er den minste hendelsesfamilien som inneholder den enkle hendelsen $\{M\}$. Finnes det en større hendelsesfamilie?

2.4. La $\Omega = \{1, 2, 3\}$. Er $\{\{1, 2\}, \emptyset\}$ en familie av mengder? Er $\{\{1, 2\}, \emptyset\}$ en hendelsesfamilie? Er $\mathcal{E} = \{\emptyset, \Omega, \{1, 2\}, \{3\}\}$ en hendelsesfamilie? La Ω være utstyrt med hendelsene i $\mathcal{E}$. Er 1 et utfall? Er $\{1\}$ en hendelse? Bruk dette til å begrunne at det kan finnes en delmengde av utfallsrommet som ikke er en hendelse.

2.5 (**T**). La $A = \{\mathcal{G} \,|\, \mathcal{G} \text{ er en hendelsesfamilie som inneholder alle intervall i } \mathbb{R}\}$. Vis at familien $\mathcal{E}_d$ av alle delmengder av $\mathbb{R}$ er inneholdt i A. La

$$\mathcal{E} = \bigcap_{\mathcal{G}\in A} \mathcal{G}$$

Er $\mathcal{E}$ en familie mengder? Er $\mathcal{E}$ en hendelsesfamilie? Er $\mathcal{E}$ den minste hendelsesfamilien som inneholder alle intervall?

2.6 (**T**). Vis at $\Omega = \mathbb{R}^3$ er et fjerde mulig utfallsrom for eksperimentet i Eksempel 2.2. Illustrer de mulige utfall som punkter i rommet og marker spesielt resultatet ω som ble observert etter tre myntkast. En *åpen boks* er en mengde på formen $B = I_1 \times I_2 \times I_3$ hvor I_j er et åpent intervall. La $\mathcal{E}$ være den minste hendelsesfamilien som inneholder alle åpne bokser. Vis at $\{\omega\}$ er en hendelse.

2.7. Resultatet av å kaste en mynt 4 ganger er $MMKM$. Definer et minste utfallsrom for eksperimentet. Husk: Hendelsene må også spesifideres! Vis at $\Omega = \{0,1\}^4$ og $\Omega = \mathbb{R}^4$ også er mulig utfallsrom for eksperimentet. Hendelser?

2.8. Vis at $\Omega = \mathbb{R}$ er et mulig utfallsrom i Eksempel 2.3. Hva er resultatet ω i det gitte eksperimentet? Hendelsesfamilie?

2.9. Se på Eksempel 2.4. Si med ord hva det betyr at hendelsen $\tilde{H} = (t_1, \infty)$ inntreffer. Vis at alle åpne intervall er hendelser i $\tilde{\Omega}$. Utfallet ω er oppgitt i eksempelet. Hva er ω_{10}? Si med ord hva hendelsen $H = (t_1, \infty) \times \cdot \times (t_n, \infty)$ betyr. Vis at alle åpne bokser er hendelser.

2.10 (**D**). Se på Eksempel 2.5. Forklar at data gitt i tabell 2.1 tilsvarer et utfall ω i utfallsrommet

$$\Omega = \mathbb{N}^k \tag{2.19}$$

med $k = 18$. Kan du estimere forventet levetid? Kan du tegne et normalisert histogram?

2.11. Gi eksempler som viser at utfallsrommet Ω kan være en endelig mengde, en tellbar mengde, eller en mengde som ikke er tellbar. Husk: Hendelsene må også defineres.

2.12. Hva er et eksperiment? Gi et eksempel og gi en begrunnelse.

2.13. La $A = \{[0, 1/4), [1/4, 1/2), [1/2, 1)\}$. Er A en familie mengder? Er A en hendelsesfamilie? La $\Omega = [0, 1)$ være utfallsrommet utstyrt med den minste hendelsesfamilien som inneholder A. Er $[1/2, 3/4)$ en hendelse?

2.14. Hva er utfallsrommet for et enkelt terningkast? Hva er utfallsrommet for gjentatt kast med terning inntil resultatet blir seks øyne? Hva er sannsynligheten for at terningen må kastes 2 ganger? Hva er sannsynligheten for at terningen må kastes n ganger?

2.15. En student som snart er ferdig med utdannelsen er innkalt til tre jobbintervju. Hun vil karakterisere et intervju som en suksess dersom det resulterer i en invitasjon til en omvisning på arbeidstedet. I motsatt fall karakteriseres intervjuet som en fiasko. Definer et utfallsrom Ω for dette eksperimentet. Hvilke utfall er inneholdt i hendelsen $A =$ "andre suksess forekommer i tredje intervju"? Finn hendelsen $B =$ "første suksess forekommer aldri".

2.16. En urne inneholder 6 brikker som er nummerert fra 1 til 6. Tre brikker trekkes fra urnen. Hvilke utfall er inneholdt i hendelsen "den nest minste verdien er 3"?

2.17. I et terningsspill (*craps*) kastes to terninger. Kasteren vinner dersom resultatet er 7 eller 11, og taper dersom resultatet er 2, 3 eller 12. Dersom resultatet er noe annet, anta 9, så fortsetter kastingen inntil resultatet er 7 (tap) eller inntil resultatet 9 fra det første kastet gjentas (vinst). Karakteriser utfallene i hendelsen "kasteren vinner med en 9".

2.18. Studentsamskipnaden viste filmen "Ny student i byen" to ganger i løpet av orienteringsdagen. Av totalt 2000 nye studenter var det 850 som så på den første forestillingen, 690 så på den andre forestillingen, mens 730 unnlot å se filmen. Hvor mange nye studenter så filmen to ganger?

2.19. La $PA = 1/3$, $PB = 1/2$ og $P(A \cup B) = 3/4$. Finn $P(A \cap B)$, $P(A^c \cup B^c)$ og $P(A^c \cap B)$.

2.20. Et eksperiment har to mulige utfall. Det ene har sannsynlighet p og det andre har sannsynlighet p^2. Finn p.

2.21. To kort trekkes i rekkefølge fra en kortstokk med 52 kort. Hva er sannsynligheten for at det andre kortet som trekkes er større enn det første kortet? Hint: Enten er kortene like store eller så er det første størst eller så er det andre størst. Disse tre hendelsene er disjunkte.

2.22. La Ω være et utfallsrom utstyrt med en hendelsesfamilie. Vis at $\emptyset$ og Ω er hendelser. Vis at et tellbart snitt av hendelser er en hendelse.

2.23. Vis at $nA = \sum_{i=1}^{n}(\omega_i \in A)$ ved å definere det som inngår. Vis at den empiriske fordelingen er en fordeling.

2.24 (**T**). La $\mathcal{E}_i$ være hendelsesfamilier i Ω. Vis at $\mathcal{F} = \mathcal{E}_1 \cap \mathcal{E}_2$ er en hendelsesfamilie. Vis at $\mathcal{G} = \cap_i \mathcal{E}_i$ er en hendelsesfamilie.

2.25. Vis at $\mathrm{P}(\emptyset) = 0$.
Vis at $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$.
Vis at $A \subset B$ gir $\mathrm{P}(A) \leq \mathrm{P}(B)$.
Vis at $\mathrm{P}(A) \leq 1$.
Vis at $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.

2.26 (**T**). Kolmogorov (1933, s.2,s.14) formulerte en teori for sannsynlighet ved en liste på seks aksiomer. De to første definerer familien av hendelser og de to neste krever at P er positiv og normalisert. De to siste aksiomene er gitt ved å kreve at P er endelig additiv og kontinuerlig: $A_1 \supset A_2 \supset \cdots$ og $\cap_i A_i = \emptyset$ impliserer $\mathrm{P}(A_i) \to 0$. Er dette ekvivalent med å anta tellbar additivitet?

Kapittel 3

Sannsynlighetstetthet

Tyche, in Greek religion, the goddess of chance, with whom the Roman Fortuna was later identified; a capricious dispenser of good and ill fortune. The Greek poet Hesiod called her the daughter of the Titan Oceanus and his consort Tethys; other writers attributed her fatherhood to Zeus, the supreme god. She was also associated with the more beneficent Agathos Daimon, a good spirit, protective of individuals and families, and with Nemesis, who, as an abstraction, represented punishment of overprosperous man and so was believed to act as a moderating influence. She was often shown winged, wearing a crown, and bearing a sceptre and cornucopia; but she also appeared blindfolded and with various devices signifying uncertainty and risk. Among her monuments was a temple at Argos, where the legendary Palamedes is said to have dedicated to her the first set of dice, which he is supposed to have invented.

Britannica Online (1997)

Fortuna, in Roman religion, goddess of chance or lot who became identified with the Greek Tyche; the original Italian deity was probably regarded as the bearer of prosperity and increase. As such she resembles a fertility deity, hence her association with the bounty of the soil and the fruitfulness of women. Frequently she was an oracular goddess consulted in various ways regarding the future. Fortuna was worshiped extensively in Italy from the earliest times. At Praeneste her shrine was a well-known oracular seat, as was her shrine at Antium. Fortuna is often represented bearing a cornucopia as the giver of abundance and a rudder as controller of destinies, or standing on a ball to indicate the uncertainty of fortune.

Britannica Online (1997)

Det er to typer fordelinger som er spesielt vanlige: Diskrete og kontinuerlige fordelinger. Sannsynlighetsfordelingene er da ofte gitt ved en tetthet.

Definisjon 3.1 (*Sannsynlighetstetthet*) En funksjon f slik at

$$P(A) = \int_A f(\omega)\,\mu(d\omega) \tag{3.1}$$

Funksjon f er da en *tetthet* for fordelingen P med hensyn på målet μ. Sannsynligheter $\mathrm{P}(A)$ kan da beregnes ved integrasjon. Slike integral drøftes mer generelt i videregående kurs i analyse (Rudin, 1987).

Vi skal se på tilfellet hvor $\mu(A) =$ antall punkter i hendelsen A. I dette tilfellet er integralet over definert ved summasjon

$$P(A) = \sum_{\omega_i \in A} f(\omega_i) \tag{3.2}$$

hvor $\omega_1, \omega_2, \ldots$ er gitte punkter. I dette tilfellet er f er en *diskret tetthet* og P er en *diskret fordeling*.

Et annet tilfelle er gitt ved at utfallsrommet er tallinjen og

$$P(A) = \int_A f(\omega)\, d\omega \tag{3.3}$$

I dette tilfellet er f en tetthet for utfallsrommet $\mathbb{R}$. Dersom f er stykkevis kontinuerlig og mengden A er et intervall, så er integralet (3.3) over gitt ved Riemann-integralet.

Et poeng med innføringen av tettheter er at det kan være enklere å spesifisere en fordeling P ved en tetthet f fremfor å spesifisere P direkte. Termen *sannsynlighetstetthet* brukes dersom det er nødvendig å understreke koplingen til sannsynligheter gitt av ligning 3.1. Det finnes andre typer tettheter. Noen eksempler er gitt ved massetetthet, ladningstetthet og befolkningstetthet.

3.1 Diskrete fordelinger

La Ω være et utfallsrom. En funksjon $f : \Omega \to [0, \infty)$ er en *diskret sannsynlighetstetthet* dersom

$$\sum_{\omega \in \Omega} f(\omega) = 1 \tag{3.4}$$

Et utfall ω er da et *atom* dersom $f(\omega) > 0$. Det kan vises at det overstående gir det finnes høyst et tellbart antall atomer $\omega_1, \omega_2, \ldots$. Summen (3.4) er derfor en sum av tellbart mange positive tall.

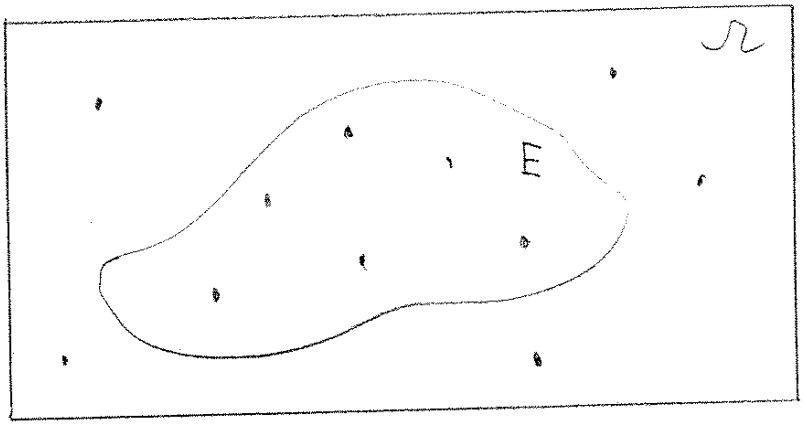
En fordeling P er en *diskret fordeling* dersom det finnes en diskret sannsynlighetstetthet f slik at

$$P(E) = \sum_{\omega \in E} f(\omega). \tag{3.5}$$

Spesielt gir dette at $P\{\omega\} = f(\omega)$ dersom $\{\omega\}$ er en hendelse. Dersom f er en diskret sannsynlighetstetthet, så kan P defineres ved formelen over. En diskret fordeling for Ω er karakterisert av atomene $\omega_1, \omega_2, \ldots$ og sannsynligheten $p_i = f(\omega_i)$ til hvert atom ω_i. Dette er illustrert i figur 3.1.

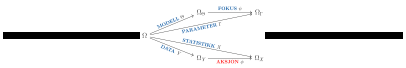
Eksempel 3.1 (To myntkast) Et mulig utfallsrom er

$$\Omega = \{MM, MK, KM, KK\} \tag{3.6}$$



Figur 3.1: Atomene til en diskret fordeling og en hendelse.

og det er rimelig å anta at alle utfallene er like sannsynlige. Dette gir $P\{\omega\} = 1/4$. Definer f ved $f(\omega) = P\{\omega\}$. Det følger at f er en diskret sannsynlighetstetthet og at $P(A) = \sum_{\omega \in A} f(\omega)$. Dermed er P en diskret fordeling.



Et utfallsrom Ω er et *diskret utfallsrom* dersom Ω er tellbar og er utstyrt med en hendelsesfamilie slik at alle enpunktsmengder $\{\omega\}$ er hendelser. Eksperimentet med to myntkast kan beskrives ved en diskret sannsynlighetstetthet fordi utfallsrommet var diskret. Dersom Ω er et diskret utfallsrom, så følger det generelt at P er diskret. Det neste eksemplet viser at P kan være diskret uten at Ω er diskret.

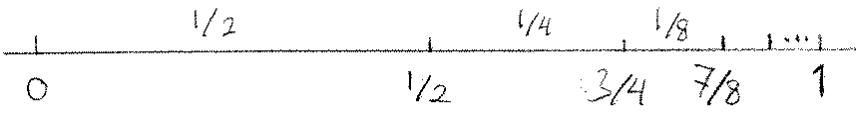
Eksempel 3.2 (Pilkast) Et mulig utfallsrom er

$$\Omega = \{BOM\} \cup \{(x, y) \,|\, x^2 + y^2 \leq R^2\} \tag{3.7}$$

Her er x, y pilens horisontale og vertikale koordinater med origo $(0, 0)$ i sentrum av blinken som har radius R. Anta at en person har kastet n ganger med resultat $\omega_1, \ldots, \omega_n$. Anta at du skal gi en vurdering av sannsynligheten for at neste pilkast gir et resultat i mengden A. En mulig vurdering for det neste pilkastet er gitt ved $\mathrm{P}(A) = nA/n$, hvor nA er antall ganger A inntraff i de foregående pilkastene. Med andre ord er nA antall j slik at ω_j er i A. Formelt kan dette skrives som $nA = \sum_{i=1}^{n} 1_A(\omega_i)$, hvor 1_A er indikatorfunksjonen til A. Tettheten er gitt ved

$$f(\omega) = \frac{n\{\omega\}}{n} = \sum_i \frac{1_{\omega_i}(\omega)}{n}. \tag{3.8}$$

Ligning (3.8) definerer den *empiriske tettheten* P_n tilsvarende observasjoner $\omega_1, \omega_2, \ldots, \omega_n$ helt generelt.



Figur 3.2: Den geometriske rekken $1/2 + 1/4 + 1/8 + \cdots = 1$.

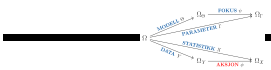
Den *uniforme diskrete fordelingen* er definert av distinkte atomer $\omega_1, \ldots, \omega_n$ og en tetthet $f(\omega_i) = 1/n$. Det antas normalt at enpunktsmengdene $\{\omega_i\}$ er hendelser. Ofte er utfallsrommet $\Omega = \{\omega_1, \ldots, \omega_n\}$, men ikke alltid. Dette er den viktigste diskrete fordelingen. I gode tilfeller kan symmetriargumenter benyttes for å konkludere at en uniform fordeling er rimelig. F eks vil vi anta at det er $1/6$ sannsynlighet for å få en 5'er i et terningkast dersom terningen er symmetrisk, dvs dersom det ikke er noen forskjell på de seks sidene på terningen. Det neste eksemplet viser at uniforme diskrete fordelinger kan benyttes til utledning av andre fordelinger.

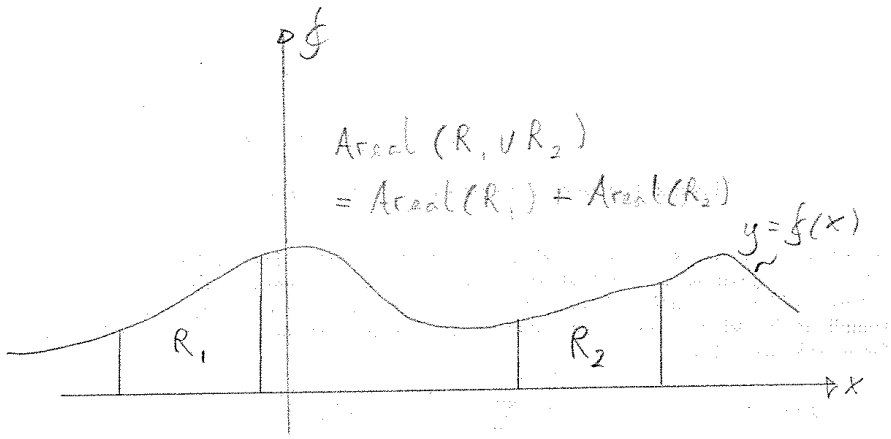
Eksempel 3.3 (Myntkast inntil resultatet er krone) Det diskrete utfallsrommet

$$\Omega = \{K, MK, MMK, \ldots\} \tag{3.9}$$

bør være selvforklarende. Vi vil nå finne en fordeling ved å finne sannsynligheten til hvert enkeltutfall. Vi antar at mynten er symmetrisk og konkluderer at $\mathrm{P}\{K\} = 1/2$. Konklusjonen $P\{MK\} = 1/4$ følger fordi hver av de fire utfallene MM, MK, KM, KK i eksperimentet hvor en mynt kastes 2 ganger er like sannsynlige ved symmetri med hensyn til kastrekkefølgen og myntens to sider. Utfallet MK i dette siste eksperimentet gir sannsynligheten for utfallet MK i Ω fordi betingelsene for dette utfallet er like i de to eksperimentene. Tilsvarende kan sannsynligheten for hvert enkeltutfall i Ω finnes ved sannsynligheten for et tilsvarende utfall i et eksperiment med 2^n like sannsynlige utfall. Dette gir generelt at $\mathrm{P}\{\omega\} = 1/2^n$ hvor n er antall myntkast tilsvarende ω. Vi kan kontrollere at P er normalisert ved $\mathrm{P}(\Omega) = 1/2 + 1/4 + 1/8 + \cdots = 1$. Dette summasjonsresultatet er illustrert i figur 3.2. Generelt er summen av en geometrisk rekke $a + ak + ak^2 + \cdots = a/(1-k)$ dersom $|k| < 1$. Vi kan nå regne ut sannsynligheten for mer kompliserte hendelser. F eks er sannsynligheten for at vi må kaste mynten et odde antall ganger før vi får krone lik $2/3$:

$$\mathrm{P}\{K, MMK, MMMMK, \ldots\} = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{2i+1} = \frac{1}{2}\sum_{i=0}^{\infty}\left(\frac{1}{4}\right)^{i} = \frac{1}{2}\frac{1}{1-1/4} = 2/3 \tag{3.10}$$





Figur 3.3: Addisjonregelen for en kontinuerlig tetthet tilsvarer addisjon av areal under grafen.

3.2 Kontinuerlige fordelinger for $\mathbb{R}$

La utfallsrommet være $\mathbb{R}$. Fordelingen P er kontinuerlig dersom

$$\mathrm{P}(A) = \int_A f(x)dx$$

Generelt defineres $\int_A f(x)dx = \int 1_A(x)f(x)dx$ hvor 1_A er indikatorfunksjonen til A. Funksjonen f er sannsynlighetstettheten til sannsynlighetsfordelingen P. Kravet $\mathrm{P}(A) \geq 0$ gir at $f \geq 0$. Normaliseringen $\mathrm{P}(\Omega) = 1$ gir normaliseringsbetingelsen

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

dvs f er normalisert. Generelt er en tetthet f for tallinjen en funksjon $f : \mathbb{R} \to [0, \infty)$ som er normalisert. Hendelsesfamilien tilsvarende P er her den minste hendelsesfamilien som inneholder alle intervall.

Det at en tetthet f definerer en fordeling er en egenskap til integralet. Den endelige addisjonregelen er illustrert i figur 3.3.

Noen tettheter er så viktige at de har fått egne navn. Vi vil nevne tre eksempler her.

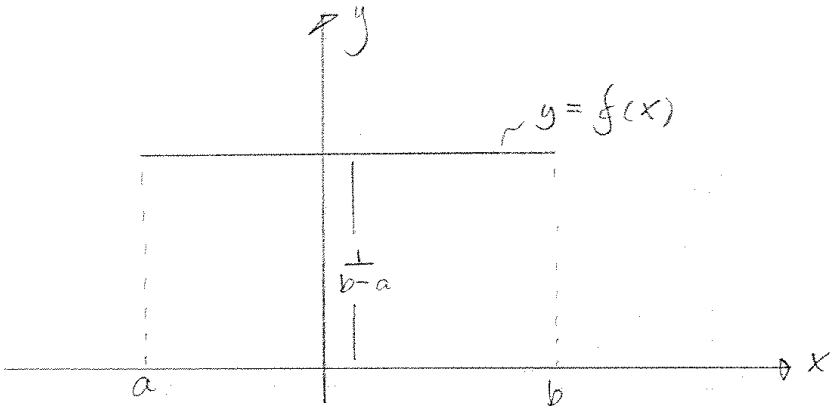
Tettheten til den *uniforme fordelingen* på $[a, b]$ er gitt ved

$$f(x) = \frac{1}{b-a}, \ a \leq x \leq b \tag{3.11}$$

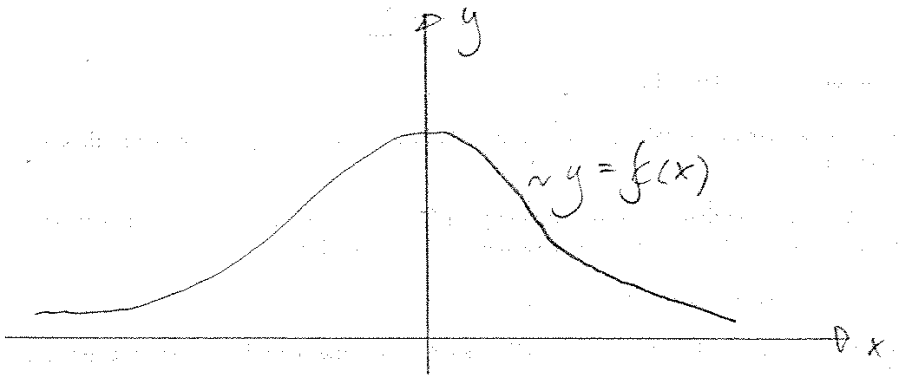
Det er her underforstått at $f(x) = 0$ når $x \notin [a, b]$. Figur 3.4 indikerer grafen til den uniforme fordelingen.

Tettheten til *normalfordelingen* er gitt ved

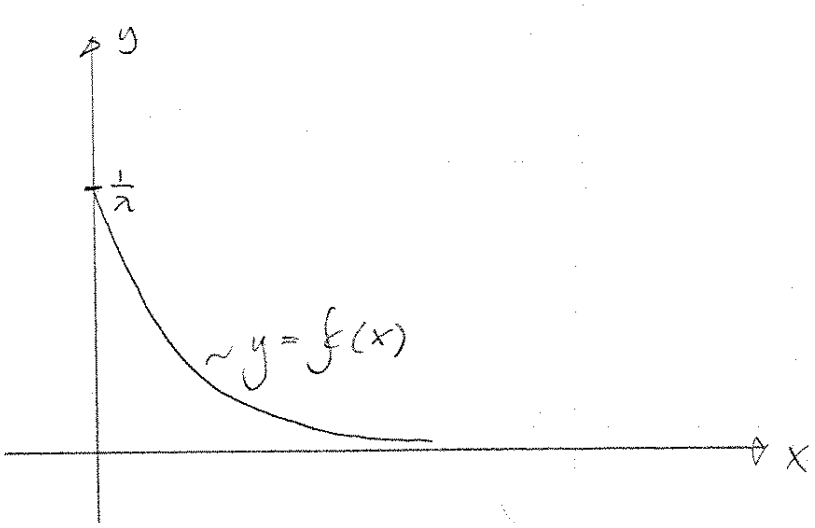
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.12}$$



Figur 3.4: Tettheten til en uniform fordeling.



Figur 3.5: Tettheten til en normalfordeling.



Figur 3.6: Tettheten til en eksponensialfordeling.

Figur 3.5 indikerer grafen til normalfordelingen.

Tettheten til *eksponensialfordelingen* er gitt ved

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, x > 0. \tag{3.13}$$

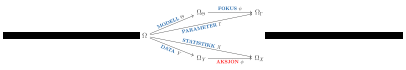
Det er her underforstått at $f(x) = 0$ når $x \leq 0$. Figur 3.6 indikerer grafen til eksponensialfordelingen.

Eksempel 3.4 (Fri veilengde) Den frie veilengden er avstanden et molekyl tilbakelegger mellom to kollisjoner med andre molekyler i en gass. Det kan argumenteres for at den frie veilengden er eksponensialfordelt med parameter λ. Vi definerer

$$\text{midlere fri veilengde } = \int x f(x) dx = \int_0^\infty x \frac{1}{\lambda} e^{-x/\lambda} dx \overset{delv.int.}{=} \lambda,$$

dvs parameteren λ er den midlere frie veilengden. Vi kan nå f eks beregne sannsynligheten for at den frie veilengden er mindre eller lik halvparten av den midlere frie veilengden:

$$P(\text{fri veilengde} \leq \lambda/2) = \int_0^{\lambda/2} \frac{1}{\lambda} e^{-x/\lambda} dx = \left[\frac{1}{\lambda} (\text{-}\lambda) \text{e}^{\text{-x}/\lambda} \right]_0^{\lambda/2} = 1 - e^{-1/2} \approx \underline{\underline{0,39}}.$$



3.3 Oppgaver

3.1. Massetetthet, ladningstetthet og befolkningstetthet er eksempel på tettheter for ulike størrelser. Hvordan vil du definere de? Hva er en sannsynlighetstetthet? Måleenheten til

en romlig massetetthet ρ er $[\rho] = \mathrm{kg\,m^{-3}}$. Hva er måleenheten til en kontinuerlig sannsynlighetstetthet? Til en diskret sannsynlighetstetthet?

3.2. Hva må f oppfylle for at ligning (3.2) skal definere en sannsynlighet P? Hva med ligning (3.3)?

3.3 (**T**). La $\omega_1, \omega_2, \ldots$ være atomene til en diskret fordeling. Hva betyr dette? Finnes det et sannsynlighetsrom slik at atomet $\{\omega_1\}$ ikke er en hendelse?

3.4. Vis at enhver delmengde av et diskret utfallsrom er en hendelse. Vis at enhver sannsynlighet på et diskret utfallsrom er diskret.

3.5. Bruk Eksempel 3.2 til å besvare følgende spørsmål. Er den empiriske fordelingen alltid uniform? Er den diskret? Er utfallsrommet til en diskret fordeling alltid diskret? Er utfallsrommet til en uniform diskret fordeling alltid diskret?

3.6. Normalfordelingen, eksponensialfordelingen og den uniforme fordelingen er alle gitt ved å spesifisere noen relle parametre. Hvilke parametre? Hvilke verdier kan parametrene ha? Hint: Hva skal til for at en funksjon er en tetthet?

3.7. La $\lambda \geq 0$. Vis at

$$f(n) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^n, n = 0, 1, 2, \ldots$$

definerer en diskret tetthet for utfallsrommet $\mathbb{R}$. Er dette utfallsrommet diskret? Hva med tilfellet $\lambda = 0$?

3.8. La f være en diskret sannsynlighetstetthet. Vis at da er P definert ved

$$P(A) = \sum_{\omega \in A} f(\omega)$$

en sannsynlighetsfordeling.

3.9. Vis at dersom utfallsrommet er diskret, så må fordelingen også være diskret. Gjelder det motsatte?

3.10. Anta at P er den uniforme fordelingen på et utfallsrom Ω. Finn tettheten til P når det er gitt at Ω inneholder 1001 elementer. Hva med tilfellet når $\Omega = \mathbb{N}$?

3.11. Vis at $1/2 + 1/4 + 1/8 + \cdots = 1$.

3.12. La k være en konstant. Finn $P(\text{Personen lever mer enn 60 år})$ når tettheten for personens alder er

$$f(t) = k \cdot t^2(100 - t)^2, 0 \leq t \leq 100.$$

Hva verdien til k?

3.13. Finn sannsynligheten for at det er mellom 50 og 100 dager mellom to gruveulykker når det er gitt at tiden mellom to gruveulykker er eksponensialfordelt med parameter $\lambda = 241$ dager.

3.14. Vis at $1_{\uplus_i A_i} = \sum_i 1_{A_i}$.

3.15. La utfallsrommet være tallinjen $\mathbb{R}$. Vis at det er en-en korrespondanse mellom kontinuerlige fordelinger og tettheter for $\mathbb{R}$.

Kapittel 4

Betinget sannsynlighet og uavhengighet

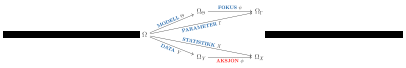
The concept of mutual independence of two or more experiments holds, in a certain sense, a central position in the theory of probability.

A.N. Kolmogorov (1933)

Uavhengighet av hendelser er et meget viktig begrep. Begrepet er tildels utgangspunktet for at sannsynlighetsteorien har utviklet seg til noe mer enn det som finnes i den grunnleggende integrasjonsteorien. To hendelser A, B som har positiv sannsynlighet er uavhengige dersom den betingede sannsynligheten til A gitt B er lik sannsynligheten til A. Med dette som utgangspunkt må vi først si litt om betingede sannsynligheter.

4.1 Betinget sannsynlighet

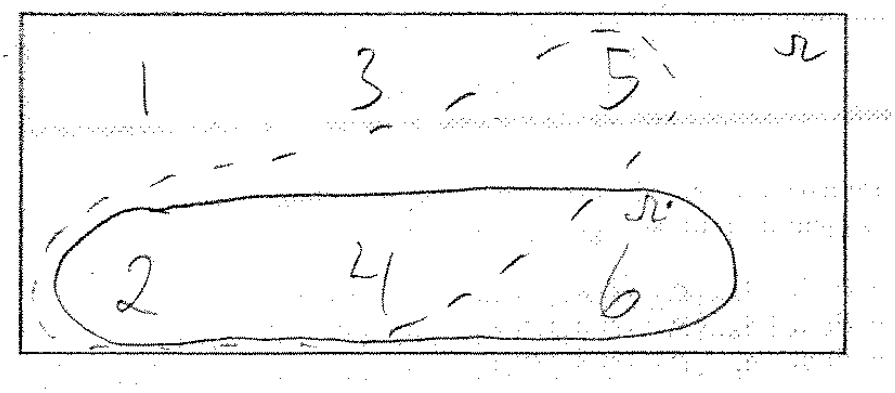
Eksempel 4.1 (Terningkast med informasjon) En terning kastes og det er oppgitt at resultatet er et partall. Noen rimelige vurderinger er da gitt ved $P\{1\} = 0$ (1 er ikke et partall.), $P\{2\} = 1/3$ (Det finnes 3 mulige partall.), og $P\{2, 4, 5\} = 2/3$ (Hendelsen inneholder 2 av de 3 mulige partallene.). Figur 4.1 kan være en hjelp til å forstå den siste vurderingen.



Anta at vi har et statistisk lovmessig eksperiment, dvs sannsynligheten for en hendelse er det samme som det relative antall ganger hendelsen inntreffer når eksperimentet gjentas. Sannsynligheten $P(A \,|\, \Omega')$ for at A inntreffer gitt at Ω' inntreffer er gitt ved å kun telle utfallene innenfor Ω'. Dette leder til

$$P(A \,|\, \Omega') = \lim \frac{n(A \cap \Omega')}{n\Omega'} = \lim \frac{\frac{n(A \cap \Omega')}{n}}{\frac{n\Omega'}{n}} = \frac{P(A \cap \Omega')}{P(\Omega')}.$$

Denne regningen motiverer en generell definisjon.



Figur 4.1: Terningkastet resulterte i at Ω' inntraff.

Definisjon 4.1 (*Betinget sannsynlighet*) Den betingede sannsynligheten til A gitt B er

$$\mathrm{P}(A \,|\, B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},\ \mathrm{P}(B) > 0. \tag{4.1}$$

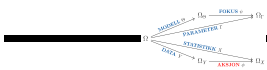
Definisjonen av den betingede sannsynligheten $\mathrm{P}(A \,|\, B)$ gjelder generelt når P er en fordeling. Når B fikseres gir dette en ny fordeling Q ved $Q(A) = \mathrm{P}(A \,|\, B)$.

Eksempel 4.2 (Kortstokk) Et kort trekkes fra en kortstokk. Hva er sannsynligheten for at det er et kløverkort gitt at det er en konge?

Svar a): Utfallsrommet består av de fire kongene, og en av disse er kløver. Sannsynligheten er dermed 1/4.

Svar b): Det opprinnelige utfallsrommet kan også brukes ved

$$P(\text{kløver} \,|\, \text{konge}) = \frac{P(\text{kløver og konge})}{P(\text{konge})} = \frac{\frac{1}{52}}{\frac{4}{52}} = \frac{1}{4}.$$

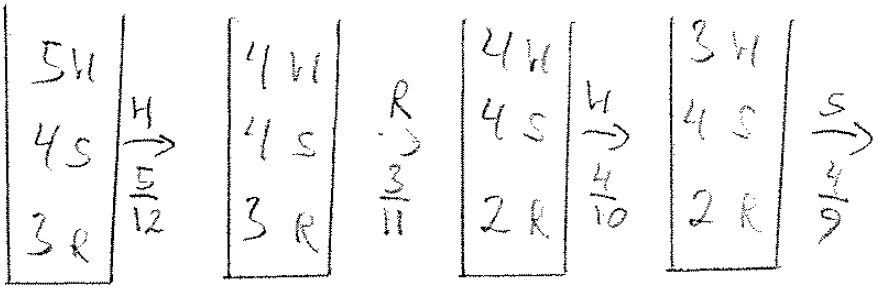


Eksempel 4.3 (Levealder) Anta at levealderen til en person er gitt ved tettheten

$$f(t) = 3 \cdot 10^{-9} \cdot t^2(100 - t)^2,\ 0 \le t \le 100.$$

Hva er sannsynligheten for at en person dør med en alder mellom 80 og 85 år gitt at levetiden er > 70 år? Denne betingede sannsynligheten kan beregnes ved

$$\begin{aligned} P(80 < \text{levetid} < 85 \,|\, \text{levetid} > 70) &= \frac{P(80 < \text{levetid} < 85 \cap \text{levetid} > 70)}{P(\text{levetid} > 70)} \\ &= \frac{P(80 < \text{levetid} < 85)}{P(\text{levetid} > 70)} = \frac{\int_{80}^{85} f(t)dt}{\int_{70}^{100} f(t)dt} \approx 0,19 \end{aligned}$$



Figur 4.2: Betingede sannsynligheter ved trekking av 4 brikker fra en urne.

Definisjonen av $\mathrm{P}(A \mid B)$ har en viktig omskrivning

$$P(A \cap B) = P(A \mid B)P(B). \tag{4.2}$$

Mer generelt kan sannsynligheten til snittet av n hendelser finnes ved produktet av n sannsynligheter ved gjentatt bruk av regneregelen over

$$\begin{aligned} P(A_n \cap \cdots \cap A_1) &= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)P(A_{n-1} \cap \cdots \cap A_1) \\ &= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)P(A_{n-1} \mid A_{n-2} \cap \cdots \cap A_1)P(A_{n-2} \cap \cdots \cap A_1) \\ &= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)P(A_{n-1} \mid A_{n-2} \cap \cdots \cap A_1) \cdots P(A_2 \mid A_1)P(A_1). \end{aligned}$$

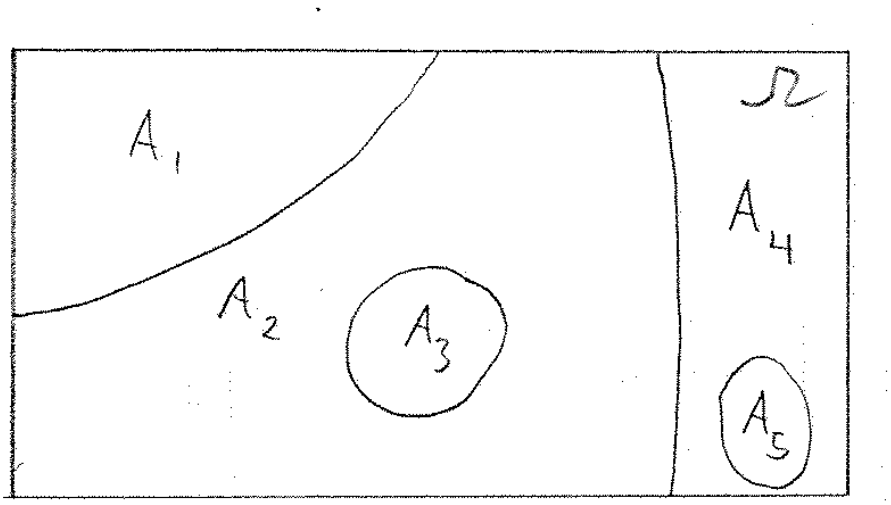
Vi tar med et eksempel på bruk av denne regneregelen.

Eksempel 4.4 (Trekking uten tilbakelegging) En urne inneholder 5 hvite brikker, 4 sorte brikke og 3 røde brikker. Hva er sannsynligheten for å trekke sekvensen (hvit, rød, hvit, sort) når 4 brikker trekkes uten tilbakelegning?

Definer hendelsene A_1 = "hvit trekkes i 1.trekning", A_2 = "rød trekkes i 2.trekning", A_3 = "hvit trekkes i 3.trekning" og A_4 = "sort trekkes i 4.trekning". Den søkte sannsynligheten er gitt ved (se figur 4.2)

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3 \cap A_4) &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)P(A_4 \mid A_1 \cap A_2 \cap A_3) \\ &= \frac{5}{12}\frac{3}{11}\frac{4}{10}\frac{4}{9} \approx 0,02. \end{aligned}$$

En familie $\{A_i\}$ av hendelser er en *partisjon* (= oppdeling) av Ω dersom $\Omega = \uplus_i A_i$. Det er underforstått i notasjonen her at hendelsene er disjunkte. En partisjon med 5 hendelser er illustrert i figur 4.3. Dersom $\{A_i\}$ er en tellbar partisjon av Ω, så følger det at



Figur 4.3: En partisjon av Ω.

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(B \cap \Omega) = \mathrm{P}(B \cap (\uplus_i A_i)) = \mathrm{P}(\uplus_i (B \cap A_i)) = \sum_i \mathrm{P}(B \cap A_i) \\ &= \sum_i \mathrm{P}(B \mid A_i) P(A_i).\end{aligned}$$

Innsetting av dette gir Bayes formel:

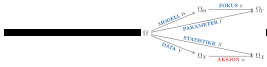
$$\mathrm{P}(A_j \mid B) = \frac{\mathrm{P}(A_j \cap B)}{P(B)} = \frac{\mathrm{P}(B \mid A_j)\mathrm{P}(A_j)}{\sum_i \mathrm{P}(B \mid A_i)\mathrm{P}(A_i)}.$$

Formelen gir sannsynligheten for A_j gitt B ved sannsynlighetene for B gitt A_i og sannsynlighetene for A_i. Legg merke til at betingelsene er snudd!

Eksempel 4.5 (Valg) Et valg mellom 1 republikaner og 3 demokrater foregår ved at 1 av demokratene velges ut først i et primærvalg. Sannsynligheten for at demokrat 1, 2, eller 3 vinner primærvalget estimeres til å være henholdsvis 35%, 40% og 25%. Hva er sannsynligheten for at republikaneren blir sittende i Kongressen dersom republikaneren med sannsynlighet 40%, 35% og 60% vinner et valg mot henholdsvis demokrat 1, 2 og 3?

La B være hendelsen at republikaneren beholder sitt sete etter valget. Hendelsene A_1, A_2, A_3 tilsvarende at henholdsvis demokrat 1, 2 eller 3 vinner primærvalget gir en partisjon av utfallsrommet. Dermed har vi

$$\mathrm{P}(B) = \sum_i \mathrm{P}(B \mid A_i)\mathrm{P}(A_i) = 0,40 \cdot 0,35 + 0,35 \cdot 0,40 + 0,60 \cdot 0,25 \approx 43\%.$$



Eksempel 4.6 (Varsellampe) Hva er sannsynligheten for at bilen du kjører har for lite olje gitt at oljelampen lyser rødt og indikerer at det er for lite olje?

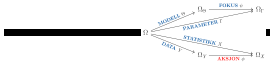
Vi antar at sannsynligheten for at oljelampen lyser ved for lite olje er 99%. Videre antar vi at oljelampen i 2% av tilfellene lyser uten grunn. Tilslutt antar vi at det er 10% sjanse for at det er for lite olje på bilen.

Sannsynligheten for at oljelampen lyser er gitt ved

$$\mathrm{P}(\text{varsel}) = \mathrm{P}(\text{varsel og nok olje})\mathrm{P}(\text{nok olje}) + \mathrm{P}(\text{varsel og lite olje})\mathrm{P}(\text{lite olje}).$$

Dette gir den søkte sannsynligheten

$$\begin{aligned}\mathrm{P}(\text{lite olje} \,|\, \text{varsel}) &= \frac{P(\text{lite olje og varsel})}{P(\text{varsel})} = \frac{P(\text{varsel} \,|\, \text{lite olje})P(\text{lite olje})}{P(\text{varsel})} \\ &= \frac{0,99 \cdot 0,10}{0,02 \cdot 0,90 + 0,99 \cdot 0,10} \approx 85\%.\end{aligned}$$



4.2 Uavhengighet

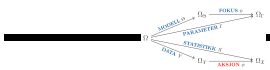
Definisjon 4.2 (*Uavhengighet*) Hendelsene i familien $\mathcal{G}$ er uavhengige dersom ulike hendelser $A_1, \ldots, A_n \in \mathcal{G}$ impliserer

$$\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1) \cdots \mathrm{P}(A_n) \tag{4.3}$$

Det viktigste tilfellet er gitt ved at hendelsene A, B er uavhengige dersom $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$, eller ekvivalent dersom $\mathrm{P}(A \,|\, B) = \mathrm{P}(A \cap B)/\mathrm{P}(B) = \mathrm{P}(A)$ i tilfellet hvor $P(B) \neq 0$.

Dersom A og B er uavhengige er sannsynligheten til A gitt B lik sannsynligheten til A.

Eksempel 4.7 (Uavhengige myntkast) Utfallsrommet tilsvarende n myntkast er gitt ved $\Omega = \{k_1 \ldots k_n \,|\, k_j = K, M\}$. Hver av de 2^n mulige utfallene er like sannsynlige ved symmetri. Dermed er $P(A) = \sum_{\omega \in A} 2^{-n}$. La $A_1 =$ "K i 1.kast" og $A_2 =$ "M i 2.kast". Da er $P(A_1) = 2^{n-1}/2^n = 1/2$ og tilsvarende er $P(A_2) = 1/2$ og $P(A_1 \cap A_2) = 1/4$. Dette gir $P(A_1 \cap A_2) = P(A_1)P(A_2)$ og det er bevist at hendelsene A_1 og A_2 er uavhengige. Det kan bevises tilsvarende at dersom $A_1, \ldots, A_n$ er hendelser tilsvarende resultatene i henholdsvis myntkast nummer $1, \ldots, n$, så er hendelsene $A_1, \ldots, A_n$ uavhengige.



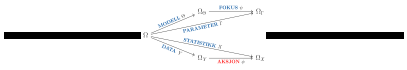
Tabell 4.1: Værmelding i forhold til tilsvarende observert vær.

Været \ Varsel	Sol	Overskyet	Regn
Sol	30 %	5 %	5 %
Overskyet	4 %	20 %	2 %
Regn	10 %	4 %	20 %

Mynteksemplet ga et eksempel hvor vi kunne bevise ved symmetri at de gjentatte forsøkene var uavhengige. I det neste eksemplet er en antagelse om uavhengighet utgangspunktet for beregning av en sannsynlighet.

Eksempel 4.8 (Uavhengige lyskryss) På vei til jobben passerer Eva 4 lyskryss. Anta at hvert av lyskryssene lyser grønt 40 sekunder i løpet av hvert minutt. Anta videre at avstanden mellom lyskryssene er så stor at hvert av lyskryssene er uavhengige. Sannsynligheten for at Eva må stoppe minst tre ganger er da

$$\begin{aligned} P\{RRRR, GRRR, RGRR, RRGR, RRRG\} &= P\{RRRR\} + 4P\{RRRG\} \\ &= (20/60)^4 + 4(20/60)^3(40/60) = 1/9. \end{aligned}$$



4.3 Oppgaver

4.1. La $\mathrm{P}(B) > 0$. Vis at Q definert ved $Q(A) = \mathrm{P}(A \,|\, B)$ er en sannsynlighet. Er det flere mulige utfallsrom som er naturlige for Q?

4.2. Tabell 4.1 gir sannsynligheter for snittet av hendelser definert av henholdsvis været og den tilsvarende værmeldingen til et meterologisk institutt. Hva er sannsynligheten for at det skal bli sol? Hva er sannsynligheten for regn dersom det var varslet sol? Hvor ofte gjetter instituttet galt? Hvordan ville du gå frem for å lage en tabell tilsvarende Tabell 4.1?

4.3. La fordelingen P være eksponensialfordelingen med parameter λ. Vis at

$$P((s+t, \infty) \,|(t, \infty)) = P(s, \infty).$$

Hva betyr dette resultatet dersom fordelingen tilsvarer tiden mellom gruveulykker?

4.4. Hva er sannsynligheten for sekvensen "sort, sort, rød, hvit, hvit" når 5 brikker trekkes uten tilbakelegning fra en urne inneholdende 6 hvite, 4 sorte og 5 røde brikker?

4.5. En gambler har tre kort; det 1. er rødt på begge sidene, det 2. er blått på begge sidene, det 3. er rødt på den ene siden og blått på den andre siden. Et kort trekkes tilfeldig og legges på bordet. Anta at kortet som vises er rødt. Gambleren vedder Kr 1000 på at den andre siden av kortet er rødt? Bør du vedde i mot?

4.6. En bøyd mynt gir krone med dobbel så stor sannsynlighet som den gir mynt. Dersom den viser krone, så trekkes en brikke fra en urne med 3 hvite og 4 røde brikker. Dersom mynten viser mynt, så trekkes en brikke fra en urne med 6 hvite og 3 røde brikker. Hva er sannsynligheten for at mynten viste mynt, når en hvit brikke ble trukket?

4.7. En sort og en hvit terning kastes. Se på hendelsene A = "odde tall på den sorte terningen", B = "odde tall på den hvite terningen" og C = "summen på terningene er odde". Vis at hver av disse hendelsene er parvis uavhengige. Er hendelsene A, B, C uavhengige?

4.8. To terninger kastes. Hva er sannsynligheten for at den ene terningen viser dobbelt så mange øyne som den andre?

4.9. Følgende problem ble stilt til Pascal av de Méré i 1654: Hvor mange ganger må to terninger kastes for at sannsynligheten for å få minst en dobbel 6'er er større enn $1/2$?

4.10. Vis at $(\uplus_i A_i) \cap B = \uplus_i (A_i \cap B)$.

4.11. La A være en hendelse slik at $P(A) > 0$. Vis at familien av alle hendelser inneholdt i A er en hendelsesfamilie i A. Vis at A blir et sannsynlighetsrom med fordeling Q når vi definerer $Q(B) = P(B)/P(A)$. Hva er forskjellen på Q og den betingede sannsynligheten gitt A?

4.12. Formuler og bevis Bayes teorem.

4.13. Vis ved å anta uavhengighet at sannsynligheten for at du må kaste en mynt 10 ganger før du får krone er lik $(1/2)^{10}$. Kan du gi et argument som ikke bygger på uavhengighet?

4.14. Skriv opp alle betingelser som skal være oppfylt for at hendelsene A_1, A_2, A_3 er uavhengige. Er hendelsene $\emptyset, \emptyset, \emptyset$ uavhengige?

4.15. Anta at A og B er uavhengige. Er da A og B^c uavhengige?

Kapittel 5

Kombinatorisk sannsynlighet

> *In the study of simple games of chance, sampling procedures, occupancy and order problems, etc., we are usually dealing with finite sample spaces in which the same probability is attributed to all points. To compute the probability of an event A we have then to divide the number of sample points in A ("favorable cases") by the total number of sample points ("possible cases"). This is facilitated by a systematic use of a few rules which we shall now proceed to review.*
>
> **W. Feller (1950)**

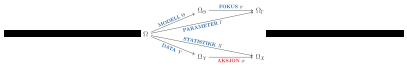
Kombinatorikk er en gammel gren av matematikken. F eks har Det gamle testamentet referanser til kombinatoriske problem. Avhandlingen *Dissertatio de arte combinatoria* (1660) skrevet av Wilhelm Leibniz kan regnes som den første faglige publikasjonen i dette feltet. Hovedproblemet i kombinatorikken er å telle antall mulige kombinasjoner. Vi vil ha nytte av kombinatorikken for f eks å få beregnet antall mulige utfall i en hendelse.

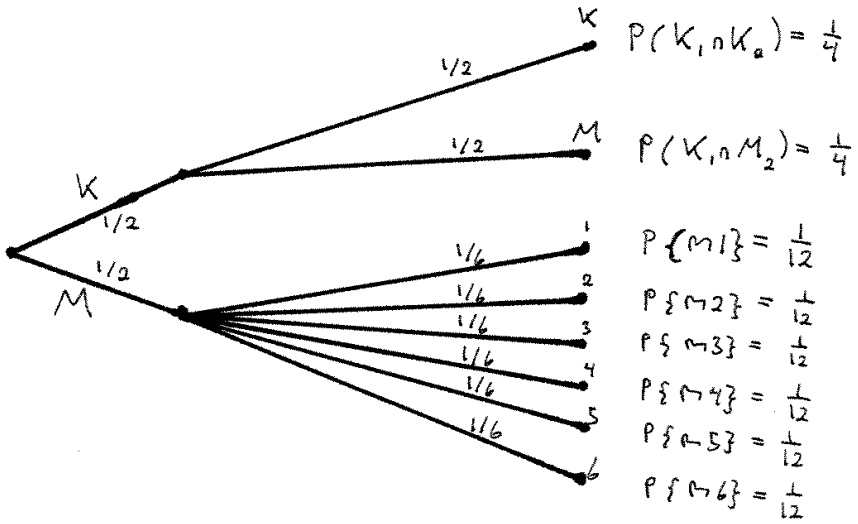
5.1 Multiplikasjonsregelen og permutasjoner

I noen tilfeller er det mulig å illustrere de mulige utfallene med et tre-diagram som i det neste eksemplet.

Eksempel 5.1 (Mynt + Terning) La et eksperiment være gitt ved at vi først kaster en mynt. Dersom resultatet er krone, så kastes mynten en gang til. Dersom resultatet var mynt, så kastes i stedet en terning. Tre-diagrammet i figur 5.1 illustrerer de mulige utfallene. Sannsynligheten til enkeltutfallene kan beregnes ved produktet av sannsynlighetene langs grenene. Sannsynligheten for f eks utfallet $m1$ er

$$P(m1) = P(m \text{ i første deleksperiment}) \cdot P(1 \text{ i andre deleksperiment} \mid m \text{ i første deleksperiment}) = \frac{1}{2}\frac{1}{6} = \frac{1}{12}.$$





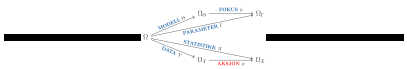
Figur 5.1: Tre-diagram for et spesielt Mynt + Terning eksperiment.

Se på et eksperiment gitt ved at vi først utfører et deleksperiment med m mulige utfall. Deretter utføres et nytt deleksperiment. Som i Eksempel 5.1 kan hvilket deleksperiment som utføres her være avhengig av resultatet i det foregående. Problemet er så å avgjøre hvor mange mulige utfall eksperimentet har. Generelt kan dette fort bli en uoversiktlig oppgave. Et tilfelle er imidlertid enkelt: Dersom det andre deleksperimentet alltid har n mulige utfall, så har eksperimentet $m \cdot n$ mulige utfall. Dette er illustrert i figur 5.2 og 5.3. Multiplikasjonsregelen er elementær, men fortjener å utheves fordi den er så viktig.

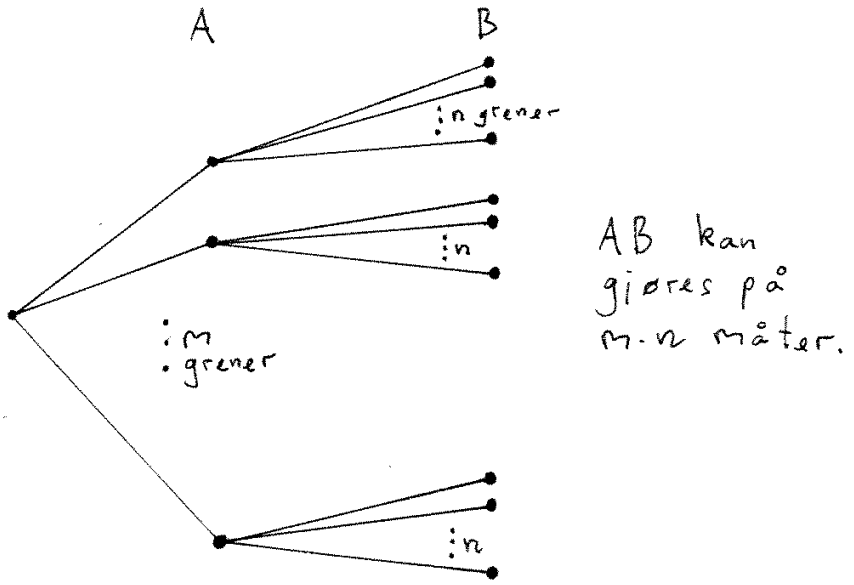
Multiplikasjonsregelen: Anta at operasjon A og B kan gjøres på hhv m og n måter. Da kan operasjonen AB (A etterfulgt av B) gjøres på $m \cdot n$ måter.

Vi valgte å formulere multiplikasjonsregelen uten referanse til sannsynlighetesteorien fordi kombinatorikken også mer generelt står på egne ben uten referanse til sannsynlighet. Ved induksjon kan det vises at operasjonen $A_1 \cdots A_n$ kan gjøres på $m_1 \cdots m_n$ måter dersom hver A_i kan gjøres på m_i måter. Dette er den generelle multiplikasjonsregelen.

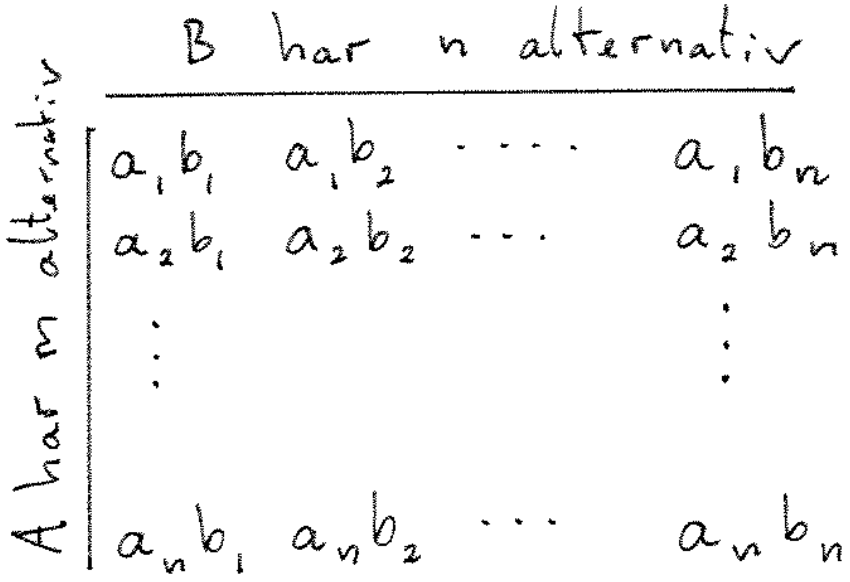
Eksempel 5.2 (Passord) Anta at et passord er på formen bokstav-bokstav-siffer-siffer. Første operasjon er her å velge en bokstav blant 29 mulige, og de resterende operasjonene er tilsvarende. Multiplikasjonsregelen gir da at antall passord blir $29 \cdot 29 \cdot 10 \cdot 10 = 67600$.



Å permutere betyr å bytte om. F eks er alle permutasjoner av *abc* gitt ved *abc*, *acb*, *bac*, *bca*, *cab*, *cba*. Dette viser at det finnes 6 permutasjoner av tre elementer. Dette kunne vi ha konkludert ved hjelp av multiplikasjonsregelen uten å ha skrevet opp alle permutasjonene.



Figur 5.2: Tre-diagram for multiplikasjonsregelen.



Figur 5.3: Matrise for multiplikasjonsregelen.

Operasjonen A_1 er å velge a, b eller c, dvs den kan gjøres på tre måter. Operasjonen A_2 er å velge en av de to gjenværende bokstavene, og operasjonen A_3 er å velge den siste bokstaven. Multiplikasjonsregelen gir da at operasjonen $A_1A_2A_3$ kan gjøres på $3 \cdot 2 \cdot 1 = 6$ måter, dvs det finnes 6 permutasjoner av 3 elementer.

Dette kan generaliseres. Vi definerer $n!$ (les: n *fakultet*) ved

$$n! = n \cdot (n-1) \cdots 1$$

og spesielt $0! = 1$. Multiplikasjonsregelen gir at det finnes $n!$ permutasjoner av n elementer $a_1, a_2, \ldots, a_n$: Operasjon A_1 er å velge første element blant n mulige. Når dette er gjort velges andre element blant de $n-1$ gjenværende element. Generelt kan operasjon A_j gjøres på $n+1-j$ måter hvor $j = 1, \ldots, n$. Spesielt kan da operasjon A_n gjøres på 1 måte fordi det bare er et element igjen å velge. Konklusjonen er da at operasjonen $A_1 \cdots A_n$ kan gjøres på $n \cdot (n-1) \cdots 2 \cdot 1$ måter. Dette betyr at vi kan velge en permutasjon på $n!$ måter.

Det finnes $n!$ mulige permutasjoner av n elementer.

La A_i være som i det foregående. Dersom vi bare gjør operasjon A_1 til og med operasjon A_k, så er resultatet en permutasjon av lengde k fra n elementer. Operasjonen $A_1 \cdots A_k$ kan gjøres på

$$n(n-1)\cdots(n+1-k) = \frac{n(n-1)\cdots(n+1-k)\cdot(n-k)\cdots 2\cdot 1}{(n-k)\cdots 2\cdot 1} = \frac{n!}{(n-k)!}$$

måter. Vi har dermed bevist at:

Det finnes $n!/(n-k)!$ permutasjoner av lengde k fra n elementer.

F eks finnes det $4 \cdot 3 = 12$ permutasjoner av størrelse 2 fra mengden $\{1, 2, 3, 4\}$. Dette kunne også vært funnet ved å liste opp alle disse, dvs $12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43$. Denne siste metoden blir fort uhåndterlig når n og m er store.

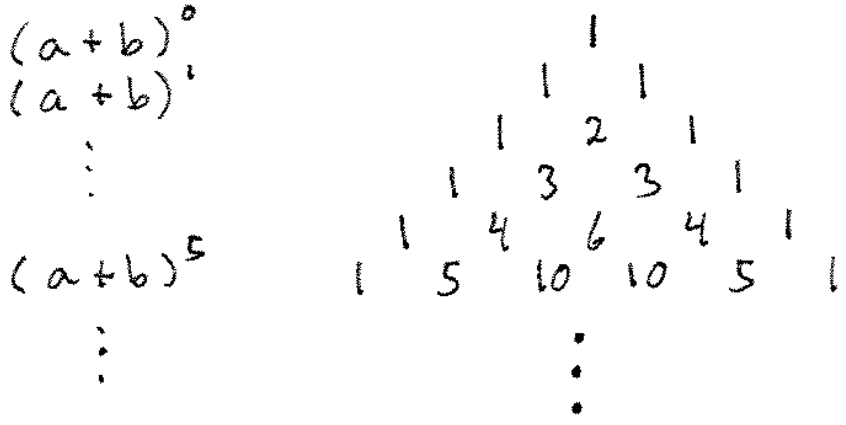
5.2 Binomialkoeffisienter og kombinatoriske sannsynligheter

Pascals trekant i figur 5.4 er gitt ved at hvert tall er gitt ved summen av tallet opp til venstre og tallet opp til høyre. Dette gjelder for alle tallene bortsett fra tallet 1 på toppen (linje 0) når alle linjene fylles ut med 0 der det ikke er skrevet noe i Pascals trekant. Ved hjelp av Pascals trekant kan en beregne koeffisientene i $(a+b)^n$. F eks gir linje 5 i Pascals trekant at

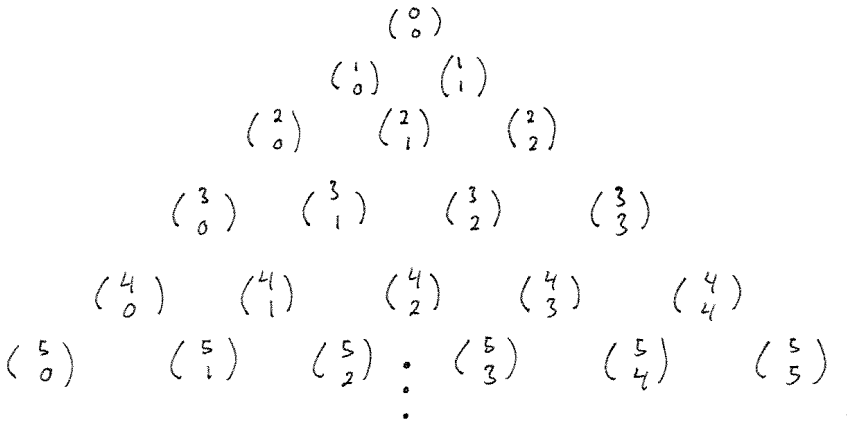
$$(a+b)^5 = b^5 + 5ab^4 + 10a^2b^3 + 10a^3b^2 + 5a^4b + a^5.$$

Tallene i Pascals trekant er *binomialkoeffisientene* definert ved

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$



Figur 5.4: Pascals trekant.



Figur 5.5: Binomialkoeffisientene i Pascals trekant.

Pascals trekant kan da skrives som i figur 5.5. Spesielt gir figur 5.4 og figur 5.5 at $\binom{5}{3} = 10$. Systemet som gir hvordan en linje i Pascals trekant kan regnes ut fra linjen over er gitt ved den kombinatoriske identiteten

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Identiteten bevises ved

$$\begin{aligned}\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-(k+1))!} \\ &= \frac{n!(k+1)}{k!(k+1)(n-k)!} + \frac{n!(n-k)}{(k+1)!(n-(k+1))!(n-k)} \\ &= \frac{n!(k+1+n-k)}{(k+1)!(n-k)!} = \frac{(n+1)!}{(k+1)!(n-k)!} = \binom{n+1}{k+1}.\end{aligned}$$

Denne kombinatoriske identiteten kan benyttes for å gi et induksjonsbevis av binomialteoremet:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

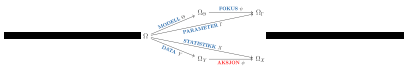
Eksempel 5.3 (Fikset mynt) La p være sannsynligheten for å få krone i et enkelt myntkast. Vi vil vise at sannsynligheten for å få m kronesider ved n myntkast er

$$\binom{n}{m} p^m (1-p)^{n-m}.$$

La Ω være mengden av alle mulige sekvenser $\omega = \omega_1 \dots \omega_n$ hvor ω_i er lik krone eller mynt. Produktregelen gir at utfallsrommet Ω har 2^n mulige enkeltutfall. Når vi antar at hvert myntkast er uavhengig av de andre, så følger det at

$$P\{\omega \,|\, \omega_1 = \cdots = \omega_m = \text{krone}, \omega_{m+1} = \cdots = \omega_n = \text{mynt}\} = p^m(1-p)^{n-m},$$

og mer generelt er dette sannsynligheten til hvert enkeltutfall med m kronesider. Vi må vise at det finnes $\binom{n}{m}$ slike enkeltutfall. Dette er det samme som at det skal finnes $\binom{n}{m}$ ikke-ordnede utvalg av størrelse m fra mengden $\{1, \dots, n\}$ hvor det at j blir trukket tilsvarer at ω_j = krone. Antall ordnede utvalg av størrelse m er lik antall permutasjoner av lengde m, dvs lik $n!/(n-k)!$. Dermed er antall ikke-ordnede utvalg lik binomialkoeffisienten $n!/[(n-k)!k!]$.



Det foregående eksemplet inneholder to resultater som fortjener å utheves.

Antall ordnede og ikke-ordnede utvalg av størrelse m fra n elementer er gitt ved hhv

$$\frac{n!}{(n-m)!} \quad \text{og} \quad \binom{n}{m}.$$

Fordelingen som ble utledet i Eksempel 5.3 er binomialfordelingen. *Binomialfordelingen* er definert ved den diskrete tettheten

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \ 0 \leq p \leq 1, \ k = 0, \ldots, n.$$

5.3 Oppgaver

5.1. Vis ved induksjon at operasjonen $A_1 \cdots A_n$ kan gjøres på $m_1 \cdots m_n$ måter dersom hver A_i kan gjøres på m_i måter.

5.2. Hvor mange ledd er det i ekspansjonen av

$$(a+b+c)(d+e+f)(x+y+u+w) \ ?$$

Hvilke av leddene aeu, cdx, bef, xvw finnes i ekspansjonen?

5.3. Stirlings formel gir at

$$n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n}.$$

Bruk Stirlings formel til å approksimere 30! ved først å beregne en approksimasjon til ln(30!). Sammenlign svaret med den eksakte verdien 265252859812268935315188480000000.

5.4. Skriv opp alle permutasjoner av elementene i mengden $\{1,2,3,4\}$. Skriv også opp alle permutasjoner av lengde 2. Hvordan stemmer disse to listene med multiplikasjonregelen?

5.5. Skriv opp alle ordnede utvalg av størrelse 4 og av størrelse 2 fra mengden $\{1,2,3,4\}$. Skriv også opp de tilsvarende ikke-ordnede utvalg. Hvordan stemmer listene med multiplikasjonregelen?

5.6. Fire menn og fire kvinner skal plasseres i en rad med 8 stoler. Hvor mange plasseringer finnes det? Hvor mange plasseringer finnes det når det skal sitte en mann i annenhver stol?

5.7. Hvor mange linjestykker kan trekkes mellom 5 punkter A, B, C, D, E i planet?

5.8. La M være en mengde. En permutasjon p av elementene i M defineres ved at det er en bijektiv funksjon $p : M \to M$. Dette betyr at p er injektiv ($m_1 \neq m_2 \Rightarrow p(m_1) \neq p(m_2)$) og surjektiv ($\{p(m) \mid m \in M\} = M$), eller med andre ord at p har en invers p^{-1} som er definert på hele M. Skriv opp alle permutasjoner p av elementene i $M = \{a,b,c\}$.

5.9. Prøv å gi definisjoner av ordnede og ikke-ordnede utvalg av størrelse n fra en generell mengde M. Vis at definisjonene er rimelige i tilfellet $n = 2$ og $M = \{1,2,3\}$.

5.10. Bruk den kombinatoriske identiteten

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

til å bevise binomialteoremet ved induksjon.

5.11. Bevis at binomialkoeffisientene er naturlige tall.

5.12. Vis at $\{1, \ldots, n\}$ har $\binom{n}{m}$ delmengder av størrelse m. Bruk dette til å gi et kombinatorisk bevis av binomialteoremet.

5.13. Vis at tettheten til binomialfordelingen er en tetthet for utfallsrommet $\mathbb{R}$.

5.14. Et spill er gitt ved at n personer kaster en mynt. Dersom en spiller får et resultat forskjellig fra alle de andre, så er denne spilleren slått ut av spillet. Hva er sannsynligheten for at en person blir slått ut?

5.15. A somewhat inebriated conventioneer finds himself in the embarrassing predicament of being unable to predetermine whether his next step will be forward or backward. What is the probability that after hazarding n such maneuvers he will have stumbled forward a distance of r steps?

Hint: La x og y være hhv antall skritt forover og bakover. Da er $x+y = n$ og $x-y = r$.

5.16. Følgende oppgave (noe modernisert her) ble stilt til Sir Isac Newton av vennen Samuel Pepys i 1693: Hva er sannsynlighetene for å få minst en 6'er når 6 terninger kastes, for å få minst to 6'ere når 12 terninger kastes, og for å få minst tre 6'ere når 18 terninger kastes?

Kapittel 6

Observatorer og tilfeldige variabler

"A random variable is a quantity whose values are determined by chance." What does that mean? The word "quantity" suggests magnitude - numerical magnitude. Ever since rigor has come to be demanded in mathematical definitions, it has been recognized that the word "variable", particularly a variable whose values are "determined" somehow or other, means in precise language a function. Accordingly a random variable is a function: a function whose numerical values are determined by chance. This means, in other words, that a random variable is a function attached to an experiment - once the experiment has been performed the value of the function is known.

P.R. Halmos (1950)

Teorien for tilfeldige variable og vektorer står sentralt i sannsynlighetsteorien. I Kolmogorovs formulering av sannsynlighetsteorien er enhver tilfeldig størrelse assosiert til et eksperiment representert ved en funksjon definert på utfallsrommet til eksperimentet. Spesielt vil et tilfeldig tall identifiseres med en funksjon som tar reelle verdier. I sannsynlighetsteorien kalles en slik funksjon for en tilfeldig variabel. Generelt gir Kolmogorovs formulering en metode for utledning av nye sannsynlighetsrom fra et underliggende sannsynlighetsrom.

6.1 Observatorer og fordelingen til en observator

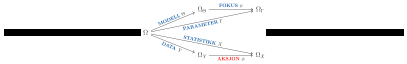
La Ω være utfallsrommet for et eksperiment. I noen situasjoner er det ikke selve utfallet ω som er interesant, men snarere en egenskap $T(\omega)$ til ufallet. Dersom f eks ω er en tilfeldig utvalgt person fra en befolkning, så er det alderen $T(\omega)$ som er relevant for å finne aldersfordelingen i befolkningen. Vi velger å diskutere et annet eksempel mer i detalj.

Eksempel 6.1 (Antall kron i myntkast med n mynter) Et enkelt myntspill er gitt ved at en person velger mynt eller krone for deretter å kaste n mynter. Dersom personen valgte krone, så vinner han alle myntene med kronesiden opp, og tilsvarende for mynt alternativet. La K være antall kron i myntkastet. Utfallsrommet $\Omega_K = \{0, 1, 2, \ldots, n\}$ gir de mulige verdiene til K. Vi vil nå vise hvordan fordelingen P_K til dette eksperimentet er gitt ved

symmetriantagelser i et tenkt underliggende eksperiment. Anta at myntene er nummerert. Da er utfallsrommet Ω mengden av de 2^n enkeltutfall som fremkommer ved at mynt 1 kan ta verdien mynt eller krone, mynt 2 kan ta verdien mynt eller krone, osv. Dersom myntene er identiske og rettferdige, så følger det at hvert enkeltutfall i dette underliggende eksperimentet er like sannsynlige. Dette gir da $P\{\omega\} = 2^{-n}$. Dersom resultatet av myntkastet er ω, så er antall kron k en funksjon K av ω, dvs $k = K(\omega)$. Antall ω som gir $k = K(\omega)$ er lik antall ikke-ordnede utvalg av størrelse k fra $\{1, \ldots, n\}$, dvs lik $\binom{n}{k}$. Direkte argument: 1.kroneside kan plasseres på n plasser, 2.kroneside kan plasseres på $n-1$ plasser osv gir $n \cdot (n-1) \cdots (n-k+1)$ muligheter, men rekkefølgen vi plasserer kronesidene i endrer ikke ω, så vi må dividere med $k!$. Dette gir

$$P_K\{k\} = P\{\omega \,|\, K(\omega) = k\} = \sum_{\substack{\omega \\ K(\omega)=k}} P\{\omega\} = \binom{n}{k} 2^{-n}.$$

Dermed har vi funnet fordelingen P_K til K. Konklusjonen er at K er binomialfordelt med parameter $p = 1/2$. Denne utledningen har vi gjort før. Det nye her er at utledningen tar utgangspunkt i funksjonen K.



I det foregående eksemplet så vi at vi kunne finne fordelingen til en egenskap ved utfallet i et eksperiment. Dette kan gjøres helt generelt. Når T er en funksjon $T : \Omega \to \Omega_T$, og A er en delmengde av Ω_T, så defineres

$$\boxed{(T \in A) = \{\omega \,|\, T(\omega) \in A\}}$$

I konvensjonell matematisk notasjon er $(T \in A) = T^{-1}(A)$. Notasjonen $(T \in A)$ har imidlertid den fordelen at den har naturlige generaliseringer til tilfeller med flere betingelser. Et eksempel er gitt ved

$$(T \in A, (U \in B \text{ eller } V \in C))$$

som er definert lik

$$\{\omega \,|\, T(\omega) \in A, (U(\omega) \in B \text{ eller } V(\omega) \in C)\}.$$

I konvensjonell matematisk notasjon er overnevnte mengde gitt ved det noe mindre leselige utrykket

$$(T^{-1}(A)) \cap ((U^{-1}(B)) \cup (V^{-1}(C))).$$

La Ω og Ω_T være utfallsrom utstyrt med hver sin hendelsesfamilie.

Definisjon 6.1 (*Observator*) En funksjon $T : \Omega \to \Omega_T$ slik at $(T \in A)$ er en hendelse i Ω for enhver hendelse A i Ω_T. Fordelingen P_T til observatoren T er

$$\mathrm{P}_T(A) = \mathrm{P}(T \in A) \tag{6.1}$$

Det er vanlig å la være å spesifisere Ω. Formuleringen "La S, T, U være observatorer" skal forstås som at det er gitt et underliggende utfallsrom slik at S, T, U alle er observatorer for dette utfallsrommet. Vi vil konvensjonelt kalle dette utfallsrommet for Ω. Mengdene Ω_S, Ω_T, Ω_U der S, T, U tar sine verdier vil vanligvis spesifiseres. La P være en fordeling for Ω og la T være en observator for Ω med verdier i Ω_T. Fordelingen P_T til observatoren T er definert ved

$$\boxed{P_T(A) = P(T \in A)}$$

Vi beviser at P_T er en fordeling for Ω_T:

1. Fordi T er en observator, så er $(T \in A)$ en hendelse for hver hendelse A. Dermed er $P_T(A) = P(T \in A) \geq 0$ veldefinert for alle hendelser A i Ω_T.

2. Fordelingen P_T er normalisert fordi $P_T(\Omega_T) = P(T \in \Omega_T) = 1$ pga $(T \in \Omega_T) = \Omega$.

3. Identiteten
$$(T \in \uplus_i A_i) = \uplus_i (T \in A_i)$$
gir $P_T(\uplus_i A_i) = P(\uplus_i (T \in A_i)) = \sum_i P_T A_i$. Dette viser at P_T oppfyller addisjonsloven.

En observator T er *enkel* dersom den tar et endelig antall verdier $t_1, \ldots, t_n$. Fordelingen til T er da gitt ved den diskrete tettheten

$$f_T(t) = P_T\{t\} = P(T = t) = P\{\omega \,|\, T(\omega) = t\}, \;\; t = t_1, \ldots, t_n.$$

Dette resultatet har da en naturlig generalisering.

Fordelingen til en diskret observator er diskret.

En *diskret observator* er en observator som tar et tellbart antall verdier.

6.2 Tilfeldige variabler

Tallinjen $\mathbb{R}$ er antagelig det utfallsrommet som benyttes oftest, direkte eller som en byggekloss. Dersom ikke annet er sagt så er det alltid underforstått at tallinjen er utstyrt med den minste hendelsesfamilien som inneholder alle intervall. La det være underforstått at et underliggende sannsynlighetsrom Ω er gitt.

Definisjon 6.2 (*Tilfeldig variabel*) Observator X med utfallsrom $\Omega_X = \mathbb{R}$

Dersom Ω er det underliggende utfallsrommet, så betyr dette spesielt at en tilfeldig variabel X er en funksjon $X : \Omega \to \mathbb{R}$.

Antagelsene i definisjonen sikrer også at sannsynligheten

$$P(a \leq X \leq b) = P_X[a, b]$$

er veldefinert for alle intervall $[a, b]$. Dette følger fordi $(a \leq X \leq b) = (X \in [a, b])$ er en hendelse i Ω fordi intervallet $[a, b]$ er en hendelse i $\mathbb{R}$. Mer generelt har vi følgende resultat.

Fordelingen til en tilfeldig variabel er en fordeling for tallinjen.

La Q være en fordeling for $\mathbb{R}$. Fordelingsfunksjonen G til Q er definert ved $G(x) = Q(-\infty, x]$. Fordelingsfunksjonen har følgende egenskaper:

1. For $x \in \mathbb{R}$ er $0 \leq G(x) \leq 1$.

2. $\lim_{x \to -\infty} G(x) = 0$ og $\lim_{x \to \infty} G(x) = 1$.

3. G er monotont voksende.

4. G er kontinuerlig fra høyre.

En fordelingsfunksjon for tallinjen er definert til å være en funksjon som har egenskapene over. Det kan bevises at dersom G er en fordelingsfunksjon, så finnes en unik fordeling Q for tallinjen som har G som fordelingsfunksjon. Utgangspunktet for beviset er sammenhengen

$$Q(a, b] = G(b) - G(a).$$

Beviset finnes i mer videregående bøker i sannsynlighetsteori eller generell integrasjonsteori. Konklusjonen er verdt å utheve.

Det er 1-1 korrespondanse mellom mengden av fordelinger og mengden av fordelingsfunksjoner for tallinjen.

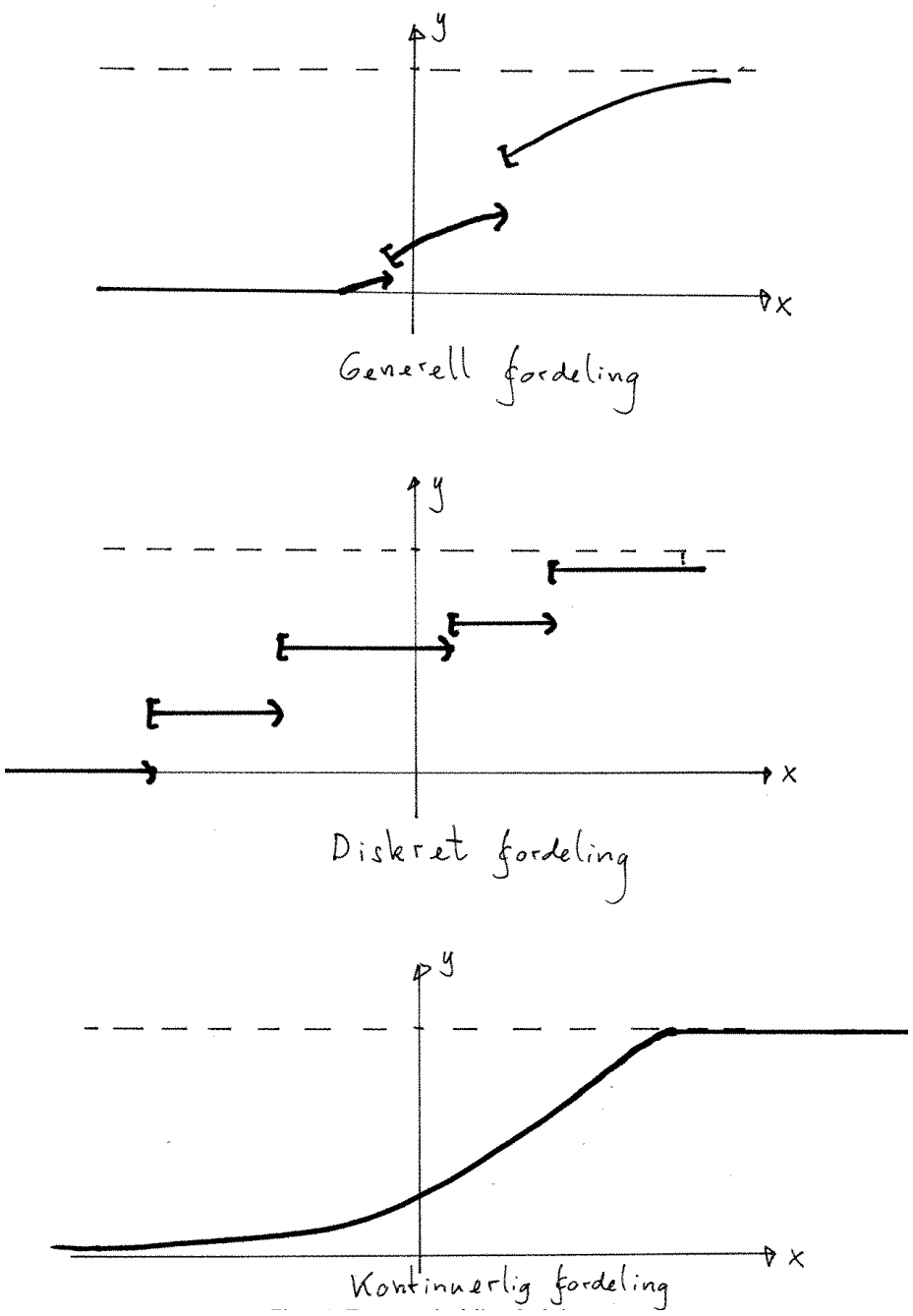
Et hovedpoeng med innføringen av fordelingsfunksjoner er at det typisk er lettere å spesifisere en fordeling ved å angi fordelingsfunksjonen fremfor å definere fordelingen direkte. Dette var også poenget med innføringen av tettheter. Fordelen med fordelingsfunksjonene er at disse er definert for enhver fordeling for tallinjen. På figur 6.1 er det skissert tre mulige fordelingsfunksjoner. Fordelingsfunksjonen F_X og tettheten f_X til en tilfeldig variabel X er ved definisjon lik hvv fordelingsfunksjonen og tettheten til fordelingen P_X til X. Dette betyr at

$$F_X(x) = P(X \leq x).$$

Dersom X har en kontinuerlig fordeling, så gir analysens fundamentalteorem at

$$\boxed{f_X = F_X'.}$$

Vi skal se at dette resultatet er meget anvendelig for beregning av tettheter til tilfeldige variable.



Figur 6.1: Tre typer fordelingsfunksjoner.

6.3 Tilfeldige vektorer og uavhengighet

La X, Y være tilfeldige variable. Da kan vi definere en observator $T = (X, Y)$ ved $T(\omega) = (X(\omega), Y(\omega))$. For hvert utfall ω er da $T(\omega)$ en vektor i vektorrommet $\mathbb{R}^2$, dvs et punkt i planet. Dette har en naturlig generalisering. En *tilfeldig vektor* er en observator som tar verdier i et vektorrom. Spesielt betyr dette at $T(\omega)$ er en vektor for alle utfall ω i Ω. Hendelsesfamilien i vektorrommet skal være valgt slik at dersom S og T er to tilfeldige vektorer med verdier i samme vektorrom, så definerer lineærkombinasjonen $\alpha S(\omega) + \beta T(\omega)$ en ny tilfeldig vektor $U = \alpha S + \beta T$. Dersom ikke annet er sagt, så er hendelsesfamilien til $\mathbb{R}^2$ gitt som den minste hendelsesfamilien som inneholder alle rektangler.

La $T_1, \ldots, T_n$ være observatorer. Da kan vi definere en observator $T = (T_1, \ldots, T_n)$ ved $T(\omega) = (T_1(\omega), \ldots, T_n(\omega))$. *Simultanfordelingen* til observatorene $T_1, \ldots, T_n$ er ved definisjon lik fordelingen til observatoren $T = (T_1, \ldots, T_n)$. Spesielt er simultanfordelingen til to tilfeldige variable X, Y gitt ved

$$P_{X,Y}(A \times B) = P(X \in A, Y \in B).$$

Produktet $A \times B$ av to mengder er definert ved

$$A \times B = \{(a, b) \,|\, a \in A, b \in B\}.$$

Spesielt er $A \times B$ et rektangel dersom A og B er intervall. Dette er illustrert i figur 6.2. Den simultane fordelingsfunksjonen $F_{X,Y}$ til X, Y er definert ved

$$F_{X,Y}(x, y) = P_{X,Y}((-\infty, x] \times (-\infty, y]) = P(X \leq x, Y \leq y).$$

La h være en funksjon definert på mengden

$$A = \{(x, y) \,|\, a \leq x \leq b,\ f(x) \leq y \leq g(x)\}$$

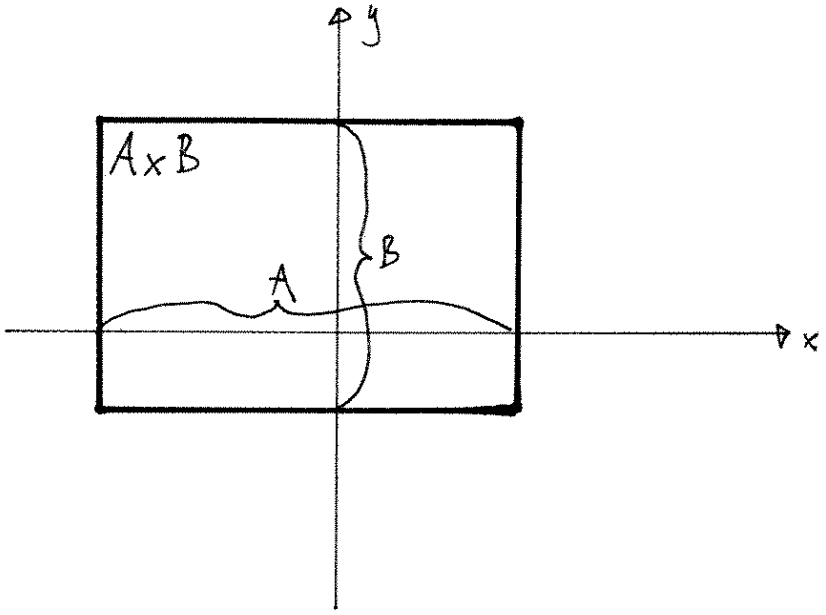
i planet. En slik mengde kalles y-enkel fordi enhver linje paralell med y-aksen snitter mengden i et intervall. Et eksempel er vist i figur 6.3. Integralet over en y-enkel mengde kan beregnes ved itterert integrasjon

$$\iint_A h(x, y)dxdy = \int_{x=a}^{b} [\int_{y=f(x)}^{g(x)} h(x, y)dy]dx.$$

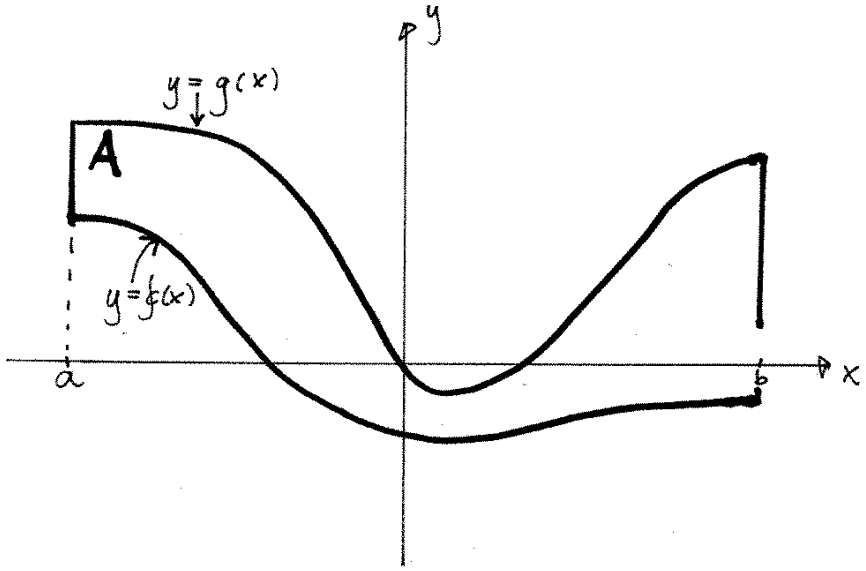
Noen mengder er både x-enkle og y-enkle og dette gir en verdifull frihet i valg av integrasjonsrekkefølge. Mer generelt kan integralet over mengder i planet beregnes ved å splitte opp mengden i mengder som er enkle.

To tilfeldige variable X, Y er *simultant kontinuerlig fordelt* dersom (X, Y) er kontinuerlig fordelt, dvs dersom det finnes en $f_{X,Y}$ slik at

$$P_{X,Y}(A) = \iint_A f_{X,Y}(x, y)dxdy.$$



Figur 6.2: Produktet av to intervall er et rektangel.



Figur 6.3: En y-enkel mengde.

Funksjonen $f_{X,Y}$ er sannsynlighetstettheten til (X, Y), og simultantettheten til X, Y. Det følger at

$$f_{X,Y} = \frac{\partial^2}{\partial x \partial y} F_{X,Y}.$$

Dersom vi kjenner fordelingen til (X, Y) kan vi finne fordelingen til X og til Y. Fordelingen til X er gitt ved

$$P_X(A) = P(X \in A, Y \in \mathbb{R}) = P_{X,Y}(A \times \mathbb{R}).$$

Fordelingen til Y finnes tilsvarende. I denne sammenhengen kalles fordelingen P_X for marginalfordelingen til X.

Dersom X, Y er simultant kontinuerlig fordelt, så er X og Y kontinuerlig fordelt. Sannsynlighetstettheten til X, den marginale tettheten til X, er gitt ved

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x, y) dy$$

Dette bevises ved derivasjon av

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \in \mathbb{R}) = \int_{t=-\infty}^{x} \int_{y=-\infty}^{\infty} f_{X,Y}(t, y) dy dt.$$

Tettheten til Y finnes tilsvarende.

Dersom $X_1, \ldots, X_n$ er tilfeldige variable, så er $T = (X_1, \ldots, X_n)$ en tilfeldig vektor i $\mathbb{R}^n$. Simultanfordelingen til $X_1, \ldots, X_n$ er $P_{X_1,\ldots,X_n} = P_T$, dvs lik fordelingen til $(X_1, \ldots, X_n)$. De foregående definisjonene for X, Y generaliseres tilsvarende.

Observatorene T_1 og T_2 er uavhengige dersom $P(T_1 \in A_1, T_2 \in A_2) = P(T_1 \in A_1)P(T_2 \in A_2)$ gjelder for alle hendelser A_1, A_2. Dette kan generaliseres. En familie $\{T_j\}$ av observatorer er en familie av *uavhengige observatorer* dersom produktregelen

$$P(T_{j_1} \in A_1, \ldots, T_{j_n} \in A_n) = P(T_{j_1} \in A_1) \cdots P(T_{j_n} \in A_n)$$

gjelder for alle distinkte $j_1, \ldots, j_n$. Følgende konsekvens er spesielt viktig:

Dersom de tilfeldige variablene $X_1, \ldots, X_n$ er uavhengige med tettheter f_{X_i}, så er

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

Vi har tidligere definert uavhengighet for hendelser. Det kan vises at familien $\{A_j\}$ av hendelser er uavhengig hvis og bare hvis familien $\{1_{A_j}\}$ av indikatorvariable er uavhengig.

Vi konkluderer med en definisjon som vil bli brukt mye i det etterfølgende. Observatorene $T_1, \ldots, T_n$ er et *tilfeldig utvalg* av størrelse n fra fordelingen Q dersom de er uavhengige og $P_{T_1} = \cdots = P_{T_n} = Q$. Dersom Q har en tetthet f, så vil vi si at $T_1, \ldots, T_n$ er et tilfeldig utvalg fra tettheten f.

6.4 Oppgaver

6.1. Finn fordelingen P_Y til observatoren Y for utfallsrommet $\Omega = \{(1,2,0),(2,1,3),(4,1,1)\}$ når $Y(x,y,z) = x+y+z$. Er Y en tilfeldig variabel?

6.2. Bevis at fordelingen P_T til en observator T er en sannsynlighetsfordeling.

6.3 (**T**). Bevis at fordelingen til en observator som tar et tellbart antall verdier er diskret. Må utfallsrommet til observatoren være diskret?

6.4. La $T : \Omega \to \Omega_T$. Vis at (i) $(T \in A^c) = (T \in A)^c$ og (ii) $(T \in \cup_i A_i) = \cup_i (T \in A_i)$.

Dette beviser at familien av mengder A slik at $(T \in A)$ er en hendelse er en hendelsesfamilie i Ω_T.

6.5. Bevis at $(T \in \uplus_i A_i) = \uplus_i (T \in A_i)$.

6.6. Vis at fordelingen til en observator som tar et tellbart antall verdier er diskret.

6.7. Hvorfor er $[0,1] \cup [10,15]$ en hendelse i utfallsrommet $\mathbb{R}$?

6.8. Anta at 5 personer, inkludert deg selv og en venn, stiller seg tilfeldig opp på en rekke. La Y være antall personer mellom deg og vennen. Finn tettheten til den tilfeldige variabelen Y.

6.9. La Y være antall ess en pokerspiller mottar blant de fem første kortene han blir tildelt. Tegn opp fordelingsfunksjonen til Y.

6.10. Hva er en tilfeldig variabel X? Hva betyr det at F_X er fordelingsfunksjonen til X?

6.11. Tegn grafen til en fordelingsfunksjon hvor den tilhørende fordelingen hverken er kontinuerlig eller diskret.

6.12. Prøv å bevise at fordelingsfunksjonen F_X til en tilfeldig variabel har egenskapene

1. For $x \in \mathbb{R}$ er $0 \leq F_X(x) \leq 1$.
2. $\lim_{x\to-\infty} F_X(x) = 0$ og $\lim_{x\to\infty} F_X(x) = 1$.
3. F_X er monotont voksende.
4. F_X er kontinuerlig fra høyre.

6.13. En tilfeldig variabel Y har en fordelingsfunksjon gitt ved

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Finn tettheten f_Y til Y. Beregn sannsynligheten $P(\frac{1}{2} < Y \leq \frac{3}{4})$.

6.14. Hva er en tilfeldig vektor? Vi skal senere definere hva som menes med et konfidensintervall J. Spesielt er et konfidensintervall et tilfeldig intervall. Prøv å gi en definisjon av hva som menes med at J er et tilfeldig intervall.

6.15. Hva menes med en x-enkel mengde? Tegn en x-enkel mengde i planet. Hva tror du er definisjonen av en z-enkel mengde i rommet?

6.16. La X være en tilfeldig variabel med kontinuerlig fordeling. Definer $Y = 2X$. Er da X, Y simultant kontinuerlig fordelt?

6.17. La en diskret tetthet være definert ved

$$f_{X,Y}(x, y) = cxy,\ (x, y) \in \{(1, 1), (2, 1), (2, 2), (3, 1)\}.$$

Finn konstanten c. Finn den marginale tettheten til X.

6.18. La X og Y være hhv antall konger og antall damer som trekkes når fire kort trekkes fra en kortstokk. Finn tettheten $f_{X,Y}$.

6.19. Se på eksperimentet gitt ved at en mynt og en terning kastes samtidig. La X være antall kronesider som vises og la Y være antall øyne som vises på terningen. Definer utfallsrommet Ω tilsvarende dette eksperimentet. Hvordan er funksjonene X og Y definert når disse skal være observatorer for Ω? Finn $F_{X,Y}(1, 2)$ og gi en tolkning av dette tallet. Hvordan vil grafen til $F_{X,Y}$ se ut?

6.20. Finn f_X og f_Y når

$$f_{X,Y} = \frac{1}{x},\ 0 < y < x,\ 0 < x < 1.$$

6.21. Vis at dersom

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n),$$

så er $X_1, \ldots, X_n$ uavhengige.

6.22. Anta at levetiden X (i timer) til en lyspære har en tetthet

$$f_X(x) = \frac{1}{1000} e^{-\frac{x}{1000}}.$$

Finn simultantettheten f_{X_1,X_2,X_3,X_4} tilsvarende levetidene til fire lyspærer innstallert i et kontor. Hva er sannsynligheten for at alle fire fortsatt lyser etter 1050 timer?

6.23. Hva betyr det at $X_1, \ldots, X_m$ er et tilfeldig utvalg fra binomialfordelingen med parametre $p = 1/2$ og $n = 10$? Finn simultantettheten til $X_1, \ldots, X_m$.

6.24. Finn konstanten k når

$$f_{X,Y}(x, y) = k,\ 0 \leq x, y \leq 1,\ x + y \leq 1$$

er simultantettheten til X, Y. Gi et geometrisk argument som viser at X og Y ikke er uavhengige.

Kapittel 7

Forventningsverdi

The relative frequency of the repetition is the "measure" of probability, just as the length of a column of mercury is the "measure" of temperature.

R. von Mises (1950)

La $x_1, \ldots, x_n$ være resultatet vi får ved å gjenta et statistisk forsøk n ganger. Selv om resulatet x_i varierer fra gang til gang, så vil noen typer eksperiment være statistisk lovmessige ved at middelverdien

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

konvergerer mot et tall μ når n vokser. Dette tallet μ er eksperimentets empiriske forventningsverdi.

Anta at eksperimentet over kan modeleres ved et tilfeldig utvalg $X_1, \ldots, X_n$ fra en fordeling P. Det kan da bevises at

$$\bar{X}(\omega) = \frac{X_1(\omega) + \cdots + X_n(\omega)}{n}$$

konvergerer mot forventningsverdien $\mathrm{E}(X)$ når n vokser. Dette konvergensteoremet kalles store talls lov. I det følgende vil forventningsverdien $\mathrm{E}(X)$ defineres og drøftes. Spesielt vil det framgå at $\mathrm{E}(X)$ ikke avhenger av ω. Ved identifiseringen $X_i(\omega) = x_i$ følger det derfor at store talls lov gir en forklaring på den observerte lovmessigheten til $\bar{x}$ ved at $\mu = \mathrm{E}(X)$.

Dersom vi lar $Y_i = 1_A(X_i)$, så gir store talls lov at det relative antall ganger ($= \bar{y}$) hendelsen ($X_i \in A$) inntreffer konvergerer mot sannsynligheten til denne hendelsen ($= \mathrm{E}(1_A(X)) = P(X \in A)$). Store talls lov kan dermed sees som en forklaring på R. von Mises tolkning av den observerte relative frekvensen som en måling av sannsynligheten til en hendelse.

7.1 Definisjon av forventningsverdi og substitusjonsformelen

La X være en tilfeldig vektor for utfallsrommet Ω utstyrt med fordelingen P. Spesielt kan X være en tilfeldig variabel.

Definisjon 7.1 (*Forventningsverdien*) $\mathrm{E}(X)$ til X er definert ved integralet

$$\mathrm{E}(X) = \int X(\omega)\, \mathrm{P}(d\omega) \tag{7.1}$$

Notasjonen $\mathrm{E}(X)$ er mao en kort skrivemåte for integralet av X med hensyn på fordelingen for utfallsrommet. Dette integralet defineres og drøftes i mer detalj i mer avanserte innføringer (Rudin, 1987). I det følgende vil vi nøye oss med å indikere hvordan det generelle integralet over er definert.

Anta at X er en enkel stokastisk vektor med verdier $x_1, \ldots, x_n$. Da er *forventningsverdien* og *integralet* definert ved

$$\mathrm{E}(X) = \int X(\omega)\, \mathrm{P}(d\omega) = \sum_i x_i f(x_i) \tag{7.2}$$

hvor $f(x_i) = \mathrm{P}(X = x_i)$ er den diskrete tettheten til X. Dermed har vi gitt mening til forventningsverdien til en enkel tilfeldig vektor. Spesielt følger det at

$$\boxed{P(A) = \mathrm{E}(1_A)}$$

Generelt kan integralet defineres ved

$$\int X(\omega) P(d\omega) = \lim \int X_i(\omega) P(d\omega),$$

hvor $X_1, X_2, \ldots$ er en følge enkle tilfeldige vektorer konstruert fra X som konvergerer mot X. Fordi hver X_i er enkel er dermed det generelle integralet definert som en grense av endelige summer.

For beregning av forventningsverdier er det spesielt to tilfeller som er viktige. Dersom P er kontinuerlig med en tetthet f, så kan det bevises at

$$\mathrm{E}[X] = \int X(\omega) f(\omega) d\omega.$$

Dersom P er diskret med tetthet f, så kan det bevises at

$$\mathrm{E}[X] = \sum X(\omega) f(\omega).$$

Disse to tilfellene viser at den generelle definisjonen av forventningsverdi generaliserer både summasjon og Riemann-integralet.

Metoden med substitusjon for beregning av Riemann-integralet har en generalisering. La P være en fordeling for utfallsrommet Ω, og la T være en observator for Ω med verdier i utfallsrommet Ω_T. La $Y = \phi(T)$ være en stokastisk vektor gitt av T og en funksjon $\phi : \Omega_T \to \Omega_Y$. Det kan da bevises at følgende formel for variabelskifte holder generelt

$$\boxed{\mathrm{E}(Y) = \int \phi(T(\omega)) P(d\omega) = \int \phi(t) P_T(dt).}$$

Et bevis av substitusjonformelen for tilfellet at P er diskret er gitt ved

$$\sum_{\omega} \phi(T(\omega))P\{\omega\} = \sum_{t} \sum_{T(\omega)=t} \phi(T(\omega))P\{\omega\} = \sum_{t} \phi(t) \sum_{T(\omega)=t} P\{\omega\} = \sum_{t} \phi(t)P(T=t).$$

Spesielt har vi benyttet at dersom P er diskret, så er P_T diskret med tetthet

$$f_T(t) = P_T\{t\} = P(T=t) = \sum_{T(\omega)=t} P\{\omega\}.$$

Konsekvensen av det generelle tilfellet er meget sentral i sannsynlighetsteorien og verdt å understreke.

Forventningsverdien til en funksjon av en observator er lik integralet av funksjonen mhp fordelingen til observatoren.

Tilfellet $T = X$ og $\phi(X) = X$ gir spesielt

$$\mathrm{E}\, X = \int x P_X(dx)$$

Formelen for substitusjon betyr at det i mange sammenhenger er unødvendig å spesifisere det underliggende sannsynlighetsrommet Ω. Det er tilstrekkelig å spesifisere fordelingen til T. Derfor er det lett å glemme hva en observator er. Vi repeterer:

En observator T er en funksjon $T : \Omega \to \Omega_T$ slik at $(T \in A)$ er en hendelse for enhver hendelse A.

Fordelingen P_T er kun én av egenskapene til T. Det er f eks umulig å avgjøre om T er uavhengig av en gitt hendelse A i Ω dersom vi bare kjenner fordelingen til T. Et annen demonstrasjon er gitt ved den sedvanlige antagelsen om at $X_1, \ldots, X_n$ er et tilfeldig utvalg. Antagelsen innebærer at alle X_i'ene har lik fordeling, men antagelsen innebærer slett ikke at variablene er like.

Vi innfører notasjonen

$$\mathrm{E}_T(\phi) = \int \phi(t) P_T(dt).$$

Her er integralet på høyresiden veldefinert fordi P_T er en fordeling. Her er mao E_T definert fra P_T på samme måte som E er definert fra P. *Substitusjonsformelen* kan da formuleres uten integraltegn.

Substitusjonsformelen: $\mathrm{E}(\phi(T)) = \mathrm{E}_T(\phi)$.

7.2 Forventningsverdier av summer og produkt

En meget viktig egenskap til forventningsverdier er at de kan tas ledd for ledd, dvs

$$\boxed{\mathrm{E}(aX + bY) = a\,\mathrm{E}\,X + b\,\mathrm{E}\,Y}$$

Egenskapen er så viktig at den fortjener et eget navn. Vi sier at operatoren E er lineær. Navnet operator forklares ved at E "opererer" på en tilfeldig vektor X, og gir en vektor $\mathrm{E}\,X$. Regneregelen over er som om vi multipliserer med E fra venstre. Dette er en av grunnene til at vi tillater oss å skrive $\mathrm{E}\,X$ fremfor å benytte funksjonsnotasjonen $\mathrm{E}(X)$.

I det overstående er det underforstått at $Z = aX + bY$ er definert ved $Z(\omega) = aX(\omega) + bY(\omega)$. For at $\mathrm{E}(aX + bY)$ skal gi mening, må det derfor bevises at Z er en tilfeldig vektor. Vi utelater beviset her.

Anta at P er diskret. Da kan vi skissere beviset av at E er lineær:

$$\mathrm{E}(aX+bY) = \sum(aX(\omega)+bY(\omega))P\{\omega\} = a\sum X(\omega)P\{\omega\}+b\sum Y(\omega))P\{\omega\} = a\,\mathrm{E}\,X+b\,\mathrm{E}\,Y.$$

En annen viktig egenskap er knyttet til uavhengighet.

Dersom X og Y er uavhengige, så gjelder produktregelen: $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

La $T = (X, Y)$ og $\phi(t) = \phi(x, y) = xy$. Uavhengighet gir $P_{X,Y}(A \times B) = P_X(A)P_Y(B)$ og en skisse av beviset av produktregelen er gitt ved

$$\begin{aligned}\mathrm{E}(XY) = \mathrm{E}\,\phi(T) = \mathrm{E}_T\,\phi = \int \phi(t)P_T(dt) = \int \phi(x,y)P_{X,Y}(dxdy)\\ = \int y[\int xP_X(dx)]P_Y(dy) = \int y\ \mathrm{E}\,X\ P_Y(dy) = \mathrm{E}\,X\ \ \mathrm{E}\,Y.\end{aligned}$$

La X være en stokastisk variabel.

Definisjon 7.2 (*Variansen*) $\mathrm{Var}(X)$ er definert ved

$$\mathrm{Var}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2.$$

Notasjonskonvensjonene $\mu_X = \mathrm{E}\,X$ og $\sigma_X = \sqrt{\mathrm{Var}\,X}$ er vanlige. Størrelsen σ_X er *standardavviket* til X.

Alternativt kan variansen beregnes ved formelen

$$\boxed{\mathrm{Var}\,X = \mathrm{E}\,X^2 - \mu_X^2.}$$

Beviset er gitt ved

$$\mathrm{Var}\,X = \mathrm{E}(X - \mu_X)^2 = \mathrm{E}(X^2 - 2X\mu_X + \mu_X^2) = \mathrm{E}\,X^2 - 2\,\mathrm{E}\,X\mu_X + \mu_X^2 = \mathrm{E}\,X^2 - \mu_X^2.$$

Generelt gjelder regneregelen

$$\boxed{\operatorname{Var} aX = a^2 \operatorname{Var} X.}$$

Dersom X og Y er uavhengige, så gir produktregelen at

$$\boxed{\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y.}$$

Vi overlater til leseren å bevise de to sistnevnte regnereglene.

Eksempel 7.1 (Middelverdi) La X være enkel med verdier $x_1, \ldots, x_n$ og $P(X = x_i) = 1/n$ (uniform fordeling). Da er forventningsverdien μ_X lik middelverdien $\overline{x}$:

$$\mu_X = \operatorname{E} X = \sum_{i=1}^{n} x_i P(X = x_i) = \frac{x_1 + \cdots + x_n}{n} = \overline{x}$$

Det fremgår av dette at forventningsverdi kan sees som en generalisering av middelverdi. Variansen beregnes tilsvarende

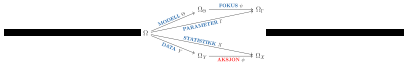
$$\sigma_X^2 = \operatorname{Var} X = \operatorname{E} X^2 - \mu_X^2 = \frac{x_1^2 + \cdots + x_n^2}{n} - \mu_X^2 = \overline{x^2} - \overline{x}^2$$

Dersom definisjonen benyttes direkte finnes

$$\operatorname{Var} X = \operatorname{E}(X - \mu_X)^2 = \frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} = \overline{(x - \overline{x})^2}$$

For vilkårlige $x_1, \ldots, x_n$ har vi altså at

$$\overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2.$$



7.3 Geometrisk tolkning av forventningsverdi og varians

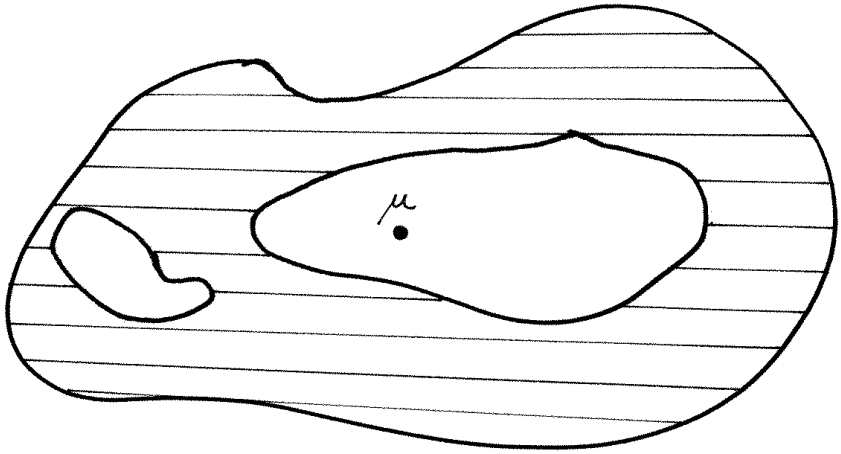
La ρ være en massetetthet i planet. Tyngdepunktet til et legeme beskrevet av denne massetettheten er definert ved

$$x_t = \frac{\int\int (x_1, x_2)\rho(x_1, x_2) dx_1 dx_2}{\int\int \rho(x_1, x_2) dx_1 dx_2}$$

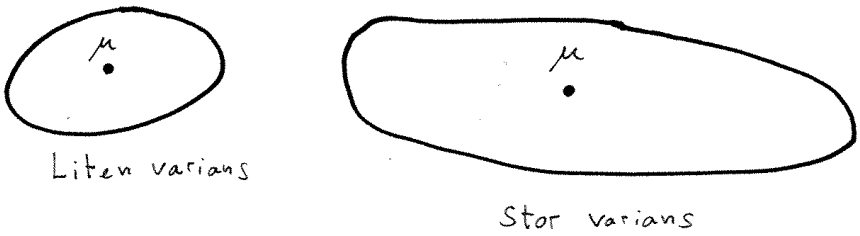
Geometrisk sier tyngdepunktet spesielt noe om hvor legemet er som på figur 7.1 For en sannsynlighetstetthet er totalsannsynligheten (jf totalmassen) lik 1. Analogien med massetetthet gir at forventningsverdien $\operatorname{E} X$ kan sees på som tyngdepunktet til fordelingen P_X.

Tilsvarende kan variansen $\operatorname{Var} X$ sees på som treghetsmomentet til fordelingen P_X om tyngdepunktet. Som på figur 7.2 følger det at variansen $\operatorname{Var} X$ er et mål på spredningen til fordelingen P_X.

Merk: $P(X = \mu_X) = 0$ er mulig, også når X er diskret fordelt. Dette kan sammenlignes med at tyngdepunktet til et legeme godt kan ligge i et hull i legemet som på figur 7.1.



Figur 7.1: Tyngdepunkt til en fordeling.



Figur 7.2: Variansen = Treghetsmoment om tyngdepunktet til en fordeling.

7.4 Beregning av forventningsverdi og varians

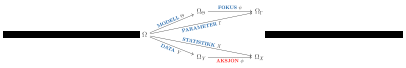
Følgende eksempler viser at beregning av forventningsverdier kan gjøres på andre måter enn ved direkte integrasjon eller summasjon.

Eksempel 7.2 (Forventningsverdi og varians til en Bernoulli variabel) La Z være *Bernoullifordelt* med parameter p, dvs $P(Z = 1) = p$ og $P(Z = 0) = (1 - p)$. Da er

$$\mathrm{E}\, Z = \sum_{z=0}^{1} zP(Z = z) = p$$

$$\mathrm{Var}\, Z = \sum_{z=0}^{1} z^2 P(Z = z) - p^2 = p(1 - p)$$

Det første integralet kunne beregnes direkte ved definisjonen fordi Z er en enkel variabel. Det andre integralet ble beregnet ved substitusjonsformelen, dvs variabelen ω ble erstattet av variabelen $z = Z(\omega)$.



Eksempel 7.3 (Forventningsverdi og varians beregnet ved variabelskifte) La X være binomialfordelt med parametre (n, p), dvs

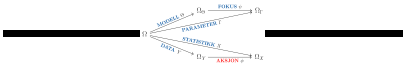
$$f_X(x) = \binom{n}{x} p^x (1 - p)^{n-x}, x = 0, \ldots, n.$$

Definer variabelen $Y = Z_1 + \cdots + Z_n$, hvor $Z_1, \ldots, Z_n$ er et tilfeldig utvalg fra en Bernoullifordeling P_Z med parameter p. Det kan da bevises at Y er binomialfordelt med parametre (n, p). Dette gir

$$\mathrm{E}\, X = \mathrm{E}\, Y = \mathrm{E}(Z_1 + \cdots + Z_n) = n\, \mathrm{E}\, Z = np$$

$$\mathrm{Var}\, X = \mathrm{Var}\, Y = \mathrm{Var}(Z_1 + \cdots + Z_n) = n\, \mathrm{Var}\, Z = np(1 - p)$$

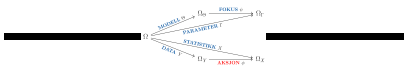
Poenget her er igjen substitusjonsformelen som gir at forventningsverdien til $\phi(X)$ er lik forventningsverdien til $\phi(Y)$ når X og Y har lik fordeling.



Eksempel 7.4 (Forventningsverdi beregnet ved normaliseringen) La X være binomialfordelt med parametre (n, p). Da er

$$\begin{aligned}\mathrm{E}\, X &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1 - p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1 - p)^{n-x} = pn \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1 - p)^{n-x} \\ &= pn \sum_{k=0}^{n} \binom{n-1}{k} p^k (1 - p)^{n-x} = pn.\end{aligned}$$

Den siste likheten følger fordi sannsynlightestettheten til en binomialfordelt variabel med parametre $(n-1,p)$ er normalisert. Dette normaliseringstrikset kan også benyttes ved beregning av enkelte integral.



Den *momentgenererende funksjonen* M_X og den *karakteristiske funksjonen* ϕ_X er definert ved

$$M_X(t) = \mathrm{E}\, e^{tX} \quad , \quad \phi_X(t) = \mathrm{E}\, e^{itX}, i = \sqrt{-1}.$$

Den karakteristiske funksjonen ϕ_X eksisterer alltid, mens eksistens av den momentgenererende funksjonen M_X avhenger av fordelingen til X. All informasjon om fordelingen til X er inneholdt i hver av disse funksjonene.

Fordelingene til to tilfeldige variable er like hvis og bare hvis de karakteristiske funksjonene er like.

Dette gir en metode for beregning av fordelinger som vi skal se eksempler på senere. Derivasjon gir

$$M_X'(0) = \left(\frac{d}{dt}\,\mathrm{E}(e^{tX})\right)_{t=0} = \left(\mathrm{E}(Xe^{tX})\right)_{t=0} = \mathrm{E}\, X.$$

Ved å derivere n ganger finnes mer generelt

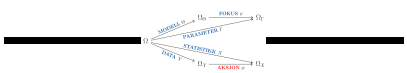
$$\boxed{\mathrm{E}\, X^n = M_X^{(n)}(0) \quad , \quad \phi_X^{(n)}(0) = \mathrm{E}(iX)^n.}$$

Eksempel 7.5 (Derivasjon av den momentgenererende funksjon) La X være binomialfordelt med parametre (n,p). Binomialteoremet gir

$$M_X(t) = \mathrm{E}\, e^{tX} = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(e^tp)^x(1-p)^{n-x} = (e^tp+1-p)^n$$

Dermed har vi enda en alternativ beregning av forventningsverdi og varians:

$$\mathrm{E}\, X = M_X'(0) = (n(e^tp+1-p)^{n-1}e^tp)_{t=0} = np, \;\; \mathrm{Var}\, X = M_X^{(2)}(0) - (np)^2 = \cdots = np(1-p)$$



7.5 Chebyshevs ulikhet

Chebyshevs ulikhet

$$\boxed{P(|X-\mu_X| \geq t\sigma_X) \leq 1/t^2}$$

gjelder generelt for enhver tilfeldig vektor X. Ulikheten gir at sannsynligheten for at utfallet i utfallsrommet Ω_X faller lenger enn t ganger standardavviket unna forventningsverdien er begrenset av $1/t^2$. Dette er dermed en presisering av i hvilken forstand standardavviket måler spredningen av utfallene.

Chebyshevs ulikhet bevises ved følgende beregning:

$$P(|X-\mu_X| \geq t\sigma_X) = \mathrm{E}\, 1_{(|X-\mu_X|\geq t\sigma_X)} \leq \mathrm{E}((\frac{X-\mu_X}{t\sigma_X})^2 1_{(|X-\mu_X|\geq t\sigma_X)}) \leq \mathrm{E}((\frac{X-\mu_X}{t\sigma_X})^2 = 1/t^2.$$

Ulikhetene over følger av at E er en monotont voksende operator, dvs

$$\boxed{Y \leq Z \;\;\Rightarrow\;\; \mathrm{E}\, Y \leq \mathrm{E}\, Z}.$$

Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra en fordeling P_X. Vi kan da beregne

$$\mathrm{E}\, \overline{X} = \frac{\mathrm{E}\, X_1 + \cdots + \mathrm{E}\, X_n}{n} = \mu_X$$
$$\mathrm{Var}\, \overline{X} = \frac{\mathrm{Var}\, X_1 + \cdots + \mathrm{Var}\, X_n}{n^2} = \frac{\sigma_X^2}{n}$$

Spesielt følger det at

$$\mathrm{Var}\, \overline{X} = \mathrm{E}(\overline{X} - \mu_X)^2 \to 0$$

når n vokser, dvs at middelverdien $\overline{X}$ i en viss forstand konvergerer mot forventningsverdien $\mathrm{E}\, X$. Denne typen konvergens kalles L^2 konvergens, og vi har bevist at

$$\boxed{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mathrm{E}\, X \;\; (L^2 - \text{ konvergens}).}$$

Dermed har vi bevist en svak versjon av store talls lov.

Innsetting av $t = \epsilon/\sigma$ gir en reformulering:

$$P(|X-\mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2}.$$

Reformuleringen av Chebyshevs ulikhet gir at vi også har en annen type konvergens

$$P(\left|\overline{X} - \mu_X\right| > \epsilon) \to 0.$$

Denne typen konvergens kalles konvergens i sannsynlighet. Konklusjonen er dermed at Chebyshevs ulikhet gir en svak versjon av store talls lov.

7.6 Oppgaver

7.1. Vis at $P(A) = \mathrm{E}\, 1_A$ ved å observere at 1_A er en enkel tilfeldig variabel.

7.2. En urne inneholder 5 røde og 4 hvite brikker. La X være antall røde brikker vi får når vi trekker 3 brikker med tilbakelegning. Finn forventningsverdien $\mathrm{E}\, X$ og variansen $\mathrm{Var}\, X$. Gi en tolkning av $\mathrm{E}\, X$. På hvilken måte er substitusjonsformelen benyttet i beregningene over?

7.3. La T være en enkel observator. Bevis substitusjonsformelen $\mathrm{E}\, \phi(T) = \mathrm{E}_T\, \phi$ ved å bruke definisjonen av integralet av en enkel tilfeldig vektor.

7.4. Anta at X er uniformt fordelt på intervallet $[a, b]$. Beregn $\mathrm{E}\, X$.

7.5. Vis at E er lineær ved å anta at P er kontinuerlig.

7.6. Ti terninger kastes. Finn forventningsverdien til summen av antall øyne.

7.7. La $X = 3X_1 + 4X_2 + 2X_3$ hvor

$$f_{X_1}(x) = \binom{6}{x}(1/3)^x(2/3)^{6-x} \quad x = 0, 1, \ldots, 6$$

$$f_{X_2}(x) = \frac{e^{-4}4^x}{x!} \quad x = 0, 1, 2, \ldots$$

$$f_{X_3}(x) = 5e^{-5x} \quad x > 0.$$

Finn $\mathrm{E}\, X$.

7.8. La X være antall kort som må deles ut før det kommer et ess. Finn $\mathrm{E}\, X$ og gi en tolkning av resultatet.

7.9. Anta at $f_{X,Y}(x, y) = e^{-x-y}$, $x, y > 0$. Finn $\mathrm{E}(X + Y)$.

7.10. Anta at $X_1, \ldots, X_n$ er uavhengige og kontinuerlig fordelte variable. Vis at $\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n)$. Vis at $\phi(X_1, X_2)$ og $\psi(X_3)$ er uavhengige tilfeldige variable.

7.11. La $f_X(x) = 2/x^3$, $x > 1$. Vis at X ikke har endelig varians.

7.12. La μ være en vektor. Definer en tilfeldig vektor X ved $X(\omega) = \mu$. Vis at X er uavhengig av enhver observator.

7.13. Anta at X og Y er uavhengige. Bevis regneregelen $\mathrm{Var}(aX + bY) = a^2\, \mathrm{Var}\, X + b^2\, \mathrm{Var}\, Y$.

7.14. La $\mathrm{E}\, X = \mu$ og la $\mathrm{Var}\, X = \sigma^2$. Vis at

$$\mathrm{E}\left(\frac{X - \mu}{\sigma}\right) = 0 \text{ og } \mathrm{Var}\left(\frac{X - \mu}{\sigma}\right) = 1.$$

7.15. La X være en enkel tilfeldig vektor med verdier $x_1, \ldots, x_n$ i rommet $\mathbb{R}^3$. Gi en tolkning av $\mathrm{E}\, X$ og $\mathrm{Var}\, X$ når fordelingen P_X tolkes som massefordelingen til et system av partikler med posisjoner $x_1, \ldots, x_n$.

7.16. La X_1, X_2 være uavhengige binomiske variable med parametre hhv (n_1, p_1) og (n_2, p_2). Finn $\mathrm{E}(4X + 6Y)$ og $\mathrm{Var}(4X + 6Y)$.

7.17. La X være uniformt fordelt på intervallet $[0, 2]$. Finn den momentgenererende funksjonen M_X. Beregn $\mathrm{E}\, X^r$ direkte og ved bruk av M_X for $r = 0, 1, \ldots$. Benytt resultatet sammen med binomialteoremet og regn ut $\mathrm{E}(X - \mathrm{E}\, X)^6$.

7.18. Anta at X er Poisson-fordelt, dvs

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \ldots$$

Finn M_X og benytt denne til å beregne $\mathrm{E}\, X$ og $\mathrm{Var}\, X$.

7.19. Anta at $f_X(x) = e^{-x}, x > 0$. Bruk Chebyshevs ulikhet til å finne en øvre skranke for sannsynligheten for at X tar en verdi mer enn 2 standardavvik unna forventningsverdien. Beregn denne sannsynligheten og kommenter resultatet.

Kapittel 8

Beregning av tettheter og betingede tettheter

This task would have been a rather hopeless one before the introduction of Lebesgue's theories of measure and integration. However, after Lebesgue's publication of his investigations, the analogies between measure of a set and probability of an event, and between integral of a function and mathematical expectation of a random variable, became apparent.

A.N. Kolmogorov (1933)

Beregningen av fordelingen til en observator for et eksperiment er et hovedproblem i sannsynlighetsteorien. I enkle tilfeller kan dette gjøres direkte ved definisjonen. Kombinatoriske argument gir typiske eksempel på dette. Generelt er dette imidlertid et vanskelig problem, og vi skal se på flere metoder. Teorien for betingede sannsynligheter kan utvides til en teori for betingede forventningsverdier. Vi skal se hvordan dette kan gjøres i noen tilfeller ved hjelp av betingede tettheter.

8.1 Funksjoner av observatorer

En observator S er *basert på* en observator T dersom S er en funksjon av T, dvs dersom det finnes en funksjon ϕ slik at

$$S(\omega) = \phi(T(\omega)).$$

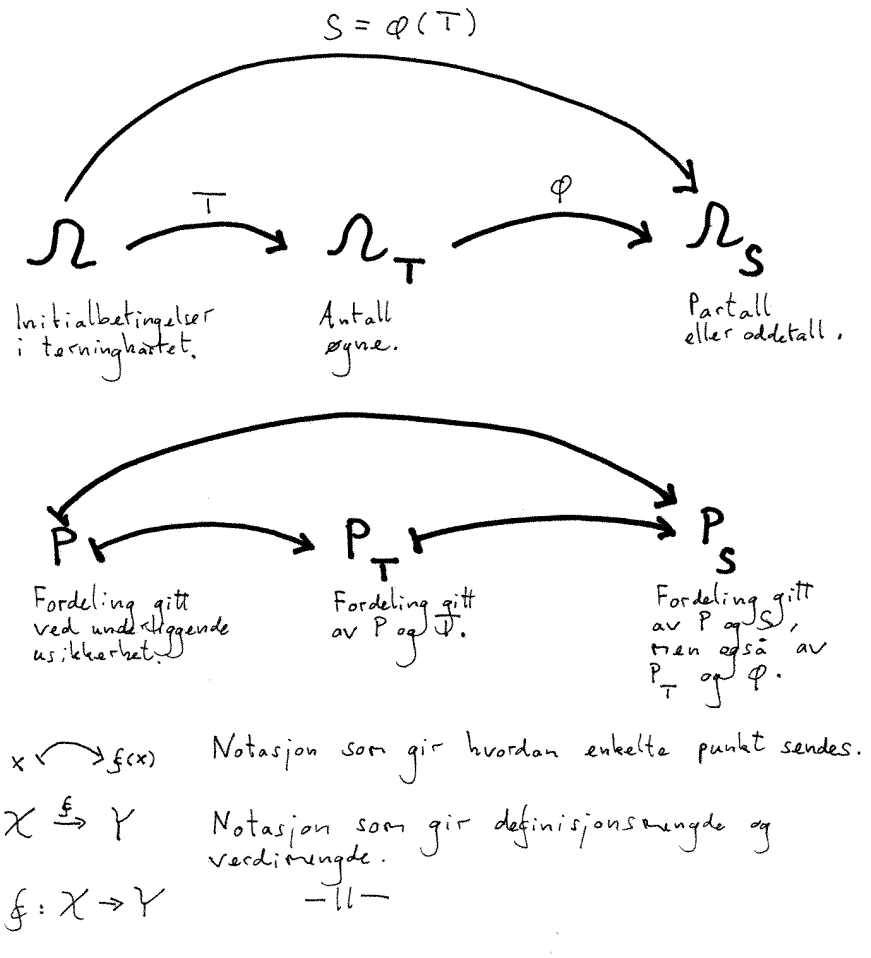
Vi benytter notasjonen $S = \phi(T)$. Fordelingen til S er gitt av fordelingen til T ved formelen

$$\boxed{P_S(A) = P_T(\phi \in A)}$$

Likheten over kan begrunnes ved

$$\begin{aligned}(S \in A) &= \{\omega \,|\, \phi(T(\omega)) \in A\} = \{\omega \,|\, T(\omega) \in \{t \,|\, \phi(t) \in A\}\} = \{\omega \,|\, T(\omega) \in (\phi \in A)\} \\ &= (T \in (\phi \in A))\end{aligned}$$

Konklusjonen er:



Figur 8.1: Funksjon av en observator.

> Fordelingen til en funksjon av en observator er gitt av funksjonen og fordelingen til observatoren.

Poenget i det overstående er at selv om P_S er definert ved hjelp av P og S, så er P_S entydig gitt av P_T og ϕ. Vi trenger derfor ikke kjenne P for å beregne P_S! Resulatetet er illustrert i figur 8.1 sammen med en antydet modell for terningkast. Resultatet tilsvarer at forventningsverdien $\mathrm{E}\,\phi(T) = \mathrm{E}_T\,\phi$ kan beregnes når E_T er kjent uten at E er kjent. Faktisk er resultatet for fordelingene et spesialtilfelle av substitusjonsformelen for forventningsverdier.

8.2 Beregning av tettheter

Anta at $S = \phi(T)$ er kontinuerlig fordelt. Da er tettheten til S gitt ved

$$f_S(s) = F_S'(s) = \frac{d}{ds} P_T(\phi \leq s).$$

Dette er metoden med fordelingsfunksjonen. Metoden med fordelingsfunksjonen er ofte den mest anvendelige metoden for å finne en tetthet.

Eksempel 8.1 (Tettheten til $aX + b$) Vi vil finne tettheten til $Y = aX + b$ når vi antar at X er kontinuerlig fordelt. Vi antar $a > 0$. Tettheten finnes ved

$$f_Y(y) = \frac{d}{dy} P(aX + b \leq y) = \frac{d}{dy} \int_{\{x \,|\, ax+b \leq y\}} f_X(x)dx = \frac{d}{dy} \int_{-\infty}^{\frac{y-b}{a}} f_X(x)dx$$

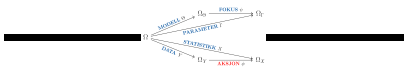
$$= \frac{d}{dy} F_X(\frac{y-b}{a}) = f_X(\frac{y-b}{a})\frac{1}{a}.$$

Tilfellet $a < 0$ kan behandles tilsvarende. Tilsammen har vi resultatet

Dersom $Y = aX + b$, så er

$$f_Y(y) = f_X(\frac{y-b}{a})\frac{1}{|a|}.$$

Spesielt gir beviset også at Y er kontinuerlig fordelt.



Eksempel 8.2 (Tettheten til $X+Y$ og konvolusjon.) Vi vil finne tettheten til $Z = X+Y$ når vi antar at X, Y er simultant kontinuerlig fordelt. Vi beregner fordelingsfunksjonen

$$F_Z(z) = P(X + Y \leq z) = \iint_{R_z} f_{X,Y}(x, y)dxdy \overset{figur\ 8.2}{=} \int_{x=-\infty}^{\infty} \left(\int_{y=-\infty}^{z-x} f_{X,Y}(x, y)dy \right) dx.$$

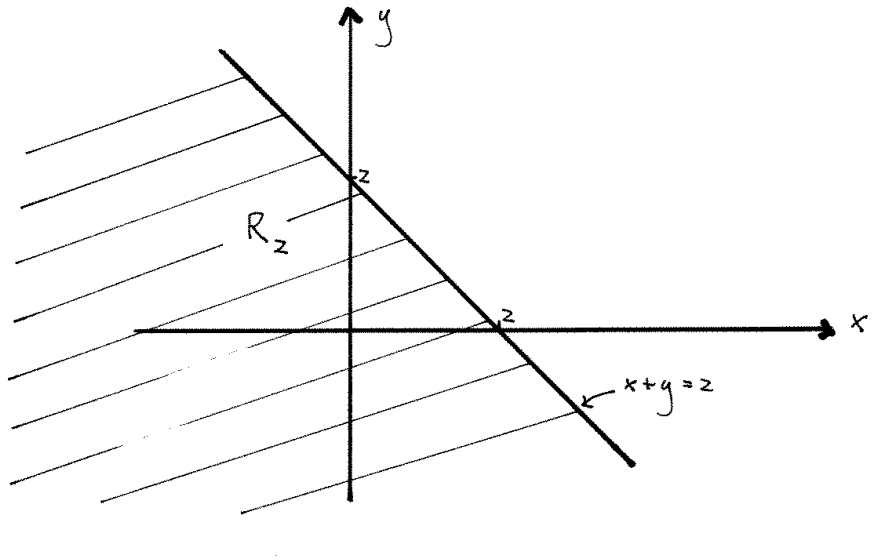
Merk: Det kan være til stor hjelp å lage en skisse av integrasjonsområdet ved beregningen av dobbelintegral. Ved derivasjon finner vi

$$f_Z(z) = F_Z'(z) = \frac{d}{dz} \int_{x=-\infty}^{\infty} \left(\int_{y=-\infty}^{z-x} f_{X,Y}(x, y)dy \right) dx = \int_{x=-\infty}^{\infty} f_{X,Y}(x, z - x)dx.$$

Dersom vi antar uavhengighet, så finner vi

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z - x)dx \overset{def}{=} (f_X \star f_Y)(z).$$

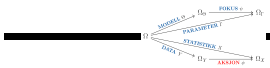
Konvolusjonsproduktet $f_X \star f_Y$ er definert av integralet over. Vi kan da formulere et viktig resultat:



Figur 8.2: Integrasjonsområdet ved konvolusjon.

Dersom X, Y er uavhengige og kontinuerlig fordelt, så er $X + Y$ kontinuerlig fordelt og tettheten er gitt ved konvolusjonen

$$f_{X+Y} = f_X \star f_Y.$$



Eksempel 8.3 (Ordningsobservatoren) Ordningsfunksjonen r er definert ved at $r(x_1, \ldots, x_n) = (t_1, \ldots, t_n)$ hvor $t_1, \ldots, t_n$ er $x_1, \ldots, x_n$ skrevet i stigende rekkefølge. Et konkret eksempel: Dersom $x = (1, 2, 1, 3, -1)$, så er $t = r(x) = (-1, 1, 1, 2, 3)$.

La $X_1, \ldots, X_n$ være tilfeldige variable. Ordningsobservatoren $T = (X_{(1)}, \ldots, X_{(n)})$ er definert ved $T = r(X_1, \ldots, X_n)$ hvor r er ordningsfunksjonen. Spesielt er da $X_{(n)} = X_{\text{maks}} = \text{maks}\{X_1, \ldots, X_n\}$ og $X_{(1)} = X_{\min} = \min\{X_1, \ldots, X_n\}$.

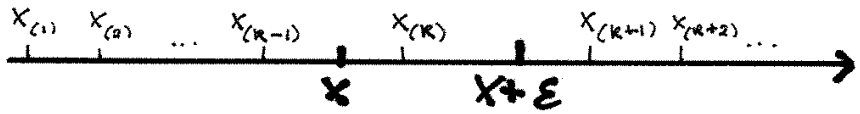
Vi vil finne marginaltetthetene til alle $X_{(i)}$, $i = 1, \ldots, n$, når vi antar at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra en kontinuerlig tetthet f. I utgangspunktet er den søkte tettheten gitt ved

$$f_{X_{(i)}}(x) = \lim_{\epsilon \to 0} \frac{P(x < X_{(i)} \leq x + \epsilon)}{\epsilon}.$$

Vi vil beregne denne grenseverdien ved å erstatte telleren med en sannsynlighet som kan beregnes kombinatorisk.

Definer variabelen

$$Y_i = \begin{cases} a & X_i \leq x, \\ b & x < X_i \leq x + \epsilon, \\ c & x + \epsilon < X_i. \end{cases}$$



Figur 8.3: Et trinomisk utfall.

Spesielt er variabelen Y_i basert på X_i. La $p_a = P(Y_i = a)$, $p_b = P(Y_i = b)$ og $p_c = P(Y_i = c)$. Videre defineres variabelen

$$N_a = \text{ antall } Y_i\text{'er som er lik } a = \#\{Y_i \,|\, Y_i = a\},$$

og N_b, N_c defineres tilsvarende. Uavhengigheten til Y_i'ene og et kombinatorisk argument gir

$$P(N_a = n_a, N_b = n_b, N_c = n_c) = {p_a}^{n_a} {p_b}^{n_b} {p_c}^{n_c} \frac{n!}{n_a! n_b! n_c!}.$$

Dette er en generalisering av det binomiske tilfellet. Vi sier at (N_a, N_b, N_c) er *trinomisk* fordelt. Vi skriver dette kort som $(N_a, N_b, N_c) \sim \text{trinomisk}(n, p_a, p_b, p_c)$.

I grensen $\epsilon \to 0$ er hendelsen $(x < X_{(k)} \leq x + \epsilon)$ lik hendelsen $(N_a = k-1, N_b = 1, N_c = n-k)$. Dette er illustrert på figur 8.3. Sannsynligheten for denne hendelsen er

$$P(N_a = k-1, N_b = 1, N_c = n-k) = {p_a}^{k-1} {p_b}^{1} {p_c}^{n-k} \frac{n!}{(k-1)!1!(n-k)!} =$$

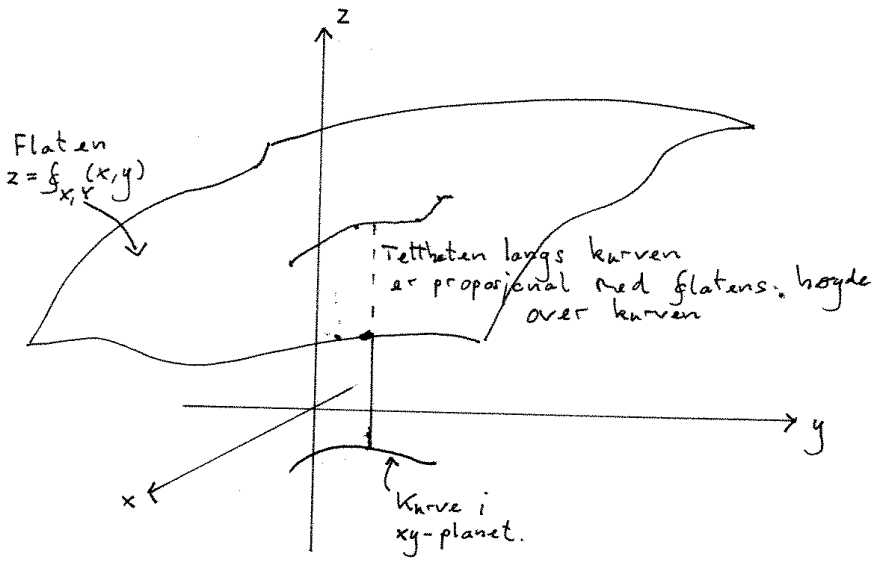
$$P(X_i \leq x)^{k-1} P(x < X_i \leq x + \epsilon)^1 P(x + \epsilon < X_i)^{n-k} \frac{n!}{(k-1)!(n-k)!} =$$

$$F(x)^{k-1} (F(x+\epsilon) - F(x))(1 - F(x+\epsilon))^{n-k} \frac{n!}{(k-1)!(n-k)!}.$$

L'Hopitals regel gir tilslutt

$$\begin{aligned} f_{X_{(k)}}(x) &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} P(x < X_{(k)} \leq x + \epsilon) \\ &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} F(x)^{k-1} (F(x+\epsilon) - F(x))(1 - F(x+\epsilon))^{n-k} \frac{n!}{(k-1)!(n-k)!} \\ &= \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} (1 - F(x))^{n-k}. \end{aligned}$$

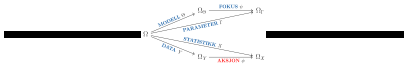
Resultatet er verdt å utheve:



Figur 8.4: Geometrisk tolkning av betinget tetthet.

La $(X_{(1)}, \ldots, X_{(n)})$ være ordningsobservatoren til et tilfeldig utvalg $X_1, \ldots, X_n$ fra en kontinuerlig fordeling med tetthet f. Da er

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} (1 - F(x))^{n-k}.$$



8.3 Betingede tettheter

Anta at X, Y har en simultan tetthet $f_{X,Y}$. Geometrisk kan da $z = f_{X,Y}(x, y)$ tolkes som en flate i rommet, og tettheten i punktet (x, y) er høyden til denne flaten som illustrert på figur 8.4. Anta så at vi kun er interessert i utfall (x, y) som ligger på en kurve. Intuitivt bør tettheten tilsvarende disse utfallene være proposjonal med høyden til flaten $z = f_{X,Y}(x, y)$ over punktene (x, y) som ligger på kurven. Spesielt kan kurven være en horisontal linje gitt ved at y er konstant.

Det foregående motiverer en definisjon. Den *betingede tettheten* $f_{X \mid Y}$ til X gitt Y er definert ved

$$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Det følger at tettheten til $X \mid Y = y$ (les: X gitt $Y = y$) er en sannsynlighetstetthet for alle y. Beviset for det kontinuerlige tilfellet er gitt ved:

1. Tettheten er større eller lik 0.

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \geq 0.$$

2. Tettheten er normalisert.

$$\int f_{X\,|\,Y}(x\,|\,y)dx = \int \frac{f_{X,Y}(x,y)}{f_Y(y)}dx = \frac{\int f_{X,Y}(x,y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1.$$

Vi har tidligere sett at produktregelen $P(A\cap B) = P(A\,|\,B)P(B)$ er nyttig. Den følgende produktregelen for tettheter er ikke mindre viktig.

Produktregelen: $f_{X,Y}(x,y) = f_{X\,|\,Y}(x\,|\,y)f_Y(y)$.

Notasjonen av indeksene kan i noen tilfeller bli for tungvint. Det er derfor konvensjonelt å utelate indeksene. Produktregelen kan f eks skrives på formen $w(x,y) = w(x\,|\,y)w(y)$, hvor det underforstås at argumentene gir hva w står for. Her er mao $w(x,y) = f_{X,Y}(x,y)$, $w(x\,|\,y) = f_{X,Y}(x\,|\,y)$ og $w(y) = f_Y(y)$.

Eksempel 8.4 (Betinget tetthet) La N være antall biler innlevert på et verksted i løpet av en dag. Vi antar at

$$f_N(n) = P(N=n) = \frac{\lambda^n}{n!}e^{-\lambda}, \;\; \lambda \geq 0, \;\; n = 0,1,\ldots.$$

Dette er tettheten til Poisson fordelingen. Vi vil finne fordelingen til antall oljeskift X i løpet av en dag.

Når det er gitt at det er levert inn n biler, så er sannsynligheten for x oljeskift gitt ved

$$f_{X\,|\,N}(x\,|\,n) = P(X=x\,|\,N=n) = \binom{n}{x}p^x(1-p)^{n-x},$$

når p er sannsynligheten for at det skal utføres oljeskift på en bil som er innlevert. Formelen gir også spesialtilfellet $P(X=0\,|\,N=0) = 1$.

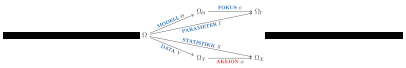
Dette gir simultanfordelingen til X, N

$$f_{X,N}(x,n) = f_{X\,|\,N}(x\,|\,n)f_N(n) = \binom{n}{x}p^x(1-p)^{n-x}\frac{\lambda^n}{n!}e^{-\lambda}.$$

Dermed kan vi regne ut tettheten til X

$$f_X(x) = \sum_n f_{X,N}(x,n) = \sum_n \binom{n}{x}p^x(1-p)^{n-x}\frac{\lambda^n}{n!}e^{-\lambda} = \frac{(\lambda p)^x}{x!}e^{-\lambda p}$$

Konklusjonen er dermed at X også er Poisson fordelt. Modellen gir at forventningsverdien til antall biler N i løpet av en dag er λ, mens forventningsverdien til antall oljeskift X i løpet av en dag er λp.



Helt generelt er det mulig å definere en betinget forventning $\mathrm{E}(X \mid Y = y)$. Vi nøyer oss med tilfellet hvor det finnes en betinget tetthet. Det kan da bevises at

$$\boxed{\mathrm{E}(\phi(X) \mid Y = y) = \int \phi(x) f_{X \mid Y}(x \mid y) dx.}$$

Et helt tilsvarende resultat gjelder for tilfellet hvor det finnes en diskret simultantfordeling for X, Y.

8.4 Oppgaver

8.1. En urne inneholder 4 røde og 3 hvite brikker. Identifiser utfallsrommet Ω og fordelingen P for eksperimentet gitt ved at to brikker trekkes fra urnen. La X være antall røde brikker som blir trukket. Finn fordelingen til X og $Y = 2X - 1$. Hvordan er funksjonen $X : \Omega \to \mathbb{R}$ definert?

8.2. Temperaturen T i en viss kjemisk reaksjon varierer fra eksperiment til eksperiment, men synes å være godt beskrevet av tettheten

$$f_T(t) = te^{-t^2/2} \quad t > 0,$$

når temperaturen er målt i grader Fahrenheit. Temperaturen kan regnes om til grader Celsius ved formelen

$$Y = \frac{5}{9}(T - 32).$$

Finn tettheten til temperaturen målt i grader Celsius.

8.3. La (X, Y) være uniformt fordelt på enhetskvadratet, dvs $f_{X,Y}(x, y) = 1,\ 0 < x, y < 1$. Finn tettheten til $X + Y$, XY og X/Y. Hint: Tegn en figur for integrasjonsområdet i beregningen av f eks $P(X + Y \leq z)$.

8.4. Anta at X er uavhengig av Y, Z. Vis at X er uavhengig av $Y + Z$.

8.5. La $X_1, \ldots, X_4$ være 4 uavhengige målinger tilsvarende at et eksperiment med tetthet

$$f(x) = \frac{1}{3}e^{-x/3} \quad x > 0$$

gjøres 4 ganger. Finn sannsynlighetene $P(X_4 \leq 5, 1)$ og $P(X_{\text{maks}} \leq 5, 1)$.

8.6. Anta at tiden du bruker på å stå i kø når du går i banken er uniformt fordelt mellom 0 og 10 minutter. Hvis du skal i banken 4 ganger i løpet av den neste måneden, hva er så sannsynligheten for at din nest lengste ventetid er mindre enn 5 minutter?

8.7. Anta at $f_{Y \mid X}(y \mid x) = \frac{2y+4x}{1+4x}$ og $f_X(x) = \frac{1}{3}(1 + 4x)$ hvor $0 < x, y < 1$. Finn marginaltettheten f_Y til Y.

8.8. La $f_{X,Y}(x,y) = 2$ for $0 < x < y < 1$. Finn den betingede sannsynligheten $P(0 < X < \frac{1}{2} \,|\, Y = \frac{3}{4}) = \mathrm{E}(1_{(0,\frac{1}{2})}(X) \,|\, Y = \frac{3}{4})$.

8.9. Bevis formelen $P_{\phi(T)}A = P_T(\phi \in A)$ ved hjelp av substitusjonsformelen for beregning av forventningsverdier.

8.10. La f være tettheten til X. Illustrer den geometriske sammenhengen mellom grafen til f og grafen til tettheten g til $X + a$.

8.11. Konvolusjonsproduktet mellom to funksjoner er definert ved integrasjon slik som i teksten. Hvordan vil du begrunne at $f \star g = g \star f$?

8.12. Benytt $(X_{\text{maks}} \leq x) = (X_1 \leq x, \ldots, X_n \leq x)$ til å beregne tettheten til X_{maks} når $X_1, \ldots, X_n$ er et tilfeldig utvalg fra en tetthet f. Sammenlign med resultatet i teksten.

8.13. Produktregelen for tettheter har en generalisering til tilfellet med n tilfeldige variable $X_1, \ldots, X_n$. Prøv å formulere og bevise en slik generalisering. Hvordan stemmer formelen når variablene er uavhengige?

8.14. Prøv å formulere og bevise en generalisering av Bayes teorem ved betingede tettheter.

8.15. Dersom Y har en kontinuerlig fordeling, så kan det vises generelt at

$$\mathrm{E}(X \,|\, Y = y) = \frac{\frac{d}{dy}\,\mathrm{E}(X 1_{(Y \leq y)})}{f_Y(y)}.$$

Vis at denne formelen gir formelen i teksten når vi antar at X, Y er simultant kontinuerlig fordelt.

8.16. La $X_1, \ldots, X_n$ være uavhengige og uniformt fordelt på intervallet $[0, 1]$. Beregn $\mathrm{E}(X_1 \,|\, X_{\text{maks}} = x)$ ved formelen i forrige oppgave.

Kapittel 9

De små talls lov: Poissonfordelingen

Die vorliegende Abhandlung stellt den ersten Versuch dar, statistischen Reihen, welche aus kleinen absoluten Zahlen bestehen, vom Standpunkte der Wahrscheinlichkeitsrechnung aus näher zu treten.

Den nåværende avhandlingen representerer det første forsøket på detaljert analyse av statistiske serier, som består av små heltall, fra sannsynlighetsregningens ståsted.

Dr. L. von Bortkewitsch (1898)

I en viss forstand er den uniforme fordelingen, normalfordelingen og poissonfordelingen de tre viktigste fordelingene i sannsynlighetsteorien. Den uniforme fordelingen er viktig av symmetrigrunner, og er koplet til den klassiske definisjonen av sannsynlighet. Klassiske sannsynligheter beregnes kombinatorisk med utgangspunkt i utfall som er like sannsynlige, dvs med utgangspunkt i en uniform modell. Spesielt leder kombinatoriske argument til binomialfordeling og hypergeometrisk fordeling.

Poissonfordelingen kalles de små talls lov. Poissonfordelingen er knyttet til mange anvendelser, f eks i forbindelse med ventetidsproblemer og hyppighet av sjeldne hendelser. Poissonfordelingen kan brukes også for studiet av hendelser som ikke er sjeldne, men betegnelsen *de små talls lov* er motivert av Bortkewitsch sin originale avhandling. Den omhandlet hendelser med liten hyppighet. Telling av forekomster gir da små heltall.

9.1 Poissonfordelingen og Poisson-punktprosesser

En tilfeldig variabel X er *poissonfordelt* med parameter $\lambda \geq 0$ dersom

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0, 1, \ldots \tag{9.1}$$

Dette noteres kort som $X \sim \text{Poisson}(\lambda)$. Tegnet "$\sim$" leses som "er fordelt som" og brukes også mer generelt. F eks betyr $X \sim Y$ at X og Y har lik fordeling.

Poissonfordelingen kan benyttes til modellering av f eks følgende situasjoner:

1. Telling av antall skjønnhetsfeil X (feil i overflaten) på en tallerken.
2. Telling av antall røde blodlegemer X i en blodprøve av angitt volum.
3. Telling av antall anrop X på en telefonsentral i løpet av en dag.
4. Telling av antall α-partikler X som sendes ut fra en radioaktiv kilde i løpet av et gitt tidsintervall.
5. Telling av antall biler X som passerer et gitt sted på veien i løpet av et gitt tidsintervall.

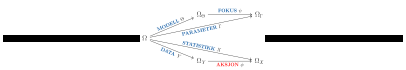
I de to første situasjonene kan usikkerheten innføres ved at en tenker seg et tilfeldig utvalg fra en stor beholdning. I de tre siste situasjonene er usikkerheten innebygget. En annen forskjell er at de tre siste situasjonene er knyttet til tid, mens de to første situasjonene er knyttet til rom.

Poissonfordelingen kan utledes som fordelingen til en observator for et underliggende sannsynlighetsrom gitt ved en Poisson-punktprosess. Vi vil ikke gjøre dette, men ser i stedet på to eksempler. Det første eksemplet er geometrisk. Det andre eksemplet er knyttet til tid og leder også til to nye fordelinger: eksponensialfordelingen og gammafordelingen.

Eksempel 9.1 (Antall skjønnhetsfeil) La X være antall skjønnhetsfeil på en tallerken som velges tilfeldig fra en stor beholdning. Det er da klart at X kan ta verdier $x = 0, 1, 2, \ldots$, men det er usikkert hvilken verdi som tas. Vi vil anta at vi kan identifisere hver skjønnhetsfeil med et punkt på tallerkenen. Vi kan tenke oss en underliggende modell for å finne fordelingen til X. Plasser tallerkenene som på figur 9.1 og tenk deg at feilene drysses tilfeldig utover tallerkenene. La $Q = (Q_1, Q_2, \ldots)$ være de tilfeldige posisjonene til feilene. Dette betyr at hver Q_i er en observator, dvs Q er et eksempel på en stokastisk prosess. Antall feil i tallerkenen som er plassert i origo er da en funksjon ϕ av posisjonene gitt ved

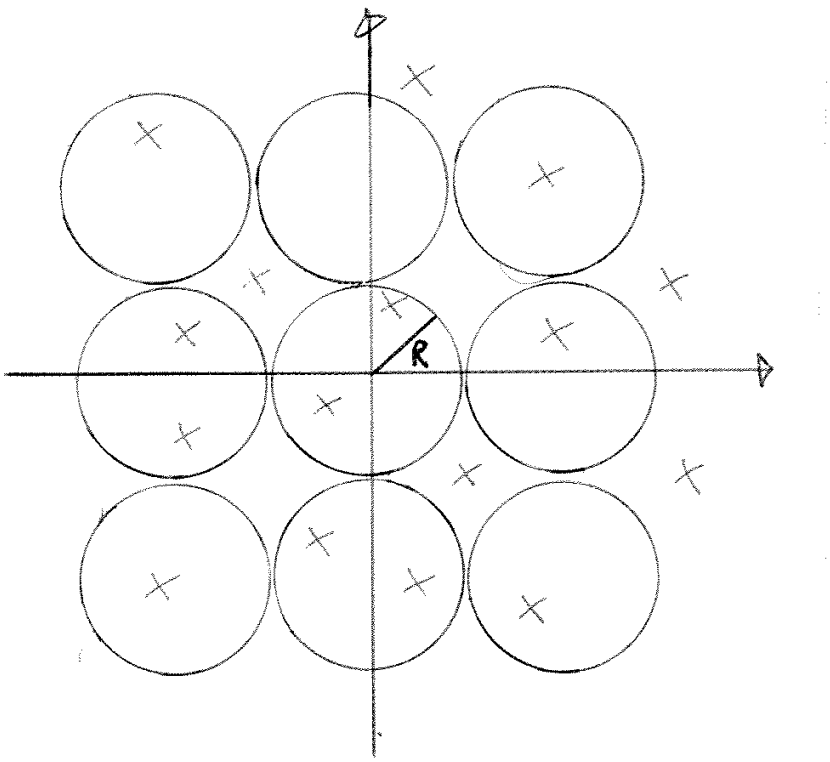
$$X = \phi(Q) = \text{ antall } n \text{ slik at } |Q_n| \leq R.$$

Modellen kan spesifiseres ved å anta at samlingen $Q_1, Q_2, \ldots$ av tilfeldige punkter i planet er en Poisson-punktprosess. Da følger det at $X \sim \text{Poisson}(\lambda\pi R^2)$. Parameteren λ angir forventet antall feil per areal. Forventet antall feil per tallerken blir $\lambda\pi R^2$, dvs proposjonal med tallerkenens areal πR^2. Vi vil ikke behandle teorien for Poisson-punktprosesser i mer detalj.



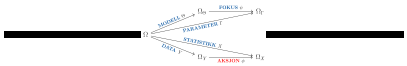
Eksempel 9.2 (Telling av biler) Anta at vi står ved en vei og skal registrere biler som passerer. Tidspunktet T_1 for første bilpassering er da usikkert. Det samme gjelder tidspunktet T_n for n'te bilpassering. Vi vil anta at antall biler

$$N_t = \#\{n \,|\, 0 < T_n \leq t\}$$



Figur 9.1: Skjønnhetsfeil tilfeldig fordelt på tallerkener. Hver feil er gitt ved et kryss.

som passerer i løpet av tidsintervallet $(0, t]$ oppfyller $N_t \sim \text{Poisson}(\lambda t)$. Parameteren λ angir forventet antall bilpasseringer per tid. Det følger at T_1 er eksponensialfordelt med parameter λ. Mer generelt følger det at T_n er gammafordelt med parametre n, λ. Vi vil utdype disse påstandene i resten av kapittelet.



9.2 Fordelinger knyttet til en Bernoulli-forsøksrekke

En tilfeldig variabel er *bernoullifordelt* med parameter p, $0 \leq p \leq 1$, dersom

$$P(X = x) = p^x(1-p)^{(1-x)} \qquad x = 0, 1. \tag{9.2}$$

Vi sier kort at $X \sim \text{B}(p)$.

Dette er det samme som at $P(X = 0) = 1 - p$ og at $P(X = 1) = p$. Spesielt vil enhver indikatorvariabel $X = 1_A$ være bernoullifordelt med $p = \text{P}(A)$.

Den momentgenererende funksjonen

$$M_X(t) = \text{E}\, e^{tX} = e^{t0}(1-p) + e^{t1}p = (1-p) + pe^t$$

gir forventningsverdi og varians

$$\text{E}\, X = M_X'(0) = (pe^t)_{t=0} = p, \;\; \text{Var}\, X = \text{E}\, X^2 - (\text{E}\, X)^2 = M_X''(0) - p^2 = p - p^2,$$

som også kunne vårt funnet enkelt direkte.

En *Bernoulli-forsøksrekke* er en rekke uavhengige like forsøk hvor det kun er to mulige utfall i hvert forsøk. En modell for en Bernoulli-forsøksrekke er gitt ved et tilfeldig utvalg $X_1, X_2, \ldots$ av Bernoulli-variable. I en suksess/fiasko modell tenker vi på resultatet $X_j = 1$ som suksess i forsøk nummer j. Resultatet 0 svarer da til fiasko. Vi skal vise at en Bernoulli-forsøksrekke gir opphav til den binomiske-, den geometriske- og den negativt binomiske fordelingen.

En tilfeldig variabel X er binomisk fordelt med parametre n, p hvor $n = 0, 1, \ldots$ og $0 \leq p \leq 1$ dersom

$$P(X = x) = \binom{n}{x} p^x(1-p)^{(1-x)} \qquad x = 0, 1, \ldots, n. \tag{9.3}$$

Vi sier kort at $X \sim \text{B}(n, p)$. Den momentgenererende funksjonen finnes ved bruk av binomialteoremet:

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x(1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} = ((1-p) + pe^t)^n.$$

La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en $\mathrm{B}(p)$ fordeling. Vi minner om at dette betyr at for alle n er $X_n \sim \mathrm{B}(p)$ og variablene er uavhengige. Det kan vises kombinatorisk at $(X_1 + \cdots + X_n) \sim \mathrm{B}(n, p)$. Et alternativ bevis er gitt ved

$$M_{X_1+\cdots+X_n}(t) = \mathrm{E}\, e^{tX_1+\cdots+tX_n} = \mathrm{E}\, e^{tX_1} \cdot \cdots \cdot \mathrm{E}\, e^{tX_n} = ((1-p) + pe^t)^n.$$

Dette er tilstrekkelig fordi en fordeling er entydig gitt av den tilhørende momentgenererende funksjonen. Vi kan ikke konkludere at X er lik summen av Bernoulli-variablene X_i, men substitusjonsteoremet gir

$$\mathrm{E}\, X = \mathrm{E}(X_1 + \cdots + X_n) = np, \;\; \mathrm{Var}\, X = \mathrm{Var}(X_1 + \cdots + X_n) = n(p - p^2),$$

fordi forventningsverdien til en tilfeldig variabel kun avhenger av variabelens fordeling.

Vi vil utheve hovedpoenget:

> Antall suksesser i n uavhengige forsøk med suksessannsynlighet p er $\mathrm{B}(n, p)$ fordelt.

En variabel N er *geometrisk fordelt* med parameter p, $0 < p \leq 1$, dersom

$$P(N = n) = p(1-p)^{n-1} \qquad n = 1, 2, \ldots. \tag{9.4}$$

Vi sier kort at $N \sim \text{Geometric}(p)$. Tilfellet $p = 1$ gir $P(N = 1) = 1$.

Den momentgenererende funksjonen finnes ved den geometriske rekken

$$M_N(t) = \mathrm{E}\, e^{tN} = \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} = pe^t \sum_{n=1}^{\infty} \left(e^t(1-p)\right)^{n-1} = \frac{pe^t}{1 - e^t(1-p)}.$$

Den momentgenererende funksjonen er kun definert for $e^t(1-p) < 1$. Funksjonen kan brukes til å finne

$$\mathrm{E}\, N = 1/p, \;\; \mathrm{Var}\, N = (1-p)/p^2.$$

Vi skal nå se at dette f eks kan tolkes som at forventet antall myntkast for å få krone er 2.

La $X_1, X_2, \ldots$ være et tilfeldig utvalg fra en $\mathrm{B}(p)$ fordeling med $p > 0$. Definer variabelen

$$N = \min\{n \mid X_n = 1\}.$$

I en suksess/fiasko modell betyr dette at N er antall forsøk som skal til for å oppnå suksess. Uavhengigheten gir

$$\begin{aligned} P(N = n) &= P(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdots P(X_{n-1} = 0) \cdot P(X_n = 1) = p(1-p)^{n-1}. \end{aligned}$$

Vi har dermed bevist at $N \sim \text{Geometric}(p)$. Dersom et forsøk gjøres hver tidsenhet, f eks et forsøk hvert minutt, så kan vi identifiserer n med et tidspunkt. Da får vi følgende reformulering:

Antall tidsenheter til suksess er geometrisk fordelt.

En variabel X er *negativt binomisk fordelt* med parametre r, p hvor $r > 0$ og $0 < p \leq 1$ dersom

$$P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x \qquad x = 0, 1, 2, \ldots . \tag{9.5}$$

Vi sier kort at $X \sim \text{NB}(r, p)$.

I tilfellet $y > 0$ og $x = 0, 1, \ldots$ er binomialkoeffisienten definert ved

$$\binom{y}{x} = \frac{y \cdot (y-1) \cdots (y-x+1)}{x} \qquad \binom{y}{0} = 1 \tag{9.6}$$

Se på en suksess/fiasko modell gitt av tilfeldig utvalgte Bernoulli-variable $X_1, X_2, \ldots$. La Y_r være antall forsøk som skal til for å oppnå r suksesser. Da gir uavhengigheten at

$$\begin{aligned} P(Y_r = y) &= P(r - 1 \text{ suksesser i } y - 1 \text{ forsøk og suksess i } y\text{'te forsøk}) \\ &= \binom{y - 1}{(y-1) - (r-1)} p^{r-1} (1-p)^{y-r} \cdot p. \end{aligned} \tag{9.7}$$

Av overstående følger det at $(Y_r - r) \sim \text{NB}(r, p)$. Ved å identifisere y med et tidspunkt som tidligere følger det at

Antall tidsenheter for å oppnå r suksesser minus r tidsenheter er negativt binomisk fordelt.

Konstruksjonen over med utgangspunkt i uavhengige Bernoulli-variable viser også at dersom $X_1, X_2, \ldots, X_r$ er et tilfeldig utvalg fra en Geometric(p) fordeling, så er

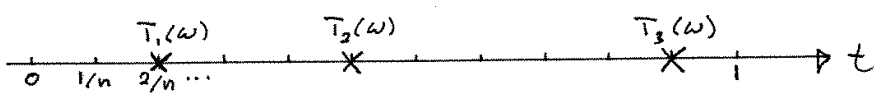
$$[(X_1 - 1) + \cdots + (X_r - 1)] \sim \text{NB}(r, p). \tag{9.8}$$

9.3 Fra binomisk fordeling til poissonfordelingen

I mange situasjoner er det ikke naturlig å anta at tiden er diskret. Et eksempel er gitt ved registrering av biler. Det er ikke naturlig å kun registrere suksesser, dvs bilpasseringer, hvert minutt. Det vil være bedre å gjøre registreringen hvert sekund. Det mest naturlige synes imidlertid å være å tillate registreringer ved vilkårlige tidspunkter. Dette leder til en grensebetrakting hvor tiden mellom registreringene går mot 0 mens forventet antall registreringer per tidsenhet holdes konstant. Det er mulig å følge denne ideen gjennom den foregående diskusjonen skritt for skritt.

Et mulig første skritt er å se på grensen tilsvarende den geometriske fordelingen. Resultatet er en overgang fra den geometriske fordelingen til *eksponensialfordelingen*. Vi skriver $X \sim \text{Exp}(\beta)$, $\beta > 0$, dersom X er kontinuerlig fordelt med tetthet

$$f_X(x) = \beta^{-1} e^{-x/\beta} \qquad x > 0. \tag{9.9}$$



Figur 9.2: De små talls lov. Når $n \to \infty$ er antallet punkter poissonfordelt. Samlingen $T_1, T_2, \ldots$ av punkter er da en Poisson-punktprosess T.

Her ser vi at β er en skalaparameter. Noen lærebøker bruker i stedet parameteren $\lambda = 1/\beta$ når eksponensialfordelingen defineres. Dette valget kan motiveres av utledningene som følger under med kopling til parameteren λ i poissonfordelingen.

Ved analogi med den geometriske fordelingen er følgende resultat ventet:

Tiden til første suksess er eksponensialfordelt.

Sammenlign dette med at du står og venter på bussen. Det er suksess når bussen kommer. Parameteren λ er lik forventet antall busspasseringer per tidsenhet. Det andre skrittet er å se på summen av r uavhengige eksponensialfordelte variable som gir tiden til r suksesser. Vi velger i stedet en annen rute, som er lettere å generalisere til flere dimensjoner.

La $X_n \sim \mathrm{B}(n, p)$. Se på X_n som antall suksesser i tidsintervallet $(0, 1]$. Vi skal nå se på hva som skjer når n vokser. Mer detaljert ser vi på en oppdeling av intervallet i n like store intervaller. I høyre endepunkt av hvert av de n intervallene plasseres et punkt med sannsynlighet p. Dersom et punkt plasseres tilsvarer dette suksess. En mulig realisering er vist på figur 9.2. Når vi antar at plasseringen av punktene er uavhengige, så følger det at totalantallet er binomisk fordelt. Det forventede antall suksesser pn skal holdes konstant, så vi setter $pn = \lambda$. Vi antar altså at

$$X_n \sim \mathrm{B}(n, \lambda/n).$$

Ved å benytte

$$e^{-\lambda} = \lim_{n\to\infty} (1 - \lambda/n)^n$$

følger det at

$$P(X_n = x) = \frac{n!}{x!(n-x)!}(\lambda/n)^x(1-\lambda/n)^{n-x} = \frac{\lambda^x}{x!}(1-\lambda/n)^n\frac{n!}{(n-x)!n^x}(1-\lambda/n)^{-x} \to \frac{\lambda^x}{x!}e^{-\lambda},$$

når $n \to \infty$. Dette er de små talls lov. Vi vil reformulere resultatet.

Vi sier at X_n konvergere mot X i fordeling dersom

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for alle x hvor $F_X(x)$ er kontinuerlig. Vi skriver kort $X_n \xrightarrow{L} X$. Bokstaven "L" står her for Lov (event. Law) i betydningen at fordelingen sees på som en lov for resultatet av et eksperiment.

De små talls lov har nå følgende formulering: La $X_n \sim \mathrm{B}(n, \lambda/n)$ og la $X \sim \mathrm{Poisson}(\lambda)$. Da vil $X_n \xrightarrow{L} X$ når n vokser. Dette er en konsekvens av at hvert av leddene i summen som gir sannsynligheten på venstresiden konvergerer mot de tilsvarende leddene på høyresiden. Vi gjentar hovedresultatet.

> La X være poissonfordelt med intensitet λ. La X_n være binomisk fordelt tilsvarende n forsøk og suksessansynlighet λ/n. Da konvergerer X_n mot X i fordeling når n vokser.

De små talls lov kalles ofte også for Poisson-approksimasjonen. Dette skyldes at binomiske sannsynligheter tilsvarende store n verdier kan beregnes ved at de er tilnærmet lik den tilsvarende Poisson-sannsynligheten. Mer generelt kan forventningsverdien til en funksjon av en binomisk variabel approksimeres. Dette følger fordi det kan vises at konvergens i fordeling er ekvivalent med at

$$\lim \mathrm{E}\,\phi(X_n) = \mathrm{E}\,\phi(X),$$

for enhver begrenset og kontinuerlig ϕ.

9.4 Fra poissonfordelingen til gammafordelingen

Anta at $X \sim \mathrm{Poisson}(\lambda)$. Poisson-approksimasjonen i foregående avsnitt rimeliggjør

$$\begin{aligned} M_X(t) &= \lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty}((1-\lambda/n) + e^t\lambda/n)^n = e^{\lambda(e^t-1)}, \\ \mathrm{E}\,X &= \lim_{n\to\infty} \mathrm{E}\,X_n = \lim_{n\to\infty} n\lambda/n = \lambda, \\ \mathrm{Var}\,X &= \lim_{n\to\infty} \mathrm{Var}\,X_n = \lim_{n\to\infty} n(\frac{\lambda}{n} - (\frac{\lambda}{n})^2) = \lambda. \end{aligned}$$

Disse resultatene kan bevises mer direkte fra fordelingen til X.

Vi har tidligere sett at den geometriske og negativt binomiske fordelingen gir en beskrivelse av hhv ventetiden til 1 suksess eller r suksesser. Utgangspunktet var da at antall suksesser i et diskretisert tidsintervall var binomisk fordelt. Vi vil nå se på det tilsvarende kontinuerlige tilfellet hvor den binomiske fordelingen er erstattet av poissonfordelingen. Vi skal vise at ventetiden til r suksesser er gammafordelt. Tilfellet $r = 1$ gir eksponensialfordelingen.

En tilfeldig variabel X er gammafordelt med formparameter $\alpha > 0$ og skalaparameter $\beta > 0$ dersom X har tetthet gitt ved

$$f_X(x) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad x > 0. \tag{9.10}$$

Vi skriver kort at $X \sim \text{Gamma}(\alpha, \beta)$. Vi ser her at β er en skalaparameter. Noen lærebøker bruker i stedet parameteren $\lambda = 1/\beta$ i definisjonen av gammafordelingen. I det følgende er det hensiktsmessig å la r betegne formparameteren i gammafordelingen.

Gammafunksjonen Γ er definert ved

$$\boxed{\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.}$$

Den viktigste egenskapen til gammafunksjonen er at den generaliserer fakultetsfunksjonen:

$$n! = \Gamma(n+1) \qquad n = 0, 1, \ldots. \tag{9.11}$$

Dette følger av at

$$\Gamma(1) = 1$$

og rekursjonsformelen

$$\Gamma(r+1) = r\Gamma(r) \qquad r > 0. \tag{9.12}$$

Rekursjonsformelen kan bevises ved en delvis integrasjon. Spesialtilfellet

$$\Gamma(1/2) = \sqrt{\pi}$$

kan benyttes sammen med rekursjonsformelen til beregning av $\Gamma(r)$ for halvtallige r.

Vi vil nå utlede fordelingen til tidspunktet T_r for suksess nummer r. La W_t være antall suksesser i intervallet $(0, t]$. Da er

$$(T_r \leq t) = (W_t \geq r).$$

I ord: At tidspunktet for r'te suksess er mindre eller lik t er ekvivalent med at det er minst r suksesser i intervallet $(0, t]$. Ved å anta $W_t \sim \text{Poisson}(\lambda t)$, så finner vi at $T_r \sim \text{Gamma}(r, 1/\lambda)$ fordi

$$\begin{aligned} f_{T_r}(t) &= \frac{d}{dt} P(T_r \leq t) = \frac{d}{dt} P(W_t \geq r) = \frac{d}{dt} \sum_{w=r}^{\infty} e^{-\lambda t} \frac{\lambda^w t^w}{w!} \\ &= \sum_{w=r}^{\infty} -\lambda e^{-\lambda t} \frac{\lambda^w t^w}{w!} + \sum_{w=r}^{\infty} e^{-\lambda t} \frac{\lambda^w w t^{w-1}}{w!} \\ &= \sum_{w=r}^{r} e^{-\lambda t} \frac{\lambda^w w t^{w-1}}{w!} = \frac{\lambda^r}{(r-1)!} e^{-\lambda t} t^{r-1} \end{aligned}$$

Antagelsen over betyr spesielt at forventet antall suksesser per tidsenhet er konstant lik λ. Vi har nå funnet den kontinuerlige beskrivelsen av ventetider:

Ventetiden til r'te suksess er gammafordelt.

Tabell 9.1: Fordelinger knyttet til poissonfordelingen.

Tolkning	Type	Verdier	Parametre	Tetthet $f_X(x)$	$M_X(t)$	E X	Var X
Suksess?	B(p)	x = 0, 1	$0 \leq p \leq 1$	$p^x(1-p)^{(1-x)}$	$(1-p)+pe^t$	p	$p(1-p)$
Antall suksesser	B(n, p)	x = 0,. . . , n	$0 \leq p \leq 1$	$\binom{n}{x}p^x(1-p)^{(1-x)}$	$((1-p)+pe^t)^n$	np	$np(1-p)$
Tid til suksess	Geometric(p)	$x = 1, 2, \ldots$	$0 < p \leq 1$	$p(1-p)^{x-1}$	$\frac{pe^t}{1-e^t(1-p)}$ $t < \ln\frac{1}{1-p}$	$1/p$	$(1-p)/p^2$
Tid til r suksesser minus r	NB(r, p)	x = 0, 1, . . .	$r > 0$ $0 < p \leq 1$	$\binom{x+r-1}{x}p^r(1-p)^x$	...	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Antall suksesser	Poisson(λ)	$x = 0, 1, \ldots$	$\lambda \geq 0$	$\frac{\lambda^x}{x!}e^{-\lambda}$	$e^{\lambda(e^t-1)}$	λ	λ
Tid til suksess	Exp($1/\lambda$)	$x > 0$	$\lambda > 0$	$\lambda e^{-\lambda x}$	$\frac{\lambda}{\lambda-t}$	$1/\lambda$	$1/\lambda^2$
Tid til r suksesser	Gamma($r, 1/\lambda$)	$x > 0$	$r, \lambda > 0$	$\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$	$\left(\frac{\lambda}{\lambda-t}\right)^r$	r/λ	r/λ^2

Det kan vises at summen av r tilfeldig valgte variable fra en Exp(β) fordeling er Gamma(r, β) fordelt. Dette kan tolkes som at fordelingen til neste suksess er eksponensialfordelt uavhengig av forhistorien. Resultatet gir også at familien av gammafordelte variable reproduserer seg selv ved at dersom $X_1 \sim \text{Gamma}(r_1, \beta)$, $X_2 \sim \text{Gamma}(r_2, \beta)$ og X_1, X_2 er uavhengige, så gjelder $X_1 + X_2 \sim \text{Gamma}(r_1 + r_2, \beta)$. En komprimert oppsummering av innholdet i dette kapitlet er gitt i tabell 9.1.

9.5 Oppgaver

9.1. Vis at en indikatorvariabel er bernoullifordelt.

9.2. Vis at en sum av uavhengige likt fordelte indikatorvariable er binomialfordelt ved et kombinatorisk argument.

9.3. Et ungt par planlegger å fortsette å få barn inntil de får en jente. Hva er deres forventede familiestørrelse? Gi en tolkning av svaret.

9.4. Vis at en geometrisk variabel X er uten hukommelse i den forstand at

$$P(X = n + k \,|\, X > n) = P(X = k).$$

9.5. Finn den momentgenererende funksjonen til en negativt binomisk variabel X. Beregn E X og Var X.

9.6. Anta at X_1, X_2 er et tilfeldig utvalg fra en NB(r, p) fordeling. Vis at $X_1 + X_2 \sim \text{NB}(2r, p)$. Gi en tolkning av resultatet ved å se på en sum av uavhengige geometriske variable.

9.7. En bildekkmaskin som er feiljustert gir et defekt bildekk med sannsynlighet $0,15$. Maskinen går hver dag inntil 3 defekte bildekk er produsert før den blir stoppet og justert.

Hva er sannsynligheten for at maskinen produserer 5 eller flere bildekk før den blir stoppet? Hvor mange bildekk vil den produsere i gjennomsnitt før den blir stoppet?

9.8. Astronomer estimerer at 100 billioner stjerner i melkeveien har minst en planet i bane rundt seg. La p være sannsynligheten for at det finnes intelligent liv i et slikt solsystem. Hvor liten kan p være for at det skal være minst 50 % sjanse for at det skal finnes intelligent liv i en de andre solsystemene i melkeveien?

9.9. Regn ut den momentgenererende funksjonen til en Poisson-variabel. Benytt denne til beregning av forventning og varians.

9.10. Anta at X er poissonfordelt med $P(X = 1) = P(X = 2)$. Finn $P(X = 4)$.

9.11. En ung dame er svært populær blant de mannlige medstudentene. I gjennomsnitt mottar hun 4 telefoner hver kveld. Hva er sannsynligheten for at antall telefoner hun mottar i morgen er mer enn standardavviket pluss gjennomsnittet?

9.12. Anta at $X \sim \text{Poisson}(\lambda)$ og at den betingede fordelingen til Y gitt $X = x$ er binomisk med parametre x, p. Vis at $Y \sim \text{Poisson}(\lambda p)$.

9.13. Anta at det i et visst land forekommer gjennomsnitlig $2,5$ flyulykker i året. Hva er sannsynligheten for at det skjer minst 4 flyulykker neste år? Hva er sannsynligheten for at tidsrommet mellom to flyulykker er mindre enn tre måneder?

9.14. Finn fordelingen til $X + Y$ når $X \sim \text{Poisson}(\lambda)$ og $Y \sim \text{Poisson}(\mu)$ er uavhengige.

9.15. Vis at $\Gamma(\frac{7}{2}) = \frac{15\sqrt{\pi}}{8}$.

9.16. Vis at tettheten til gammafordelingen er normalisert.

9.17. Beregn den momentgenererende funksjonen til gammafordelingen.

9.18. Anta at $X_1, \ldots, X_n$ er uavhengige. Vis at gammafordelingen reproduserer seg selv ved at $X_i \sim \text{Gamma}(r_i, \beta)$ gir $\sum_i X_i \sim \text{Gamma}(\sum_i r_i, \beta)$. Gi en tolkning av resultatet ved å se på en sum av uavhengige eksponensialfordelte variable.

9.19. Anta at en værstasjon på Svalbard har 3 vindmålere, hvor en er i drift og to er reserve. Hver av vindmålerne har en levetid som er eksponensialfordelt med forvented levetid på 1000 timer. Finn fordelingen til tiden det tar før alle er defekte. Hva er den forventede levetiden til vindmålingssystemet?

Kapittel 10

Normalfordelingen: Loven for målefeil

The function just found cannot, it is true, express rigorously the probabilities of the errors: for since the possible errors are in all cases confined within certain limits, the probability of errors exceeding those limits ought always to be zero, while our formula always gives some value. However, this defect, which every analytical function must, from its nature, labor under, is of no importance in practice, because the value of our function decreases so rapidly, when ϵ/σ has acquired a considerable magnitude, that it can safely be considered as vanishing. Besides, the nature of the subject never admits of assigning with absolute rigor the limits of error.

C.F. Gauss (1809)

Det overstående sitatet er fra side 259 i C.H. Davi (1857) sin oversettelsen av avhandlingen til Gauss om planetenes baner rundt solen: *Theory of the motion of the heavenly bodies moving about the sun in conic sections*. Funksjonen det refereres til er sannsynlighetstettheten til normalfordelingen. Gauss utledet fordelingen som en modell for målefeil. Han drøftet problemet med å bestemme en planetbane basert på et vilkårlig stort antall observasjoner. Dette gir et overbestemt ligningssystem. Gauss løste dette med utgangspunkt i normalfordelingen og innførte samtidig minste kvadrats metode og en utgave av rimelighetsmetoden. Dette er viktige statistiske metoder som vi skal se på senere.

Dersom $X_1, \ldots, X_n$ er et tilfeldig utvalg fra en fordeling som har forventningsverdi μ og varians σ^2, så er summen $X_1 + \cdots + X_n$ tilnærmet normalfordelt. Mer presist gjelder

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{L} Z \sim \mathrm{N}(0,1)$$

Dette er det viktigste grenseteoremet i statistikken. Resultatet kalles sentralgrenseteoremet, som er en litt freidig oversettelse av navnet G.Polya ga teoremet i sin artikkel *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem* i Mathematischen Zeitschrift (8), 1920. Teoremet forklarer meget generelt hvorfor summen av mange uavhengige tilfeldige variable har en tendens til å være godt beskrevet av normalfordelingen. Teoremet ble bevist av Laplace allerede i 1810.

10.1 Egenskaper til normalfordelingen

En tilfeldig variabel X er normalfordelt med parametre μ, σ hvor $-\infty < \mu < \infty$ og $\sigma > 0$ dersom

$$P(X \leq x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \int_{-\infty}^{z} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \qquad -\infty < x < \infty. \tag{10.1}$$

Vi skriver kort at $X \sim \mathrm{N}(\mu, \sigma^2)$.

For videre beregninger trenger vi

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Dette vises ved omforming til et dobbeltintegral som kan beregnes ved polarkoordinater $x = r\cos\theta$, $y = r\sin\theta$:

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy &= \int_{y=-\infty}^{\infty} (\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx) e^{-\frac{y^2}{2}} dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \iint_{\mathbb{R}^2} e^{-\frac{r^2}{2}} dA = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-\frac{r^2}{2}} r dr d\theta = 2\pi(-e^{-r^2/2})|_0^{\infty} = 2\pi\end{aligned}$$

Anta at Y er standard normalfordelt, dvs anta at $Y \sim \mathrm{N}(0,1)$. Da følger det at normaliseringen er i orden

$$P(-\infty < Y < \infty) = \frac{1}{\sqrt{2\pi}} \frac{1}{1} \int_{-\infty}^{\infty} e^{-\frac{(y-0)^2}{2\cdot 1^2}} dy = 1.$$

Den momentgenererende funksjonen finnes ved

$$M_Y(t) = \mathrm{E}\, e^{tY} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-t)^2 - t^2}{2}} dy = e^{\frac{t^2}{2}}$$

Dermed kan forventning og varians finnes

$$\mathrm{E}\, Y = M_Y'(0) = e^{t^2/2} t \Big|_{t=0} = 0, \ \mathrm{Var}\, Y = (M_Y''(0))^2 - (\mathrm{E}\, Y)^2 = e^{t^2/2}(t^2+1) \Big|_{t=0} = 1$$

Vi vil nå vise at $Y \sim N(0,1)$ gir at $X = \sigma Y + \mu \sim N(\mu, \sigma^2)$. I regningen under foretas variabelskiftet $u = \sigma y + \mu$:

$$P(X \leq x) = P(\sigma Y + \mu \leq x) = P(Y \leq \frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du$$

Dette viser at $X \sim N(\mu, \sigma^2)$. Resultatet er viktig fordi det gjør det mulig å benytte tabulerte verdier for standard normalfordelingen til å beregne tilsvarende verdier for generelle normalfordelinger. Her vil vi benytte resultatet for å finne den momentgenererende funksjonen til en generell $Z \sim N(\mu, \sigma^2)$:

$$M_Z(t) = \mathrm{E}\, e^{tZ} = \mathrm{E}\, e^{tX} = \mathrm{E}\, e^{t\sigma Y + t\mu} = e^{t\mu} M_Y(t\sigma) = e^{t\mu + \frac{t^2\sigma^2}{2}}.$$

Vi kan også enkelt vise at valget av symboler for parametrene i normalfordelingen er velbegrunnet

$$\mathrm{E}\, Z = \mathrm{E}\, X = \mathrm{E}(\sigma Y + \mu) = \mu,\ \mathrm{Var}\, Z = \mathrm{Var}\, X = \mathrm{Var}(\sigma Y) + \mathrm{Var}(\mu) = \sigma^2.$$

Vi har foreløpig ikke tillatt tilfellet $\sigma = 0$. Vi utvider den tidligere definisjonen ved å skrive $Z \sim \mathrm{N}(\mu, 0)$ dersom $P(Z = \mu) = 1$. Resultatet $\sigma Y + \mu \sim N(\mu, \sigma^2)$ er da et spesialtilfelle av følgende:

> Dersom X, Y er uavhengige og normalfordelt, så er lineærkombinasjonen $\alpha X + \beta Y$ normalfordelt.

Anta at $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, og at X, Y er uavhengige. Vi beviser at $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Beviset følger ved å se på den momentgenererende funksjonen

$$M_{X+Y}(t) = \mathrm{E}\, e^{t(X+Y)} \overset{uavh}{=} \mathrm{E}\, e^{tX}\, \mathrm{E}\, e^{tY} = e^{t\mu_X + \frac{t^2\sigma_X^2}{2}} e^{t\mu_Y + \frac{t^2\sigma_Y^2}{2}} = e^{t(\mu_X\mu_Y) + \frac{t^2(\sigma_X^2+\sigma_Y^2)}{2}}.$$

10.2 Bevis av det viktigste grenseteoremet i statistikken

Generelt er den normaliserte utgaven Y av en tilfeldig variabel X gitt ved

$$Y := (X - \mathrm{E}\, X)/\sqrt{\mathrm{Var}\, X}.$$

Egenskapene $\mathrm{E}\, Y = 0$ og $\mathrm{Var}\, Y = 1$ begrunner navnet. En variabel X har en normalisert utgave dersom den har endelig forventning og varians.

Det viktigste grenseteoremet i statistikken er:

> La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en fordeling. Anta at den normaliserte utgaven Z_n av $X_1 + \cdots + X_n$ eksisterer. Da vil
>
> $$\lim_{n\to\infty} P(Z_n \le z) = P(Z \le z), \quad \text{hvor } Z \sim \mathrm{N}(0, 1)$$

En reformulering av det overstående er gitt ved å si at Z_n konvergerer i fordeling mot Z, eller enda kortere $Z_n \xrightarrow{L} Z$.

Det kan vises at dersom $M_{Z_n}(t) \to M_Z(t)$, så vil $Z_n \xrightarrow{L} Z$. Dette resulatet vil vi ta som gitt, og vår oppgave er da å bevise at $M_{Z_n}(t) \to e^{t^2/2}$.

Vi har $\mathrm{E}(X_1 + \cdots + X_n) = n\mu_X$ og $\mathrm{Var}(X_1 + \cdots + X_n) = n\sigma_X^2$. Den normaliserte utgaven av summen blir da

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu_X}{\sqrt{n}\sigma_X}$$

Dette gir

$$M_{Z_n}(t) = \mathrm{E}\, e^{tZ_n} = \mathrm{E}\, e^{t\frac{X_1+\cdots+X_n-n\mu_X}{\sqrt{n}\sigma_X}} = \mathrm{E}\, e^{\frac{t}{\sqrt{n}}\frac{X_1-\mu_X}{\sigma_X}+\cdots+\frac{t}{\sqrt{n}}\frac{X_n-\mu_X}{\sigma_X}} \overset{uavh}{=} \mathrm{E}\, e^{\frac{t}{\sqrt{n}}\frac{X_1-\mu_X}{\sigma_X}} \cdots \mathrm{E}\, e^{\frac{t}{\sqrt{n}}\frac{X_n-\mu_X}{\sigma_X}}$$
$$= (\mathrm{E}\, e^{\frac{t}{\sqrt{n}}W})^n.$$

Oppgaven er dermed redusert til å bevise at

$$\left(\mathrm{E}\, e^{\frac{t}{\sqrt{n}}W}\right)^n \to e^{t^2/2}.$$

når det antas at $\mathrm{E}\, W = 0$ og $\mathrm{Var}\, W = 1$.

En mulighet er å bruke rekkeutviklingen $e^x = 1 + x + x^2/2 + \cdots$ til eksponensialfunksjonen:

$$(\mathrm{E}\, e^{\frac{t}{\sqrt{n}}W})^n = (\mathrm{E}(1 + \frac{t}{\sqrt{n}}W + \frac{t^2}{2n}W^2 + \cdots))^n = (1 + 0 + \frac{t^2}{2n}1 + \cdots)^n \to e^{t^2/2}.$$

I det overstående kan det by på problemer å begrunne de enkelte overgangene.

Et bevis via to gangers bruk av L'Hopitals regel er en annen mulighet:

$$\lim_{n\to\infty} \ln(\mathrm{E}\, e^{\frac{t}{\sqrt{n}}W})^n = \lim_{n\to\infty} \frac{\ln(\mathrm{E}\, e^{\frac{t}{\sqrt{n}}W})}{1/n} = \lim_{x\to 0} \frac{\ln(\mathrm{E}\, e^{xtW})}{x^2}$$
$$= \lim_{x\to 0} \frac{(\mathrm{E}\, e^{xtW}tW)/(\mathrm{E}\, e^{xtW})}{2x} = \lim_{x\to 0} \frac{1}{2}\frac{\mathrm{E}(e^{xtW}t^2W^2)(\mathrm{E}\, e^{xtW}) - (\mathrm{E}\, e^{xtW}tW)^2}{(\mathrm{E}\, e^{xtW})^2} = t^2/2$$

I bevisene over er det stilltiende antatt at alle de momentgenererende funksjonenen finnes. Et mer generelt gyldig bevis er gitt ved å ta utgangspunkt i den karakteristiske funksjonen $\phi_Z(t) = M_Z(it)$ i stedet. Begge bevisene vi ga kan modifiseres til dette tilfellet.

I beregningene er det benyttet et ombytte av henholdsvis summasjon og derivasjon med forventningsverdi. Begge disse ombyttene kan begrunnes, men det kreves mer verktøy enn vi kan gi i denne framstillingen.

10.3 Khi-kvadrat-, Fisher- og Student-t-fordelingene

Vi vil se på tre fordelinger som er avledet fra normalfordelingen. Disse fordelingene vil vi senere bruke i statistikkdelen av denne fremstillingen.

Vi gir her bare en oppsummering av hvordan disse fordelingene fremkommer.

La Z være standard normalfordelt. Da vil $Y = Z^2$ være khi-kvadratfordelt med 1 frihetsgrad. La $Z_1, \ldots, Z_n$ være et tilfeldig utvalg fra en $N(0,1)$ fordeling. Da er

$$Y = Z_1^2 + \cdots + Z_n^2$$

khi-kvadratfordelt med n frihetsgrader. Vi skriver kort $Y \sim {\chi^2}_n$.

La $U \sim {\chi^2}_m$, $V \sim {\chi^2}_n$ og anta at U, V er uavhengige. Da er variabelen $F = (U/m)/(V/n)$ Fisher-fordelt med m og n frihetsgrader. Vi skriver kort $F \sim \mathrm{F}(m, n)$.

La $Z \sim N(0,1)$, $V \sim {\chi^2}_n$ og anta at Z, V er uavhengige. Da er variabelen $T = Z/\sqrt{V/n}$ Student-t-fordelt med n frihetsgrader. Vi skriver kort at $T \sim \mathrm{t}(n)$.

La $K = m + \beta T$ hvor $T \sim \mathrm{t}(1)$. Da er K Cauchyfordelt med lokaliseringsparameter m og skalaparameter β. Vi skriver kort at $K \sim \mathrm{C}(m, \beta)$.

Metoden med fordelingsfunksjonen og metoden med den momentgenererende funksjonen kan brukes for å finne tetthetene til ${\chi^2}_n$-, $\mathrm{F}(m,n)$-, $\mathrm{t}(n)$- og $\mathrm{C}(m,\beta)$-fordelingene.

10.4 Oppgaver

10.1. Beregn integralet

$$I = \frac{1}{\sqrt{2\pi}} \int_{1.25}^{2.50} e^{-\frac{t^2}{2}} dt$$

ved bruk av tabell.

10.2. La $X \sim \mathrm{B}(100, 0.4)$. Beregn $P(38 < X < 43)$ ved hjelp av sentralgrenseteoremet. Hvorfor kan sentralgrenseteoremet benyttes? Kan Poisson-approksimasjonen benyttes her?

10.3. Flyselskapene A og B gir identisk service på to flyavganger med samme avgangstid, dvs sannsynligheten for at en passasjer velger selskap A er 1/2. Anta at selskapene konkurerer om en kundemasse på 400 potensielle passasjerer. Selskap A har en kapasitet på 230 passasjerer, og selger en bilett til enhver som vil kjøpe. Hva er sannsynligheten for at selskap A får for mange passasjerer?

10.4. Diameteren til styrestangen i en viss sportsbil må være mellom 1.480 og 1.500 cm (inklusive endepunktene) for å være brukbar. Fabrikken produserer med en diameter som er normalfordelt med forventning 1.495 cm og standardavvik 0.005 cm. Hvor mange prosent blir ubrukelige?

10.5. La X være en tilfeldig variabel med momentgenererende funksjon

$$M_X(t) = e^{3t+8t^2}.$$

Beregn $P(-1 \leq X \leq 9)$ ved bruk av tabell.

10.6. Blodtrykket (systolisk) til 18 år gamle kvinner er normalfordelt med forventning 120 mm Hg og standardavvik 12 mm Hg. Hva er sannsynligheten for at en tilfeldig valgt 18 år gammel kvinne har et blodtrykk over 150 ? Mellom 110 og 130?

10.7. La $X_1, \ldots, X_9$ være et tilfeldig utvalg fra en $\mathrm{N}(2,4)$ fordeling. La $Y_1, \ldots, Y_4$ være et uavhengig tilfeldig utvalg fra en $\mathrm{N}(1,1)$ fordeling. Beregn $P(\overline{X} \geq \overline{Y})$.

10.8. Hvor stort må et tilfeldig utvalg fra en normalfordeling være for at utvalgets forventning $\overline{X}$ med sannsynlighet større enn 90% skal ligge mindre enn et standardavvik unna fordelingens forventning μ? Mindre enn to standardavvik unna?

10.9. En heis i idretts-seksjonen ved et statsuniversitet i USA skal ha en kapasitet på 2400 pounds. Anta at 10 fotballspillere går på heisen i 20 etasje. Hva er sannsynligheten for at det vil mangle 10 spillere ved neste trening når vekten til fotballspillerne ved universitetet er normalfordelt med forventning 220 pounds og standardavvik 20 pounds?

10.10. En student bruker minibankkortet 100 ganger mellom hver månedlig utskrift fra banken. I stedet for å notere de eksakte beløpene, så avrunder han beløpet til nærmeste 10 krone. La Y_j være avrundingsfeilen tilhørende transaksjon nummer j, og anta at Y_j er uniformt fordelt i intervallet $[-5, 5]$.

Benytt sentralgrenseteoremet til å approksimere sannsynligheten for at avrundingsfeilen blir større (i absoluttverdi) enn 50 kroner i løpet av en måned.

Sammenlign resultatet med det som finnes ved bruk av Chebyshevs ulikhet.

10.11. Anta at en terning kastes 100 ganger. Estimer sannsynligheten for at summen av resultatene er større enn 370.

10.12. La $\{X_j\}$ være et tilfeldig utvalg av størrelse n fra en Poisson(λ) fordeling. Vis at da er

$$X_1 + \cdots + X_n \sim \text{Poisson}(n\lambda).$$

Begrunn at en Poisson-variabel med stor parameter er tilnærmet normalfordelt.

10.13. Et elektronikkfirma mottar gjennomsnittlig 50 bestillinger i uken på en elektronisk brikke. Anta at antall bestillinger er poissonfordelt. Approksimer (ved sentralgrenseteoremet) sannsynligheten for at et lager på 60 brikker er for lite for en uke.

10.14. Anta at $X \sim {\chi^2}_n$. Vis at $\mathrm{E}\, X = n$ og $\mathrm{Var}\, X = 2n$ når det er gitt at dette er sant for $n = 1$.

10.15. La $Z_1, \ldots, Z_4$ være et tilfeldig utvalg fra N(0, 1)-fordelingen. Anta at

$$\sum_j Z_j^2$$

beskriver resultatet av et eksperiment. Når dette eksperimentet gjentas 100 ganger, hvor mange ganger venter du at resultatet ligger mellom 2 og 3?

10.16. Vis at dersom $X \sim \mathrm{F}(m, n)$, så er $1/X \sim \mathrm{F}(n, m)$.

10.17. Kvantilen $t_{\alpha,\nu}$ er definert ved $P(T > t_{\alpha,\nu}) = \alpha$ hvor $T \sim \mathrm{t}(\nu)$. Finn $t_{0.05,24}$ i en tabell.

10.18. Anta at $T \sim \mathrm{t}(\nu)$. Vis at $-T \sim \mathrm{t}(\nu)$. Hva kan du konkludere om tettheten til T?

10.19. Dersom X, Y er uavhengige og normalfordelt, så er lineærkombinasjonen $\alpha X + \beta Y$ normalfordelt. Bruk dette til å bevise at $\sigma Y + \mu \sim N(\mu, \sigma^2)$ når det antas at Y er standard normalfordelt.

10.20. Prøv å gi et bevis for sentralgrenseteoremet i et mer generelt tilfelle. Hint: Behold regningen i beviset som er gitt, men modifiser antagelsene slik at variablene i summen kan ha ulike fordelinger.

Del III

Statistikk

Kapittel 11

Estimering

Most studies and experiments, scientific or industrial, large scale or small, produce data whose analysis is the ultimate object of the endeavor. Mathematical statistics deals with situations in which the data can be thought of as the outcome of a random experiment.

P. J. Bickel og Kjell A. Doksum (1977)

I do not know of any explicit statement of the properties, consistency, efficiency and sufficiency, which may characterize estimates prior to my 1922 paper. I had noted the functional peculiarity of sufficiency in an earlier paper, 1920, in the Monthly Notices of the Astronomical Society.

R.A.Fisher (1940)

Much modern work may be considered as stemming from a very remarkable paper of Fisher (1922) which gave, among other things, the idea of sufficiency. Fisher's later work on the foundations of statistical inference stressed likelihood and conditionality and his views on the bases are set out in his last book (Fisher, 1973); see also particularly his papers of 1925 and 1934.

D.R. Cox and D.V. Hinkley (1974)

The earliest example of confidence intervals appears to occur in the work of Laplace (1812), who points out how an (approximate) probability statement concerning the difference between an observed frequency and a binomial probability p can be inverted to obtain an associated interval for p. Other examples can be found in the work of Gauss (1816), Fourier (1826), and Lexis (1875). However, in all these cases, although the statements made are formally correct, the authors appear to consider the parameter as the variable which with the stated probability falls in the fixed confidence interval. The proper interpretation seems to have been pointed out for the first time by E. B. Wilson (1927). About the same time two examples of exact confidence statements were given by Working and Hotelling (1929) and Hotelling (1931). A general method for obtaining exact confidence bounds for a real-valued parameter in a continuous distribution was proposed by Fisher (1930), who however later disavowed this interpretation of his work.

E. L. Lehmann (1986)

En sannsynlighetsmodell kan benyttes til beregning av sannsynligheter knyttet til observasjonene i et eksperiment. Utgangspunktet for modellen er antagelser om en underliggende usikkerhet. Det antas at usikkerheten i eksperimentet er gitt ved en kjent fordeling. I en statistisk analyse er problemstillingen snudd. Med utgangspunkt i observasjonene er hensikten å trekke konklusjoner om en underliggende modell. En statistisk modell er gitt ved en antagelse om at den underliggende usikkerheten er beskrevet av en av flere mulige fordelinger. Familien av mulige fordelinger indekseres av en parametermengde, dvs til hver parameter svarer det en fordeling.

Forventningsverdien til resultatet av et eksperiment blir da en funksjon av modellparameteren. Forventningsverdien er dermed en parameter. En type statistisk konklusjon er gitt ved et estimat for forventningsverdien beregnet ved hjelp av gitte data. Estimatet er med andre ord en funksjon av de gitte data. Den tilhørende estimatoren er dermed en statistikk. Fordelingen til en estimator kan benyttes til en vurdering av om estimatoren er god. En estimator sies f.eks. å være forventningsrett dersom forventningsverdien er lik estimanden. Dette er et eksempel på en egenskap til en estimator som er ønskelig i noen situasjoner. Du kan kanskje tenke deg til andre ønskelige egenskaper?

Hensikten med en estimator er å beregne et estimat. Estimatet er vår gjetning på en ukjent størrelse i modellen. Nøyaktigheten til estimatoren sørger for at i en tenkt gjentagelse av eksperimentet, så vil de tilsvarende beregnede estimatene ligge tett samlet i nærheten av den ukjente størrelsen. Konsistensen til estimatoren sørger for at estimatet beregnet ved hjelp av et økende antall eksperiment konvergerer mot den ukjente estimanden. Spesielt svarer dette intuitivt til at dersom fordelingen til eksperimentet er kjent, dvs som om eksperimentet er gjentatt uendelig mange ganger, så gir estimatoren en riktig beregningsformel.

Tilstrekkelighet er kanskje det viktigste nye begrepet ved overgangen fra en sannsynlighetsmodell til en statistisk modell. Intuitivt er en observator tilstrekkelig for et eksperiment dersom verdiene $t_1, t_2, \ldots$ til observatoren inneholder samme informasjon om eksperimentet som verdiene $\omega_1, \omega_2, \ldots$ som eksperimentet gir. Informasjon om eksperimentet betyr

her informasjon om parameteren i den statistiske modellen for eksperimentet. I praktiske eksempler kan ω tilsvare 100 tall, mens t er gitt ved 2 tall. Det sier seg selv at en slik datareduksjon kan gjøre analysen svært mye enklere.

I enkle tilfeller finnes det et naturlig estimat for en ukjent størrelse i en statistisk modell. F eks er naturlige estimatorer for en fordelings momenter gitt av et tilfeldig utvalgs momenter. Dette motiverer momentmetoden hvor et estimat finnes ved å kreve likhet mellom fordelingens første momenter og dataenes momenter. Momentmetoden gir ofte en estimator uten for mye regnearbeid.

Rimelighetsmetoden regnes for å være teoretisk bedre fundert. Utgangspunktet er rimelighetsfunksjonen til de gitte dataene, dvs tettheten til observatoren sett på som funksjon av parameteren i modellen. Et mulig prinsipp i en statistisk analyse er gitt ved at enhver konklusjon kun skal avhenge av rimelighetsfunksjonen til de gitte data. Spesielt må da et estimat være en funksjon av rimelighetsfunksjonen. Dette er spesielt tilfelle for en rimelighetsestimator, som ved definisjon er gitt ved å velge den parameterverdien som maksimaliserer rimelighetsfunksjonen. Et rimelighetsestimat er dermed gitt ved at det maksimaliserer rimeligheten, dvs sannsynligheten, til de gitte observasjonene. Intuitivt synes dette å være et godt valg.

Fremfor å gjette på verdien til en ukjent størrelse kan det være hensiktsmessig å gjette på et område som inneholder den ukjente størrelsen. Når denne gjetningen skal basere seg på en observator, dvs være en funksjon av en observator, så ender vi opp med at gjetningen må være gitt av en tilfeldig mengde. Hvert utfall i den tenkte gjentagelsen av eksperimentet gir en mengde, dvs et områdeestimat for estimanden.

Dersom områdeestimatet inneholder estimanden i minst 95% av gjentagelsene i det lange løp, så sies områdeestimatoren å være en 95% områdeestimator. Ekvivalent sier vi at estimatoren har et konfidensnivå lik 95%. Når nivået er gitt er det ønskelig at det beregnede områdeestimatet er så lite som mulig.

Vi vil se på flere metoder for å utlede og beregne områdeestimatorer. De fleste metodene gir et områdeestimat som et intervall, dvs ved en intervallestimator.

11.1 Beslutningsteori

Beslutningsteori gir et av flere mulige utgangspunkt for begrunnelse av metoder som brukes i anvendt statistikk. Vi ser på et klassisk eksempel før den generelle teorien presenteres.

Eksempel 11.1 (Atomubåten Scorpion) Den amerikanske atomubåten Scorpion ble savnet våren 1968 (Larsen and Marx, 2012, s.333). Den siste meldingen fra Scorpion ble mottatt 21. mai fra en posisjon nær øygruppen Azorene. Azorene er en portugisisk øygruppe i Atlanterhavet som vist i Figur 11.1. Den amerikanske marinen satte inn en stor søkeaksjon langs østkysten av USA, men uten suksess. De søkte deretter eksperthjelp av John Craven. Han laget en søkestrategi med utgangspunkt i sannsynlighetsteori, og dette resulterte i at Scorpion ble funnet i oktober 1968. Besetningen på 99 menn var da omkommet.

Craven tok som utgangspunkt at Scorpion ikke hadde nådd østkysten, men var å finne nær Azorene. Trinn 1 i søkestrategien var å dele søkeområdet i n kvadratiske områder. En

Figur 11.1: Azorene er en øygruppe langt ute i Atlanterhavet.

Tabell 11.1: Områdesannsynligheter i en leteaksjon.

Område	1	2	3	4	5	6	7
Sannsynlighet (%)	22	24	3	24	17	3	7

gruppe veterankommandanter for ubåter ble rådført, og dette ga en vurdering av sannsynligheten for at Scorpion var i hvert enkelt område. Disse sannsynlighetene er ikke lett tilgjengelige, men vi vil tenke oss at de er som gitt i Tabell 11.1. Ut fra dette er du kanskje enig i at et første søk bør gjøres i område 2 eller område 4, og dette var også strategien valgt av Craven. Valget her synes opplagt, men hvordan skal det egentlig begrunnes?

En teoretisk begrunnelse er gitt ved å innføre en *tapsfunksjon*. La γ være det området som inneholder Scorpion, og la $\hat{\gamma} = 2$ være området som det ble besluttet å lete i. Tapet ved denne beslutningen antas å være

$$l = \ell(\gamma, \hat{\gamma}) = \begin{cases} c & \text{når } \hat{\gamma} \neq \gamma \\ 0 & \text{når } \hat{\gamma} = \gamma \end{cases} \tag{11.1}$$

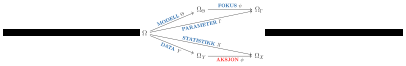
Her er $c > 0$ en tenkt stor konstant assosiert med tapet av mannskapet. Valget av c kan synes problematisk, men valget har ingen konsekvens i analysen videre her.

Problemet nå er at γ er ukjent, så hvordan skal valget av $\hat{\gamma}$ begrunnes? Vi vil anta at usikkerheten er beskrevet av en stokastisk variabel Γ med diskret tetthet g gitt av Tabell 11.1. *Risikoen* til beslutningen $\hat{\gamma}$ er definert ved forventet tap

$$r = \mathrm{E}(\ell(\Gamma, \hat{\gamma})) = \sum_{\gamma} \mathrm{P}(\Gamma = \gamma)\, \ell(\gamma, \hat{\gamma}) = c \cdot \mathrm{P}(\Gamma \neq \hat{\gamma}) = c \cdot (1 - g(\hat{\gamma})) \tag{11.2}$$

Her følger det da at risikoen er minimalisert ved å velge $\hat{\gamma}$ slik at sannsynligheten $g(\hat{\gamma})$ er maksimalisert. Vi velger derfor $\hat{\gamma} = \mathrm{argmax}(g)$. Dette gir en begrunnelse for å velge $\hat{\gamma}$ lik

2 eller 4. Her er $\hat{\gamma}$ et *estimat* av den ukjente parameteren γ. Den minimale risikoen blir $r = c \cdot (1 - 24\%) = c \cdot 76\%$.



Det foregående eksempelet illustrerer et generelt prinsipp i beslutningsteori.

Definisjon 11.1 (*Risiko*) Risikoen er forventningsverdien til tapet. En beslutning er optimal dersom den minimaliserer risikoen.

Et estimat av en ukjent parameter er en beslutning. I Eksempel 11.1 svarer dette til en beslutning om å lete i område $\hat{\gamma} = 2$. Det finnes også andre typer beslutninger som vi skal se på senere, men i dette kapittelet ser vi på estimering. Eksempelet viser at en optimal beslutning ikke trenger å være unik. Estimatet $\hat{\gamma} = 2$ av γ er en optimal beslutning, men estimatet 4 er også en optimal beslutning.

Anta så mer generelt at vi ønsker et estimat $\hat{\gamma}$ av en ukjent parameter γ med usikkerhet beskrevet av en observator Γ. En naturlig kandidat motivert av Eksempel 11.1 er gitt ved $\hat{\gamma} = \text{argmax}(g)$ i tilfellet hvor $\Gamma \sim g$. Andre naturlige kandidater er gitt ved forventningsverdien $\hat{\gamma} = \text{E}(\Gamma)$ eller medianen $\hat{\gamma} = \text{M}(\Gamma)$. Det følgende teoremet gir at dette er optimale estimatorer tilsvarende ulike tapsfunksjoner.

Teorem 11.1. En optimal estimator $\hat{\gamma}$ til $\gamma = \Gamma(\omega)$ tilsvarende tapet l er

1. $\hat{\gamma} = \text{argmax}(g)$ hvor $\Gamma \sim g$ når $l = 1 - \delta(\gamma - \hat{\gamma})$.
2. $\hat{\gamma} = \text{E}(\Gamma)$ når $l = |\gamma - \hat{\gamma}|^2$.
3. $\hat{\gamma} = \text{M}(\Gamma)$ når $l = |\gamma - \hat{\gamma}| - |\gamma - \gamma_0|$.

Begrunnelsen for estimatoren $\hat{\gamma} = \text{argmax}(g)$ krever i utgangspunktet en diskret tetthet for Γ og er gitt som i ligning (11.2) fra tilfellet $c = 1$. Her er $\delta(x) = (x = 0)$ den diskrete *deltafunksjonen*. Multiplikasjon av tapet med en vilkårlig positiv konstant gir ingen endring i beslutningsregelen.

I tilfellet med en kontinuerlig fordeling med tetthet g, så kan regelen $\hat{\gamma} = \text{argmax}(g)$ begrunnes ved å se på en følge av tapsfunksjoner tilsvarende en diskretisering. Resultatet tilsvarer tapsfunksjonen over, men deltafunksjonen er da *Dirac-deltafunksjonen*. Den er definert ved at den oppfyller identiteten

$$\int \delta(\gamma - \hat{\gamma}) g(\gamma)\, d\gamma = g(\hat{\gamma}) \tag{11.3}$$

og dette gir regelen $\hat{\gamma} = \text{argmax}(g)$ også i dette tilfellet. Både Dirac-deltafunksjonen og den diskrete deltafunksjonen betegnes som deltafunksjonen. Leseren må gi riktig tolkning ut fra sammenhengen.

Deltafunksjonen er ikke en funksjon, men en generalisert funksjon. Intuitivt er den en funksjon med verdi ∞ i 0, verdi 0 ellers, og med egenskapen

$$\int \delta(\gamma)\, d\gamma = 1 \tag{11.4}$$

Det finnes ingen funksjon med denne egenskapen. Generaliserte funksjoner kan defineres ved følger av funksjoner. Deltafunksjonen kan identifiseres med følgen δ_n av sannsynlighetstettheter tilsvarende normalfordelinger $\mathrm{N}(0, \sigma_n^2)$ hvor variansen $\sigma_n^2 \to 0$. Generaliserte funksjoner er viktige i teorien for differensialligninger og mer generelt i matematikk slik som her, men vi må nøye oss med denne korte presentasjonen.

Begrunnelsen for estimatoren $\hat{\gamma} = \mathrm{E}(\Gamma) = \mu$ følger fra risikoen

$$r = \mathrm{E}\,|\Gamma - \hat{\gamma}|^2 = \mathrm{E}\,|(\Gamma - \mu) + (\mu - \hat{\gamma})|^2 = \mathrm{Var}(\Gamma) + |\mu - \hat{\gamma}|^2 \tag{11.5}$$

som er minimalisert ved $\hat{\gamma} = \mu$. Risikoen r er da

$$r = \mathrm{Var}(\Gamma) = \mathrm{E}\,|\Gamma - \mu|^2 \tag{11.6}$$

Det er dermed mindre risiko dersom variansen er liten. Den resulterende usikkerheten i Γ vil typisk være et resultat av hvordan data er innsamlet og hva slags type data som er innsamlet. Design av eksperiment kan brukes som en samlebetegnelse for disse valgene. Konklusjonen her er da at et design hvor variansen til Γ blir så liten som mulig er å foretrekke. Intuitivt er dette rimelig fordi det tilsvarer at fordelingen til Γ har liten spredning. Designet har dermed resultert i liten usikkerhet.

En kvadratisk tapsfunksjon er sjelden realistisk, men den kan være en god approksimasjon nær minimum til en mer realistisk tapsfunksjon. Den brukes mye. En begrunnelse er at den gir enkle analytiske beregninger slik som illustrert over. En annen begrunnelse er gitt ved store talls lov som gir at forventningsverdien er grensen av empirisk middel ved tilfeldig trekning fra fordelingen. Estimatoren tilsvarende kvadrattap gir dermed et intuitivt rimelig estimat for verdien. Estimatet er i tillegg unik dersom det finnes.

Den siste tapsfunksjonen i Teorem 11.1 fører til medianen som den optimale estimatoren. Dette følger umiddelbart ved at vi definerer *medianen* ved

$$m = \mathrm{M}(\Gamma) = \operatorname*{argmin}_{\gamma} \mathrm{E}(|\Gamma - \gamma| - |\Gamma - \gamma_0|) \tag{11.7}$$

Medianen definert i ligning (11.7) avhenger ikke av valget av referanseparameteren γ_0. Den er innført fordi

$$||\Gamma - \gamma| - |\Gamma - \gamma_0|| \leq |\gamma - \gamma_0| < \infty \tag{11.8}$$

sikrer at både tapet l og risikoen r er begrenset. Ulikheten over følger av ulikheten $||a| - |b|| \leq |a - b|$, som er en konsekvens av trekantulikheten på formen $|a| = |(a - b) + b| \leq |a - b| + |b|$. Dersom $\mathrm{E}\,|\Gamma| < \infty$, så er medianen gitt enklere ved

$$m = \mathrm{M}(\Gamma) = \operatorname*{argmin}_{\gamma} \mathrm{E}(|\Gamma - \gamma|) \tag{11.9}$$

Dersom Γ er en kontinuerlig stokastisk variabel med kumulativ fordeling F, så er medianen $m = F^{-1}(0.5) = \gamma_{0.5}$. Notasjonen γ_α betegner her α *fraktilen* til Γ, dvs $\mathrm{P}(\Gamma \leq \gamma_\alpha) = \alpha$. Påstanden her er en konsekvens av det mere generelle

Teorem 11.2. Fraktilen $\hat{\gamma} = \gamma_\alpha$ er et optimal estimat til $\gamma = \Gamma(\omega)$ når tapet er

$$l = v(\gamma - \hat{\gamma}) - v(\gamma - \gamma_0), \quad v(x) = \begin{cases} a|x| & \text{når } x \le 0 \\ b|x| & \text{når } x > 0 \end{cases} \tag{11.10}$$

hvor $a, b > 0$ og $\alpha = a/(a+b)$.

Tilfellet $a = b = 1$ gir medianen som et spesialtilfelle. Vi vil bevise Teorem 11.2 med tilleggsantagelsene $\mathrm{E}\,|\Gamma| < \infty$ og en antagelse om at Γ har kontinuerlig fordeling. Den første antagelsen gjør at vi kan se på den enklere tapsfunksjonen $l = v(\gamma - \hat{\gamma})$ fordi andre ledd ikke avhenger av $\hat{\gamma}$. Vi antar også at $\gamma_\alpha > 0$. Da følger:

$$\begin{aligned} \mathrm{E}\, v(\Gamma - \gamma_\alpha) &= \mathrm{E}\Big\{[\Gamma \le \gamma_\alpha] a(\cancel{\gamma_\alpha} - \Gamma) + [\Gamma > \gamma_\alpha] b(\Gamma - \cancel{\gamma_\alpha})\Big\} \\ &\le \mathrm{E}\Big\{[(\Gamma \le 0) + \cancel{(0 < \Gamma \le \gamma_\alpha)}] a(0 - \Gamma) + \\ &\qquad [(\Gamma > 0) - \cancel{(0 < \Gamma \le \gamma_\alpha)}] b(\Gamma - 0)\Big\} \\ &= \mathrm{E} v(\Gamma) \end{aligned}$$

Den første likheten følger fordi $\alpha a = (1-\alpha)b$ og gir kanselleringen som antydet. Ulikheten følger etter neste kansellering fordi begge ledd er negative. Ulikheten over gjelder generelt for en vilkårlig fraktil til en vilkårlig stokastisk variabel. Dette gir

$$\mathrm{E}\, v(\Gamma - \gamma_\alpha) = \mathrm{E}\, v((\Gamma - \gamma) - (\gamma_\alpha - \gamma)) \le \mathrm{E}\, v(\Gamma - \gamma) \tag{11.11}$$

fordi $\gamma_\alpha - \gamma$ er α fraktilen til $\Gamma - \gamma$. Tilsammen beviser dette at $\hat{\gamma} = \operatorname{argmin}_\gamma \mathrm{E}\, v(\Gamma - \gamma) = \gamma_\alpha$. Det generelle beviset av Teorem 11.2 følger ved modifikasjon av argumentet her, men vi velger i stedet å se på et eksempel.

Eksempel 11.2 (Kjøpmannen Bør) Dette oppdiktede eksempelet omhandler kjøpmannen Bør Børson Jr. som holdt til i den oppdiktede Olderdalen i Sør-Trøndelag. Bør driver bygdas landhandleri og bestiller wienerbrød fra Johan Glein som drev Størens Bageri. Hvor mange wienerbrød skal Bør bestille til 17.mai feiringen i 1911?

Bør feilberegnet stygt siste 17.mai og gikk tom for wienerbrød allerede før klokken 12. Tapet var på $a_1 = 0.5$ kr per wienerbrød som kunne ha blitt solgt. Bør tror at minst 200 flere kunne blitt solgt, og har hatt mareritt om dette tapet på 100 kr med jevne mellomrom i snart et år.

Bør besluttet derfor å rådføre seg med sin tremenning Caspar Wessel som var oppkalt etter sin tipp-tippoldefar Caspar Wessel. Caspar synes dette var et interessant problem. Som sin tipp-tippoldefar var Caspar rask til å lage nye matematiske modeller og begrep når det var nødvendig. Han laget følgende tapsfunksjon etter litt fundering og diskusjon med Bør

$$l = \ell(\gamma, \hat{\gamma}) = \begin{cases} a(\gamma - \hat{\gamma}) & \text{når } \hat{\gamma} \le \gamma \\ b(\hat{\gamma} - \gamma) & \text{når } \gamma < \hat{\gamma} \end{cases} \tag{11.12}$$

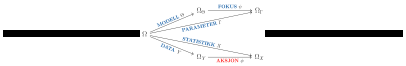
hvor $a = a_1 N$, $b = b_1 N$, $b_1 = 0.05$ kr og $N = 1000$ kunder. Her er γN lik antallet wienerbrød som kan selges ut fra kundegrunnlaget på $N = 1000$. Året før valgte Bør å bestille $\hat{\gamma} N = 200$ wienerbrød, men tror han burde ha bestilt $\gamma N = 400$. Caspar spør om Bør er helt sikker på dette. Bør går med på at det er usikkert, men at antallet i hvert fall må være mellom 350 og 450.

Med denne informasjonen gikk Caspar tilbake til arbeidsrommet for å fundere litt til. Caspar hadde lest de siste publikasjonene til matematikerne Karl Pearson og Corrado Gini om $\text{Beta}(s, f)$ fordelingen. Caspar antar derfor at usikkerheten i γ er gitt ved en stokastisk variabel $\Gamma \sim \text{Beta}(s, f)$. Han tolker informasjonen gitt av Bør til å begrunne $\mu = \mathrm{E}\,\Gamma = 40\% = s/(s+f)$ og $\sigma^2 = \mathrm{Var}\,\Gamma = (5\%/2)^2 = (sf)/[(s+f)^2(s+f+1)]$. Løsning av disse ligningene ga parametrene $s \approx 153.2$ og $f \approx 229.8$.

I Caspars modell inngikk nå tapet $l\ell(\gamma, \hat{\gamma})$ og usikkerheten i Γ var bestemt ved $\Gamma \sim \text{Beta}(s, l)$. Hvordan skulle så det optimale antall wienerbrød $N\hat{\gamma}$ bestemmes. Etter litt fundering konkluderer Caspar at han må velge $\hat{\gamma}$ slik at forventet tap $r = \mathrm{E}\,\ell(\Gamma, \hat{\gamma})$ må minimaliseres. Her er alt kjent, så det gjelder "bare" å bestemme $\hat{\gamma}$ som minimaliserer denne risikoen r.

Etter en dag kommer Caspar frem til, som i Teorem 11.2 at fraktilen $\hat{\gamma} = \gamma_\alpha$ er en optimal løsning hvor $\alpha = a/(a+b) \approx 90.91\%$. Etter timevis med arbeid finner Caspar at løsningen er $\hat{\gamma} = \gamma_\alpha \approx 43.35\%$ og sier til Bør at han må bestille 434 wienerbrød. Bør smiler fornøyd og sier at han allerede har bestilt 450 basert på fjorårets erfaringer og egne vurderinger.

Etter 17.mai feiringen dette året satt Bør igjen med 36 wienerbrød og et tap på $l = 36 * b_1 \approx 1.8$ kr. Det ideelle valget hadde vært å bestille $N\gamma = 414$ wienerbrød som hadde gitt $l = 0$. Caspar sitt forslag hadde gitt et tap $l = 20 * b_1 \approx 1$ kr. Bør likte dårlig å ha gitt fra seg 0.8 kr, men resulatet var vesentlig bedre enn året før.



Medianen til $\gamma_1, \ldots, \gamma_n$ er definert som medianen til Γ med den empiriske fordelingen $\mathrm{P}(\Gamma \in A) = \sum_i (\gamma_i \in A)/n$. Dette betegnes også som den *empiriske medianen* til disse data. Empirisk forventning og empirisk varians defineres tilsvarende fra den empiriske fordelingen. Mer generelt gir empirisk fordeling en definisjon av en empirisk størrelse tilsvarende enhver egenskap som kan defineres fra en fordeling.

Tapsfunksjonene i Teorem 11.1 krever ikke en antagelse om at utfallsrommet $\Omega_\Gamma = \mathbb{R}$. De gir mening direkte når Ω_Γ er et vektorrom og $|\gamma - \gamma_0|$ er lengden til vektoren $\gamma - \gamma_0$. Utledningen som gir variansen som det minste mulige kvadrattapet er gyldig direkte når lengden $|\gamma|$ til en vektor γ er gitt fra et indreprodukt ved $|\gamma|^2 = \gamma^* \gamma$. Mer generelt kan lengden $|\gamma - \gamma_0|$ erstattes av en avstand $d(\gamma, \gamma_0)$. Den resulterende forventningsverdien kan da kalles *Frechet-forventning* og den resulterende median kan kalles *geometrisk-median*. Disse er nyttige for eksempel i tilfellet hvor Ω_Γ er *sirkelen* $S = \{(\gamma_1, \gamma_2) \,|\, \gamma_1^2 + \gamma_2^2 = 1\}$, *sfæren* $S_2 = \{(\gamma_1, \gamma_2, \gamma_3) \,|\, \gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1\}$, *torusen* $T = S \times S$, eller andre mer eksotiske geometriske objekt. Avstanden mellom to punkter er her definert som lengden til korteste

kurve mellom punktene inneholdt i Ω_Γ. Statistikk for retninger gir S eller S_2 som naturlige utfallsrom.

11.2 Statistiske modeller kontra sannsynlighets modeller

Anta at i en urne med 100 kuler, så er 37 hvite. Sannsynlighetene for de forskjellige mulige trekningene kan beregnes ved å spesifisere en sannsynlighetsmodell. F eks kan en slik modell begrunne at sannsynligheten for å trekke to kuler som er hvite er lik $(37/100) \cdot (36/99)$.

Anta så at antall hvite kuler er en ukjent størrelse θ. To kuler trekkes tilfeldig og resultatet er to hvite kuler. En statistisk modell kan settes opp for å trekke konklusjoner om θ med utgangspunkt i observasjonene. Vi skal se på konklusjoner av tre typer:

Punktestimering: Med utgangspunkt i observasjonene beregner vi et estimat $w = \hat{\theta}$ for den θ-verdien som passer best med observasjonene, f eks $w = 100 \cdot (1+1)/2$. Dette er punktestimering, og konklusjonen er at vi gjetter på at $\theta = 100$. Det kan gis mange argument for at dette er en dårlig gjetning, men det hele kommer an på den gitte konkrete situasjonen og tilgjengelig informasjon.

Intervallestimering: Vi finner et estimat s for et område hvor vi finner de θ verdiene som passer best med observasjonene, f eks $s = \{50, 51, \ldots, 100\}$. I tilfellet hvor θ er et reelt tall, så velger vi oftest å se på et område gitt ved et intervall $s = [a, b]$. Dette er intervallestimering, og konklusjonen i eksemplet er at vi gjetter på at $\theta \in [50, 100]$.

Hypotesetesting: Utgangspunktet er en gitt hypotese. Hypotesen kan f eks være at det er minst 50 hvite kuler. Konklusjonen er enten at hypotesen forkastes eller at den aksepteres, dvs det er bare to mulige konklusjoner. Dersom hypotesen aksepteres i eksemplet, så er konklusjonen at den gitte hypotesen $\theta \in \{50, 51, \ldots, 100\}$ er sann.

Definisjon 11.2 (*En statistisk modell*) er gitt ved en familie $\{\mathrm{P}^\theta \mid \theta \in \Omega_\Theta\}$ av fordelinger for et utfallsrom Ω.

Modellparametermengden Ω_Θ er mengden av mulige verdier til modellparameteren θ. Dette er en generalisering av den tilsvarende sannsynlighetsmodell hvor vi kun så på én fordeling P for Ω. Modellen er fullstendig spesifisert dersom vi for hver parameter θ har spesifisert fordelingen P^θ. Vi gjentar hovedpoenget

Forskjellen på en statistisk modell i forhold til en sannsynlighetsmodell er gitt ved innføringen av modellparametermengden og den tilhørende familien av mulige fordelinger. En statistisk modell brukes til å trekke konklusjoner om en mulig fordeling ved hjelp av data, mens en sannsynlighetsmodell brukes til å trekke konklusjoner om de mulige observasjonene når en fordeling er gitt.

Fordi utfallsrommet Ω er beholdt ved overgangen til en statistisk modell, så kan definisjonene av de sentrale begrepene i sannsynlighetsmodellen overføres direkte. Spesielt er en observator T en funksjon $T : \Omega \to \Omega_T$. Fordelingen P_T^θ til T er gitt ved

$$P_T^\theta(A) = P^\theta(T \in A).$$

Resultatet er en familie $\{P_T^\theta\}$ av fordelinger for Ω_T. I en statistisk modell har dermed hver observator en familie av mulige fordelinger.

11.3 Punktestimering

La $\{P^\theta\}$, $\theta \in \Omega_\Theta$, være en statistisk modell. En estimator W for parameteren $\gamma = \psi(\theta)$ er en observator for Ω med verdier i en mengde Ω_W som inneholder parametermengden Ω_Γ. Det som skiller W fra en generell observator er altså tilleggsantagelsen $\Omega_W \supset \Omega_\Gamma$.

Eksempel 11.3 (Estimat for θ i en uniform fordeling) Anta at $X = (X_1, \ldots, X_n)$ hvor X_1, …, X_n er et tilfeldig utvalg fra den uniforme fordelingen på intervallet $[0, \theta]$. Vår oppgave er å finne en estimator for θ. Uavhengigheten (tilfeldig utvalg!) gir at fordelingen P_X^θ til X er gitt av simultantettheten

$$f_X(x; \theta) = f_{X_1}(x_1; \theta) \cdots f_{X_n}(x_n; \theta).$$

Den uniforme fordelingen er gitt ved tettheten

$$f_Y(y; \theta) = \frac{1}{\theta};\ \theta > 0;\ 0 \le y \le \theta.$$

Antagelsen om at vi har et utvalg fra den uniforme fordelingen betyr at $f_{X_i} = f_Y$ for alle i. Denne siste antagelsen kunne også vært skrevet som $X_i \sim \mathrm{U}[0, \theta]$. Fordelingen P_X^θ er dermed fullstendig spesifisert av antagelsene. Parametermengden i denne modellen er $\Omega_\Theta = \{\theta \,|\, \theta > 0\}$.

En mulig estimator for θ er gitt ved

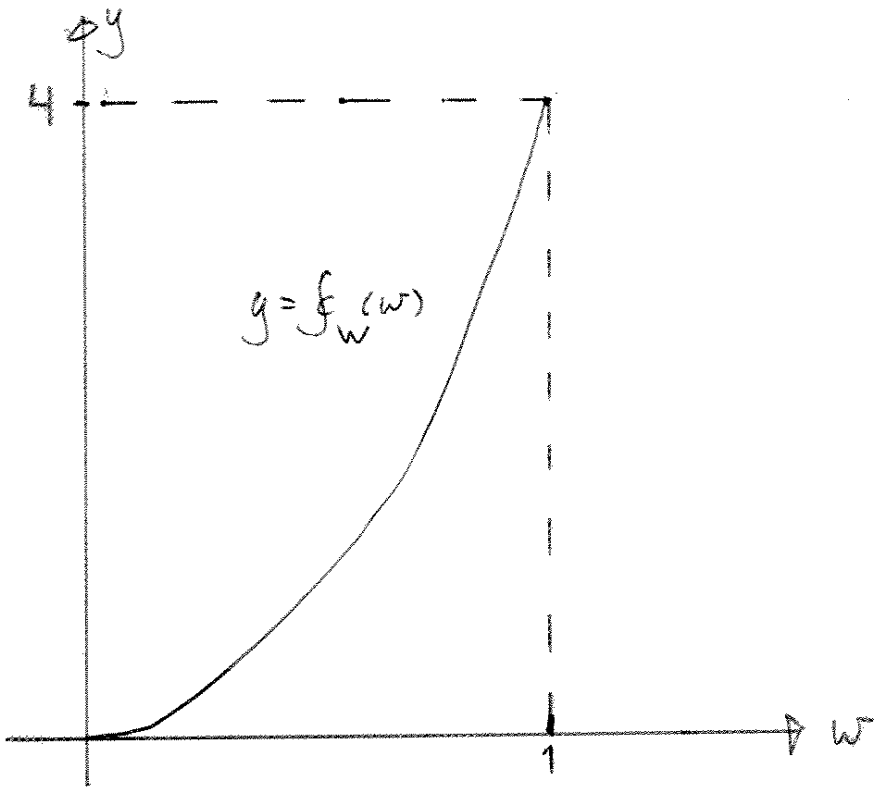
$$W = \mathrm{maks}\{X_1, \ldots, X_n\}.$$

Her er $W = \phi(X)$ hvor $\phi(x) = \mathrm{maks}\{x_1, \ldots, x_n\}$. Dermed er W en funksjon av X, dvs W er basert på X. Estimatoren er en statistikk.

Dersom utfallet er ω, så er estimatet lik $w = W(\omega) = \phi(X(\omega))$. Når utfallet $x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$ er gitt, så kan estimatet $w = \mathrm{maks}\{x_1, \ldots, x_n\}$ beregnes. Poenget med dette siste er at beregningen kan gjøres uten at utfallet ω er kjent. Fordi W er basert på X kan w beregnes når x er gitt.

La $X_{(1)}, \ldots, X_{(n)}$ være ordningsobservatorene. Vi har da at $W = X_{(n)}$. Vi har tidligere funnet fordelingen til hver enkelt $X_{(i)}$. Spesielt er fordelingen til W gitt ved tettheten

$$f_W(w) = f_{X_{(n)}}(w) = n f_Y(w) F_Y(w)^{n-1}$$



Figur 11.2: Tettheten til observatoren W ved $n = 4$ og $\theta = 1$

Her er $F_Y(w) = \int_0^w f_Y(y)dy = w/\theta$ for $0 \leq w \leq \theta$. Dette gir

$$f_W(w) = \frac{n}{\theta^n} w^{n-1};\ \theta > 0;\ 0 \leq w \leq \theta.$$

Vi kan skissere tettheten til W for hver θ og n verdi. Et eksempel er gitt på figur 11.2. Spesielt gir en skisse av tettheten et inntrykk av hvordan fordelingen er lokalisert i forhold til verdien θ som skal estimeres.

Et numerisk lokaliseringsmål er gitt av forventningsverdien til W

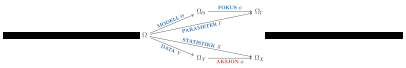
$$\mathrm{E}^\theta W = \int_0^\theta w \cdot \frac{n}{\theta^n} w^{n-1} = \frac{n}{(n+1)\theta^n} w^{n+1} \Big|_0^\theta = \frac{n}{n+1}\theta.$$

Fordi $\mathrm{E}^\theta W \neq \theta$ sier vi at W ikke er forventningsrett. Vi kan lage en estimator V som er forventningsrett:

$$V = \frac{n+1}{n} W = \frac{n+1}{n} \,\text{maks}\{X_1, \ldots, X_n\}$$

gir

$$\mathrm{E}^\theta V = \frac{n+1}{n} \mathrm{E}^\theta W = \frac{n+1}{n} \frac{n}{n+1}\theta = \theta.$$



Mer generelt er vi være interessert i en estimator for en funksjon $\tau(\theta)$ av parameteren θ.

La Ω_Θ være parametermengden i en statistisk modell. En estimand τ er en funksjon $\tau : \Omega_\Theta \to \Omega_\tau$. En estimand er mao det samme som en parameter.

Et eksempel på en estimand er gitt ved $\tau(\theta) = \theta$, hvor $\Omega_\tau = \Omega_\Theta$. Et annet eksempel er gitt ved forventningsverdien til en tilfeldig variabel, dvs $\mu = \mathrm{E}\, X$. Notasjonen betyr at $\mu(\theta) = \mathrm{E}^\theta X$ for alle θ, og $\Omega_\mu = \mathbb{R}$. Denne konvensjonen tilsvarer at likhet $f = g$ mellom funksjoner betyr at $f(x) = g(x)$ for alle x i definisjonsområdet $D_f = D_g$. Konvensjonen med å utelate θ vil bli brukt også i det følgende.

En estimator er en observator. La $W : \Omega \to \Omega_W$ være en observator og la $\tau : \Omega_\Theta \to \Omega_\tau$ være en estimand. Da er W en estimator for estimanden τ dersom $\Omega_\tau \subset \Omega_W$.

Kravet $\Omega_\tau \subset \Omega_W$ er et meget svakt krav, men det gir dog en sammenheng mellom estimatoren og estimanden. Det er et svakt krav fordi W ikke trenger å ta alle verdier i Ω_W, og τ trenger ikke ta alle verdier i Ω_τ. F eks kan $\Omega_\tau = (0, \infty)$, $\Omega_W = \mathbb{R}$ og $W(\omega) = -10$, dvs estimatoren gir alltid estimatet -10. Denne estimatoren bommer alltid på estimanden τ som i dette tilfellet alltid er positiv.

Forventningsretthet er et kriterium som sikrer at en estimator "sikter på blinken".

Estimatoren W er *forventningsrett* for estimanden τ dersom

$$\boxed{\mathrm{E}\, W = \tau}$$

I tilfellet $W = (W_1, \ldots, W_n)$ for tilfeldige variable W_i betyr dette at $\mathrm{E}^\theta W_j = \tau_j(\theta)$ for alle θ og j. Det antas i det overstående at Ω_W er et vektorrom slik at forventningsverdien $\mathrm{E}\, W$ er definert.

Det er ønskelig at en estimator er forventningsrett fordi dette sikrer at estimatorens fordeling er lokalisert omkring verdien som skal estimeres. Vi har følgende reformulering

> En estimator er forventningsrett dersom tyngdepunktet til estimatorens fordelingen er lik estimanden.

I analysen av en gitt statistikk $x_1, \ldots, x_n$ er det vanlig å anta at statistikken er en realisering av et tilfeldig utvalg, dvs at

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$$

hvor $X_1, \ldots, X_n$ er et tilfeldig utvalg. Forventningen og variansen til utvalget er to meget brukte estimatorer.

Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg. Utvalgets forventningsverdi er

$$\boxed{\overline{X} = \frac{X_1 + \cdots + X_n}{n}}$$

og for $n \geq 2$ er utvalgets varians

$$\boxed{S^2 = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}.}$$

Både $\overline{X}$ og S^2 er tilfeldige variable. De er basert på utvalget $X_1, \ldots, X_n$ og estimatene $\overline{x}$ og s^2 kan dermed beregnes når en statistikk $x_1, \ldots, x_n$ er gitt. Forventningsverdien μ og variansen σ^2 til fordelingen $P^{\theta}_{X_j}$ er begge funksjoner av θ, dvs de er estimander. Dermed er $\overline{X}$ og S^2 estimatorer for hhv μ og σ^2.

Grunnen til at vi dividerer med $n-1$ i definisjonen av S^2 er gitt av det generelle resultatet:

Et tilfeldig utvalgs forventningsverdi og varians er forventningsrette estimatorer for hhv forventningsverdi og varians.

Beviset for $\overline{X}$ bruker ikke antagelsen om uavhengighet

$$\mathrm{E}\,\overline{X} = \mathrm{E}\,\frac{X_1 + \cdots + X_n}{n} = \frac{\mathrm{E}\,X_1 + \cdots + \mathrm{E}\,X_n}{n} = \frac{\mu + \cdots + \mu}{n} = \mu.$$

Uavhengigheten gjør at vi også kan beregne variansen

$$\mathrm{Var}\,\overline{X} = \mathrm{Var}\,\frac{X_1 + \cdots + X_n}{n} = \frac{\mathrm{Var}\,X_1 + \cdots + \mathrm{Var}\,X_n}{n^2} = \frac{1}{n}\sigma^2.$$

Merk: Variansen til estimatoren $\overline{X}$ går mot null når utvalgets størrelse vokser. Dette tolkes som at presisjonen til estimatoren er bedre jo større utvalget er.

Beviset for at S^2 er forventningsrett er litt mer komplisert og bygger på formelen

$$\boxed{S^2 = \frac{n}{n-1}\left(\overline{X^2} - \overline{X}^2\right).}$$

Denne beregningsformelen gir ofte også den enkleste måten å beregne S^2. Beviset av beregningsformelen følger av identiteten

$$\overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2$$

som er et spesialtilfelle av beregningsformelen $\mathrm{Var}\,Z = \mathrm{E}\,Z^2 - (\mathrm{E}\,Z)^2$ for variansen. I det overstående er

$$\overline{X^2} = (X_1^2 + \cdots + X_n^2)/n.$$

Beregningsformelen gir følgende resultat.

Et utvalgs forventningsverdi og varians er en funksjon av summen og kvadratsummen til utvalget.

Likheten $\operatorname{Var} Z = \operatorname{E} Z^2 - (\operatorname{E} Z)^2$ gir oss også at $\operatorname{E} X_i^2 = \operatorname{Var} X_i + (\operatorname{E} X_i)^2 = \sigma^2 + \mu^2$, som gir

$$\operatorname{E} \overline{X^2} = \frac{(\sigma^2 + \mu^2) + \cdots + (\sigma^2 + \mu^2)}{n} = \sigma^2 + \mu^2.$$

Forventningsverdien til $\overline{X}^2$ finnes ved

$$\operatorname{E} \overline{X}^2 = \operatorname{Var} \overline{X} + (\operatorname{E} \overline{X})^2 = \frac{1}{n}\sigma^2 + \mu^2.$$

Beviset av at S^2 er forventningsrett kan dermed konkluderes ved innsetting av de foregående resultater i beregningsformelen for S^2

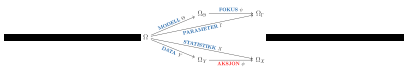
$$\begin{aligned}\operatorname{E} S^2 &= \operatorname{E} \frac{n}{n-1}[\overline{X^2} - \overline{X}^2] = \frac{n}{n-1}[\operatorname{E} \overline{X^2} - \operatorname{E} \overline{X}^2] \\ &= \frac{n}{n-1}[\sigma^2 + \mu^2 - \frac{1}{n}\sigma^2 - \mu^2] = \frac{n}{n-1}[\sigma^2 - \frac{1}{n}\sigma^2] = \sigma^2.\end{aligned}$$

Eksempel 11.4 (Estimering av forventning og varians til en Bernoulli fordeling) La X_1, …, X_n være et tilfeldig utvalg fra en $\mathrm{B}(p)$-fordeling. Vi antar at $n \geq 2$. Fordelingen har forventningsverdi p og varians $p(1-p)$. Både forventningsverdien og variansen er funksjoner av parameteren p og er dermed estimander. Identifiser dette med en suksess/fiasko modell, dvs hendelsen $(X_j = 1)$ tolkes som suksess i forsøk j. La Y være antall suksesser. Da følger det at Y/n er en forventningsrett estimator for parameteren p. Dette er en konsekvens av at p er forventningsverdien til fordelingen og av at $Y = \sum X_i$.

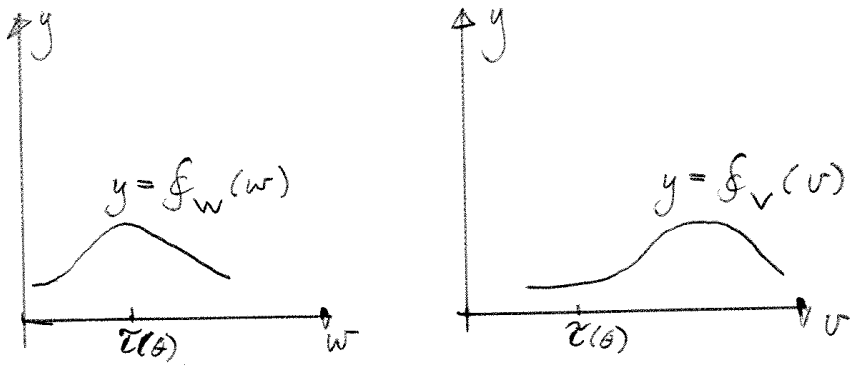
En annen forventningsrett estimator for p er gitt av X_1. Estimatoren X_1 er ikke basert på Y: Selv om vi har gitt antall suksesser y, så kan vi ikke generelt avgjøre om forsøk nummer 1 var en suksess eller ikke. Både estimatoren Y/n og X_1 er imidlertid basert på $X = (X_1, \ldots, X_n)$. I følge kritererier som skal diskuteres senere vil det følge at X_1 er en dårligere estimator enn Y/n.

Vi kan også finne en forventningsrett estimator for variansen som er basert på Y. Fordi Bernoulli-variablene oppfyller $X_i^2 = X_i$ finner vi at utvalgets varians er basert på Y:

$$S^2 = \frac{n}{n-1}[\overline{X^2} - \overline{X}^2] = \frac{n}{n-1}[\overline{X} - \overline{X}^2] = \frac{n}{n-1}\frac{Y}{n}(1 - \frac{Y}{n}).$$



Det finnes også andre posisjonsmål for en fordeling, f eks medianen. Dette betyr at forventningsretthet bare er et av mange mulige mål for at posisjonen til estimatoren stemmer overens med verdien som skal estimeres.



Figur 11.3: Fordelingen til estimatoren W har bedre posisjon enn fordelingen til V i forhold til $\tau(\theta)$.

11.4 Nøyaktigheten til en punktestimator

Vi begynner med en oppsummering av de viktigste definisjonene fra forrige kapittel. Utgangspunktet er en statistisk modell $\{P^\theta\}$ med en parametermengde Ω_Θ. En estimator W er en observator. Anta at $\tau : \Omega_\Theta \to \Omega_\tau \subset \Omega_W$. Da er W en estimator for τ. Anta at utfallet er ω. Da gir estimatoren W estimatet $w = W(\omega)$. Estimatoren W er forventningsrett (unbiased) dersom $\mathrm{E}^\theta W = \tau(\theta)$ for alle θ.

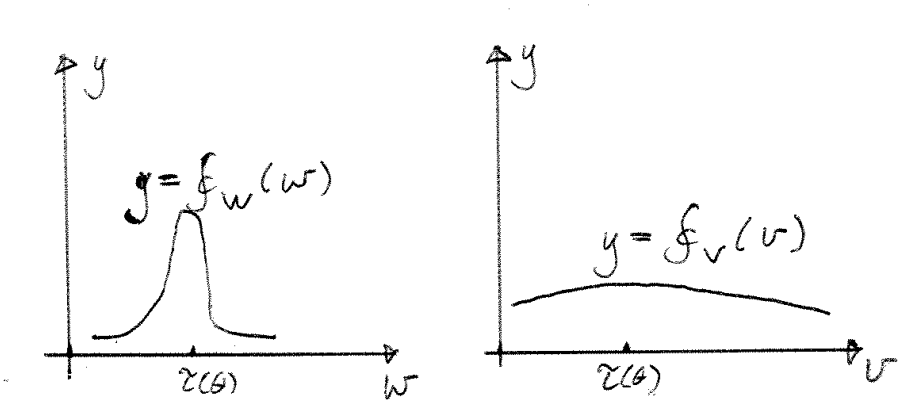
Anta at V, W er tilfeldige variable med tettheter som på figur 11.3. På figurene fremgår det at W's fordeling er lokalisert omkring $\tau(\theta)$, mens dette ikke gjelder V's fordeling. Dermed vil vi konkludere at W er en bedre estimator for $\tau(\theta)$ enn V. Dette er et geometrisk argument.

En analytisk begrunnelse er gitt ved at vi foretrekker en forventningsrett estimator fremfor andre estimatorer som ikke er forventningsrette. En estimator W er jo forventningsrett for τ dersom τ er tyngdepunktet til W's fordeling.

> Tenk et øyeblikk på W som om den beskriver resulatet av at vi skyter på blink. Vi forsøker å treffe τ. At W er forventningsrett kan sammenlignes med at vi sikter på riktig punkt.

Anta at V, W er tilfeldige variable med tettheter som på figur 11.4. Her foretrekkes W fremfor V fordi fordelingen til W har mindre spredning enn fordelingen til V. Dette kan reformuleres som at W er mer nøyaktig enn V. Vi vil bruke variansen til W som et mål for spredningen til W.

> Tenk igjen på W som om den beskriver resulatet av at vi skyter på blink etter τ. At W har liten varians kan sammenlignes med at de fleste skuddene treffer i nærheten av blinken.



Figur 11.4: Fordelingen til estimatoren W har mindre spredning enn fordelingen til V.

Anta at V og W er forventningsrette estimatorer for τ. Da er W nøyaktigere enn V dersom

$$\mathrm{Var}^\theta W < \mathrm{Var}^\theta V \text{ for alle } \theta.$$

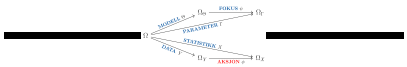
Den relative nøyaktigheten til W i forhold til V er $\mathrm{Var}\, W/\mathrm{Var}\, V$.

Her er det, som alltid i en statistisk modell, underforstått at Var avhenger av parameteren θ. Konvensjonen med å unnlate å skrive opp θ avhengigheten brukes også i andre sammenhenger. Begrepet "relativ effisiens" er synonymt med relativ nøyaktighet. Alternativt benyttes også "relativ effektivitet".

Eksempel 11.5 (Mer nøyaktig estimator) Anta at $Y_i \sim \mathrm{Exp}(1/\theta)$ og at $Y_1, \ldots, Y_n$ er uavhengige. Her er $\mathrm{E}\, Y_i = \theta$. Dette gir at $W = \overline{Y}$ og $V = Y_1$ er forventningsrette for θ. Vi beregner variansene ved

$$\mathrm{Var}\, V = \mathrm{Var}\, Y_1 = \theta^2,\ \mathrm{Var}\, W = \mathrm{Var}\, \overline{Y} = \mathrm{Var}\, \frac{Y_1 + \cdots\ Y_n}{n} = \frac{1}{n^2} \cdot n\theta^2 = \frac{1}{n}\theta^2.$$

Dette gir at $\mathrm{Var}\, W < \mathrm{Var}\, V$ når $n \geq 2$, dvs W er mer nøyaktig enn V når $n \geq 2$.



Vi velger følgende definisjon:

En estimator W for τ er en *beste estimator* dersom W er forventningsrett og $\mathrm{Var}\, W \leq \mathrm{Var}\, V$ for alle forventningsrette estimatorer V.

En beste estimator er dermed en forventningsrett estimator som har minst like liten varians som alle andre forventningsrette estimatorer. Som tidligere skal ulikheten $\mathrm{Var}\, W \leq \mathrm{Var}\, V$ tolkes som at

$$\mathrm{Var}^\theta W \leq \mathrm{Var}^\theta V \text{ for alle } \theta.$$

11.5 Konsistente estimatorer

La W_n være en estimator for τ. Da er W_n en *konsistent estimator* dersom

$$W_n \xrightarrow{P} \tau(\theta) \text{ for alle } \theta.$$

Notasjonen $\xrightarrow{P}$ betyr konvergens i sannsynlighet. Dette betyr at for alle $\epsilon > 0$ vil

$$P(|W_n - \tau| < \epsilon) \to 1 \text{ når } n \to \infty.$$

I det overstående er både P og τ avhengig av θ og det er underforstått at påstanden skal gjelde for alle θ. En estimator W_n er mao konsistent for τ dersom W_n konvergerer i sannsynlighet mot τ.

Dersom W_n er forventningsrett for τ, så gir Tsjebysjevs ulikhet at

$$P(|W_n - \tau| < \epsilon) > 1 - \frac{\operatorname{Var} W_n}{\epsilon^2}.$$

Dermed har vi resultatet:

Dersom W_n er forventningsrett og $\operatorname{Var} W_n \to 0$, så er W_n konsistent.

Det finnes eksempler som viser at det motsatte ikke er tilfelle: W_n kan være konsistent uten at $\operatorname{Var} W_n \to 0$,

Et meget viktig eksempel på en konsistent estimator er gitt av et tilfeldig utvalgs forventningsverdi $\overline{X}$. Her er $W_n = \overline{X}$ hvor n er utvalgets størrelse. Det følger at W_n er en konsistent estimator for forventningsverdien $\mu = \mathrm{E}^\theta X_i$ dersom variansen $\sigma^2 = \mathrm{Var}^\theta X_i < \infty$. Begrunnelsen er at $\operatorname{Var} W_n = \sigma^2/n \to 0$. Resultatet er viktig i forhold til tolkningen av sannsynlighetsteorien og fortjener et eget navn.

Store talls lov (svak versjon) sier at et tilfeldig utvalgs forventningsverdi $\overline{X}$ er en konsistent estimator for forventningsverdien.

11.6 Tilstrekkelige observatorer

R. A. Fisher ga følgende definisjon av tilstrekkelighet:

En observator er tilstrekkelig dersom den inneholder all relevant informasjon.

Dette er ikke en matematisk definsjon. Formuleringen gir rom for flere mulige presiseringer. Vi velger følgende:

Anta at X er en observator med tetthet $f_X(x;\theta)$. Observatoren T er en *tilstrekkelig observator* for θ i forhold til X dersom $T = \phi(X)$ og

$$f_X(x;\theta) = g(x) \cdot h(\phi(x);\theta).$$

Idé: g er uavhengig av θ, så denne er gitt uten at θ er kjent. All informasjon om θ-avhengigheten er i funksjonen h. Konklusjoner om θ finnes via konklusjoner om h, og denne avhenger kun av x gjennom $\phi(x)$.

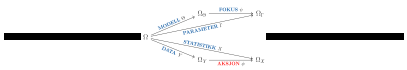
All informasjon om θ inneholdt i X er dermed også inneholdt i $T = \phi(X)$.

Mer generelt defineres T til å være tilstrekkelig for θ i forhold til X dersom fordelingen til X gitt T er uavhengig av parameteren θ. Det kan vises at dette er i samsvar med definisjonen vi valgte. Den elementære definisjonen av betinget sannsynlighet er imidlertid ikke generell nok til at vi kan gjennomføre beviset her. Begrepet "suffisient" er synonymt med tilstrekkelighet.

Eksempel 11.6 (Tilstrekkelig observator for p) Vi skal vise at antall suksesser Y i uavhengige Bernoulli-forsøk $X_1, \ldots, X_n$ er en tilstrekkelig observator for suksess-sannsynligheten p. Vi har at

$$\begin{aligned} f_X(x;p) &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_1}(1-p)^{1-x_1} = p^{\sum x_i}(1-p)^{n-\sum x_i} \\ &= p^y(1-p)^{n-y} = g(x)h(\phi(x);p), \end{aligned}$$

hvor $g(x) = 1$, $y = \phi(x) = \sum x_i$, og $h(y;p) = p^y(1-p)^{n-y}$. Dermed har vi bevist at $Y = \phi(X)$ er tilstrekkelig.



Eksempel 11.7 (Tilstrekkelig observator for μ og for (μ, σ^2)) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en $N(\mu, \sigma^2)$-fordeling.

Anta at σ er kjent. Dette betyr at parametermengden er

$$\Omega_\Theta = \{\mu \,|\, -\infty < \mu < \infty\}.$$

Vi skal finne en tilstrekkelig estimator for X og μ. Tettheten til X er

$$\begin{aligned} f_X(x;\mu) &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \overset{!}{=} \left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\right)^n e^{-\frac{n\overline{x^2}-2\mu n\overline{x}+n\mu^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\right)^n e^{-\frac{n\overline{x^2}}{2\sigma^2}} \cdot e^{-\frac{-2\mu n\overline{x}+n\mu^2}{2\sigma^2}}. \end{aligned}$$

Overgangen $\overset{!}{=}$ følger ved at

$$\sum(x_i-\mu)^2 = \sum x_i^2 - 2\mu\sum x_i + n\mu^2 = n\overline{x^2} - 2\mu n\overline{x} + n\mu^2.$$

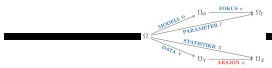
Vi har faktorisert f_X i en faktor som er uavhengig av μ og i en faktor som kun avhenger av x gjennom $\overline{x}$. Dermed kan vi konkludere at utvalgets forventning $\overline{X}$ er en tilstrekkelig observator for μ når variansen σ^2 er kjent.

Anta så at både variansen og forventningsverdien er ukjente, dvs parametermengden er

$$\Omega_\Theta = \{(\mu, \sigma^2) \,|\, \sigma > 0, -\infty < \mu < \infty\}.$$

Da er $T = (\overline{X}, S^2)$ en tilstrekkelig observator. Beviset følger av foregående regning når vi i tillegg bruker at

$$S^2 = \frac{n}{n-1}\left(\overline{X^2} - \overline{X}^2\right) \text{ som gir } \overline{X^2} = \frac{n-1}{n}S^2 + \overline{X}^2.$$



Anta at parameteren i en statistisk modell er på formen $\theta = (\theta_1, \theta_2)$. Når θ_2 er kjent betyr dette at parameterrommet er redusert fra Ω_Θ til

$$\Omega_1 = \{\theta_1 \,|(\theta_1, \theta_2) \in \Omega_\Theta\}.$$

Vi har et generelt tilstrekkelighetsprinsipp:

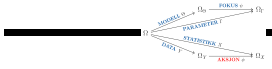
Dersom T er en tilstrekkelig observator, så skal estimatoren være basert på T.

Det er viktig å merke seg at overstående skal gjelde for alle tilstrekkelige observatorer.

Eksempel 11.8 (Diskvalifisering av en estimator og datareduksjon) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en B(p)-fordeling. Da er $\overline{X}$ og X_1 begge forventningsrette estimatorer for p basert på det tilfeldige utvalget. Estimatoren $\overline{X}$ er en tilstrekkelig observator, så denne inneholder all informasjon om p. Estimatoren X_1 diskvalifiseres fordi X_1 ikke er basert på $\bar{X}$. Derfor foretrekker vi estimatoren $\overline{X}$ fremfor X_1.

Enkeltobservasjonen $\overline{x}$ (antall suksess delt på n) inneholder dermed like mye informasjon om p som de n observasjonene $x_1, \ldots, x_n$. Dette viser at en passelig valgt tilstrekkelig observator kan lede til **datareduksjon**. Observatoren $X = (X_1, \ldots, X_n)$ er også tilstrekkelig. Vi velger å trekke slutninger om p med utgangspunkt i $\bar{x}$ fremfor x fordi det er enklere med bare ett tall.

Det kan virke utrolig at all informasjonen er samlet i et tall. Forklaringen er antagelsen om at observasjonene er gitt ved en så enkel modell. Dersom det f eks ikke er uavhengighet mellom X_i'ene, så er fortsatt X tilstrekkelig, men statusen til $\overline{X}$ er uklar.



Idéen om tilstrekkelighet er en av de mest elegante i statistikken. Vi har sett på to viktige konsekvenser. Den ene konsekvensen er en metode som generelt diskvalifiserer enkelte estimatorer. Den andre konsekvensen er at det i noen tilfeller er mulig å oppnå en kraftig reduksjon av datamengden som skal analyseres. Begge konsekvensene er imidlertid avhengig av **antagelsen** om riktig statistisk modell, og denne antagelsen kan være gal!

11.7 Metoder for å finne estimatorer

Så langt har vi definert hva en estimator er, og sett på noen egenskaper en estimator bør ha. Nå skal vi presentere to metoder som kan brukes til å finne estimatorer. Vi starter med en repetisjon av noen sentrale begrep.

En estimator er en observator W. Når dataene fra eksperimentet foreligger, så benyttes estimatoren til å beregne estimatet w.

Spesielt har estimatoren W en tilhørende fordeling. Denne kan brukes til å vurdere om estimatoren sikter riktig og om estimatene har liten spredning. Dette vurderes kvantitativt ved å se på om W er forventningsrett, $\mathrm{E}\, W = \tau$, og ved å se på om variansen $\mathrm{Var}\, W$ er liten.

I tillegg har vi diskutert konsistens, $W_n \xrightarrow{P} \tau$, og tilstrekkelighet. En observator $T = \phi(X)$ er tilstrekkelig for θ i forhold til X dersom vi har en faktorisering

$$f_X(x;\theta) = g(x)h(t;\theta), \quad \text{hvor } t = \phi(x).$$

Begrepet ga:

1. En metode for å diskvalifisere enkelte estimatorer.

2. I enkelte tilfeller en kraftig reduksjon av datamengden.

Et tilfeldig utvalgs forventning og varians er forventningsrette estimatorer for forventningen og variansen. Dette gjelder uansett hvilken statistisk modell som er valgt. Estimatorer for andre estimander vil typisk avhenge av modellen.

Vi skal se på momentmetoden (*method of moments*) og rimelighetsmetoden (*maximum likelihood method, sannsynlighetsmaksimeringsmetoden*). Momentmetoden gir oss en estimator ved løsning av et ligningssystem. Den kan brukes når f eks rimelighetsmetoden blir for komplisert.

11.8 Momentestimatoren

La en statistisk model $\{P^\theta\}$ være gitt. Anta at $x_1, \ldots, x_n$ er resultatet av en trekning fra fordelingen til X.

Momentestimatet $\tilde{\tau}$ for estimanden $\tau : \Omega_\Theta \to \Omega_\tau$ er gitt ved at $\tilde{\tau} = \tau(\tilde{\theta})$ hvor

$$\boxed{\mathrm{E}^{\tilde{\theta}}\, X^j = \overline{x^j} \qquad j = 1, \ldots, k.}$$

Ved løsning av ligningene velges k, hvis det er mulig, slik at ligningssystemet har en unik løsning. Det typiske er at $\theta = (\theta_1, \ldots, \theta_m)$, hvor hver θ_j er et reelt tall, og man velger $k = m$.

En reformulering av momentmetoden er gitt ved:

Momentestimatoren finnes ved å kreve at fordelingens første momenter er lik verdien til et tilfeldig utvalgs første momenter.

Med $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$ betyr dette f eks at oppgaven er redusert til å løse to ligninger med to ukjente

$$\mathrm{E}^{\tilde{\theta}_1, \tilde{\theta}_2}\, X = \frac{1}{2}(x_1 + x_2) \qquad \mathrm{E}^{\tilde{\theta}_1, \tilde{\theta}_2}\, X^2 = \frac{1}{2}({x_1}^2 + {x_2}^2).$$

Eksempel 11.9 (Momentmetoden) Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra tettheten

$$f_X(x; \theta) = (\theta + 1)y^\theta \qquad 0 < x < 1.$$

Vi skal finne momentestimatoren for $\theta > -1$. Fordi vi bare har et ukjent tall trenger vi bare 1.moment, dvs forventningsverdien

$$\mathrm{E}^\theta\, X = \int_0^1 x(\theta + 1)x^\theta dx = \frac{\theta + 1}{\theta + 2} x^{\theta+2} \Big|_0^1 = \frac{\theta + 1}{\theta + 2}$$

Vi trenger bare bruke den første momentligningen:

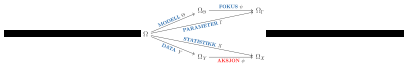
$$\mathrm{E}^\theta\, X = \frac{\theta + 1}{\theta + 2} = \overline{x} \qquad \text{som gir}$$

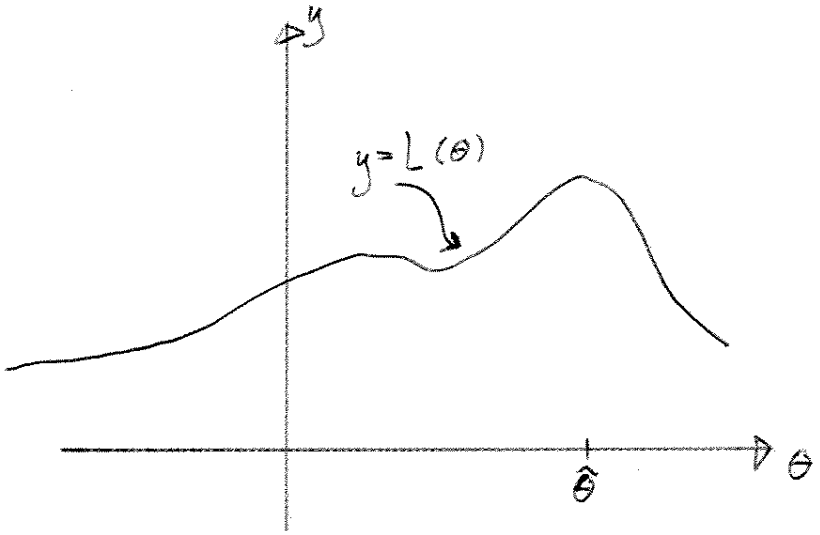
$$\theta + 1 = \overline{x}(\theta + 2),\ \theta(1 - \overline{x}) = 2\overline{x} - 1,\ \theta = \frac{2\overline{x} - 1}{1 - \overline{x}}.$$

Momentestimatoren er dermed

$$W = \frac{2\overline{X} - 1}{1 - \overline{X}}.$$

Merk: Metoden ga en estimator uten mye regning. I utgangspunktet var det vanskelig å gjette på mulige kanditater.





Figur 11.5: Rimelighetsestimat

11.9 Rimelighetsestimatoren

La $\{P^\theta\}, \theta \in \Omega_\Theta$, være en statistisk modell. Anta at observatoren X har en tetthet f_X. Denne tettheten er avhengig av θ fordi fordelingen P_X^θ til X er avhengig av θ.

Rimelighetsfunksjonen $L_X^x : \Omega_\Theta \to \mathbb{R}$ til verdien x for observatoren X er definert ved

$$L_X^x(\theta) = f_X(x; \theta).$$

Notasjonen $L_X(\theta; x)$ brukes også. I de fleste tilfeller vil vi forenkle notasjonen ved å utelate en eller begge av indeksene X og x.

Det typiske tilfellet er at X er en vektor med diskret eller kontinuerlig fordeling. Betegnelsen "likelihood funksjonen" (the likelihood function) benyttes synonymt.

La x være en gitt observasjon. Da er $\hat{\tau}$ et *rimelighetsestimat* for estimanden $\tau : \Omega_\Theta \to \Omega_\tau$ dersom $\hat{\tau} = \tau(\hat{\theta})$ hvor $\hat{\theta}$ er et globalt maksimalpunkt for rimelighetsfunksjonen L^x, dvs dersom

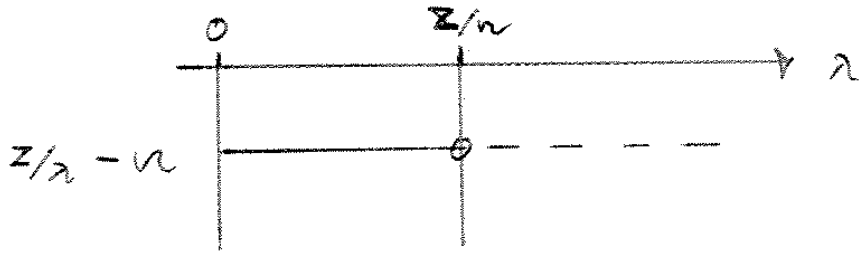
$$L^x(\theta) \leq L^x(\hat{\theta}) \qquad \text{for alle } \theta.$$

Den intuitive ideen er gitt ved:

> Vi velger den parameterverdien som gjør observasjonen så rimelig, dvs sannsynlig, som mulig.

Et eksempel er gitt i figur 11.5.

Et rimelighetsestimat $\hat{\tau}$ avhenger av observasjonen x, dvs $\hat{\tau} = \phi(x)$. En rimelighetsestimator for τ er da gitt ved $\phi(X)$.



Figur 11.6: Fortegnsskjema

Rimelighetsmetoden vil oftest være vanskeligere å gjennomføre enn momentmetoden. Spesielt er dette tilfellet når $\Omega_\Theta \subset \mathbb{R}^k$ og dimensjonen k er stor.

Advarsel: I litteraturen brukes noen ganger $\hat{\theta}$ også som betegnelse på rimelighetsestimatoren.

Eksempel 11.10 (Rimelighetsestimator) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra tetthetten

$$f(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad x = 0, 1, 2, \ldots$$

hvor $\lambda \geq 0$. Dette betyr spesielt at vi antar $X_i \sim \text{Poisson}(\lambda)$. Rimelighetsfunksjonen er

$$\begin{aligned} L(\lambda) = f_X(x;\lambda) &\overset{uavh.}{=} f_{X_1}(x_1;\lambda)\cdots f_{X_n}(x_n;\lambda) = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} \\ &= \frac{1}{x_1!\cdots x_n!}\cdot\lambda^z e^{-\lambda n} \quad \text{hvor } z = x_1 + \cdots + x_n. \end{aligned}$$

Merk: Regningen over viser også at $Z = X_1 + \cdots + X_n$ er en tilstrekkelig observator.

Rimelighetsestimatet kan nå finnes ved å maksimalisere L. Fordi ln er strengt voksende er dette ekvivalent med å maksimalisere $\ln L$. Pga faktoriseringen over ser vi at det er tilstrekkelig å maksimalisere $\ln(\lambda^z e^{-\lambda n})$:

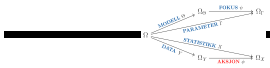
$$0 = \frac{d}{d\lambda}(z\ln\lambda - \lambda n) = z/\lambda - n \quad \text{gir } \lambda = z/n.$$

Fortegnsskjemaet for den deriverte i figur 11.6 gir at det kritiske punktet z/n er et maksimalpunkt. Konklusjonen er at rimelighetsestimatoren faller sammen med utvalgets forventning

$$\overline{X} = Z/n.$$

Følgende observasjoner er viktige:

1. Det kan lønne seg å maksimalisere $\ln L$.
2. Det kan lønne seg å faktorisere L i en θ-avhengig og en θ-uavhengig del. Da finnes også en eventuell tilstrekkelig observator.
3. Dersom T er en tilstrekkelig observator, så vil rimelighetsestimatoren automatisk være basert på T.



Eksempel 11.11 (Rimelighetsestimator for μ, σ^2) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en $N(\mu, \sigma^2)$-fordeling. Her er $\Omega_\Theta = \{(\mu, \sigma^2) \,|\, \mu \in \mathbb{R},\ \sigma > 0\}$. Uavhengigheten gir rimelighetsfunksjonen

$$L(\mu, \sigma^2) \overset{uavh}{=} \left(\frac{1}{\sqrt{2\pi}}\right)\left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2}(\overline{x^2} - 2\mu\overline{x} + \mu^2)}.$$

Vi må maksimalisere

$$u(\mu, \sigma^2) = \ln\left(\left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2}(\overline{x^2} - 2\mu\overline{x} + \mu^2)}\right) = -\ln(\sigma^2)\ n/2 - \frac{n}{2\sigma^2}(\overline{x^2} - 2\mu\overline{x} + \mu^2).$$

Dette er en funksjon av to variable, og en metode er å løse ligningssystemet

$$\frac{\partial u}{\partial \mu} = 0, \quad \frac{\partial u}{\partial \sigma^2} = 0.$$

I vårt spesielle tilfelle er det imidlertid klart (hvorfor?) at maksimalpunktet må være slik at $\overline{x^2} - 2\mu\overline{x} + \mu^2$ er minimalisert. Regningen

$$\overline{x^2} - 2\mu\overline{x} + \mu^2 = \overline{x^2} - \overline{x}^2 + (\overline{x} - \mu)^2$$

gir at minimalpunktet er $\mu = \overline{x}$. Oppgaven er dermed redusert til å minimalisere

$$-\ln(\sigma^2)\ n/2 - \frac{n}{2\sigma^2}(\overline{x^2} - \overline{x}^2)$$

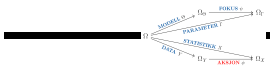
Ved $t = \sigma^2$ og derivasjon følger

$$0 = \frac{d}{dt}\left(-\ln(t)\ n/2 - \frac{n}{2t}(\overline{x^2} - \overline{x}^2)\right) = -1/t \cdot n/2 + t^{-2}\frac{n}{2}(\overline{x^2} - \overline{x}^2) \qquad |\cdot t^2$$

$$0 = -tn/2 + \frac{n}{2}(\overline{x^2} - \overline{x}^2)$$

Det kritiske punktet er dermed $\sigma^2 = \overline{x^2} - \overline{x}^2$ og et fortegnsskjema gir at dette er et maksimalpunkt. Ved sammenligning med beregningsformelen for utvalgets varians S^2 kan vi konkludere at rimelighetsestimatoren for (μ, σ^2) er

$$W = \left(\overline{X}, \frac{n-1}{n}S^2\right)$$



11.10 Bayes-estimatoren

Momentestimatoren og rimelighetsestimatoren er eksempel på metoder som ofte gir enkle kandidater for estimatorer. De resulterende estimatorene kan imidlertid være teoretisk *ubrukelige* ved at det kan finnes alternative estimatorer som er uniformt bedre og i tillegg strikt bedre for noen modeller. Bayes-estimatoren er imidlertid alltid teoretisk *brukelig*[1] Vi vil her nøye oss med å gi oppskriften for å finne en Bayes-estimator $\hat{\gamma}$ for en parameter $\gamma = \psi(\theta)$.

Første steg er å oppfatte den statistiske modellen $\{\mathrm{P}_Y^\theta\}$ for data Y som en betinget sannsynlighet:

$$\mathrm{P}^\theta(Y \in A) = \mathrm{P}(Y \in Y \,|\, \Theta = \theta) \tag{11.13}$$

Andre steg er å anta en kjent fordeling P_Θ for modellen Θ. Dette tolkes som en vurdering av usikkerheten i modellen θ før data er observert. Tilsammen er dette da en spesifikasjon av fordelingen til (Y, Θ). Tredje steg er å beregne posterior fordelingen når data y er observert. Denne er definert ved $\mathrm{P}_\Theta^y(B) = \mathrm{P}(\Theta \in B \,|\, Y = y)$. Fjerde og siste steg er å bestemme $\hat{\gamma}$ ved at den minimaliserer Bayes posterior risiko

$$r^y = \mathrm{E}^y \, \ell(\Gamma, \hat{\gamma}) \tag{11.14}$$

Dette siste steget er som i vår innledende diskusjon om tapsfunksjoner. Dersom tapet $\ell(\gamma, \hat{\gamma}) = (\gamma - \hat{\gamma})^2$, så følger det at Bayes estimatoren er

$$\hat{\gamma} = \mathrm{E}^y \, \Gamma \tag{11.15}$$

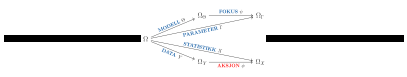
Eksempel 11.12 (Beta prior) Anta at vi har observert s suksesser og f fiaskoer i en Bernoulli forsøksrekke med suksesssannsynlighet θ. En Beta(α, β) prior for suksessen gir da posterior tetthet

$$\pi(\theta \,|\, y) \propto \theta^s (1-\theta)^f \theta^{\alpha-1} (1-\theta)^{\beta-1} \tag{11.16}$$

som gir at posterior fordeling er Beta$(s + \alpha, f + \beta)$. Bayes-estimatoren er

$$\hat{\theta} = \mathrm{E}^y \, \Theta = \frac{s + \alpha}{s + f + \alpha \; beta} \tag{11.17}$$

Denne er bare lik empirisk middel $s/(s+f)$ dersom $\alpha = \beta = 0$. Dette er *Haldane* prior. Det er ikke en sannsynlighetsfordeling, og illustrerer at nyttige estimatorer kan finnes ved å tillate prior fordelinger som ikke er sannsynlighetsfordelinger.



[1] 'admissible' er den engelse termen for dette.

Det foregående eksempel gir også den generelle oppskriften når fordelingene er gitt ved tettheter. Begrunnelsen er gitt ved regningen

$$\pi(\theta \,|\, y) = \frac{f(y \,|\, \theta)\pi(\theta)}{f(y)} \propto f(y \,|\, \theta)\pi(\theta) \tag{11.18}$$

Marginaltettheten $f(y)$ er kun en normaliseringsfaktor og kan i mange tilfeller finnes ved å først gjenkjenne posteriortettheten som en kjent tetthet.

11.11 Intervallestimering og konfidensintervall

La $\{P^\theta\}$, $\theta \in \Omega_\Theta$, være en statistisk modell for utfallsrommet Ω. En punktestimator W for en estimand $\tau : \Omega_\Theta \to \Omega_\tau$ er en observator $W : \Omega \to \Omega_W$ hvor $\Omega_\tau \subset \Omega_W$.

Punktestimatoren gir en gjetning på verdien til estimanden med utgangspunkt i observasjonene. En områdeestimator gir en gjetning på et område, dvs en mengde, som skal inneholde verdien til estimanden. Vi velger følgende definisjon:

La $\Omega_{(V)}$ være en mengde. Anta at $V(\omega) \subset \Omega_{(V)}$ for alle utfall ω. Da er V en områdeestimator i $\Omega_{(V)}$ dersom

$$(x \in V) \stackrel{def}{=} \{\omega \,|\, x \in V(\omega)\}$$

er en hendelse for alle punkter x i $\Omega_{(V)}$.

Hvert utfall ω gir en mengde $V(\omega)$ inneholdt i $\Omega_{(V)}$. Usikkerheten i ω gir en tilsvarende usikkerhet for $V(\omega)$. Dette betyr at V kan oppfattes som en tilfeldig mengde. Antagelsen "$(x \in V)$ er en hendelse" i definisjonen sikrer at sannsynligheten $P(x \in V)$ for at V inneholder et fiksert punkt x er definert.

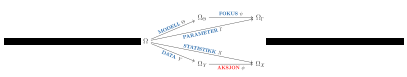
La τ være en estimand med verdier i Ω_τ. En *områdeestimator* V i $\Omega_{(V)}$ er en områdeestimator for τ dersom $\Omega_\tau \subset \Omega_{(V)}$.

Når eksperimentet er gjort, så brukes V til å regne ut områdeestimatet $v = V(\omega)$ for τ. Konklusjonen er at vi gjetter på at $\tau \in v$.

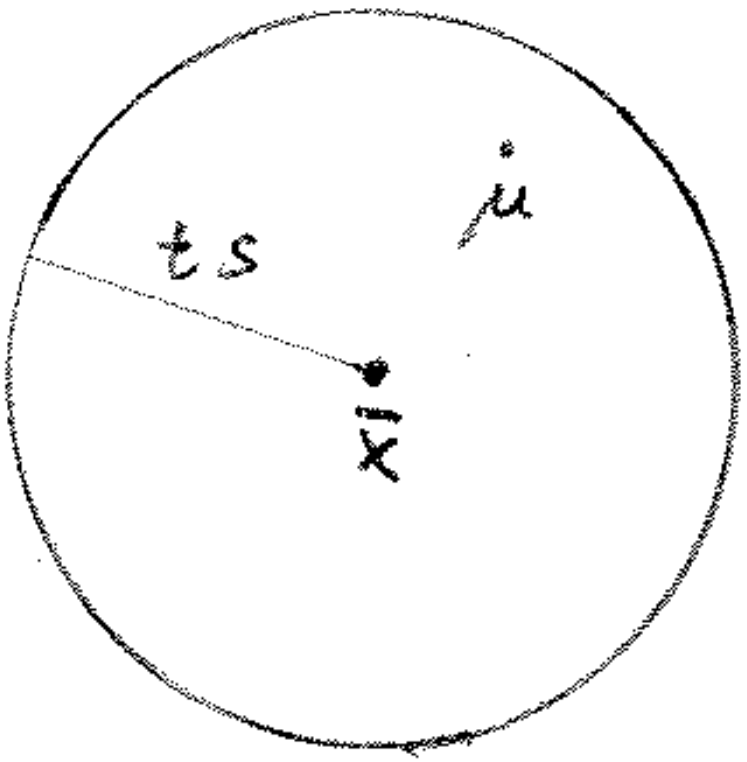
Eksempel 11.13 (Mengdeestimator) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en fordeling. En områdeestimator V for fordelingens forventningsverdi μ er gitt av

$$V = \{y \,|\, \left|y - \overline{X}\right| \leq tS\},$$

hvor $\overline{X}$ og S^2 er utvalgets forventning og varians, og t er et tall. Et mulig mengdeestimat er vist i figur 11.7.



La V være en områdeestimator. Dersom $V(\omega)$ er et lukket intervall for alle ω, så er V en *intervallestimator*.



Figur 11.7: Mengdeestimat for $\mu = (\mu_1, \mu_2)$

Vi beviser at et tilfeldig intervall $V = [X, Y]$, der X, Y er tilfeldige variable, er en intervallestimator. Dette beviset ved:

$$(w \in V) = (w \in [X, Y]) = (X \leq w \leq Y) = (X \leq w \text{ og } w \leq Y) = (X \leq w) \cap (w \leq Y),$$

som er en hendelse når X, Y er tilfeldige variable.

Metodene vi ser på tar som utgangspunkt at vi kan se på dataene som resultatet av et statistisk eksperiment. Grovt sett vurderes en metode til å være god dersom den gir et godt resultat i de fleste av de tenkte gjentagelsene av eksperimentet. F eks er en punktestimator god dersom den gir et estimat i nærheten av estimanden i de fleste av eksperimentene. Tilsvarende er en områdeestimator god dersom den gir et estimert området som inneholder estimanden i de fleste av eksperimentene. Dette motiverer en definisjon.

La α være et reelt tall. En områdeestimator V for en estimand τ er en konfidensmengde for τ med konfidensnivå α dersom

$$P(\tau \in V) \geq \alpha.$$

I så fall er V en $\alpha 100\%$ områdeestimator for τ.

I det overstående krever vi ikke likhet fordi venstresiden avhenger av parameteren via P og τ, mens konfidensnivået α er uavhengig av parameteren.

> Ved gjentagelse av et statistisk eksperiment vil et $\alpha 100\%$ områdeestimat inneholde estimanden i minst $\alpha 100\%$ av tilfellene i det lange løp.

Det er ønskelig at konfidensnivået er så høyt som mulig. Det vanlige er å sammenligne områdeestimatorer med et gitt nivå, og blant disse foretrekkes den som gir det minste området, hvis en slik finnes.

La $\alpha' \leq \alpha$. Da vil et konfidensområde V med nivå α også være et konfidensområde med nivå α' fordi

$$P(\tau \in V) \geq \alpha \geq \alpha'.$$

Dette betyr at det svarer uendelig mange konfidensnivå til et konfidensområde. Dette motiverer en definisjon.

Konfidenskoeffisienten α til en områdeestimator V for en estimand τ er den største nedre skranken for mengden

$$\{P^\theta(\tau(\theta) \in V) \mid \theta \in \Omega_\Theta\}.$$

Definisjonen betyr spesielt at V har et konfidensnivå α dersom α er mindre eller lik konfidenskoeffisienten. Intuitivt er konfidenskoeffisienten det beste konfidensnivået til områdeestimatoren. Til en områdeestimator finnes det mange konfidensnivå, men kun én konfidenskoeffisient. I praksis kan ofte konfidenskoeffisienten finnes som minimalverdien til funksjonen $g : \Omega_\Theta \to \mathbb{R}$, som er definert ved

$$g(\theta) = P^\theta(\tau(\theta) \in V).$$

Advarsel: Konvensjonene for hva som menes med konfidensnivå og konfidenskoeffisient varierer fra forfatter til forfatter.

11.12 Metoder for å finne konfidensintervall

La W være en punktestimator for τ. En intuitivt rimelig intervallestimator for τ er da gitt ved

$$V = [W - X, W + Y],$$

hvor X og Y er tilfeldige variable som tar positive verdier. En mulighet er å la $X = Y = a$. Konstanten a kan bestemmes av det ønskede konfidensnivået. Tsjebysjevs ulikhet gir en variant av dette.

Eksempel 11.14 (Konfidensområde fra Tsjebysjevs ulikhet) La W være en forventningsrett estimator for τ. Anta at $s > 0$ er en konstant slik at

$$\operatorname{Var}^\theta W \leq s^2$$

for alle θ. Da gir Tsjebysjevs ulikhet at

$$P(|W - \tau| < ts) > 1 - \frac{\operatorname{Var} W}{t^2 s^2} \geq 1 - \frac{s^2}{t^2 s^2} = 1 - \frac{1}{t^2}.$$

Konklusjonen er at

$$V = \{w \,|\, |W - w| \leq ts\}$$

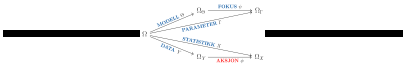
er en $(1 - \frac{1}{t^2})100\%$ områdeestimator for τ. Dette følger fordi

$$(\tau \in V) = (|W - \tau| \leq ts).$$

Spesielt finnes et 90% konfidensområde ved å velge t slik at $1 - \frac{1}{t^2} = 0,90$.

Vi ønsker å få et områdeestimat som er så lite som mulig, dvs vi må velge s så liten som mulig under betingelsen $\operatorname{Var}^\theta W \leq s^2$. Spesielt ser vi at dersom W er en nøyaktig estimator, dvs dersom variansen er liten, så vil metoden gi et lite områdeestimat.

Utledningen over gjelder uansett hvilken fordeling vi har. Når fordelingstypen er gitt kan vi bruke bedre ulikheter enn Tsjebysjevs ulikhet.



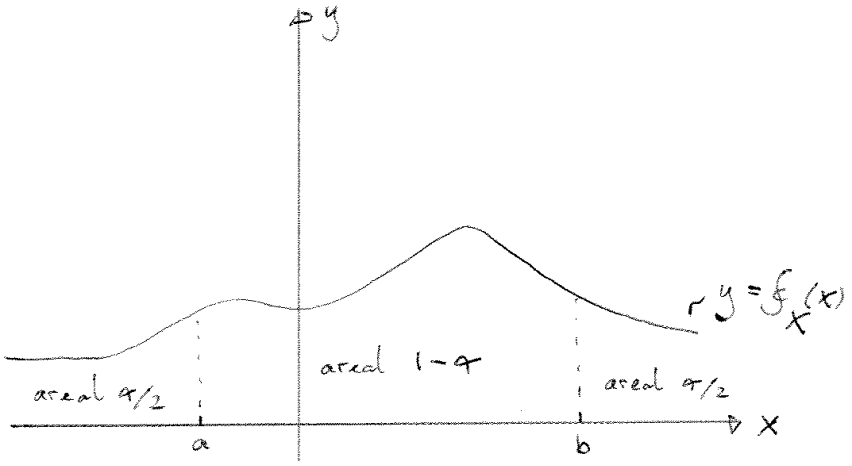
I noen tilfeller kan et konfidensintervall finnes enklere ved å finne to ensidige konfidensintervall. Anta at

$$P(W_- \leq \tau) \geq \alpha_-, \qquad P(W_+ \geq \tau) \geq \alpha_+, \qquad P(W_- \leq W_+) = 1.$$

Da følger det at

$$P(W_- \leq \tau \leq W_+) = 1 - P(W_- \geq \tau) - P(W_+ \leq \tau) \geq \alpha_- + \alpha_+ - 1,$$

dvs $[W_-, W_+]$ er et konfidensintervall med nivå $\alpha_- + \alpha_+ - 1$.



Figur 11.8: $P(a \leq X \leq b) = 1 - \alpha$ når $P(a > X) = P(X > b) = \alpha/2$.

I det følgende vil vi beskrive metoden med intervall-invertering. Metoden gir alltid en intervallestimator med ønsket nivå med utgangspunkt i en tilfeldig variabel.

La X være en tilfeldig variabel. I noen tilfeller kan en finne en $a(\theta)$ og en $b(\theta)$ slik at

$$P(a \leq X \leq b) \geq 1 - \alpha,$$

ved å finne $a \leq b$ slik at

$$P(a > X) = P(X > b) \leq \alpha/2.$$

Dersom X har en kontinuerlig fordeling, så finnes alltid en unik a og en unik b ved å kreve likhet. Tallet b er da ikke annet enn $\alpha/2$ kvantilen til fordelingen til X. Dette er illustrert i figur 11.8. Generelt finnes alltid en unik største a og en unik minste b slik at ulikhetene holder. Fordelingen til X avhenger av parameteren θ, så a og b må nødvendigvis avhenge av θ.

Den doble ulikheten

$$a(\theta) \leq x \leq b(\theta)$$

gir generelt et områdeestimat

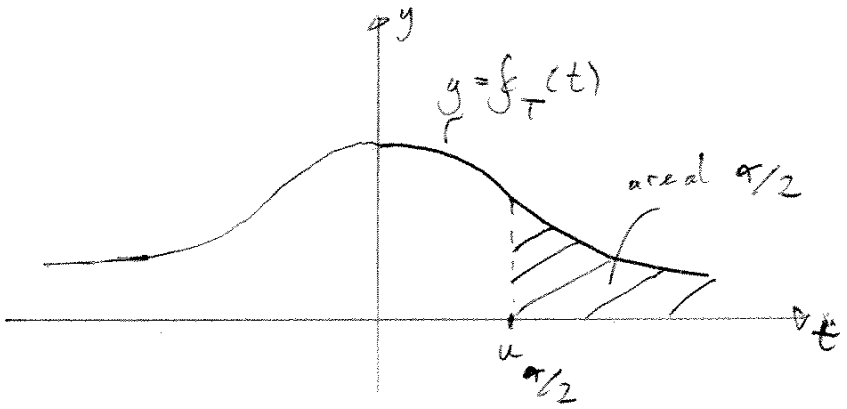
$$v = \{\theta \,|\, a(\theta) \leq x \leq b(\theta)\}.$$

Dersom a, b er valgt slik som over, så følger det at den tilsvarende estimatoren V har konfidensnivå lik $1 - \alpha$ fordi

$$(\theta \in V) = (a(\theta) \leq X \leq b(\theta)).$$

Dersom $\Omega_\Theta \subset \mathbb{R}$ finnes også et unikt minste intervallestimat som inneholder v.

En ulempe med den foregående metoden er at den generelt resulterer i mye regnearbeid fordi fordelingen til X avhenger av θ. Denne observasjonen motiverer en definisjon.



Figur 11.9: Kvantilen $u_{\frac{\alpha}{2}}$ til standard normalfordelingen.

La $\{P^\theta\}$, $\theta \in \Omega_\Theta$, være en statistisk modell. En pivot observator T er en funksjon som gir en observator T^θ for hver parameterverdi θ slik at fordelingen til T^θ er uavhengig av θ.

Det neste eksemplet viser at metoden med intervall-invertering kan forenkles ved bruk av en pivot variabel.

Eksempel 11.15 (Normalfordeling med kjent varians) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra $N(\mu, \sigma^2)$-fordelingen med σ kjent. Dette betyr at $\Omega_\Theta = \{\mu \,|\, \mu \in \mathbb{R}\}$. Da er μ den eneste parameteren.

Vi vet at $\overline{X}$ er en forventningsrett estimator for μ. Videre er $\operatorname{Var}\overline{X} = \frac{1}{n}\sigma^2$. Fordi en sum av normalfordelte variable er normalfordelt følger det at $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$. Dette betyr at

$$T^\mu := \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1),$$

dvs T er en pivot variabel. Dette gir

$$P(T > u_{\frac{\alpha}{2}}) = \frac{\alpha}{2},$$

hvor $\frac{\alpha}{2}$ kvantilen $u_{\frac{\alpha}{2}}$ til standard normalfordelingen kan finnes i tabell. Fra figur 11.9 følger det at

$$P(|T| \le u_{\frac{\alpha}{2}}) = 1 - \alpha.$$

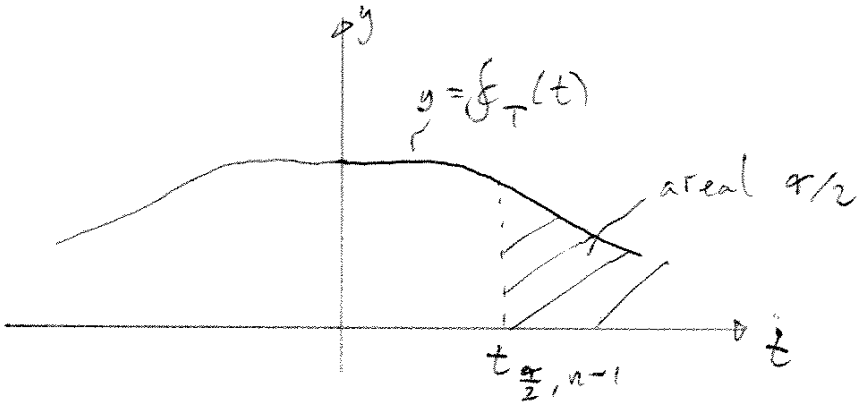
Vi har følgende ekvivalente ulikheter

$$|T| \le u_{\frac{\alpha}{2}}, \quad \left|\frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}\right| \le u_{\frac{\alpha}{2}}, \quad \left|\overline{X} - \mu\right| \le \frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \quad \overline{X} - \frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} \le \mu \le \overline{X} + \frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}.$$

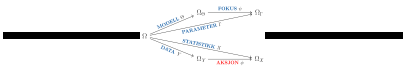
Konklusjonen er dermed at

$$\boxed{\left[\overline{\mathrm{X}}\text{-}\frac{\mathrm{u}_{\frac{\alpha}{2}}\sigma}{\sqrt{\mathrm{n}}},\ \overline{\mathrm{X}} + \frac{\mathrm{u}_{\frac{\alpha}{2}}\sigma}{\sqrt{\mathrm{n}}}\right]}$$

er en $(1-\alpha)100\%$ intervallestimator for μ.



Figur 11.10: Kvantilen $t_{\frac{\alpha}{2},n-1}$ til $\mathrm{t}(n-1)$-fordelingen.



Eksempel 11.16 (Intervallestimator for μ når σ^2 er ukjent) La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen. Dette betyr at

$$\Omega_\Theta = \{(\mu, \sigma^2) \,|\, \mu \in \mathbb{R}, \sigma > 0\}.$$

Vi skal finne et $(1-\alpha)100\%$-konfidensintervall for μ. Utgangspunktet er at (godta dette!)

$$\frac{(\overline{X}-\mu)\sqrt{n}}{S} \sim \mathrm{t}(n-1).$$

Dette gir en pivot variabel som kun avhenger av parameteren (μ, σ^2) gjennom estimanden μ! Deretter kan vi gjøre som i forrige eksempel:

$$P(\frac{(\overline{X}-\mu)\sqrt{n}}{S} > t_{\frac{\alpha}{2},n-1}) = \frac{\alpha}{2},$$

hvor $\frac{\alpha}{2}$ kvantilen $t_{\frac{\alpha}{2}}$ til $\mathrm{t}(n-1)$-fordelingen kan finnes i tabell. Fra figur 11.10 følger det at

$$P(\left|\frac{(\overline{X}-\mu)\sqrt{n}}{S}\right| \leq t_{\frac{\alpha}{2},n-1}) = 1-\alpha.$$

Vi har følgende ekvivalente relasjoner

$$\left|\frac{(\overline{X}-\mu)\sqrt{n}}{S}\right| \leq t_{\frac{\alpha}{2},n-1}, \quad \left|\overline{X}-\mu\right| \leq \frac{t_{\frac{\alpha}{2},n-1}S}{\sqrt{n}}, \quad \overline{X} - \frac{t_{\frac{\alpha}{2},n-1}S}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{t_{\frac{\alpha}{2},n-1}S}{\sqrt{n}}.$$

Vi konkluderer at

$$\boxed{[\overline{X} - \frac{t_{\frac{\alpha}{2},n-1}S}{\sqrt{n}}, \overline{X} + \frac{t_{\frac{\alpha}{2},n-1}S}{\sqrt{n}}]}$$

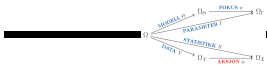
er en $(1-\alpha)100\%$ intervallestimator for μ.

Når dataene er gitt, kan vi beregne intervallestimatet

$$[\overline{x} - \frac{t_{\frac{\alpha}{2},n-1}s}{\sqrt{n}}, \overline{x} + \frac{t_{\frac{\alpha}{2},n-1}s}{\sqrt{n}}].$$

Dette estimatet vil oftest være et større intervall enn det vi fikk når σ var kjent. Dette er ikke uventet fordi både μ og σ^2 var ukjente i dette siste tilfellet.

t$(n-1)$-fordelingen ligner på normalfordelingen, men grafen til tettheten er litt flatere. Dette gir som resultat at $t_{\frac{\alpha}{2},n-1} > u_{\frac{\alpha}{2}}$. Dette gir en alternativ forklaring på hvorfor estimatet med ukjent varians oftest er større. Når n vokser vil t$(n-1)$-fordelingen nærme seg $N(0,1)$-fordelingen. Spesielt vil $\lim t_{\frac{\alpha}{2},n-1} = u_{\frac{\alpha}{2}}$.



11.13 Konfidensintervall: Flere eksempler

Vi starter med en repetisjon av de viktigste begrepene. En intervallestimator V er gitt ved to tilfeldige variable A, B ved $V = [A, B]$. Dette er et α-konfidensintervall for τ dersom

$$P^{\theta}(A \leq \tau(\theta) \leq B) \geq \alpha \text{ for alle } \theta.$$

I så fall er α konfidensnivået til V. Konfidenskoeffisienten til V er den største α slik at at ligningen er oppfylt.

Vi vil peke på to kriterier for vurdering av konfidensintervall.

1. Vi ønsker til et gitt konfidensnivå å finne et konfidensintervall som er så lite som mulig.
2. Dersom T er en tilstrekkelig observator, så skal V være basert på T.

Merk: Dette siste punktet skal gjelde for alle tilstrekkelige observatorer. Det siste punktet er strengt tatt kun en gjentagelse av det generelle tilstrekkelighetsprinsippet som gjelder for enhver estimator, og mer generelt for enhver statistisk slutning.

Vi har ikke beskrevet noen metode som generelt gir et konfidensintervall for en generell estimand, men snarere pekt på noen muligheter.

Intuitivt er en god mulighet å se på konfidensintervall med utgangspunkt i et intervall sentrert om en forventningsrett estimator. Lengden til intervallet kan bestemmes av det ønskede konfidensnivået. Vi har sett at Tsjebysjevs ulikhet kan brukes i denne forbindelsen ved å velge intervall-lengden proposjonal med en øvre skranke for standardavviket til estimatoren. Spesielt ser vi da at en nøyaktig estimator gir opphav til et nøyaktig, dvs lite,

konfidensintervall. Dette er mest av teoretisk interesse. I konkrete modeller kan en bruke bedre ulikheter som gir tilsvarende bedre intervallestimatorer.

La det være gitt at et $(1-\alpha)$-områdeestimat for parameteren θ skal finnes med utgangspunkt i en tilfeldig variabel X. Generelt kan vi da finne a og b slik at

$$P(a \leq X \leq b) \geq 1 - \alpha.$$

Dersom X har kontinuerlig fordeling kan til og med likhet oppnås ved å bestemme a og b ved

$$P(X < a) = P(X > b) = \alpha/2.$$

Her er det klart at a og b er funksjoner av parameteren θ. Et $(1-\alpha)$-områdeestimat for θ er da gitt ved

$$V = \{\theta \,|\, a(\theta) \leq X \leq b(\theta)\}.$$

Dette er en generelt anvendelig metode, men valget av variabelen X er ikke opplagt. Et konfidensområde for en generell estmand τ er gitt av

$$\tau(V) = \{\tau(\theta) \,|\, \theta \in V\},$$

men det er ikke gitt at dette er det beste valget. Denne metoden refererer vi til som intervall-inverteringsmetoden.

En tredje mulighet er gitt ved en variasjon av intervall-inverteringsmetoden. Den tilfeldige variabelen X erstattes av en pivotvariabel Y, dvs en familie $\{Y^\theta\}$ av tilfeldige variable slik at fordelingen til Y^θ er uavhengig av parameteren θ. Et $(1-\alpha)$-områdeestimat for θ er da gitt ved

$$V = \{\theta \,|\, a \leq Y^\theta \leq b\},$$

hvor a og b velges slik at

$$P(a \leq Y \leq b) = 1 - \alpha.$$

Poenget er her at fordelingen til Y er uavhengig av parameteren, så a og b kan velges uavhengig av parameteren. Et annet poeng er at en i noen tilfeller kan finne en pivotvariabel som avhenger av parameteren θ gjennom estimanden τ, dvs $Y^\theta = Z^{\tau(\theta)}$. I så fall sier vi at Z er en pivotvariabel for τ. Et $(1-\alpha)$-områdeestimat for τ er da gitt ved

$$V = \{t \,|\, a \leq Z^t \leq b\}.$$

Eksempel 11.17 (Intervall-invertering) La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Da er $W = \frac{n+1}{n} Y_{maks}$ forventningsrett for θ. I tillegg er W en tilstrekkelig observator. Tettheten er

$$f_W(w) = \frac{n^{n+1}}{(n+1)^n \theta^n} w^{n-1}.$$

Vi vil finne et konfidensintervall med nivå $\alpha = 90\%$ ved å finne a, b slik at $P(a \leq W \leq b) = \alpha$.

Merk: $[a, b]$ er ikke det søkte konfidensintervallet!

Vi forenkler litt ved å finne a, b slik at $P(W < a) = P(W > b) = (1-\alpha)/2 = 5\%$:

$$\begin{aligned} P(W < a) &= \int_0^a f_W(w)dw \\ &= \int_0^a \frac{n^{n+1}}{(n+1)^n\theta^n} w^{n-1}dw = \frac{n^{n+1}}{(n+1)^n\theta^n}\frac{a^n}{n} = 5\% \\ \Rightarrow a(\theta) &= \sqrt[n]{5\%}\frac{n}{n+1}\theta \end{aligned}$$

$$\begin{aligned} P(W > b) &= 1 - \int_0^b f_W(w)dw = 5\% \\ \Rightarrow b(\theta) &= \sqrt[n]{95\%}\frac{n}{n+1}\theta \end{aligned}$$

Vi inverterer intervallet, dvs løser den dobble ulikheten mhp parameteren θ

$$\begin{aligned} a \le W &\le b \\ \sqrt[n]{5\%}\frac{n}{n+1}\theta \le W &\le \sqrt[n]{95\%}\frac{n}{n+1}\theta \\ \frac{n}{n+1}\frac{W}{\sqrt[n]{95\%}} \le \theta &\le \frac{n}{n+1}\frac{W}{\sqrt[n]{5\%}}. \end{aligned}$$

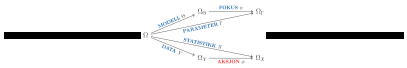
Konklusjonen er at

$$V = [\frac{n}{n+1}\frac{W}{\sqrt[n]{95\%}}, \frac{n}{n+1}\frac{W}{\sqrt[n]{5\%}}]$$

er et konfidensintervall med nivå 90%.

Merk: V er en estimator. Når vi setter inn verdier fra et utført eksperiment, så finnes intervallestimatet

$$v = [\frac{n}{n+1}\frac{w}{\sqrt[n]{95\%}}, \frac{n}{n+1}\frac{w}{\sqrt[n]{5\%}}].$$

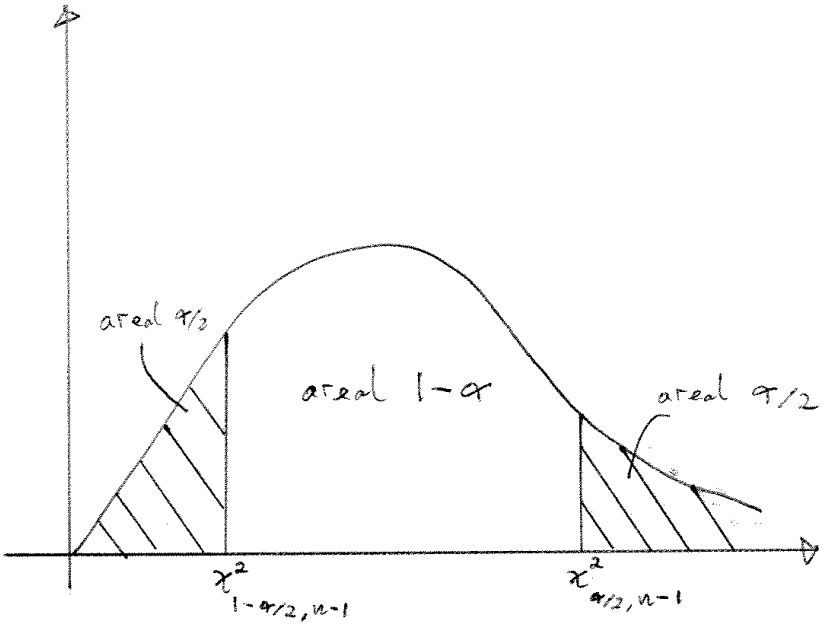


Eksempel 11.18 (Pivot variabel) La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $N(\mu, \sigma^2)$-fordelingen. Vi vil finne et $(1-\alpha)$-konfidensintervall for σ^2. I følge tilstrekkelighetsprinsippet vil en god estimator nødvendigvis være en funksjon av $\overline{Y}$ og S^2 fordi $T = (\overline{Y}, S^2)$ er en tilstrekkelig observator. Her er

$$\frac{(n-1)S^2}{\sigma^2}$$

en pivotvariabel for σ^2. Fordelingen er uavhengig av (μ, σ^2) og gitt ved

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2{}_{n-1}\,.}$$



Figur 11.11: Kvantiler i khi-kvadrat fordelingen.

Beviset av dette utelates her. Dette gir

$$P\left(\frac{(n-1)S^2}{\sigma^2} \geq {\chi^2}_{\frac{\alpha}{2},n-1}\right) = \frac{\alpha}{2}.$$

Kvantilen ${\chi^2}_{\frac{\alpha}{2},n-1}$ til khi-kvadrat fordelingen finnes i tabell. Fra figur 11.11 følger det at

$$P\left({\chi^2}_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq {\chi^2}_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

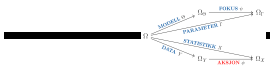
Vi løser den ene ulikheten mhp σ^2:

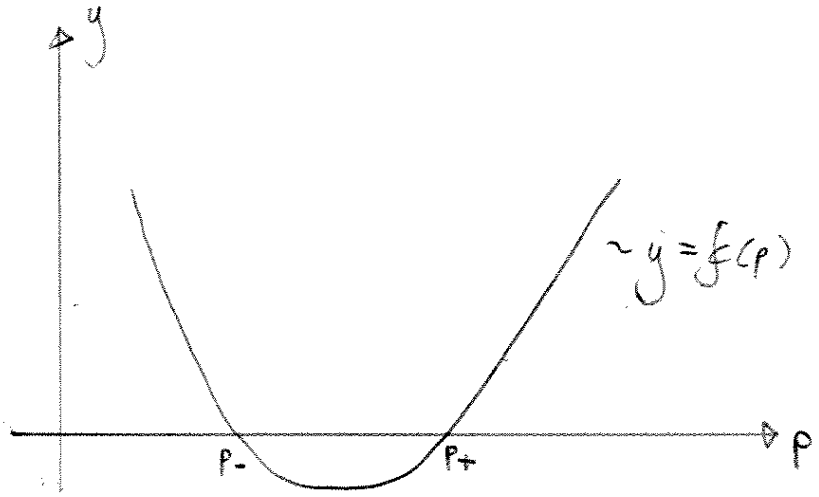
$$\frac{(n-1)S^2}{\sigma^2} \leq {\chi^2}_{\frac{\alpha}{2},n-1}, \quad \frac{(n-1)S^2}{{\chi^2}_{\frac{\alpha}{2},n-1}} \leq \sigma^2.$$

Dette gjøres tilsvarende med den andre ulikheten. Konklusjonen er at vi har funnet $(1-\alpha)$-konfidensinterallet

$$V = [\frac{(n-1)S^2}{{\chi^2}_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{{\chi^2}_{1-\frac{\alpha}{2},n-1}}]$$

for σ^2. Om μ er gitt blir løsningen den samme.





Figur 11.12: Intervallestimat for p.

Eksempel 11.19 (Tilnærmede konfidensintervall) La Y være $\mathrm{B}(n,p)$ fordelt. Vi vil finne et konfidensintervall for p. Vi tar utgangspunkt i den normaliserte variabelen

$$\frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}},$$

som har forventning 0 og varians 1. Sentralgrenseteoremet ($Y \sim X_1 + \cdots + X_n$!) gir tilnærmingen

$$P\left(-u_{\frac{\alpha}{2}} \leq \frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{\frac{\alpha}{2}}\right) \doteq 1-\alpha$$

for store n. Vi finner et tilnærmet konfidensintervall ved å ta utgangspunkt i at variabelen har $N(0,1)$ fordeling.

Et intervallestimat finnes ved å løse ulikheten

$$\frac{\left|\frac{y}{n}-p\right|}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{\frac{\alpha}{2}}$$

med hensyn på p. Kvadrering og ordning gir

$$p^2\left(1+\frac{{u_{\frac{\alpha}{2}}}^2}{n}\right) - p\left(\frac{2y}{n}+\frac{{u_{\frac{\alpha}{2}}}^2}{n}\right) + \left(\frac{y}{n}\right)^2 \leq 0.$$

Figur 11.12 gir at nullpunktene $p_\pm$ til 2.grads polynomet $f(p)$ på venstresiden gir inter-

vallestimatet $v = [p_-, p_+]$. Løsningene er

$$p_\pm = \frac{\left(\frac{y}{n} + \frac{{u_{\frac{\alpha}{2}}}^2}{2n}\right) \pm \frac{u_{\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\frac{y}{n}(1-\frac{y}{n}) + \frac{{u_{\frac{\alpha}{2}}}^2}{4n}}}{1 + \frac{{u_{\frac{\alpha}{2}}}^2}{n}}$$

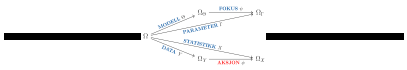
Vi har antatt at n er stor så vi kan approksimere en gang til ved å neglisjere leddene

$$\frac{{u_{\frac{\alpha}{2}}}^2}{2n}, \quad \frac{{u_{\frac{\alpha}{2}}}^2}{4n}, \quad \frac{{u_{\frac{\alpha}{2}}}^2}{n}.$$

Dette gir det tilnærmede konfidensintervallestimatet

$$v = [\frac{y}{n} - u_{\frac{\alpha}{2}}\sqrt{\frac{y}{n}(1-\frac{y}{n})\frac{1}{n}}, \frac{y}{n} + u_{\frac{\alpha}{2}}\sqrt{\frac{y}{n}(1-\frac{y}{n})\frac{1}{n}}]$$

Formelen over er lik den vi har funnet tidligere for normalfordelingen. Den kjente σ^2 er erstattet av estimatet $y/n \cdot (1 - y/n) \cdot 1/n$.



11.14 Oppgaver

11.1. Vis at tapet $l = c(1 - \delta(\gamma - \hat{\gamma}))$ i ligning (11.1). Bruk dette til å begrunne at en optimal estimator for $\gamma = \Gamma(\omega)$ er gitt ved $\hat{\gamma} = \text{argmax}(g)$ når $\Gamma \sim g$ hvor g er en diskret tetthet. Er estimatet unikt? Finnes det alltid?

11.2. Deltafunksjonen kan identifiseres med følgen δ_n av sannsynlighetstettheter tilsvarende normalfordelinger $\text{N}(0, \sigma_n^2)$ hvor variansen $\sigma_n^2 \to 0$. Skisser grafene til funksjonene δ_n. Beregn $\lim_{n\to\infty} \delta_n(x)$.

11.3. Dirac-deltafunksjonen δ er definert ved at den oppfyller ligning (11.3). La $\Gamma \sim g$ hvor g er en kontinuerlig tetthet. Vis at en optimal estimator for $\gamma = \Gamma(\omega)$ er gitt ved $\hat{\gamma} = \text{argmax}(g)$ når tapet er gitt ved $l = 1 - \delta(\gamma - \hat{\gamma})$.

11.4. Gi en begrunnelse for den siste likheten i ligning (11.5) ved å regne ut leddene i $|(\Gamma - \mu) + (\mu - \hat{\gamma})|^2$.

11.5. La $\Gamma \sim \text{C}(m, \beta)$. La tapet være $l = (\gamma - \hat{\gamma})^2$. Vis at det ikke finnes noen optimal estimator for γ. Finnes det alternative tapsfunksjoner som gir at $\hat{\gamma} = m$ er et optimalt estimat.

11.6. Vis at medianen definert i ligning (11.7) ikke avhenger av valget av referanseparameteren γ_0.

11.7. Vis at dersom $\text{E}\,|\Gamma| < \infty$, så er medianen gitt enklere ved $\text{M}(\Gamma) = \text{argmin}_\gamma \text{E}(|\Gamma - \gamma|)$.

11.8. Finn median og empirisk forventningsverdi til observerte data $1, 2, 3$, til data $1, 2, 3, 4$, og til data $1, 10, 7, 1, 3$. Hva er det optimale estimatet $\hat{\gamma}$ tilsvarende tapsfunksjonen $l = c(1 - \delta(\gamma - \hat{\gamma}))$ hvor usikkerheten i γ er gitt av den empiriske fordelingen i disse tre tilfellene?

11.9. Finn medianen til $U \sim \mathrm{U}(0, 2)$, $V = 5U + 7$ og $W = \exp(V)$. Kan du finne en generell regneregel for medianen som kan brukes til løsning av disse tre oppgavene?

11.10. Gjelder $\mathrm{M}(X + b) = \mathrm{M}(X) + b$, $\mathrm{M}(aX + b) = a\,\mathrm{M}(X) + b$ og $\mathrm{M}(X + Y) = \mathrm{M}(X) + \mathrm{M}(Y)$ når X og Y er stokastiske vektorer? Hva hvis M erstattes av E?

11.11. Vis at dersom Γ er en kontinuerlig stokastisk variabel med kumulativ fordeling F, så er medianen $m = F^{-1}(0.5) = \gamma_{0.5}$.

Hint: Vis først at $\mathrm{E}\,|W - w_{0.5}| \leq \mathrm{E}\,|W|$.

11.12. Vis at $\gamma_\alpha - \gamma$ er α fraktilen til $\Gamma - \gamma$ når γ_α er α fraktilen til Γ.

11.13. Gi en begrunnelse og tolkning av a og b i tapsfunksjonen l i ligning (11.12). Gi en intuitiv begrunnelse for at $\hat{\gamma} > \gamma_{0.5}$ når $a > b$. Hvorfor er forskjellen på $\hat{\gamma}$ og $\gamma_{0.5}$ såpass liten i Eksempel 11.2.

11.14. En retning γ i planet kan identifiseres med en enhetsvektor γ. La $\gamma_1, \ldots, \gamma_n$ være observerte retninger. Den sirkulære middelverdien $\overline{\gamma}$ er definert som retningen til vektoren $\sum_i \gamma_i$. La fire observerte retninger være gitt ved vinklene $\theta = (0°, 350°, 3°, 357°)$. Tegn opp de 4 observasjonene som punkter på enhetssirkelen. Er $\overline{\theta}$ et godt mål på en midlere retning til observasjonene? Regn ut sirkulær middelverdi, empirisk median og empirisk forventning til de fire observerte retningene.

11.15. Anta at i en urne med 100 kuler, så er 37 hvite. Sannsynligheten for å trekke to kuler som er hvite er lik $(37/100) \cdot (36/99)$.

Begrunn dette på 3 måter ved å sette opp 3 forskjellige sannsynlighetsmodeller. Eventuelle tilfeldige variable som benyttes skal spesifiseres.

Hint: Produktregel for betingede sannsynligheter, ordnet utvalg og ikke-ordnet utvalg.

11.16. Anta at to kuler trekkes tilfeldig fra en urne med 100 kuler. Resultatet er to hvite kuler.

Sett opp en fullstendig spesifisert statistisk modell og estimer antall hvite kuler i urnen.

Finn variansen til estimatoren.

11.17. Hva er forskjellen (og likheten) mellom områdeestimering og hypotesetesting?

11.18. Forklar hva som menes med at forventningsverdien $\mathrm{E}\,X$ til en tilfeldig variabel X er en estimand τ. Er i så fall X en forventningsrett estimator for τ?

11.19. La $Y_1, \ldots, Y_4$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. En estimator for θ er gitt ved $W = Y_{\text{maks}}$. Finn sannsynligheten for at W ikke er lenger enn $1/8$ unna θ. Hva er sannsynligheten når $\theta = 1$?

11.20. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{Exp}(1/\theta)$-fordelingen. En estimator for θ er gitt av $W = nY_{\min}$. Finn tettheten f_W til W.

11.21. Ved et sykehus nær et lager for radioaktivt avfall fra et kjernekraftverk så registreres antall pasienter per år som får diagnosen kronisk leukemi. Resultatet de siste 12 årene er 2, 1, 0, 0, 1, 1, 1, 4, 1, 0, 2, 3. Anta at antall forekomster per år er beskrevet av en poissonfordeling. Estimer den relevante parameteren.

11.22. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra en fordeling med forventningsverdi μ. Under hvilke betingelser er estimatoren

$$W = \phi(Y_1, \ldots, Y_) = \sum_{j=1}^{n} a_j Y_j$$

forventningsrett for μ?

11.23. Anta at 14, 10, 18, 21 er resultatet av en trekning fra en $\mathrm{U}(0, \theta)$-fordeling. Finn en forventningsrett estimator W for θ basert på den tredje ordningsobservatoren $Y_{(3)}$. Hva er verdien til W for utvalget som ble trukket? Gi et eksempel som viser at estimatoren W ikke er en god estimator.

11.24. La $X \sim \mathrm{U}(0, \theta)$ og definer $\tau(\theta) = \theta^2$. Finn en forventningsrett estimator W for τ.
Hint: La W være basert på X^2.

11.25. En estimator W_n er asymptotisk forventningsrett for τ dersom

$$\lim_{n\to\infty} \mathrm{E}^{\theta}\, W_n = \tau(\theta).$$

La $W_n = Y_{\text{maks}}$ hvor $Y_1, \ldots, Y_n$ er et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Er W_n forventningsrett for τ når $\tau(\theta) = \theta$? Er W_n asymptotisk forventningsrett?

11.26. Fremfor å kreve forventningsretthet, så kan en estimators fordeling sies å være riktig sentrert dersom estimatorens median er lik estimanden. En slik estimator kalles median-forventningsrett. Estimatoren $[(n+1)/n]Y_{\text{maks}}$ er forventningsrett for θ når $Y_1, \ldots, Y_n$ er et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Er den median-forventningsrett?

11.27. En estimator W er en beste lineær forventningsrett estimator for τ dersom

1. W er en lineær funksjon av observasjonene $X_1, \ldots, X_n$.
2. $\mathrm{E}\, W = \tau$.
3. $\mathrm{Var}\, W \leq \mathrm{Var}\, W^*$ for alle estimatorer W^* som oppfyller de to overnevnte krav.

Vis at et tilfeldig utvalgs forventning er en beste lineær forventningsrett estimator for fordelingens forventningsverdi.

Hint: La $W = a_1 X_1 + \cdots + a_n X_n + c$. Vis at $c = 0$. Partiellderiver mhp a_j'ene for å vise at $a_j = 1/n$.

11.28. Bevis at

$$\overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2$$

11.29. Bevis at S^2 er en forventningsrett estimator for σ^2.
Hva er antagelsene i beviset?

11.30. Bevis at $X^2 = X$ dersom X er en Bernoulli-variabel.

11.31. Hva betyr det at W er en observator for utfallsrommet Ω?
Hva betyr det at W er en estimator?
Hva betyr det at τ er en estimand?
Hva betyr det at W er en estimator for τ?
Hva betyr det at W er en estimator for τ basert på observatoren $(X_1, \ldots, X_n)$?

11.32. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra Exp$(1/\lambda)$-fordelingen. La $W_1 = \overline{Y}$, $W_2 = Y_1$ og $W_3 = nY_{min}$. Vis at $W_1 \sim$ Gamma$(n, n/\lambda)$. Vis at W_i'ene er forventningsrette estimatorer for λ. Finn de relative nøyaktighetene til W_i'ene. Hvilken estimator er best?

11.33. Vi har diskutert to mulige estimatorer X_1 og $\overline{X}$ for den binomiske parameteren p. Sammenlign disse.

11.34. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra U$(0, \theta)$-fordelingen. Vis at Y_{maks} er en konsistent estimator for θ. Anta at $\theta = 1$. Hvor stor må n være for at $P(|Y_{\text{maks}} - \theta| < 0,04) \geq 0,95$? Er Y_1 en konsistent estimator?

11.35. Vis at et tilfeldig utvalgs forventning er en konsistent estimator for forventningsverdien.

11.36. La $X \sim$ Exp$(1/\lambda)$ og la $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra fordelingen til $X + \theta$. Vis at Y_{min} er en tilstrekkelig observator for θ når $\lambda = 1$.
Er $\overline{Y}$ en tilstrekkelig observator for λ? Hva betyr dette??

11.37. La X_1, X_2, X_3 være uavhengige Bernoulli-variable med $P(X_j = 1) = p$. Er $X_1 + 2X_2 + 3X_3$ en tilstrekkelig observator for p i forhold til X_1, X_2, X_3? I forhold til $\sum_j X_j$?

11.38. Gi et eksempel som viser at bruken av en tilstrekkelig observator kan gi datareduksjon. Betyr dette at de opprinnelige dataene er overflødige?

11.39. Anta at W_n konvergerer i sannsynlighet mot τ. Følger det da at W_n er en konsistent estimator for τ? Er W_n forventningsrett?

11.40. Hva betyr det at W er en beste estimator for τ?

11.41. La $\{P^\theta\}$ være en statistisk modell hvor alle fordelingene P^θ er diskrete.
Vis at dersom T er tilstrekkelig for θ i forhold til X, så er fordelingen $P_{X\,|\,T}$ til X gitt T uavhengig av θ.
Vis at dersom fordelingen $P_{X\,|\,T}$ til X gitt T er uavhengig av θ, så er T tilstrekkelig for θ i forhold til X.

11.42. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra Exp$(1/\theta)$-fordelingen. En estimator for θ er gitt av $W_1 = nY_{\text{min}}$. En annen estimator er gitt ved utvalgets forventning, dvs $W_2 = \overline{Y}$. Beregn Var W_1 og Var W_2. Hvilken estimator er mest presis?
Hint: Et av integralene som må beregnes kan beregnes ved integralet i definisjonen av gammafunksjonen.

11.43. En kriminolog søker igjennom FBI arkiver for å undersøke forekomsten av en sjelden type fingeravtrykk. Antallet forekomster i 6 arkiver, hver med 100 000 fingeravtrykk, ble funnet til å være hhv 3, 0, 3, 4, 2, 1. Estimer den relevante parameteren med momentmetoden. Er estimatoren forventningsrett? Begrunn valget av statistisk modell.

11.44. Finn momentestimatorene for μ, σ^2 basert på et tilfeldig utvalg av størrelse n fra $N(\mu, \sigma^2)$-fordelingen. Er estimatoren for μ konsistent?

11.45. Vis at $\overline{X^k}$ er en momentestimator for $\mathrm{E}\, Y^k$ når $X_1, \ldots, X_n$ er et tilfeldig utvalg fra fordelingen til Y. En del av oppgaven er å spesifisere eventuelle forutsetninger.

11.46. La $y_1, \ldots, y_n$ være resultatet av en trekning fra en $\mathrm{B}(p)$-fordeling. Finn rimelighetsestimatet $\hat{p}$ for parameteren p. Er rimelighetsestimatoren for variansen forventningsrett?

11.47. En ung student fører statistikk over hvor mange kvinner han må invitere på en forestilling på kino i forhold til hvor mange som aksepterer. Fem kinobesøk var resultatet av å måtte invitere hhv 3, 6, 4, 2, og 9 kvinner. Gi en begrunnelse for at statistikken kan sees som resultatet av en trekning fra en geometrisk fordeling. Kommenter antagelsene i modellen. Estimer forventningsverdien og variansen til fordelingen ved utvalgets varians, ved momentmetoden, og ved rimelighetsmetoden.

11.48. La T være en tilstrekkelig observator for parameteren θ i forhold til observatoren X. Vis at rimelighetsestimatoren for θ er basert på T. Er estimatoren basert på X?

11.49. I løpet av en 8 dagers periode er avgangstidene til en buss observert til å være 5:15, 5:21, 5:14, 5:23, 5:29, 5:17, 5:15, 5:18. Anta at avgangstiden er uniformt fordelt i intervallet $[\theta_1, \theta_2]$. Estimer θ_1 og θ_2. Beregn også rimelighetsestimatet for den forventede avgangstiden.

11.50. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra tettheten

$$f(y \,|\, \theta) = (\theta + 1)y^\theta \qquad 0 < y < 1$$

Finn momentestimatoren og rimelighetsestimatoren for θ. Beregn de tilhørende estimatene når resultatet av en trekning er 0.5, 0.4 og 0.9.

Beregn også estimatorer og estimat for fordelingens forventning og varians ved de to metodene.

11.51. La $f(x) = 1/2$ for $x = \theta$ og la $g(x) = 1$ for $\theta < x < \theta + 1/2$. Definer $f(x), g(x)$ lik 0 for andre x verdier. En fordeling Q for tallinjen er da definert ved

$$Q(A) = \sum_{x \in A} f(x) + \int_A g(x)dx.$$

Vis at fordelingens forventnig er $\mu = \theta + 1/8$. Tegn grafen til fordelingens fordelingsfunksjon F.

Estimer θ ved momentmetoden basert på verdier fra en tilfeldig trekning fra fordelingen.

Fordelingen Q har en diskret del og en kontinuerlig del. Approksimer den kontinuerlige delen med g_n gitt ved

$$g_n(x) = 1/(2n) \qquad x = \theta + 1/(2n), \theta + 2/(2n), \ldots, \theta + 1/2.$$

Bruk denne approksimasjonen til å finne en rimelighetsestimator for θ basert på et tilfeldig utvalg. Hva skjer i grensen $n \to \infty$?

11.52. Hensikten med denne oppgaven er å gi en metode, betingingsmetoden, som med utgangspunkt i en forventningsrett estimator gir en forventningsrett estimator basert på en tilstrekkelig observator. I noen tilfeller gir metoden en beste estimator. Metoden virker generelt, men vi ser kun på det diskrete tilfellet.

La $\{P^\theta\}$ være en statistisk modell hvor alle fordelingene P^θ er diskrete. Anta at W er en forventningsrett estimator for τ.

Vis at dersom T er en tilstrekkelig observator for θ, så er $\mathrm{E}^\theta(W \mid T = t)$ et estimat for τ basert på observasjonen t.

Vis at estimatoren over er forventningsrett for τ.

La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra en $\mathrm{B}(p)$-fordeling. Vis at X_1 er en forventningsrett estimator for p. Vis at $T = X_1 + \ldots + X_n$ er en tilstrekkelig observator for p. Bruk betingingsmetoden til å finne en forventningsrett estimator basert på T for p.

11.53. La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Vis at X_{maks} er en tilstrekkelig observator for θ.

Dersom $X_{\text{maks}} = t$, så er $X_1 = t$ med sannsynlighet $1/n$ og ellers uniformt fordelt over $(0, t)$ med den gjenværende sannsynligheten $(n-1)/n$. Begrunn dette.

Benytt den forventningsrette estimatoren $2X_1$ til å finne en forventningsrett estimator for θ basert på T ved hjelp av betingingsmetoden.

11.54. Anta at Caspar samler inn data for å korrigere prior fordelingen han kom frem til i eksempelet med Bør Børson Jr.. La dataene være at 14 svarte at de ville kjøpe wienerbrød mens 18 svarte at de ikke ville i en liten spørreundersøkelse blant potensielle kunder. Bruk regningen i Eksempel 11.12 til å oppdatere prioren som fulgte mer direkte fra Bør sin vurdering. Endrer dette konklusjonen noe særlig?

11.55. La $\overline{y} = 2.5\text{m}$ være observert middelverdi ved 11 gjentatte målinger av en lengde. La $\sigma = 1\text{mm}$ være kjent målefeil til måleinstrumentet, og anta at enkeltmålingene er normalfordelt med middel lik den sanne lengden μ. Finn Bayes-estimatoren $\hat{miu}$ når det antas en prior fordeling $\mathrm{N}(2\text{m}, 100^2\text{m}^2)$. Er dette en rimelig prior fordeling her?

11.56. La $y_1, \ldots, y_n$ være et tilfeldig utvalg fra $\Gamma(\alpha, \beta)$. Anta at α er kjent, men at $\beta = 1/\lambda$ er ukjent. La λ ha en prior $\Gamma(a, b)$. Finn posterior fordeling til λ.

11.57. Gi et talleksempel som viser at intervallestimatet beregnet ved antatt ukjent varians for forventningsverdien i en normalfordeling kan bli smalere enn intervallestimatet beregnet ved antatt kjent varians.

11.58. Hva tror du er definisjonen av hva som menes med et ensidig konfidensintervall? Vis hvordan to ensidige konfidensintervall kan gi et konfidensintervall med ønsket nivå.

11.59. Vi har tidligere gitt en generell definisjon av hva en estimator er. Gi en begrunnelse for at en intervallestimator V er en estimator ved å identifisere et utfallsrom Ω_V slik at V tar verdier i Ω_V. Identifiser en hendelsesfamilie i Ω_V.

11.60. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Finn en 90% intervallestimator ved å bruke metoden med intervall-invertering på Y_{maks}. Bruk også metoden på $2Y_1$, og kommenter resultatet.

11.61. Vis at $(\tau \in V)$ er en hendelse når V er en områdeestimator for estimanden τ. Avhenger denne hendelsen av parameteren i modellen?

11.62. Gi en begrunnelse for kravet $\Omega_\tau \subset \Omega_{(V)}$ i definisjonen av at V er en områdeestimator for τ.

11.63. La $X \sim \mathrm{N}(\mu, \sigma^2)$. Finn h slik at $P(\mu - h < X < \mu + h) = 0.50$.

11.64. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen. Presiser hvordan μ kan oppfattes som en estimand i denne modellen.

11.65. Bevis at

$$V = \{y \,|\, \left|y - \overline{X}\right| \leq tS\}$$

gir en områdeestimator V for en fordelings forventningsverdi μ. Her er V basert på et tilfeldig utvalg fra fordelingen.

Finn konfidenskoeffisienten til V når fordelingen er bernoullifordelingen med parameter p.

11.66. Finn $t_{0.15,24}$-kvantilen til Student-t-fordelingen. Sammenlign med den tilsvarende kvantilen for normalfordelingen.

11.67. Beregn $P(-0.883 < T < 1.383)$ når T er Student-t fordelt med 9 frihetsgrader. Hvordan er fordelingen til T definert?

11.68. De ansatte ved 15 sykehus deltok i en undersøkelse vedrørende bieffekter til en gitt medisinering. Ved de 15 sykehusene var resultatet at hhv 5.8, 5.3, 4.5, 3.9, 4.6, 5.4, 7.9, 8.2, 6.9, 5.7, 4.6, 6.3, 8.4, 4.6, 7.3 prosent av pasientene fikk bieffekter etter en slik medisinering.

Beregn et 95% konfidensintervall for den virkelige middelverdien representert av disse tallene.

Spesifiser den statistiske modellen med en begrunnelse.

Beregn et 95% konfidensintervall ved hjelp av Tsjebysjevs ulikhet ved å anta at variansen er begrenset av 2 ganger utvalgets varians.

11.69. La $\overline{X}$ være forventningsverdien til et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen. Beregn konfidenskoeffisienten til intervallestimatoren

$$V = [\overline{X} - \frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \overline{X} + \frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}]$$

for forventningen μ når variansen σ er kjent.

Er V en intervallestimator når σ er ukjent?

11.70. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen hvor σ^2 er kjent. Utled en $(1 - \alpha) \cdot 100\%$ intervallestimator for μ. Hvor stor må n være for at lengden til konfidensintervallet skal være mindre enn L?

11.71. Resultatene av et eksperiment er 4.3, 5.8, 8.4, 3.7, 5.2 og 5.1. Det antas at variansen er 1.50 og at tallene kan sees som en trekning fra en normalfordeling. Beregn et 90% og et 50% konfidensintervall for forventningsverdien til fordelingen.

11.72. La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Finn et 90% konfidensintervall for θ basert på estimatoren $(n+1)Y_{\min}$. Finn også et basert på Y_{maks}. Hvilken intervallestimator er best?

11.73. Kan det finnes informasjon i et punktestimat som "forsvinner" i et intervallestimat?

11.74. Utbrudd av matforgiftninger er ofte resultatet av bakterier i salater. I 1967 undersøkte helserådet i New York 220 tunfisk salater markedsført av ulike leverandører og utsalg. Totalt ble 179 funnet å være uegnet i følge deres kriterium. Finn et 95% konfidensintervall for p, dvs den virkelige andelen av dårlige tunfisk salater i New York. Gi en begrunnelse for antagelsene i modellen.

11.75. Anta at 3 uavhengige Bernoulli-forsøk resulterte i 1 suksess. Beregn et 95% konfidensintervall for suksess-sannsynligheten ved hjelp av en eksakt intervallestimator.

11.76. For et tilfeldig utvalg fra en $\mathrm{N}(\mu, \sigma^2)$-fordeling kan det vises at utvalgets forventning $\overline{Y}$ og utvalgets varians S^2 er uavhengige observatorer.

Bruk dette til å begrunne at

$$P(\frac{\left|\overline{Y} - \mu\right|}{\sigma}\sqrt{n} < z_{\alpha/2},\ \chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}) = (1-\alpha)^2.$$

Benytt overstående til å finne et $100 \cdot (1-\alpha)$ konfidensområde i planet for parameteren $\theta = (\mu, \sigma^2)$.

11.77. Yrkespiloter risikerer å få svekket hørsel pga støy i cockpiten. Lydnivået målt for 18 kommersielle flygninger ble målt til å være hhv 74, 77, 80, 82, 82, 85, 80, 75, 75, 72, 90, 87, 73, 83, 86, 83, 83, 80 (dB). Finn et 95% konfidensintervall for forventningen μ og for variansen σ^2 ved å anta at støyen er normalfordelt. Finn et 95% konfidensområde for (μ, σ^2). Tegn en figur som gir området.

11.78. En tilfeldig trekning fra en $\mathrm{N}(\mu, \sigma^2)$ fordeling gir resultatet 249, 254, 243, 268, 253, 269, 287, 241, 273, 306, 303, 280, 260, 256, 278, 344, 304, 283, 310.

Estimer μ og σ^2.

Finn 95% konfidensintervall for μ og σ^2.

Begrunn at et 95% konfidensintervall for σ finnes ved å ta kvadratroten av intervallgrensene for 95% konfidensintervallet for σ^2.

11.79. En tilfeldig trekning fra en $\mathrm{B}(10, p)$ fordeling ga resultatet 3. Finn et eksakt 95% konfidensintervall for p.

Sammenlign dette med det approksimative 95% konfidensintervallet som følger ved tilnærming med normalfordeling.

Formuler grenseteoremet som er relevant i dette tilfellet, og forklar hvorfor det kan brukes her.

Prøv å gi en definisjon av hva som menes med et approksimativt 95% konfidensintervall.

11.80. Stikkordene "Tsjebysjevs ulikhet", "intervall-invertering", "pivotvariabel" og "approksimativt konfidensintervall" antyder fire mulige metoder for å finne konfidensintervall. Gi en kortfattet fremstilling, med talleksempel og utledning, av disse metodene.

For tilfellet "approksimativt konfidensintervall" skal du lage et eksempel hvor Poisson-approksimasjonen benyttes.

Det er en fordel om alle de fire metodene blir benyttet på det samme tallmaterialet.

11.81. Er tilsrekkelighet et relevant begrep i forbindelse med intervallestimering?

Begrunn svaret så godt du kan.

11.82. Bevis at

$$\tau(V) = \{\tau(\theta) \,|\, \theta \in V\}$$

er et konfidensområde med nivå α for en generell estmand τ dersom V er et konfidensområde med nivå α for θ.

Kapittel 12

Hypotesetesting

I chose the term 'null hypothesis' without particular regard for its etymological justification but by analogy with a usage, formerly and perhaps still current among physicists, of speaking of a null experiment, or a null method of measurement, to refer to a case in which a proposed value is inserted experimentally in the apparatus and the value is corrected, adjusted, and finally verified, when the correct value has been found; because the set-up is such, as in the Wheatstone Bridge, that a very sensitive galvanometer shows no deflection when exactly the right value has been inserted.
The governing consideration physically is that an instrument made for direct measurement is usually much less sensitive than one which can be made to kick one way or the other according to wether too large or too small a value has been inserted.

R. A. Fisher (1958)

I en gitt statistisk modell kan en hypotese, dvs en antagelse om en ukjent størrelse, identifiseres med en delmengde av parametermengden. Dette er analogt med at en hendelse identifiseres med en delmengde av utfallsrommet.

En hypotesetest basert på en observator er gitt ved at nullhypotesen forkastes hvis og bare hvis observatoren gir en verdi i det kritiske området. Beslutningsregelen i en hypotesetest kan dermed identifiseres med en observator og en hendelse, dvs det kritiske området, i verdiområdet til observatoren.

Koeffisienten til en hypotesetest er den minste øvre skranken for sannsynligheten for å feilaktig forkaste nullhypotesen. I et statistisk lovmessig eksperiment betyr dette at når en test med koeffisient 5% benyttes på gjentagelser av eksperimentet, så vil nullhypotesen forkastes feilaktig i ikke mer enn 5% av tilfellene i det lange løp.

12.1 Terningen på tiltalebenken

Så langt har vi konkludert analysen med et estimat, dvs enten et punktestimat eller et intervallestimat. Estimatet er vår gjetning på verdien til en ukjent størrelse. Intervallet er

vår gjetning på et område som inneholder den ukjente størrelsen.

Metodene er mulig å vurdere ved hjelp av sannsynlighetsteori. F eks vil et 90%-konfidensintervall være slik at sannsynligheten for at intervallet inneholder den ukjente størrelsen er minst 90%. Her er det viktig å merke seg at det er intervallet som er den tilfeldige størrelsen.

I en hypotesetest er problemstillingen enklere, dvs det finnes bare to mulige svar. En test gir en prosedyre for å avgjøre om en hypotese er riktig eller gal på grunnlag av de observerte data. Igjen kan sannsynlighetsteorien benyttes til å vurdere egenskapene til testen. Vi konkretiserer diskusjonen med et eksempel.

Eksempel 12.1 (Hypotese i terningkast) Anta at vi kaster en terning 18 ganger og registrerer at 12 av kastene resulterer i odde antall øyne. På dette grunnlaget er det nærliggende å tro at det kanskje er mer sannsynlig å få et odde antall øyne enn et like antall øyne i et terningkast. For å ta stilling til dette kan vi sette opp en statistisk modell.

La Y være antall terningkast som gir et odde antall øyne når en terning kastes 18 ganger. Da er det rimelig å anta at Y er en tilfeldig variabel med verdier i $\Omega_Y = \{0, 1, \ldots, 18\}$. Dersom vi antar at terningkastene er uavhengige og at

$$p = P(\text{odde antall øyne i et enkeltkast}),$$

så følger det at

$$Y \sim \mathrm{B}(18, p).$$

I utgangspunktet er det grunn til å anta at terningen er symmetrisk. Vi formulerer dette som at nullhypotesen er gitt ved $p = 1/2$. Den alternative hypotesen er at det er mer sannsynlig å få et odde antall øyne enn et like antall øyne i et terningkast, dvs at $p > 1/2$.

Parameterrommet i vår modell er dermed $\Omega_\Theta = \{p \,|\, 1/2 \leq p \leq 1\}$. En vanlig notasjon for det foregående er gitt ved

$$\begin{aligned} H_0 &: \quad p = 1/2, \\ H_1 &: \quad p > 1/2. \end{aligned} \tag{12.1}$$

Vi ser at hypotesene er gitt ved en partisjon av parameterrommet, dvs

$$\Omega_\Theta = \Omega_0 \uplus \Omega_1 = \{1/2\} \uplus (1/2, 1].$$

Dette gjelder generelt.

> I en statistisk modell er hypotesene i en hypotesetest spesifisert ved en oppdeling av parameterrommet i en del tilsvarende nullhypotesen og en del tilsvarende den alternative hypotesen.

La Y betegne antall terningkast av 18 som gir et odde antall øyne. Intuitivt synes det rimelig å forkaste nullhypotesen $p = 1/2$ dersom eksperimentet gir en uventet høy Y verdi.

Mer konkret er en mulig beslutningsregel for denne hypotesetesten gitt ved at H_0 forkastes dersom $Y \geq y^*$. Verdien y^* kalles den kritiske verdien til testobservatoren Y.

Vi velger y^* så stor at hendelsen

$$(Y \geq y^*), \quad \text{dvs forkastning av } H_0,$$

er svært lite sannsynlig når H_0 er sann. Et mulig valg er gitt ved å velge y^* slik at

$$P(H_0 \text{ forkastes} \,|\, H_0 \text{ er riktig}) = P(Y \geq y^* \,|\, p = 1/2) \leq 5\%.$$

I så fall er testens signifikansnivå lik 5%. Dersom testen gir forkastning av nullhypotesen, dvs dersom $y \geq y^*$, så sier vi at den alternative hypotesen er bevist "hevet over enhver tvil". Signifikansnivået er en presisering av hva som menes med "hevet over enhver tvil".

I et statistisk lovmessig eksperiment vil en hypotesetest med nivå α feilaktig forkaste nullhypotesen i ikke mer enn $\alpha \cdot 100\%$ av tilfellene i det lange løp.

For testen med $y^* = 11$ finner vi at

$$P(H_0 \text{ forkastes} \,|\, H_0 \text{ er riktig}) = \sum_{y=11}^{18} \binom{18}{y} (\frac{1}{2})^y (\frac{1}{2})^{18-y} \approx 0.24,$$

og vi sier at signifikanskoeffisienten til testen er 24%. Distinksjonen mellom signifikansnivå og koeffisient for en test tilsvarer distinksjonen mellom konfidensnivå og koeffisient for et konfidensintervall. Enhver $\alpha \geq 0.24$ kvalifiserer som et nivå for testen, og vi har spesielt bevist at testen har signifikansnivå 25%. Fordi vi observerte $y = 12$ ser vi at denne testen vil gi forkastning av nullhypotesen, men et 25% nivå kan vel neppe sies å tilsvare at H_1 er bevist "hevet over enhver tvil".

Tenk et øyeblikk på terningen som den tiltalte i en rettsak. Hypotesen H_1 er tiltalen og nullhypotesen H_0 tilsvarer at terningen er uskyldig. Et nivå på 25% betyr at rettsystemet tillater at terningen blir uskyldig dømt i opptil 25% av tilfellene. Et slikt rettssystem vil neppe oppfattes som rettferdig av folk flest.

Vi ledes derfor til å se på en test med en større kritisk verdi y^*. For $y^* = 15$ finner vi at

$$P(H_0 \text{ forkastes} \,|\, H_0 \text{ er riktig}) = \sum_{y=15}^{18} \binom{18}{y} (\frac{1}{2})^y (\frac{1}{2})^{18-y} \approx 0.004,$$

dvs signifikanskoeffisienten er lik 4 ‰ . Dette gir at testen har et nivå på 5 ‰ , som er for lite til å være av praktisk interesse. Denne testen gir at nullhypotesen beholdes fordi $y = 12$ ikke er inneholdt i mengden $\{15, 16, 17, 18\}$, dvs i testens kritiske område.

Tenk igjen på terningen som den tiltalte i en rettsak. Et nivå på 5 ‰ fører til at det skal svært mye til for at terningen skal bli dømt. Selv med en skriftlig tilståelse, fingeravtrykk og 10 øyevitner er det ikke gitt at terningen blir dømt. Et slikt rettssystem vil heller ikke oppfattes som rettferdig av folk flest.

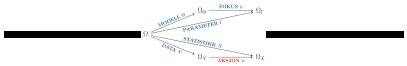
Testen med $y^* = 13$ har et nivå på 5% fordi

$$P(H_0 \text{ forkastes} \,|\, H_0 \text{ er riktig}) = \sum_{y=13}^{18} \binom{18}{y} (\frac{1}{2})^y (\frac{1}{2})^{18-y} \approx 0.048,$$

som f eks finnes ved oppslag i en statistisk tabell. Dette signifikansnivået synes rimelig. Fordi observasjonen $y = 12$ ikke er med i det kritiske området $\{13, \ldots, 18\}$ har vi følgende konklusjon:

Med et signifikansnivå på 5% forkaster denne testen ikke nullhypotesen.

Spesifiser alltid signifikansnivået i konklusjonen på en hypotesetest.



12.2 Hypotese, forkastningsregel og signifikansnivå

La $\{P^\theta\}$, $\theta \in \Omega_\Theta$, være en statistisk modell for utfallsrommet Ω.

Nullhypotesen er gitt av en delmengde Ω_0 av parameterrommet Ω_Θ. Den alternative hypotesen er gitt av komplementmengden $\Omega_1 = \Omega_\Theta \setminus \Omega_0$.

Tradisjonelt betegner H_0 nullhypotesen og H_1 den alternative hypotesen. Nullhypotesen H_0 er riktig hvis og bare hvis $\theta \in \Omega_0$. En hypotese kan dermed identifiseres med en delmengde av parameterrommet. Dette kan sammenlignes med at en hendelse identifiseres med en delmengde av utfallsrommet.

En forkastningsregel er gitt ved en testobservator og et kritisk område for testobservatoren. En testobservator W er en observator. Et kritisk område K_W for W er en hendelse i verdiområdet til observatoren. Dersom eksperimentet gir at $W(\omega) \in K_W$, så forkastes nullhypotesen.

En forkastningsregel kan også kalles en beslutningsregel, men begrepet beslutningsregel kan benyttes mer generelt. F eks er beslutningsregelen tilsvarende en intervallestimator for en estimand τ gitt ved at dersom eksperimentet gir et intervallestimat $[x, y]$, så er konklusjonen at vi beslutter at $x \leq \tau \leq y$.

Kort fortalt er en beslutningsregel i en hypotesetest gitt ved en observator $W : \Omega \to \Omega_W$ og en hendelse $K_W \subset \Omega_W$. Dersom hendelsen $(W \in K_W)$ inntreffer, så forkastes nullhypotesen.

En *hypotesetest* er fullstendig spesifisert av nullhypotesen, den alternative hypotesen og en forkastningsregel.

I definisjonen av en hypotesetest inngår kun utfallsrommet og parameterrommet for eksperimentet, mens en mulig statistisk lovmessighet ikke spiller noen rolle. Familien $\{P^\theta\}$ av de mulige fordelingene kan benyttes ved en vurdering av en gitt test.

La W og K_W være hhv testobservatoren og det kritiske området til en hypotesetest hvor nullhypotesen er gitt av parameterdelmengden Ω_0. Testen har *signifikansnivå* α dersom

$$P^\theta(W \in K_W) \leq \alpha \text{ for alle } \theta \in \Omega_0.$$

Signifikanskoeffisienten til testen er lik den minste av de α verdiene som oppfyller ulikheten.

Konvensjonen med å utelate parameteren θ er vanlig også i denne sammenhengen. Vi vedtar at ulikheten

$$P(H_0 \text{ forkastes} \,|\, H_0) \leq \alpha$$

betyr eksakt det samme som foregående ulikhet. Denne notasjonen er intuitivt rimelig og leses som

"sannsynligheten for at H_0 forkastes gitt H_0 er mindre enn eller lik α".

Her kan det og bemerkes at testen gir at hendelsen $(W \in K_W)$ er identisk med hendelsen (H_0 forkastes).

I praksis kan ofte signifikanskoeffisienten til en test finnes som maksimalverdien til funksjonen $f : \Omega_0 \to \mathbb{R}$ definert ved

$$f(\theta) = P^\theta(W \in K_W).$$

Nullhypotesen er *enkel* dersom den er gitt ved en enpunktsmengde, dvs dersom

$$\Omega_0 = \{\theta_0\}.$$

I dette spesielle tilfellet er signifikanskoeffisienten lik funksjonsverdien $f(\theta_0)$.

12.3 Utledning av en test med et gitt nivå

I mange situasjoner er det ønskelig å utlede en test med et gitt signifikansnivå α. Vi gir et enkelt eksempel på hvordan dette kan gjøres.

Anta at testen skal baseres på en tilfeldig variabel W med et kritisk område av typen

$$K_W = \{w \,|\, w \geq w^*\}.$$

Dersom vi videre antar at W er kontinuerlig fordelt gitt nullhypotesen, og at W kun har én mulig fordeling gitt nullhypotesen, så er den kritiske verdien $w^* = \tilde{w}$, hvor $\tilde{w}$ er gitt av

$$P(W \geq \tilde{w} \,|\, H_0) = \alpha.$$

Dermed er den kritiske verdien ikke annet enn α kvantilen til fordelingen til W. Dette er illustrert i figur 12.1.

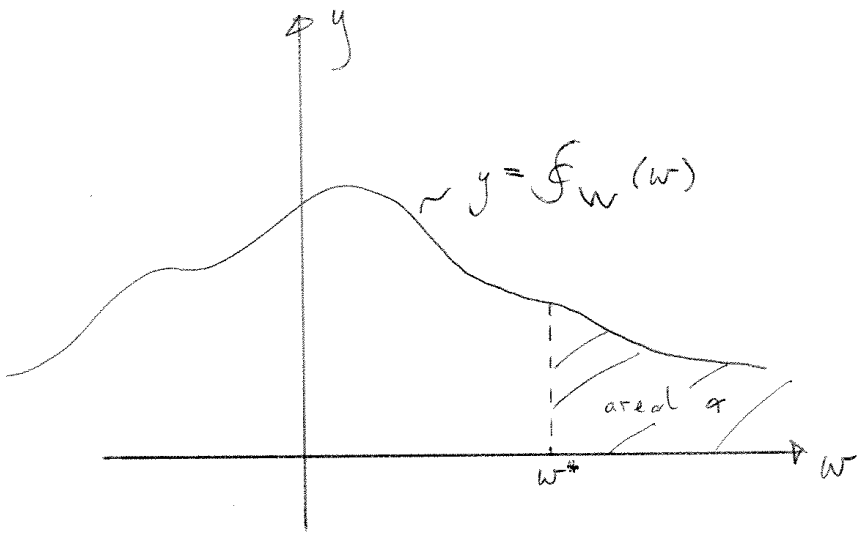
Mer generelt gir ligningen en $\tilde{w}(\theta)$ for hver θ som er mulig under nullhypotesen. Dersom vi velger

$$w^* \geq \tilde{w}(\theta) \text{ for alle } \theta,$$

så følger det at

$$P^\theta(W \geq w^*) \leq P^\theta(W \geq \tilde{w}(\theta)) = \alpha, \quad \text{for alle } \theta \in \Omega_0.$$

En unik w^* kan finnes som den minste øvre skranke for $\tilde{w}(\theta)$. I praksis kan ofte w^* finnes som maksimalverdien til funksjonen $\tilde{w}$.



Figur 12.1: Kritisk område $w \geq w^*$ for en test med signifikanskoeffisient α.

I eksemplene så langt har forkastningsområdet vært på formen

$$[w^*, \infty).$$

I terningeksemplet så var dette intuitivt motivert. I andre eksempler kan det naturlige forkastningsområdet ha en annen form, f eks

$$[w^*, \infty),\ [w_1, w_2],\ \{w \,|\, |w - w_0| \geq w_r\} \text{ etc.}$$

12.4 Styrkefunksjonen og rimelighetsmetoden

Enhver beslutningsregel basert på tilfeldige størrelser gir feil beslutning med en viss sannsynlighet. Regelen er statistisk god dersom sannsynligheten for feil er liten.

I en hypotesetest er det to mulige feil. Regelen kan forkaste nullhypotesen selv om denne er riktig, eller regelen kan beholde nullhypotesen selv om alternativet er riktig. Disse feilene kalles hhv type-1 og type-2 feil. Nivået til en hypotesetest begrenser sannsynligheten for type-1 feil, og velges typisk lik 5% eller 1% fordi sannsynligheten for type-1 feil ønskes liten.

Når nivået α er gitt ønsker en å minimalisere sannsynligheten for type-2 feil. Denne sannsynligheten avhenger imidlertid av parameteren i den statistiske modellen, og for parameterverdier i nærheten av nullhypotesen er denne sannsynligheten omtrent lik $1 - \alpha$. Dette betyr at det typisk er umulig å ha en liten sannsynlighet for både type-1 og type-2 feil.

Styrkefunksjonen til en test gir sannsynligheten for at nullhypotesen forkastes som funksjon av parameteren. Verdien til styrkefunksjonen for parameterverdier tilsvarende

den alternative hypotesen gir teststyrken. Dette betyr at teststyrken er lik 1 minus sannsynligheten for type-2 feil. En test er uniformt bedre enn en annen dersom den har større teststyrke, dvs dersom den har mindre sannsynlighet for type-2 feil.

La l_0 være en minste øvre skranke for verdiene til rimelighetsfunksjonen tilsvarende nullhypotesen, og la l være en minste øvre skranke for alle verdiene til rimelighetsfunksjonen. Forholdet l_0/l er den relative rimeligheten til nullhypotesen. Den relative rimeligheten avhenger av rimelighetsfunksjonen, som igjen er gitt av en observasjon. Rimelighetsobservatoren er den tilhørende tilfeldige variabelen gitt ved at observasjonene er tilfeldige. Rimelighetstesten er gitt ved å forkaste nullhypotesen dersom den relative rimeligheten til nullhypotesen er mindre enn eller lik den kritiske rimeligheten. Den kritiske rimeligheten kan velges slik at testen får riktig nivå.

Vi starter med en rask oppsummering av tidligere hovedpunkter. En hypotese H_0 er gitt ved en delmengde $\Omega_0 \subset \Omega_\Theta$. En beslutningsregel er gitt ved en testobservator W og et kritisk område $K_W \subset \Omega_W$. H_0 forkastes dersom $w = W(\omega) \in K_W$. Denne testen har signifikansnivå α dersom

$$P(H_0 \text{ forkastes} \,|\, H_0) = P^\theta(W \in K_W \,|\, \theta \in \Omega_0) \leq \alpha.$$

Dersom eksperimentet er statistisk lovmessig betyr dette at testen feilaktig forkaster H_0 i ikke mer enn $\alpha \cdot 100\%$ av tilfellene i det lange løp.

Den alternative hypotesen har spillt en liten rolle så langt. Vi skal se at den alternative hypotesen kommer frem fra kulissene når vi skal sammenligne hypotesetester med et gitt nivå.

Vi har heller ikke gitt noen generell metode for utledning av hypotesetester. Dette skal vi rette på ved å presentere rimelighetstesten. Rimelighetstesten har samme status for hypotesetesting som rimelighetsestimatoren har for estimering.

Dersom en test feilaktig forkaster nullhypotesen, så er dette en *type-1 feil*. Med denne definisjonen kan definisjonen av nivået til en test reformuleres ved at testen har nivå α hvis og bare hvis

$$P(\text{type-1 feil}) \leq \alpha.$$

Dermed har vi:

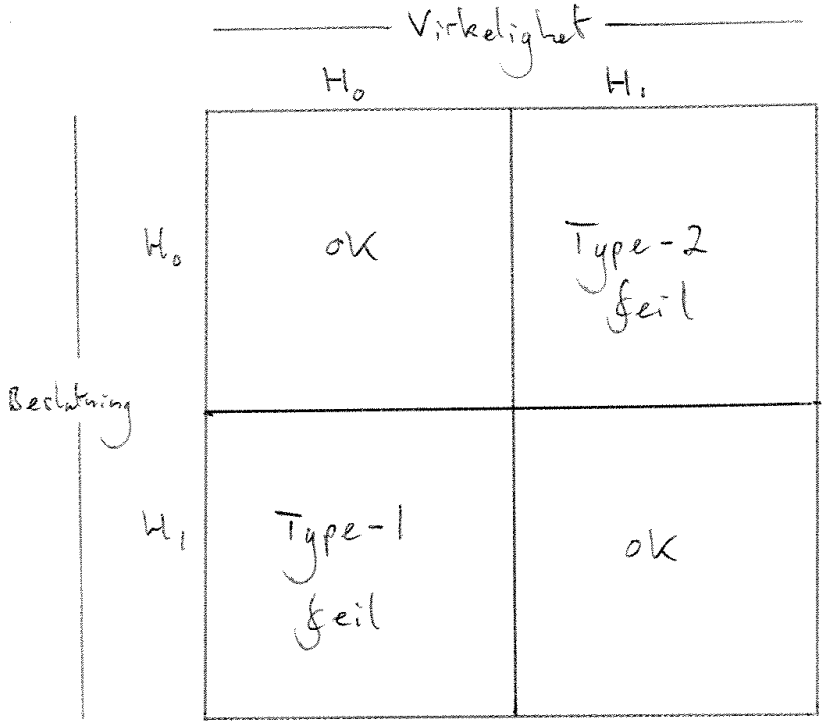
Nivået til en hypotesetest begrenser sannsynligheten for å gjøre en type-1 feil.

Det finnes en annen mulig feil. Dersom en test ikke forkaster nullhypotesen, selv om den alternative hypotesen er riktig, så er dette en *type-2 feil*. De mulige feiltypene er illustrert i figur 12.2.

Notasjonen indikerer en usymmetri mellom nullhypotesen og den alternative hypotesen. En type-1 feil sees på som mer alvorlig enn en type-2 feil. Dette kommer til uttrykk ved at vi først kontrollerer sannsynligheten for type-1 feil ved å velge ønsket nivå.

Sannsynligheten for type-2 feil er gitt ved

$$P(\text{type-2 feil}) = P(H_0 \text{ forkastes ikke} \,|\, H_1) = P^\theta(W \notin K_W \,|\, \theta \in \Omega_1).$$



Figur 12.2: Klassifisering av feil i en hypotesetest.

Gitt et signifikansnivå, dvs gitt en begrensning på sannsynligheten for å gjøre type-1 feil, så ønsker vi at sannsynligheten for type-2 feil skal være så liten som mulig. Dette kan benyttes til sammenligning av tester med et gitt signifikansnivå. Tradisjonelt gjøres dette ved hjelp av testenes styrkefunksjoner.

La $\{P^\theta\}$, $\theta \in \Omega_\Theta$, være en statistisk modell. La videre W og K_W være testobservatoren og det kritiske området for en hypotesetest. *Styrkefunksjonen*

$$\beta : \Omega_\Theta \to [0, 1]$$

defineres ved

$$\beta(\theta) = P^\theta(W \in K_W).$$

Styrkefunksjonen gir dermed sannsynligheten for å forkaste nullhypotesen som en funksjon av parameteren. Symbolet γ benyttes også i noen tilfeller for å betegne styrkefunksjonen. Vi har her valgt β fordi dette synes mest tradisjonelt, og symbolet "β" ligner på symbolet "p" som kan tas som en indikasjon på den engelske betegnelsen "power-function".

Styrkefunksjonen β inneholder all informasjon som skal til for en vurdering av testen. F eks har testen et nivå α hvis og bare hvis

$$\beta(\theta) \leq \alpha \text{ for alle } \theta \in \Omega_0,$$

og koeffisienten til testen er den minste øvre skranken for mengden $\{\beta(\theta) \,|\, \theta \in \Omega_0\}$.

Teststyrken $\beta(\theta)$ er definert for alle parameterverdier θ tilsvarende den alternative hypotesen, dvs for $\theta \in \Omega_1$. Dette er det samme som at teststyrken er restriksjonen av styrkefunksjonen β til Ω_1. Vi bemerker at teststyrken er gitt ved

$$P^\theta(W \in K_W \,|\, \theta \in \Omega_1) = 1 - P^\theta(W \notin K_W \,|\, \theta \in \Omega_1) = 1 - P(\text{type-2 feil}).$$

Teststyrken er dermed lik 1 minus sannsynligheten for type-2 feil. Dermed er det klart at det er ønskelig med så stor teststyrke som mulig for å minimalisere sannsynligheten for type-2 feil. Dette motiverer et generelt sammenligningskriterium for tester.

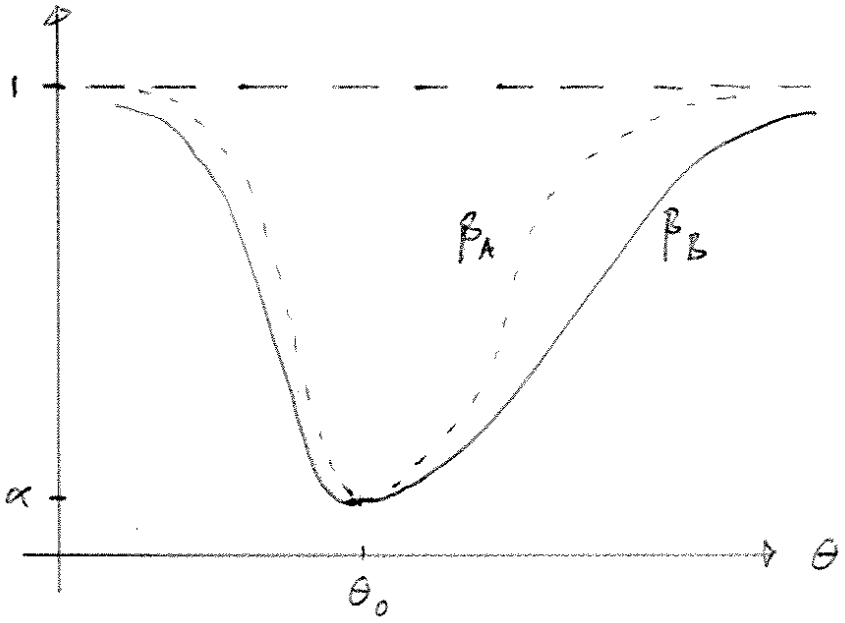
La A og B betegne to tester med likt signifikansnivå og la β_A og β_B være de tilhørende styrkefunksjonene. Testen A er en *uniformt bedre test* enn testen B dersom

$$\beta_A(\theta) \geq \beta_B(\theta) \text{ for alle } \theta \in \Omega_1.$$

Adverbet "uniformt" er her en understrekning av at ulikheten skal gjelde for alle parameterverdier tilsvarende den alternative hypotesen, og ikke bare for en enkelt parameterverdi. Definisjonen kan reformuleres ved bruk av definisjonen av teststyrke:

> En test er uniformt bedre enn en annen test dersom den har større teststyrke.

Figur 12.3 viser et eksempel på to mulige styrkefunksjoner. Som regel er styrkefunksjonen kontinuerlig. Dette gir spesielt at teststyrken tvinges til å være liten for parameterverdier i nærheten av nullhypotesens parameterverdier. Dette er forklaringen på at vi ikke kan klare å kontrollere både sannsynligheten for type-1 og type-2 feil.



Figur 12.3: Styrkefunksjonene viser at test A er uniformt bedre enn test B.

12.5 Rimelighetstesten

Anta som før at en statistisk modell med parameterrom Ω_Θ er gitt, og la Ω_0 være parametermengden tilsvarende nullhypotesen H_0. La L være rimelighetsfunksjonen tilsvarende en observasjon x.

Rimeligheten $\lambda(x)$ til hypotesen H_0 gitt observasjonen x er definert lik

$$\lambda(x) = \frac{\sup\{L(\theta) \,|\, \theta \in \Omega_0\}}{\sup\{L(\theta) \,|\, \theta \in \Omega_\Theta\}}.$$

Her er telleren $\sup\{L(\theta) \,|\, \theta \in \Omega_0\}$ lik det minste tallet l_0 som oppfyller

$$L(\theta) \leq l_0 \text{ for alle } \theta \in \Omega_0,$$

og tilsvarende for nevneren. Generelt betegner $\sup M$ den minste øvre skranken for mengden M.

Anta at rimelighetsestimatet $\hat{\theta}$ finnes, dvs anta at det finnes et globalt maksimalpunkt $\hat{\theta}$ for L. Anta at $\hat{\theta}_0$ er rimelighetsestimatet tilsvarende at parametermengden Ω_Θ er erstattet av parametermengden Ω_0. Da følger det at den relative rimeligheten er

$$\boxed{\lambda(x) = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}}.$$

Metodene vi har brukt for å bestemme rimelighetsestimat er pga dette også viktige i forbindelse med konstruksjon av en rimelighetstest.

La $\lambda(x)$ være den relative rimeligheten til hypotesen H_0 gitt observasjonen x av verdien til observatoren X. *Rimelighetsobservatoren* Λ til hypotesen H_0 basert på X er definert ved

$$\Lambda = \lambda(X).$$

Intuitivt synes det rimelig å forkaste nullhypotesen dersom den har liten relativ rimelighet. Dette motiverer definisjonen av den generelle rimelighetstesten.

Fikser et tall λ^* og la Λ være rimelighetsobservatoren til nullhypotesen basert på observatoren X. En rimelighetstest med kritisk rimelighet λ^* er gitt ved å forkaste nullhypotesen dersom $\Lambda \leq \lambda^*$. Denne entydige testen er *rimelighetstesten* for nullhypotesen basert på X.

Den kritiske rimeligheten λ^* kan bestemmes slik at testen får et ønsket nivå α. Når α er gitt betyr dette at λ^* velges så stor som mulig slik at det kritiske området $[0, \lambda^*]$ blir så stort som mulig. Dette betyr at λ^* kan velges lik den største nedre skranke for mengden av $1-\alpha$ kvantiler tilsvarende mengden av de mulige fordelinger til rimelighetsobservatoren Λ.

Det vanligste tilfellet er at en rimelighetstest baseres på et tilfeldig utvalg, dvs X er spesielt en vektor med dimensjon lik størrelsen til utvalget. Formuleringen av rimelighetstesten stiller imidlertid ingen krav til observatoren X. Et godt valg vil være å velge en observator X som er tilstrekkelig for parameteren i forhold til observasjonene. Det kan vises at en rimelighetstest basert på en tilstrekkelig observator er ekvivalent med rimelighetstesten basert på observasjonene selv. Dette viser at en rimelighetstest automatisk oppfyller tilstrekkelighetsprinsippet.

Eksempel 12.2 (Rimelighetstest for uniform fordeling) La $Y_1, \ldots, Y_n$ være et tilfeldig utvalg fra $\mathrm{U}(0, \theta)$-fordelingen. Vi vil teste nullhypotesen $\theta = \theta_0$ mot alternativet $\theta < \theta_0$. Parameterrommet er da

$$\Omega_\Theta = \{\theta \,|\, \theta \leq \theta_0\}.$$

Nullhypotesen er enkel og gitt ved

$$\Omega_0 = \{\theta_0\}.$$

Rimelighetsfunksjonen til observasjonene $y_1, \ldots, y_n$ er

$$L(\theta) = f(y_1 \,|\, \theta) \cdots f(y_n \,|\, \theta) = 1/\theta^n \qquad 0 \leq y_j \leq \theta,$$

og $L(\theta) = 0$ ellers. Rimelighetsestimatet $\hat{\theta}_0$ for θ tilsvarende parametermengden Ω_0 er trivielt gitt ved

$$\hat{\theta}_0 = \theta_0.$$

Rimelighetsestimatet $\hat{\theta}$ for θ tilsvarende parametermengden Ω_Θ er gitt ved

$$\hat{\theta} = y_{\text{maks}},$$

fordi dette er et maksimalpunkt for L. Tilsammen finner vi den relative rimeligheten

$$\lambda(y) = \frac{1/\theta_0^n}{1/y_{\text{maks}}^n} = \left(\frac{y_{\text{maks}}}{\theta_0}\right)^n \qquad y_{\text{maks}} \leq \theta_0.$$

Dersom $y_{\text{maks}} > \theta_0$, så er $\lambda(y) = 0$ fordi dette gir $L(\theta_0) = 0$. Notasjonen over er konsistent med konvensjonen å kun spesifisere funksjonen der verdien er ulik 0, slik vi også gjør for tettheter til tilfeldige variable.

Rimelighetsobservatoren er dermed gitt ved

$$\Lambda = \left(\frac{Y_{\text{maks}}}{\theta_0}\right)^n \cdot 1_{(Y_{\text{maks}} \leq \theta_0)}.$$

Med en kritisk rimelighet $\lambda^* \leq 1$ finner vi at sannsynligheten for å feilaktig forkaste nullhypotesen er

$$P(\Lambda \leq \lambda^*) = P(Y_{\text{maks}} \leq \theta_0 \sqrt[n]{\lambda^*}) = P(Y_1 \leq \theta_0 \sqrt[n]{\lambda^*}) \cdots P(Y_n \leq \theta_0 \sqrt[n]{\lambda^*}) = \lambda^*.$$

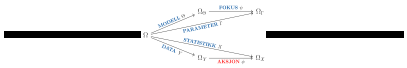
Konklusjonen er dermed at koeffisienten til testen er lik den kritiske rimeligheten!

En ekvivalent α nivå test med Y_{maks} som testobservator har et kritisk område

$$[0, \theta_0 \sqrt[n]{\alpha}].$$

Dette følger fordi

$$(0 \leq \Lambda \leq \alpha) = (0 \leq Y_{\text{maks}} \leq \theta_0 \sqrt[n]{\alpha}).$$



Eksempel 12.3 (Test for forventning med ukjent varians) Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen. Parameterrommet er

$$\Omega_\Theta = \{(\mu, \sigma^2) \,|\, \mu \in \mathbb{R}, \sigma \geq 0\}.$$

Vi vil teste $H_0 : \mu = \mu_0$ mot alternativet $H_1 : \mu \neq \mu_0$.

Nullhypotesen er gitt ved

$$\Omega_0 = \{(\mu, \sigma^2) \,|\, \mu = \mu_0, \sigma \geq 0\}.$$

Denne nullhypotesen er ikke enkel, dvs ikke gitt av et enkelt punkt i parameterrommet. Geometrisk er nullhypotesen gitt ved en stråle slik som indikert i figur 12.4.

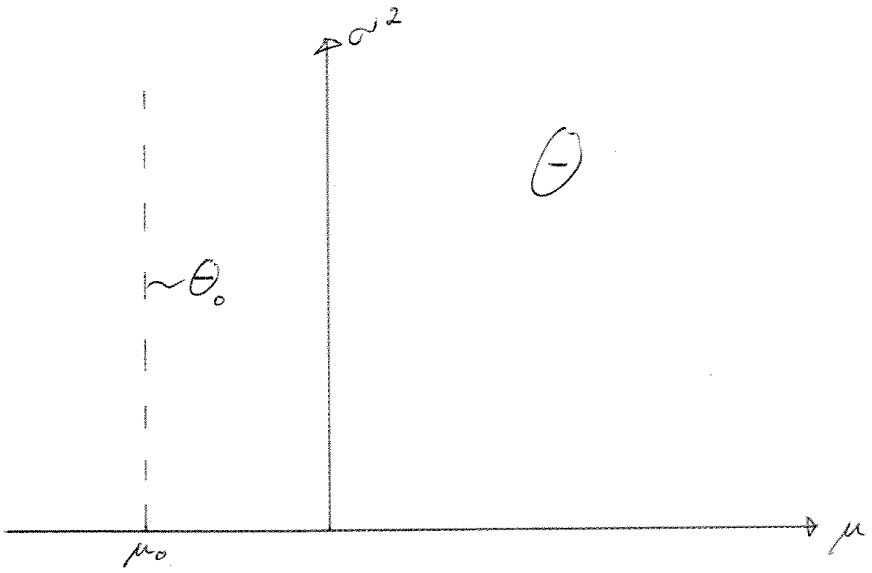
Vi vil bruke rimelighetsmetoden til konstruksjon av en test. Den relative rimeligheten til H_0 er

$$\lambda(x) = \frac{\text{“rimeligheten til } H_0\text{”}}{\text{“total rimelighet”}} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})},$$

hvor $\hat{\theta}_0$ og $\hat{\theta}$ er rimelighetsestimatene tilsvarende Ω_0 og Ω_Θ.

Beregning av $\hat{\theta}_0$:

$$L = f_X(x \,|\, \theta) = f(x_1 \,|\, \theta) \cdots f(x_n \,|\, \theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2}\left(\frac{x_1-\mu_0}{\sigma}\right)^2 - \cdots - \frac{1}{2}\left(\frac{x_n-\mu_0}{\sigma}\right)^2} \tag{12.2}$$



Figur 12.4: Nullhypotesen er ikke enkel.

$$\begin{aligned}
\ln L &= \frac{n}{2}\ln\Big(\frac{1}{2\pi\sigma^2}\Big) - \frac{1}{2\sigma^2}\big((x_1-\mu_0)^2+\cdots+(x_n-\mu_0)^2\big) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\big(n\overline{x^2} - 2\mu_0 n\overline{x} + n\mu_0^2\big) \\
&\overset{t=\sigma^2}{=} -\frac{n}{2}(\ln(2\pi)+\ln(t)) - \frac{n}{2t}\big(\overline{x^2} - 2\mu_0\overline{x} + \mu_0^2\big)
\end{aligned} \tag{12.3}$$

$$0 = \frac{d}{dt}\ln L = -\frac{n}{2}t^{-1} + t^{-2}\frac{n}{2}\big(\overline{x^2} - 2\mu_0\overline{x} + \mu_0^2\big)$$

$$\sigma^2 = t = \overline{x^2} - 2\mu_0\overline{x} + \mu_0^2 = \overline{x^2} - \overline{x}^2 + (\overline{x}-\mu_0)^2$$

Regningen over gir rimelighetsestimatet

$$\hat{\sigma}_0^2 = \overline{x^2} - \overline{x}^2 + (\overline{x}-\mu_0)^2$$

for variansen når forventningsverdien er kjent. Verdien til rimelighetsfunksjonen blir

$$l_0 = L(\hat{\theta_0}) = \Big(\frac{1}{2\pi}\Big)^{n/2}\Big(\frac{1}{\overline{x^2} - \overline{x}^2 + (\overline{x}-\mu_0)^2}\Big)^{n/2} e^{-n/2}$$

En tilsvarende regning gir rimelighetsestimatet

$$\hat{\theta} = (\overline{x}, \overline{x^2} - \overline{x}^2)$$

når både forventningsverdi og varians er ukjent. Verdien til rimelighetsfunksjonen blir

$$l = L(\hat{\theta}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\overline{x^2} - \overline{x}^2}\right)^{n/2} e^{-n/2}$$

Den relative rimeligheten til H_0 blir da

$$\begin{aligned}\lambda(x) = l_0/l &= \left(\frac{\overline{x^2} - \overline{x}^2}{\overline{x^2} - \overline{x}^2 + (\overline{x} - \mu_0)^2}\right)^{n/2} \\ &= \left(\frac{\overline{x^2} - \overline{x}^2 + (\overline{x} - \mu_0)^2}{\overline{x^2} - \overline{x}^2}\right)^{-n/2} = \left(1 + \frac{(\overline{x} - \mu_0)^2}{\overline{x^2} - \overline{x}^2}\right)^{-n/2} = \left(1 + \left(\frac{\sqrt{n}}{\sqrt{n-1}} \frac{\overline{x} - \mu_0}{s}\right)^2\right)^{-n/2},\end{aligned} \tag{12.4}$$

hvor

$$s^2 = \frac{n}{n-1}(\overline{x^2} - \overline{x}^2)$$

er utvalgets varians.

Gitt nullhypotesen $\mu = \mu_0$ så er variabelen

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S}$$

$\mathrm{t}(n-1)$-fordelt. Rimelighetsobservatoren

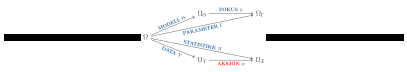
$$\Lambda = \lambda(X) = \left(1 + \frac{1}{n-1}T^2\right)^{-n/2}$$

har dermed også en fordeling som er uavhengig av parameteren $\theta \in \Omega_0$! En rimelighetstest med koeffisient α er dermed gitt ved å velge den kritiske rimeligheten λ^* lik $1-\alpha$ kvantilen til fordelingen til Λ.

En ekvivalent test med T som testobservator er imidlertid mer praktisk fordi kvantilene til T er tabulert. Forkastning ved $\Lambda \leq \lambda^*$ er ekvivalent med forkastning ved $|T| \geq t^*$ ved passende valg av t^*. Figur 12.5 gir at vi får en test med koeffisient α ved å forkaste nullhypotesen når

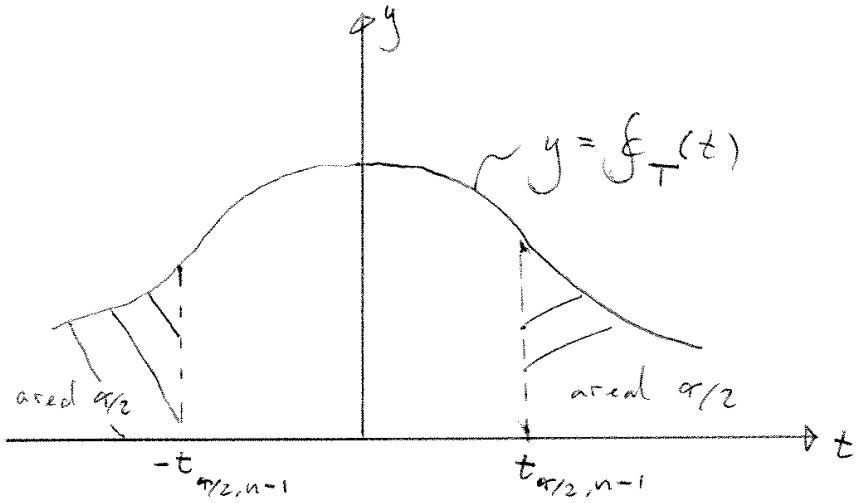
$$|T| \geq t_{\alpha/2,n-1},$$

hvor $t_{\alpha/2,n-1}$ er $\alpha/2$ kvantilen til $\mathrm{t}(n-1)$-fordelingen.

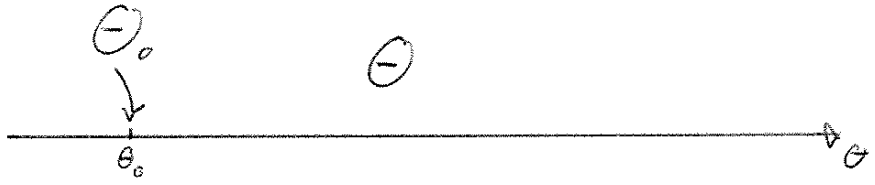


Eksempel 12.4 (Test for forventning med kjent varians) Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen hvor variansen σ^2 er kjent. Parameterrommet er

$$\Omega_\Theta = \{\mu \,|\, \mu \in \mathbb{R}\} = \mathbb{R}.$$



Figur 12.5: $P(|T| \geq t_{\alpha/2,n-1}) = \alpha/2 + \alpha/2 = \alpha$.



Figur 12.6: Nullhypotesen er enkel.

Vi vil teste $H_0 : \mu = \mu_0$ mot alternativet $H_1 : \mu \neq \mu_0$.

Nullhypotesen er enkel og gitt ved

$$\Omega_0 = \{\mu_0\}.$$

Geometrisk er nullhypotesen gitt ved et punkt slik som indikert i figur 12.6.

Ved beregning av rimelighetsestimatene følger det at den relative rimeligheten er:

$$\lambda(x) = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{\exp\left[-\frac{n}{2\sigma^2}(\overline{x^2} - 2\mu_0\overline{x} + \mu_0^2)\right]}{\exp\left[-\frac{n}{2\sigma^2}(\overline{x^2} - 2\overline{x}^2 + \overline{x}^2)\right]} = e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu_0)^2}.$$

Fordi

$$\lambda(x) \leq \lambda^*$$

er ekvivalent med

$$\left|\sqrt{n}(\overline{x} - \mu_0)/\sigma\right| \geq u^*$$

for passende u^* følger det at vi kan benytte

$$U = \sqrt{n}(\overline{X} - \mu_0)/\sigma$$

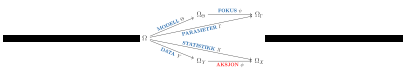
som testobservator. Fordelen med U fremfor rimelighetsobservatoren Λ er at

$$U \sim \mathrm{N}(0, 1)$$

under antagelsen $\mu = \mu_0$. Med testobservator U og kritisk område

$$\{u \,|\, |u| \geq u_{\alpha_2}\}$$

følger det at vi har en test med koeffisient α når $u_{\alpha/2}$ er $\alpha/2$ kvantilen til standard normalfordelingen.



Eksempel 12.5 (Test for varians) Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra $\mathrm{N}(\mu, \sigma^2)$-fordelingen. Parameterrommet er

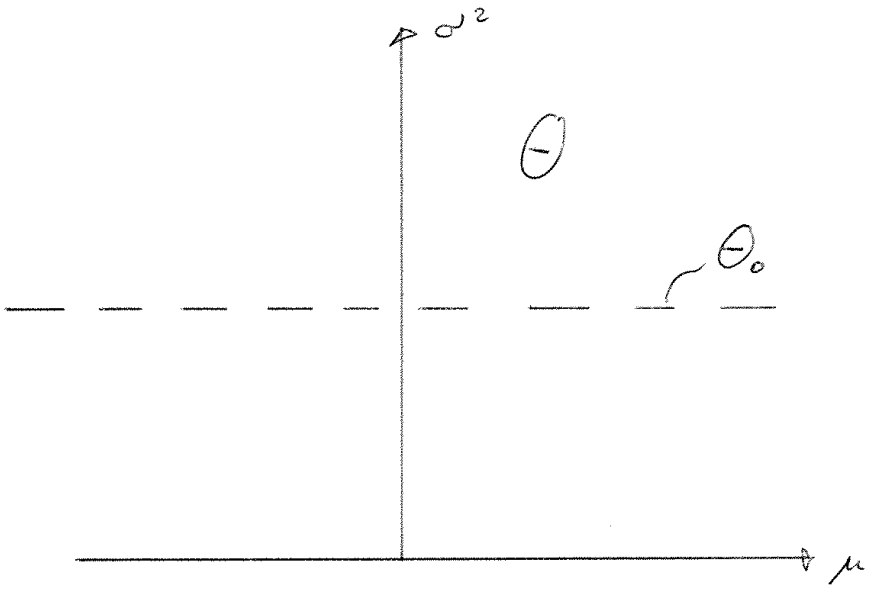
$$\Omega_\Theta = \{(\mu, \sigma^2) \,|\, \mu \in \mathbb{R}, \sigma \geq 0\}.$$

Vi vil teste $H_0 : \sigma^2 = {\sigma_0}^2$ mot alternativet $H_1 : \sigma^2 = {\sigma_0}^2$.

Nullhypotesen er gitt ved

$$\Omega_0 = \{(\mu, \sigma^2) \,|\, \mu \in \mathbb{R}, \sigma^2 = {\sigma_0}^2\}.$$

Denne nullhypotesen er ikke enkel, dvs ikke gitt av et enkelt punkt i parameterrommet. Geometrisk er nullhypotesen gitt ved en linje slik som indikert i figur 12.7.



Figur 12.7: Nullhypotesen er gitt ved en linje i parameterrommet.

Rimelighetsestimatene for θ tilsvarende Ω_0 og Ω_Θ er gitt ved

$$\hat{\theta}_0 = (\overline{x}, {\sigma_0}^2) \qquad \hat{\theta}_0 = (\overline{x}, \overline{x^2} - \overline{x}^2).$$

Den relative rimeligheten til nullhypotesen blir dermed

$$\lambda(x) = \frac{\left(\frac{1}{{\sigma_0}^2}\right)^{n/2} \exp\left(-\frac{n}{2}\frac{\overline{x^2}-\overline{x}^2}{{\sigma_0}^2}\right)}{\left(\frac{1}{\overline{x^2}-\overline{x}^2}\right)^{n/2} \exp\left(-\frac{n}{2}\frac{\overline{x^2}-\overline{x}^2}{\overline{x^2}-\overline{x}^2}\right)} = \left(\frac{\overline{x^2}-\overline{x}^2}{{\sigma_0}^2}\right)^{n/2} e^{-\frac{n}{2}\left(\frac{\overline{x^2}-\overline{x}^2}{{\sigma_0}^2}+1\right)}$$

Rimelighetsobservatoren blir

$$\Lambda = \lambda(X) = Z^{n/2} e^{-\frac{n}{2}(Z+1)},$$

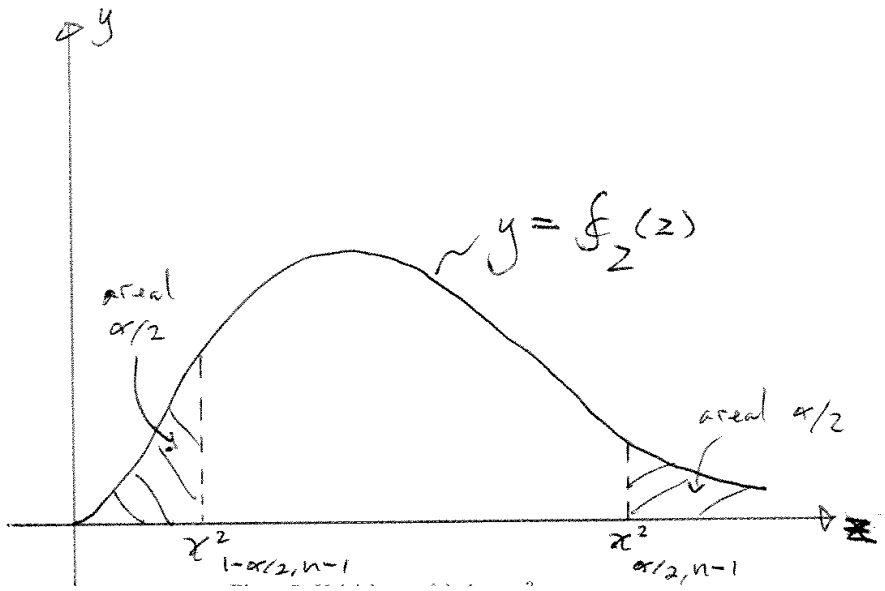
hvor

$$Z = \frac{\overline{X^2} - \overline{X}^2}{{\sigma_0}^2}.$$

Under nullhypotesen $\sigma^2 = {\sigma_0}^2$ følger det at

$$Z \sim \chi^2(n-1),$$

dvs spesielt er fordelingen til Z uavhengig av $\theta \in \Omega_0$. Dette gir at fordelingen til rimelighetsobservatoren Λ også er uavhengig av $\theta \in \Omega_0$! En test med koeffisient α er dermed gitt ved å velge den kritiske rimeligheten lik $1-\alpha$ kvantilen til Λ. Kvantilene til Λ kunne godt ha vært tabulert for forskjellige n, men dette er ikke vanlig i mindre tabellverk.



Figur 12.8: Kritisk område i en χ^2-test.

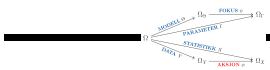
Et alternativ er å la Z være testobservatoren. Det er klart at Λ er liten dersom Z er liten (pga $Z^{n/2}$ faktoren) eller dersom Z er stor (pga $\exp -\frac{n}{2}(Z+1)$ faktoren). Dette motiverer å velge et kritisk område på formen $[0,a] \cup [b,\infty)$. Dette gir en test med koeffisient α dersom

$$P_Z([0,a] \cup [b,\infty)) = \alpha.$$

Konvensjonelt velges det kritiske området slik at hvert av intervallene får like stor sannsynlighet, dvs det kritiske området velges lik

$$[0, \chi^2_{1-\alpha/2,n-1}] \cup [\chi^2_{\alpha/2,n-1}, \infty).$$

Dette er illustrert i figur 12.8 . Denne χ^2-testen har koeffisient α, men den er ikke ekvivalent med rimelighetstesten, dvs det er ikke en rimelighetstest.



12.6 Oppgaver

12.1. Definisjonen av et kritisk område K_W for W er at det er en hendelse i verdiområdet til observatoren. Prøv å gi en begrunnelse for at definisjonen velges slik.

Hint: Hva med $(W \in K_W)$?

12.2. Gi en begrunnelse for at i et statistisk lovmessig eksperiment vil en hypotesetest med nivå α feilaktig forkaste nullhypotesen i ikke mer enn $\alpha \cdot 100\%$ av tilfellene i det lange løp. Hva betyr det at koeffisienten til testen er lik α?

12.3. La $Y \sim \mathrm{B}(18, p)$. Konstruer en test av hypotesen $p = 1/2$ mot alternativet $p > 1/2$ med signifikansnivå 10%. Det kritiske området skal være på formen $K_Y = \{y^*, \ldots, 18\}$ hvor y^* er så liten som mulig. Beregn også signifikanskoeffisienten til denne testen. Hva er konklusjonen på testen dersom vi observerer $y = 12$?

Hint: En statistisk tabell kan være nyttig.

12.4. La $Y \sim \mathrm{B}(n, p)$.

Anta at $n = 12$. Finn en 5% forkastningregel for å teste $H_0 : p = 1/2$ mot $H_1 : p > 1/2$.

Anta at $n = 14$. Finn en 5% forkastningregel med testobservator $W = n - Y$ for å teste $H_0 : p = 3/5$ mot $H_1 : p < 3/5$.

Anta at $n = 20$. Finn en 1% forkastningregel for å teste $H_0 : p = 0{,}3$ mot $H_1 : p > 0{,}3$.

12.5. Gjør oppgave 6.2.3 i Larsen-Marx.

12.6. Gjør oppgave 6.2.4 i Larsen-Marx.

12.7. Gjør oppgave 6.2.5 i Larsen-Marx.

12.8. Gjør oppgave 6.2.7 i Larsen-Marx.

12.9. Gjør oppgave 6.2.8 i Larsen-Marx.

12.10. Gjør oppgave 6.2.9 i Larsen-Marx.

12.11. Gjør oppgave 6.2.10 i Larsen-Marx.

12.12. Gjør oppgave 6.2.11 i Larsen-Marx.

12.13. Anta at nullhypotesen er enkel. Det er gitt at en hypotesetest skal basere seg på en tilfeldig variabel W med et kritisk område av typen $K_W = [a, b]$. Hvordan vil du bestemme $[a, b]$ slik at testen får en konfidenskoeffisient lik α når du antar at W er kontinuerlig fordelt? Hva med tilfellet når W er diskret fordelt? Hvilken rolle spiller den alternative hypotesen her?

12.14. La p være parameteren tilsvarende en Bernoulli forsøksrekke. Anta at $H_0 : p = p_0$ blir testet mot alternativet $H_1 : p > p_0$.

Dersom H_0 forkastes med et nivå 0,05, vil da H_0 nødvendigvis forkastes i en test med et nivå 0,01?

Dersom H_0 forkastes med et nivå 0,01, vil da H_0 nødvendigvis forkastes i en test med et nivå 0,05?

Begrunn svarene!

12.15. Oppgave 6.3.3 i Larsen-Marx.

12.16. Oppgave 6.3.4 i Larsen-Marx.

12.17. La $X \sim \mathrm{U}(0, \theta)$ være testobservatoren for en test med et kritisk område

$$[0,\ 0{,}1] \cup [1{,}9,\ \infty).$$

Begrunn at dette er et rimelig valg av kritisk område. Finn signifikanskoeffisienten α tilsvarende nullhypotesen $\theta = 2$. Skisser styrkefunksjonen β for denne testen når den alternative hypotesen er gitt ved $\theta \neq 2$. Hva er teststyrken for $\theta = 2{,}5$?

12.18. En urne inneholder 10 brikker. Vi ønsker å teste

$$H_0: \text{ halvparten av brikkene er hvite}$$

mot alternativet

$$H_1: \text{ mer enn halvparten av brikkene er hvite.}$$

Testen utføres ved å trekke tre brikker uten tilbakelegning og H_0 forkastes dersom minst to av disse er hvite.

Sett opp en statistisk modell hvor parameteren θ er antall hvite brikker i urnen. Finn parameterrommet tilsvarende denne testen. Finn signifikanskoeffisienten til testen. Finn teststyrken $\beta(\theta)$ for $\theta = 6$ og $\theta = 7$. Skisser styrkefunksjonen til testen.

12.19. Oppgave 6.3.8 i Larsen-Marx.

12.20. Oppgave 6.3.9 i Larsen-Marx.

12.21. La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra Geometric(p)-fordelingen. Finn rimelighetsobservatoren Λ for å teste hypotesen $p = p_0$ mot alternativet $p \neq p_0$. Kan du konstruere en test med nivå α?

12.22. La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra Exp(λ)-fordelingen. Finn rimelighetsobservatoren W for å teste hypotesen $\lambda = \lambda_0$ mot alternativet $\lambda \neq \lambda_0$. Kan du konstruere en test med nivå $\alpha = 0{,}05$?

12.23. La $X_1, \ldots, X_n$ være et tilfeldig utvalg fra N($\mu, 1$)-fordelingen. Finn rimelighetsobservatoren Λ for å teste hypotesen $\mu = \mu_0$ mot alternativet $\mu \neq \mu_0$. Kan du konstruere en test med nivå $\alpha = 0{,}01$?

12.24. Anta at observatoren T er tilstrekkelig for parameteren θ i forhold til observatoren X. Bevis at rimelighetsobservatoren basert på T er lik rimelighetsobservatoren basert på X.

Hint: Bruk faktoriseringen av rimelighetsfunksjonen som tilstrekkeligheten gir.

12.25. Finn rimelighetsobservatoren Λ basert på Y_{maks} fra et tilfeldig utvalg av størrelse n fra U($0, \theta$)-fordelingen tilsvarende en test av hypotesen $\theta = \theta_0$ mot alternativet $\theta < \theta_0$.

Kommenter resultatet i forhold til eksemplet i teksten.

La $\theta_0 = 1$ og bestem den kritiske rimeligheten slik at testen har koeffisient lik 0,05. Skisser styrkefunksjonen β for denne testen.

Erstatt Y_{maks} med Y_1 i det foregående. Skisser styrkefunksjonen γ for denne testen.

Er den første testen uniformt bedre enn den siste testen? Er den første testen bedre enn den siste på andre vis?

12.26. Vis ved et talleksempel at χ^2-testen ikke er en rimelighetstest.

Mer konkret: Finn utfall x og y slik at den relative rimeligheten til H_0 oppfyller $\lambda(x) \leq \lambda(y)$, mens χ^2-testen forkaster nullhypotesen for y, men ikke for x.

Hint: Beregn verdiene til den relative rimeligheten tilsvarende intervallendepunktene i χ^2-testen for et passende eksempel.

12.27. Bruk sentralgrenseteoremet til å begrunne at testene vi har gitt for tilfeldig utvalg fra normalfordelingen gir approksimative tester dersom vi tar et tilstrekkelig stort utvalg fra en vilkårlig fordeling med endelig varians.

Begrunn de tre tilfellene: test av forventning med ukjent varians, test av forventning med kjent varians, test av varians hver for seg.

12.28. Oppgave 7.5.4 i Larsen-Marx.

12.29. Oppgave 7.5.5 i Larsen-Marx.

12.30. Oppgave 7.5.6 i Larsen-Marx.

12.31. Oppgave 7.5.9 i Larsen-Marx.

12.32. Oppgave 7.5.10 i Larsen-Marx.

12.33. Oppgave 7.5.13 i Larsen-Marx.

12.34. Oppgave 7.5.14 i Larsen-Marx.

12.35. Oppgave 7.5.15 i Larsen-Marx.

12.36. Oppgave 7.5.17 i Larsen-Marx.

12.37. Oppgave 7.7.1 i Larsen-Marx.

12.38. Oppgave 7.7.2 i Larsen-Marx.

12.39. Oppgave 7.7.3 i Larsen-Marx.

12.40. Oppgave 7.7.7 i Larsen-Marx.

12.41. Oppgave 7.7.5 i Larsen-Marx.

12.42. Oppgave 7.7.9 i Larsen-Marx.

Del IV

Tillegg

Tillegg A

Matematikk

Det antas at leseren er fortrolig med de vanligste reelle funksjoner og sammenhengene mellom derivasjon og integrasjon slik det presenteres i innføringskurs i analyse. Det inkluderer også bruk av mengdeteori med eksempler fra mengder i planet og i rommet og naturlige tall. Noe av dette er raskt repetert i dette tillegget.

A.1 Litt analyse

Eksponensialfunksjonen exp kan defineres ved

$$y = \exp(x) = \mathrm{e}^{\mathrm{x}} = 1 + \mathrm{x} + \frac{\mathrm{x}^2}{2!} + \cdots \tag{A.1}$$

Alternativt kan den defineres som den inverse funksjonen til logaritmefunksjonen ln

$$x = \ln(y) = \int_1^y \frac{dt}{t} \tag{A.2}$$

Et tredje alternativ er via ligningen

$$\exp' = \exp, \quad \exp(0) = 1 \tag{A.3}$$

Et fjerde alternativ er funksjonalligningen

$$\exp(a+b) = \exp(a)\exp(b) \tag{A.4}$$

Et femte alternativ er ved å utvide den elementære potensfunksjonen

$$\exp(n) = \prod_{i=1}^{n} \mathrm{e} \tag{A.5}$$

fra heltall n til rasjonale tall n/m og videre til reelle tall x ved approksimasjon. Utvidelse videre til komplekse argument kan gjøres ved

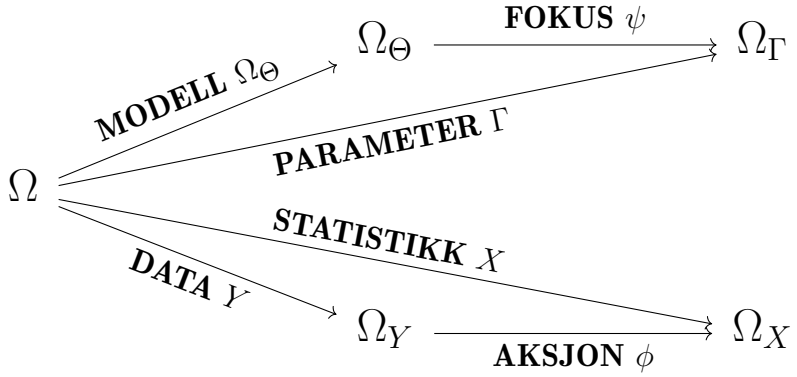
$$\exp(a + \mathbf{i}b) = \exp(a)\left[\cos(b) + \mathbf{i}\sin(b)\right] \tag{A.6}$$

men relasjonene over holder direkte for komplekse tall. Et unntak er definisjonen via logaritmefunksjonen som krever noe mer arbeid.

A.2 Funksjoner og mengder

Begrepene *definisjonsmengde* og *verdimengde* til en funksjon antas kjent. Spesielt antas det også at disse begrepene er kjent for funksjoner hvor definisjonsmengde og verdimengde er delmengder av vektorrom. I den sammenhengen kalles ofte funksjonen en *transformasjon* eller en *avbildning*. I boken her skilles det ikke på disse begrepene og de brukes synonymt. Kjært barn har mange navn!

I statistikk og sannsynlighetsteori er det generelle funksjonsbegrepet viktig. Intuitivt er en funksjon noe som sender en ting over til en annen ting. Det er derfor vanlig å bruke pilnotasjonen $f : \mathcal{D} \to \mathcal{Y}$ for en *funksjon* f. Funksjonen f er definert ved at for hver x i *definisjonsmengden* $\mathcal{D}$, så er $y = f(x)$ med i *kodomenet* $\mathcal{Y}$. Notasjonen $x \mapsto f(x)$ er vanlig.



Notasjonen $\mathcal{D}(f)$ brukes for å betegne definisjonsmengden til funksjonen f. Notasjonen $\mathcal{R}(f)$ brukes for å betegne *rekkevidden* til funksjonen f. Betegnelsen *verdimengde* er også vanlig i bruk. Definisjonen er $\mathcal{R}(f) = \{f(x) \,|\, x \in \mathcal{D}(f)\}$. I definisjonen av en funksjon må også *domenet* $\mathcal{X}$ spesifiseres. Dette er en mengde $\mathcal{X}$ slik at $\mathcal{D}(f) \subset \mathcal{X}$. Dersom domenet $\mathcal{X}$ til en funksjon f ikke er spesifisert, så vil det antas at $\mathcal{X} = \mathcal{D}(f)$.

Konvensjonene over varierer noe fra forfatter til forfatter. Spesielt er det noen forfattere som ikke skiller på begrepene *domene* og *definisjonsmengde*. Konvensjonen vi bruker her gjør at mengden $\mathcal{M}$ av funksjoner med domene $\mathcal{X}$ og kodomene $\mathcal{Y}$ er veldefinert. Med $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ følger det da at både exp og ln er inneholdt i $\mathcal{M}$.

Den *inverse funksjonen* f^{-1} er definert dersom ligningen $y = f(x)$ har en unik løsning x for alle $y \in \mathcal{R}(f)$ og da er $f^{-1}(y) = x$. Videre er $\mathcal{D}(f^{-1}) = \mathcal{R}(f)$, $\mathcal{R}(f^{-1}) = \mathcal{D}(f)$ og domene og kodomene bytter roller på samme måte.

Det er vanlig å la samme symbol betegne ulike funksjoner. Funksjonene $\exp : \mathbb{R} \to \mathbb{R}$, og $\exp : \mathbb{C} \to \mathbb{C}$ er ulike funksjoner, men samme symbol exp brukes ofte for begge.

Abstrakt mengdeteori står sentralt i Kolmogorovs formulering av sannsynlighetsteorien. I det følgende vil vi ikke gjøre noe forsøk på å bygge opp mengdeteorien aksiomatisk, men vil ta som gitt at leseren har en viss fortrolighet med manipulering med mengder. Hovedinnholdet her blir da å fastlegge noen notasjons- og språklige konvensjoner. Noen av disse er allerede brukt over.

Paranteser av type $\{\cdots\}$ benyttes for å beskrive mengder. Noen eksempler på endelige mengder er gitt ved $\{a, b, c\}$ og $\{1, 2, \ldots, n\}$. Den tomme mengden $\emptyset$ er også endelig, den

har ingen elementer. Vi har $\emptyset = \{\omega \,|\, \omega \neq \omega\}$. Symbolet $|$ leses i denne sammenhengen som *slik at* og brukes generelt til definisjon av mengder. Mengden av de naturlige tall betegnes med $\mathbb{N} = \{1, 2, 3, \ldots\}$, mengden av heltall betegnes med $\mathbf{Z} = \{\ldots, -1, 0, 1, \ldots\}$, og mengden av de reelle tall betegnes med $\mathbb{R} = \{x \,|\, -\infty < x < \infty\}$. De rasjonale tallene er $\mathbb{Q} = \{m/n \,|\, m, n \in \mathbf{Z}\}$.

En mengde I er *tellbar* dersom den er i en-en korrespondanse med en endelig mengde eller mengden av de naturlige tall. Dette betyr at vi kan liste opp alle elementene i I ved $i_1, i_2, \ldots$. Dette burde forklare navnet. Bortsett fra mengden $\mathbb{R}$ av reelle tall er dermed alle overnevnte mengder tellbare.

Vi skriver $a \in A$ dersom a er et element i mengden A. Symbolet $\in$ leser vi som *er et element i*. Symbolet $\notin$ leser vi tilsvarende som *er ikke et element i*. To eksempler er gitt ved $b \in \{a, b, c\}$ og $d \notin \{a, b, c\}$.

En mengde A er en delmengde av en mengde B dersom alle elementer i A også er elementer i B. Vi benytter notasjonen $A \subset B$ for dette. Vi har da at $a \in A \Rightarrow a \in B$. Tegnet $\subset$ leses som *er en delmengde av*. Tegnet $\Rightarrow$ leses som *medfører at*. Tre eksempler er gitt ved $\{1, 100, 14\} \subset \mathbb{N}$, $\emptyset \subset \mathbb{R}$ og $\{a, b, c\} \subset \{a, b, c\}$. Merk: Den vanligste metoden for å bevise at $A = B$ er gitt ved å vise at $A \subset B$ og $B \subset A$.

Anta at A_ι er en mengde for hvert element ι i en mengde I. Da er $\{A_\iota\}$, $\iota \in I$, en *indeksert familie* av mengder. Et eksempel er gitt ved $A_\iota = [\iota, \infty)$ for $\iota \in \mathbb{R}$, $\{A_\iota\} = \{[\iota, \infty)\}$, $\iota \in \mathbb{R}$.

Unionen og snittet av mengdene i familien $\{A_i\}$ er gitt ved henholdsvis

$$\bigcup_{i \in I} A_i = \{\omega \,|\, \exists i,\ \omega \in A_i\} \quad \text{og} \quad \bigcap_{i \in I} A_i = \{\omega \,|\, \forall i,\ \omega \in A_i\}.$$

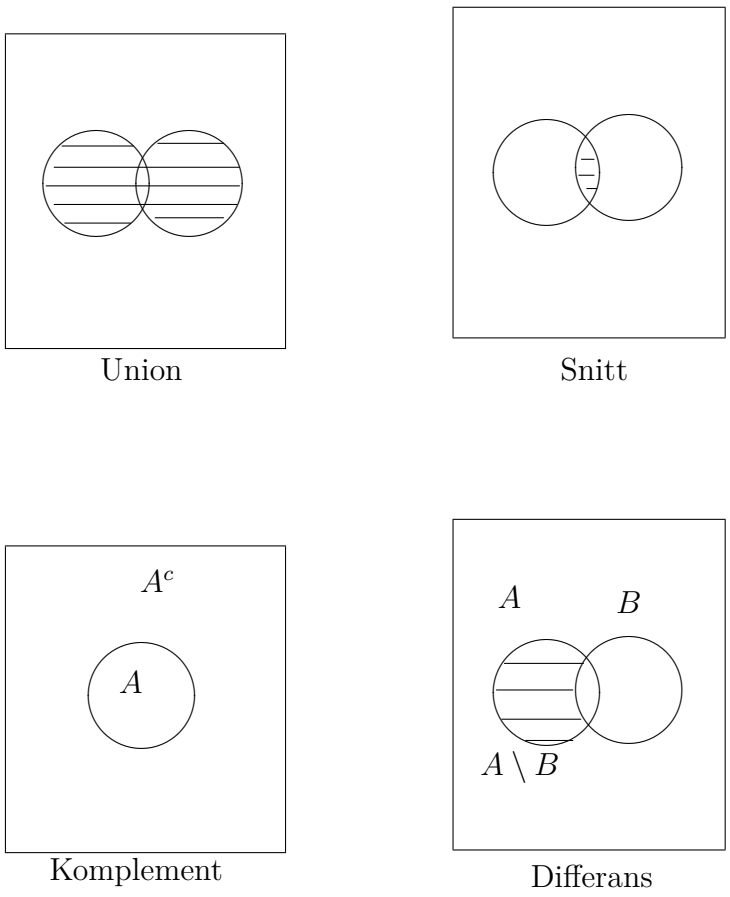
Tegnet $\exists$ leses som *det eksisterer en*. Tegnet $\forall$ leses som *for alle*. Unionen er dermed mengden av elementer i Ω som er inneholdt i minst en A_i og snittet er mengden av elementer som er inneholdt i alle A_i. Her har vi antatt at alle mengdene A_i er delmengder av en mengde Ω. To eksempler er gitt ved $\cup_{\iota \in \mathbb{R}}[\iota, \infty) = \mathbb{R}$ og $\cap_{\iota \in \mathbb{R}}[\iota, \infty) = \emptyset$. I noen sammenhenger er det underforstått hva indeksmengden I er og union og snitt skrives kortere på formen $\cup_i A_i$ eller $\cap_i A_i$.

Komplementet til A i Ω er gitt ved $A^c = \{\omega \,|\, \omega \notin A\}$. Komplementet A^c til A er mengden av alle elementer i Ω som ikke er elementer i A. F.eks. gir $\Omega = \{1, 2, 3, 4\}$, $A = \{1, 3\}$ at $A^c = \{2, 4\}$.

Symbolet $\setminus$ brukes for differansen mellom mengder. *Differansen* mellom A og B er mengden av alle elementer som er inneholdt i A men ikke i B. Formelt noteres dette som $A \setminus B = \{\omega \,|\, \omega \in A, \omega \notin B\}$. Det følger at $A \setminus B = A \cap B^c$ og at $A^c = \Omega \setminus A$. Et eksempel er gitt ved $\mathbb{R} \setminus \mathbb{N} = (-\infty, 0] \cup (\cup_{n=1}^{\infty}(n-1, n))$.

To mengder A og B sies å være *disjunkte* dersom de ikke har noen felles elementer, dvs. dersom $A \cap B = \emptyset$. F.eks. er mengdene gitt ved intervallene $(0, 1)$ og $(1, 2)$ disjunkte.

En union av to disjunkte mengder noteres som $A \uplus B$. Operasjonen $\uplus$ er kun definert for disjunkte mengder. Når vi skriver $A \uplus B$, så er det underforstått at mengdene A, B er disjunkte. Denne konvensjonen er sammenlignbar med at vi ofte skriver $f(x)$ for verdien til en funksjon f i et punkt x uten å eksplisitt skrive ned at vi antar at x er et element i



Figur A.1: Venn-diagram.

definisjonsmengden til f. Notasjonen $\uplus_i A_i$ benyttes tilsvarende for unionen av en vilkårlig familie disjunkte mengder. Dette siste betyr at $i \neq j \Rightarrow [A_i \cap A_j = \emptyset]$. Et eksempel er gitt ved

$$\mathbb{R} = \biguplus_{t \in \mathbb{R}} \{t\}$$

Relasjoner mellom mengder og operasjoner på mengder som gjelder generelt vil også måtte gjelde for delmengder av et rektangel i planet. Dette gir muligheten til illustrasjon av sammenhenger som ellers kan virke uoversiktlige i Venn-diagram. Noen eksempler er gitt i figur A.1.

Tillegg B

Statistikk på datamaskin

Mange av figurene i boken er laget ved *R kode* som finnes i dette tillegget. Venables and Ripley (2002) gir en mye mer fullstendig introduksjon til R kommandoer. Koden som er vedlagt kan for eksempel kjøres i *RStudio*, i *ESS* pakken i Emacs, eller ved å klippe inn koden i et vindu i en nettleser som har åpnet https://rdrr.io/snippets/ (testet i desember 2019). Noen av oppgavene i boken krever bruk av datamaskin, og eksemplene her kan være til hjelp.

B.1 Sannsynligheter fra fordelinger

```
# Sannsynlighet for X < nAge naar X er normalfordelt
nAge=25 # Alder
nMu=31.01 # Forventningsverdi
nSigma=5.05 # Standardavvik
nPT = pnorm(nAge, nMu, nSigma) # Kumulativ sannsynlighet
print(nPT)
```

B.2 Histogram og fordeling

```
# install.packages("MASS")
library(MASS)

aMorR = rep(c(17,18.5,22.5,27.5,32.5,37.5,42.5,47.5),
  c(56, 366, 5456, 17748, 20106, 9336, 1848, 154))

aFarR = rep(c(17,18.5,22.5,27.5,32.5,37.5,
  42.5,47.5,52.5,57.5,62.5),
  c(21, 96, 2536, 12189, 19025, 12701, 5163,1809, 565, 185, 82))

# Fordeling for mor # truehist gir ikke "add"?
hist(aMorR, breaks=c(14,18,20,25,30,35,40,45,50),
  xlim=c(10,70), ylim=c(0,0.09),
  col=rgb(1,0,0,0.5), xlab="Alder",
```

```
  ylab="Tetthet", main="Fordeling␣til␣alder" )

# Fordeling for far
hist(aFarR, breaks=c(14,18,20,25,30,35,40,45,50,55,60,65),
  xlim=c(10,70), col=rgb(0,0,1,0.5), add=T)

# Add legend
legend("topright", legend=c("Mor","Far"), col=c(rgb(1,0,0,0.5),
  rgb(0,0,1,0.5)), pt.cex=2, pch=15 )

# Plott av tetthet
xfit<-seq(10,70,length=100)
yfit<-dnorm(xfit,mean=mean(aMorR),sd=sd(aMorR))
lines(xfit, yfit, col=rgb(1,0,0,0.5), lwd=2)
yfit<-dnorm(xfit,mean=mean(aFarR),sd=sd(aFarR))
lines(xfit, yfit, col=rgb(0,0,1,0.5), lwd=2)
```

Tillegg C

Faginformasjon 1997

Tillegg D

Eksamen desember 1997

Tillegg E

Eksamen juni 1997

Bibliografi

Davis, D. J. (1952). An Analysis of Some Failure Data. *Journal of the American Statistical Association 47*(258), 113–150.

Fisher, R. A. (1973). *Statistical Methods and Scientific Inference*. Hafner press.

Høyland, A. (1988). *Sannsynlighetsregning Og Statistisk Metodelære (Bind 1-2)*. Tapir.

Kolmogorov, A. (1933). *Foundations of the Theory of Probability*. Chelsea (1956).

Larsen, R. J. and M. L. Marx (2012). *An Introduction to Mathematical Statistics and Its Applications* (5 edition ed.). Boston: Pearson.

Løvås, G. (2018). *Statistikk for universiteter og høgskoler*. Universitetsforlaget.

Proschan, F. (1963, August). Theoretical Explanation of Observed Decreasing Failure Rate. *Technometrics 5*(3), 375–383.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.

Sverdrup, E. (1964). *Lov Og Tilfeldighet (Bind 1-2)*. Universitetsforlaget.

Taraldsen, G. (2006). Instrument resolution and measurement accuracy. *Metrologia 43*, 539–544.

Taraldsen, G. (2011, May). Analysis of rounded exponential data. *Journal of Applied Statistics 38*(5), 977–986.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (4 ed.). Statistics and Computing. New York: Springer-Verlag.

von Bortkiewicz, L. (1898). *Das Gesetz der kleinen Zahlen*. Leipzig : B.G. Teubner.

Register

åpen boks, **22**

atom, **26**
avbildning, **172**

Betinget sannsynlighet, **34**
Binomialfordelingen, **47**
binomialkoeffisientene, 44

definisjonsmengden, **172**
Differansen, **173**
disjunkte, **173**
diskret fordeling, **26**
diskret sannsynlighetstetthet, **26**
diskret tetthet, **26**
diskret utfallsrom, **27**
domenet, **172**

Eksperiment, **19**
eksponensialfordelingen, **31**
empirisk middel, **6**
empirisk standardavvik, **6**
empirisk varians, **6**
empiriske fordelingen, **21**
empiriske tettheten, **27**
ESS, 175

fakultet, 44
familie av mengder, **15**
fordeling, **20**
forventningsverdi, **5**
funksjon, **172**

gjentas i det lange løp, **21**
graf, 5
Grafene er statistikker, 5

hendelse, **15**
hendelsesfamilie, **15**
Hyppigheten, **21**

indeksert familie, 173
inntruffet, **16**
inverse funksjonen, **172**

kakediagram, 5
kodomenet, **172**
Komplementet, **173**

Loven for de små tall, 8

modell, **21**

normalfordeling, **5**
normalfordelingen, **29**
normaliserte histogram, **4**

Observator, **50**

Parameter, **6**
parameter, 6
partisjon, **35**

R kode, 175
rekkevidden, **172**
relative hyppigheten, **5**, **21**
RStudio, 175

sanne fordelingen, **21**
Sannsynlighet, **19**
sannsynlighetsfordeling, **20**
sannsynlighetskvote, **v**
sannsynlighetsmål, **20**
Sannsynlighetsmodell, **21**
sannsynlighetsrom, **20**
Sannsynlighetstetthet, **25**

sannsynlighetstetthet, **26**
Simultanfordelingen, **54**
sjanseeksperiment, **21**
standardavvik, **5**
Statistikk, **3**
statistikk, 3, 5
statistikkfaget, **3**
statistisk eksperiment, **21**
statistisk lovmessig eksperiment, **21**
statistisk modell, 5
stokastiske variabler, 5
suffisient observator, **v**

tellbar, **173**
tellbar additivitet, **20**
tetthet, **26**
Tilfeldig variabel, **51**
tilfeldig vektor, **52**
transformasjon, **172**

Uavhengighet, **37**
uniforme diskrete fordelingen, **28**
uniforme fordelingen, **29**
utfall, **15**
Utfallsrom, **16**
utfallsrom for et eksperiment, **16**

verdimengde, **172**