

Ny og bedre versjon 2019

MAT110

Statistikk 1

Kompendium 2019, del 1

Per Kristian Rekdal
Bård-Inge Pettersen



Høgskolen i Molde
Vitenskapelig høyskole i logistikk

Innhold

I	Grunnleggende teori	11
1	Sannsynlighetsteori	13
1.1	Sannsynlighetsmodell	14
1.1.1	Motivasjon	14
1.1.2	Utfallsrommet	17
1.1.3	Mengdelære	19
1.1.4	Begivenheter - delmengder av utfallsrommet	20
1.1.5	Sannsynlighetsmodell	29
1.1.6	Fundamentale setninger i sannsynlighetsteori	30
1.1.7	Diskret sannsynlighetsmodell	45
1.1.8	Uniform sannsynlighetsmodell	50
1.1.9	Kombinatorikk	56
1.2	Betinget sannsynlighet	63
1.2.1	Multiplikasjonssetningen	70
1.2.2	Bayes lov	72
1.2.3	Sannsynlighetstrær	78
1.2.4	Oppsplitting av Ω	81
1.2.5	Uavhengighet	85
2	Stokastiske variabler, forventning og varians	95
2.1	Stokastiske variabler	96
2.1.1	Diskret VS kontinuerlig stokastiske variabler	97
2.2	Forventning og varians	103
2.2.1	Forventning	103
2.2.2	Varians	106
2.2.3	Kovarians	113
2.2.4	Noen regneregler	124
3	Diskrete stokastiske fordelinger	147
3.1	Den binomiske fordelingen	148
3.1.1	Forventningsverdi	164
3.1.2	Varians	166
3.2	Den hypergeometriske fordelingen	172
3.2.1	Forventning og varians	180
3.3	Sammenheng mellom $\text{Hyp}[N, M, n]$ og $\text{Bin}[n, p]$	184
3.3.1	Forventningsverdi	185
3.3.2	Varians	185

3.4	Poissonfordelingen	188
3.4.1	Forventning og varians	194
4	Kontinuerlige stokastiske fordelinger og CLT	198
4.1	Normalfordelingen (kontinuerlig)	198
4.1.1	Standardisering	202
4.1.2	Sammenhengen mellom $P(Z \leq z)$ og $G(z)$	205
4.1.3	Diskret vs kontinuerlig fordeling: en viktig forskjell	219
4.1.4	Standardavvik σ og %-vis areal	220
4.2	Oversikt: Bin, Hyp, Poi og N	233
4.3	Sentralgrensesetningen	236
4.4	Diskrete fordelinger $\rightarrow$ normalfordeling	253
4.5	Sammenheng: Bin, Hyp, Poi og N	263
4.6	Sum av uavhengige stokastiske variabler	265
II	Statistisk inferens	275
5	Beskrivende statistikk	277
5.1	Populasjon og utvalg	278
5.2	Statistiske mål (èn variabel)	283
5.2.1	Lokaliseringsmål	283
5.2.2	Spredningsmål	288
5.3	Statistiske mål (to variabler)	302
5.4	Grafisk fremstilling av data	315
6	Estimering og estimatorer	317
6.1	Estimering, estimator og estimand	318
6.1.1	Det finnes 2 typer estimering	319
6.1.2	Egenskaper ved en <u>gode</u> estimator	319
6.1.3	Utvalg vs populasjon	320
6.2	Punktestimering	321
6.2.1	Forventingsrette estimatorer	321
6.3	Intervallestimering	327
6.3.1	Standardisering	331
6.3.2	Sammenheng mellom $G(z_{\alpha/2})$ og α	333
6.4	student-t fordelingen	341
7	Hypotesetesting	349
7.1	Motivasjon	349
7.2	Generelt om hypotesetesting	349
7.3	P -verdi	349
8	Regresjonsanalyse	350
8.1	Introduksjon	351
8.2	Lineære sammenhenger	352
8.3	Teoretisk modell vs estimert modell	356
8.4	Residual og SSE	357

8.5	Minste kvadraters regresjonslinje	360
8.6	Forklaringsstyrke og SST	365
A	Mengdelære	374
B	Kombinatorikk	380
B.1	Koblinger	381
B.2	4 situasjoner (endelig populasjon)	385
B.3	Binomialkoeffisienten	393
B.4	Kombinatoriske sannsynligheter	394

Forord

Dette er del 1 (av 2) av kompendiet i emnet “*MAT110 Statistikk 1*” ved Høgskolen i Molde.

Forelesningene vil i all hovedsak følge dette kompendiet. Kompendiet er stort sett i samsvar med læreboken:

Statistikk for økonomifag, Jan Ubøe, Gyldendal akademisk, 5. utg., 2015

Selv om pensum defineres ut fra kompendiet, forelesningene og øvingene så anbefales det å ha læreboken i tillegg. Læreboken behandler flere detaljer enn kompendiet. Læreboken kan også være nyttig som oppslagsverk senere i studiet og i arbeidslivet etter endt studium. Eldre utgaver av læreboken kan benyttes. Men kun årets versjon av kompendiet kan benyttes da dette oppdateres fra år til år.

Studentene ved Høgskolen i Molde som skal ha dette kurset er blant annet studenter innenfor følgende studieretninger:

- økonomi
 - regnskap og revisjon
 - økonomi og administrasjon
- logistikk
 - logistikk og supply chain management
 - petroleumslogistikk

Både eksemplene i dette kompendiet og de tilhørende øvingsoppgavene preges av disse disiplinene. Det betyr at eksemplene i kompendiet og tematikken i øvingsoppgavene er i stor grad motivert ut fra disse studieretningene. Dette for å illustrere relevansen av statistikk i disse sammenhengene.

Det er også en formelsamling i dette emnet laget av foreleser. Formlene i denne formelsamlingen er stort sett de formlene som er markert med **rød skrift** og ramme rundt i dette kompendiet. Studentene oppfordres til å bruke formelsamlingen aktivt når øvingsoppgaver skal løses. Formelsamlingen sammen med godkjent kalkulator er tillatte hjelpemidler under eksamen.

Alle forelesningene blir tatt opp på **video**.

Alle disse videoene finnes på Høgskolen i Molde sin åpne kursportal:

www.himoldeX.no
--

Legg også merke til at videoer og kursmateriell fra tidligere år ligger helt åpent på himoldeX.no. Blant annet finnes komplette sett av forelesningsvideoene fra 2017 og 2018 finnes på denne åpne kursplattformen.

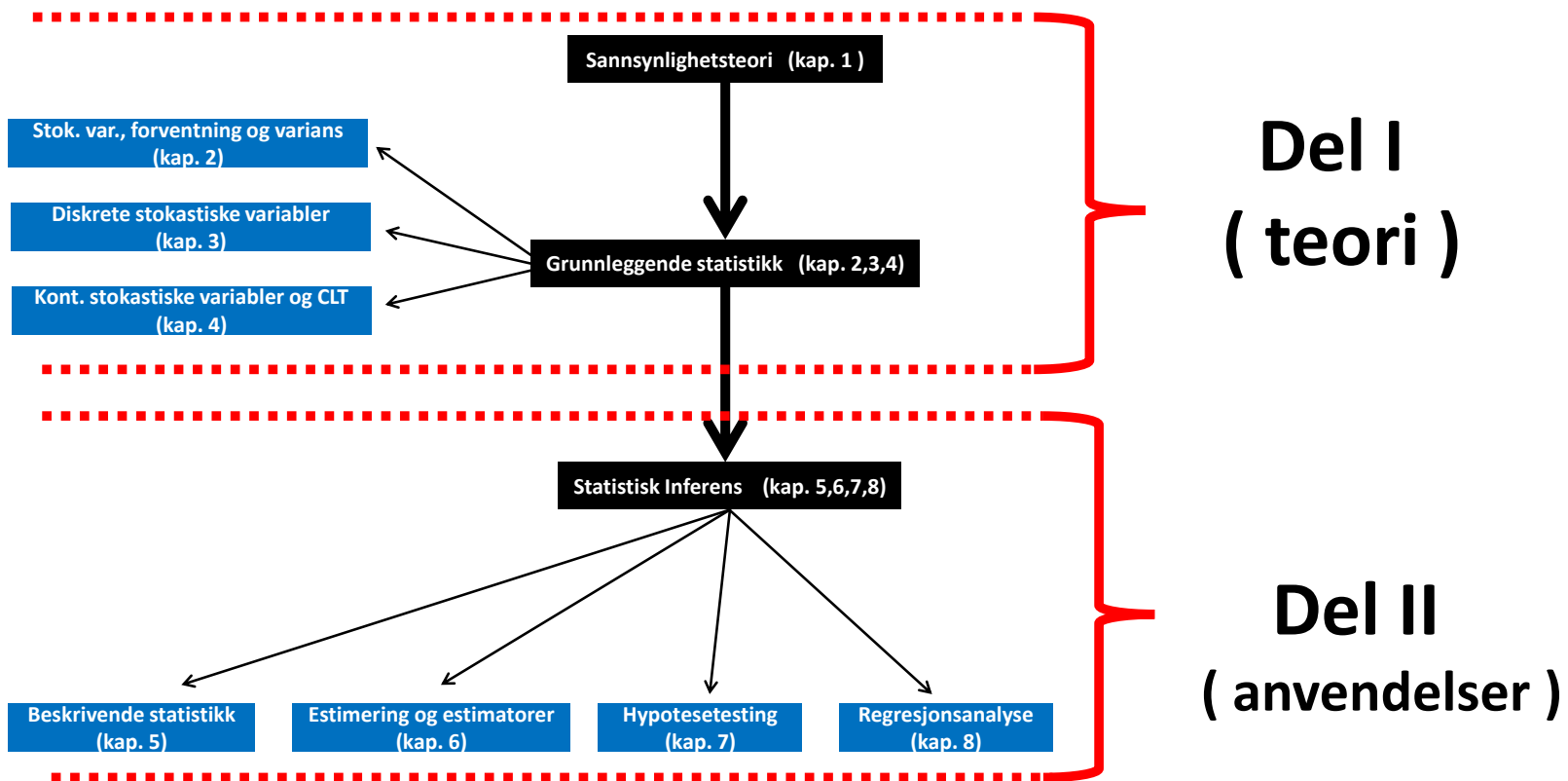
Alt er gratis!



Per Kristian Rekdal

Bård-Inge Pettersen

Copyright © Høgskolen i Molde, januar 2019.



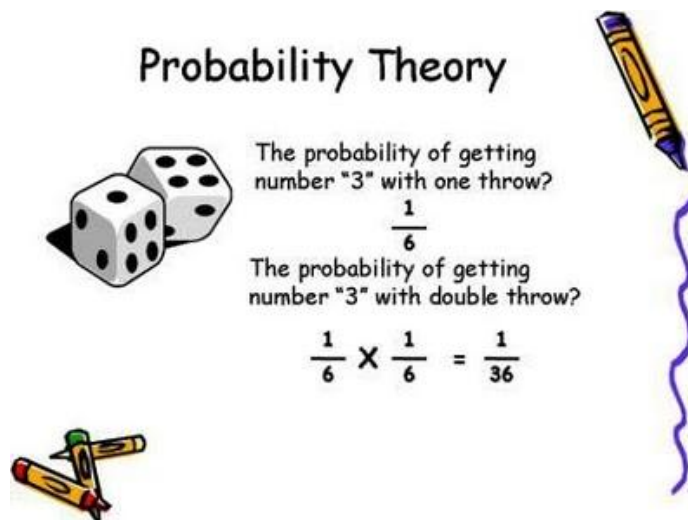
Figur 1: Oversikt over kompendiet.

Del I

Grunnleggende teori

Kapittel 1

Sannsynlighetsteori



Figur 1.1: Sannsynlighetsteori.

1.1 Sannsynlighetsmodell

1.1.1 Motivasjon

”Sannsynlighet” er et nyttig og ofte brukt begrep, men kan bli tolket på flere måter. Som et eksempel på dette les følgende dialog mellom vennen til en alvorlig syk pasient og en sykesøster, hvor pasienten har akkurat fått et nytt potensielt livreddende legemiddel: ¹

VENN:	Søster, hva er sannsynligheten for at legemiddelet fungerer?
SYKESØSTER:	Jeg håper det fungerer, vi får se i morgen.
VENN:	Ja, men hva er sannsynligheten for at middelet virker?
SYKESØSTER:	Hvert tilfelle er forskjellig, vi må bare vente.
VENN:	Men av 100 tilfeller, hvor mange hadde det fungert på?
SYKESØSTER:	Jeg sa det har forskjellig effekt!!! Jeg vet ikke! (irritert)
VENN:	Hva hvis du måtte tippe om det fungerer eller ikke? (gira).
SYKESØSTER:	Jeg ville tippe at det fungerer! (blidere)
VENN:	Så kan du gi meg en odds på 1:2 på at det virker eller ikke?
SYKESØSTER:	For en syk tanke!! Du kaster bort tiden min! (svært provosert)



Figur 1.2: Sykesøster.

¹En odds på 1:2 på at middelet virker eller ikke, betyr at for hver krone satset på at det ikke virker, får du 2 tilbake dersom dette viser seg å stemme.

I dialogen bruker vennen begrepet om ”sannsynlighet” for å diskutere en usikker situasjon. Vennen ser bare to utfall av legemiddelet:

$$u_1 : \text{ legemidlet fungerer} \quad (1.1)$$

$$u_0 : \text{ legemidlet fungerer ikke} \quad (1.2)$$

enten fungerer det eller ikke. Deretter assosierer han en sannsynlighet for hvert av utfallene:

$$p_1 = \text{ sannsynligheten for at middelet fungerer} \quad (1.3)$$

$$p_0 = \text{ sannsynligheten for at middelet ikke fungere} \quad (1.4)$$

Utfordringen i dialogen er at vennen ikke kjenner disse sannsynlighetene. Han spør derfor sykesøsteren om et

$$\text{estimat på } p_1 \quad (1.5)$$

dvs. et estimat for sannsynligheten for at legemiddelet fungerer. Vennen forsøker på forskjellige måter å fremskaffe et estimat fra sykesøsteren:

Estimat 1: (frekvenssannsynlighet)

Når vennen spør hvor mange av 100 tilfeller middelet ville ha virket på, henvender han seg til den historiske erfaringen sykesøsteren har vedrørende legemiddelet. Han håper hun har en formening om hvor frekvent legemiddelet fungerer gjennom gjentatt bruk på tidligere '.

Denne sannsynligheten kalles **frekvenssannsynligheten**.

Siden sykesøsteren ikke svarer, kan han anta at hun har liten erfaring med legemiddelet, f.eks. fordi det er et relativt nytt legemiddel eller at sykesøsteren ikke har stor erfaring med det.

Estimat 2: (subjektiv sannsynlighet)

Siden vennen ikke får et estimat via historiske data (frekvenssannsynligheten), appellerer han så til sykesøsterens **subjektive** formening om hva sannsynligheten er. Han spør om hun **tror** det fungerer eller ikke, som i prinsippet betyr om sannsynligheten er større enn eller mindre enn 50 %.

Siden sykesøsteren svarer ja på dette spørsmålet, kan vennen konkludere med at sannsynligheten for at middelet 50 % eller større.

Estimat 3: (risikovillighet)

Vennen vet nå at sannsynligheten *minst* er 50%, men han kunne tenke seg et mer nøyaktig estimat. Siste strategi er å få et estimat i forhold til hvor *risikovillig* sykesøsteren er ift. hvorvidt legemiddelet vil virke eller ikke. Han gir henne derfor en odds 1:2 på at det virker eller ikke.

Sykesøsteren *ikke* er villig til å satse på at legemiddelet virker til en odds på 1:2. En slik odds svarer til en sannsynlighet på $\frac{2}{3}$. Derfor kan vennen konkludere at sykesøsteren mener at sannsynligheten er *lavere* enn $\frac{2}{3}$ for at legemiddelet virker.

Konklusjon:

Sannsynligheten for at legemiddelet virker ligger mellom 0.5 og 0.667.

1.1.2 Utfallsrommet

I forrige avsnitt 1.1.1 ble det snakket løst om de to utfallene om hvorvidt legemiddelet fungerer eller ikke og deres assosierte sannsynligheter. I dette avsnittet defineres hva som menes med en *sannsynlighetsmodell*. En sannsynlighetsmodell har som hensikt å fange opp det *usikre* aspektet ved et gitt *eksperiment*.

Et *eksperiment* kan f.eks. være

1. å ta et legemiddel og observere utfallet
2. å slå en mynt og observere utfallet
3. å slå en mynt om igjen og om igjen inntil krone vises første gang og observere antall kast som måtte brukes
4. å kaste en terning og observere utfallet
5. å kaste 2 terninger 1 gang og observere utfallet
6. å spille en fotballkamp og observere sluttstillingen
7. å måle temperaturen ute en gitt dag

Alle utfallene ved disse eksperimentene har usikkerhet knyttet til seg.



Figur 1.3: Eksperiment.

Definisjon: (utfallsrom)

Resultatet av et gitt eksperiment E har et visst antall *mulige utfall*.
Denne mengden av mulige utfall kalles utfallsrommet til E :²

$$\Omega_E = \left\{ \text{mengden av } \underline{\text{alle}} \text{ mulige utfall ved eksperiment } E \right\} \quad (1.6)$$

Dersom det er underforstått hvilket eksperiment det er snakk om, brukes bare Ω .



²Den greske bokstaven Ω kalles “omega”.

1.1.3 Mengdelære

Siden en definerer de mulige utfallene av et eksperiment som en *mengde*, er det viktig med en liten introduksjon til mengdelære.

Se tillegg A.

1.1.4 Begivenheter - delmengder av utfallsrommet

Motiverende eksempel: (utfallsrom, sannsynlighet og begivenheter)

Et eksperiment som har kun to utfall, som f.eks. ”å slå en mynt”, er en svært enkel situasjon og gir derfor ikke det totale bildet av en sannsynlighetsmodell. Men eksperimentet ”å kaste to terninger” er et godt eksempel på hva en sannsynlighetsmodell bør innebære.



Figur 1.4: Blå og rød terning.

Utfallsrommet for to kast med en terning kan defineres som: ³

$$\Omega = \left\{ \begin{array}{l} (1,1) , (1,2) , (1,3) , (1,4) , (1,5) , (1,6) \\ (2,1) , (2,2) , (2,3) , (2,4) , (2,5) , (2,6) \\ (3,1) , (3,2) , (3,3) , (3,4) , (3,5) , (3,6) \\ (4,1) , (4,2) , (4,3) , (4,4) , (4,5) , (4,6) \\ (5,1) , (5,2) , (5,3) , (5,4) , (5,5) , (5,6) \\ (6,1) , (6,2) , (6,3) , (6,4) , (6,5) , (6,6) \end{array} \right\} \quad (1.7)$$

³Anta at terningene er merket slik at første terning er blå og andre terning er rød.

Dette eksperimentet har 36 utfall, ett for hver kombinasjon av terningene. For dette eksperimentet gir det mening å spørre om situasjoner som omhandler mer enn sannsynligheten for selve enkeltutfallene: ⁴

mer enn enkeltutfall

Spørsmål :

S_{Ω} = "hva er sannsynligheten for å få *ett* av utfallene i utfallsrommet?"

S_1 = "hva er sannsynligheten for å få *minst* en 1er?"

S_2 = "hva er sannsynligheten for å *ikke* få noen 1ere?"

S_3 = "hva er sannsynligheten for at *summen* av terningene er lik 7?"

S_4 = "hva er sannsynligheten for å få minst en 1er *og* at summen av terningene er lik 7?"

S_5 = "hva er sannsynligheten for å få minst en 1er *eller* at summen av terningene er lik 7?"

⁴Hvert utfall har sannsynligheten $\frac{1}{36}$, gitt at alle terningene er perfekt balanserte, dvs. ingen av terningene viser en tendens mot en av sidene.

Spørsmål S_Ω : (”hva er sannsynligheten for å få *ett* av utfallene i utfallsrommet?”)

Sannsynligheten for å få minst ett av utfallene betyr at ethvert kast må få en kombinasjon som finnes i utfallsrommet Ω . Hvis dette ikke er tilfelle, betyr det at utfallsrommet selv ikke er veldefinert. Utfallsrommet skal inneholde alle mulige utfall fra eksperimentet, dvs: ⁵

$$P(\Omega) = 1 \quad (1.8)$$

At sannsynligheten for å få et utfall fra utfallsrommet er lik 1, er en fundamental egenskap ved en sannsynlighetsmodell.

Spørsmål S_1 : (”hva er sannsynligheten for å få *minst* en 1er?”)

For å svare på spørsmål S_1 må en bestemme hvilke utfall inneholder minst en 1er:

$$S_1 = \left\{ \begin{array}{l} (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6) \\ (2, 1) \\ (3, 1) \\ (4, 1) \\ (5, 1) \\ (6, 1) \end{array} \right\} \quad (1.9)$$

Siden S_1 er en *delmengde* av Ω bestående av **11 mulige utfall** hvorav hvert utfall har sannsynlighet $\frac{1}{36}$, sier *intusjonen* at sannsynligheten for å minst få en 1er er gitt ved: ⁶

$$P(S_1) = \frac{11}{36} \quad (1.10)$$

⁵”*P*” er en forkortelse for det engelske ordet *probability* som betyr *sannsynlighet* på norsk.
 $P(\Omega)$ kan leses som: ”sannsynligheten for at et av utfallene fra eksperimentet er med i mengden Ω , er lik 1”.

⁶Denne intusjonen er korrekt.

Spørsmål S_2 : (”hva er sannsynligheten for å *ikke* få noen 1ere?”)

For å svare på spørsmål S_2 må en bestemme hvilke utfall *ikke* inneholder en 1er:

$$S_2 = \left\{ \begin{array}{l} (2, 2) , (2, 3) , (2, 4) , (2, 5) , (2, 6) \\ (3, 2) , (3, 3) , (3, 4) , (3, 5) , (3, 6) \\ (4, 2) , (4, 3) , (4, 4) , (4, 5) , (4, 6) \\ (5, 2) , (5, 3) , (5, 4) , (5, 5) , (5, 6) \\ (6, 2) , (6, 3) , (6, 4) , (6, 5) , (6, 6) \end{array} \right\} \quad (1.11)$$

Siden S_2 er en *delmengde* av Ω bestående av **25 mulige utfall** hvorav hvert utfall har sannsynlighet $\frac{1}{36}$, sier *intusjonen* at sannsynligheten for å ikke få noen 1ere er gitt ved: ⁷

$$P(S_2) = \frac{25}{36} \quad (1.14)$$

⁷Legg merke til at:

S_2 er den *komplementære* mengden til S_1 , dvs S_2 inneholder alle utfallene som *ikke* er med i S_1 :

$$S_2 = \overline{S_1} = \Omega - S_1 \quad (1.12)$$

Sannsynligheten tilhørende lign.(1.12) er:

$$\underline{\underline{P(S_2)}} = P(\overline{S_1}) = P(\Omega - S_1) = P(\Omega) - P(S_1) = \overbrace{1 - \frac{11}{36}}^{\text{komplement}} = \underline{\underline{\frac{25}{36}}} \quad (1.13)$$

At sannsynligheten for at et utfall er i den komplementære mengden er lik 1 minus sannsynligheten for at utfallet er i mengden, er en fundamental egenskap ved en sannsynlighetsmodell og kalles *komplementsetningen*. Vi kommer tilbake til denne setningen senere i dette kapitlet.

Spørsmål S_3 : (”hva er sannsynligheten for at *summen* av terningene er lik 7?”)

For å svare på spørsmål S_3 må en bestemme hvilke utfall hvor summen av terningene er lik 7:

$$S_3 = \left\{ \begin{array}{l} (1, 6) , \\ (2, 5) , \\ (3, 4) , \\ (4, 3) , \\ (5, 2) , \\ (6, 1) \end{array} \right\} \quad (1.15)$$

Siden S_3 er en *delmengde* av Ω bestående av 6 mulige utfall hvorav hvert utfall har sannsynlighet $\frac{1}{36}$, sier *intusjonen* at sannsynligheten for å få et kast hvor summen av terningene er lik 7 er gitt ved: ⁸

$P(S_3) = \frac{6}{36} = \frac{1}{6} \quad (1.16)$
--

⁸Denne intuisjonen er korrekt.

Spørsmål S_4 : (”hva er sanns. for å få minst en 1er *og* at summen av terningene er lik 7?”)

For å svare på spørsmål S_4 må en bestemme hvilke utfall hvor summen av terningene er lik 7 *og* hvor en av terningene har en 1er:

$$S_4 = \left\{ \begin{array}{l} (1, 6) , \\ (6, 1) \end{array} \right\} \quad (1.17)$$

Siden S_4 er en *delmengde* av Ω bestående av 2 mulige utfall hvorav hvert utfall har sannsynlighet $\frac{1}{36}$, sier *intusjonen* at sannsynligheten for å få et kast hvor summen av terningene er lik 7 *og* hvor en av terningene har en 1er er gitt ved: ⁹

$$P(S_4) = \frac{2}{36} = \frac{1}{18} \quad (1.19)$$

⁹Legg merke til at:

$$S_4 = S_1 \cap S_3 \quad (1.18)$$

Når en slår sammen to situasjoner til en ved hjelp av *og*, betyr det med andre ord at en tar *snittet* av enkeltsituasjonene.

Spørsmål S_5 : (”hva er sanns. for å få minst en 1er *eller* at summen av terningene er lik 7?”)

For å svare på spørsmål S_5 må en bestemme hvilke utfall hvor summen av terningene er lik 7 *eller* at en av terningene viser en 1er:

$$S_5 = \left\{ \begin{array}{l} (1, 1) , (1, 2) , (1, 3) , (1, 4) , (1, 5) , (1, 6) , \\ (2, 1) , (2, 5) , \\ (3, 1) , (3, 4) , \\ (4, 1) , (4, 3) , \\ (5, 1) , (5, 2) , \\ (6, 1) \end{array} \right\} \quad (1.20)$$

Siden S_3 er en *delmengde* av Ω bestående av 15 mulige utfall hvorav hvert utfall har sannsynlighet $\frac{1}{36}$, sier *intusjonen* at sannsynligheten for å få et kast hvor summen av terningene er lik 7 *eller* at en av terningene viser en 1er er gitt ved: ¹⁰

$$P(S_5) = \frac{15}{36} = \frac{5}{12} \quad (1.23)$$

¹⁰Legg merke til at:

$$S_5 = S_1 \cup S_3 \quad (1.21)$$

Sannsynligheten tilhørende lign.(1.21) er:

$$\underline{\underline{P(S_5)}} = P(S_1 \cup S_3) = P(S_1) + P(S_3) - P(S_1 \cap S_3) = \frac{11}{36} + \frac{6}{36} - \frac{2}{36} = \frac{15}{36} = \underline{\underline{\frac{5}{12}}} \quad (1.22)$$

Denne egenskapen er en fundamental egenskap ved sannsynlighetsmodeller og kalles *den generelle addisjonssetningen*. En må trekke ifra snittet, for hvis ikke, teller en med snittet *to* ganger. (I dette tilfellet ville en ha telt med kastene (1,6) og (6,1) dobbelt opp). Vi kommer tilbake til denne setningen senere i dette kapitlet.

Konklusjon: (begivenheter)

Ved et gitt eksperiment E med tilhørende utfallsrom Ω , er det *ikke* nok å kjenne til sannsynligheten for hvert utfall. Man er også interessert i sannsynligheten for at utfallet av eksperimentet er med i en

$\underbrace{\text{delmengde } S \text{ av utfallsrommet } \Omega}_{\text{flere enkeltutfall}}$

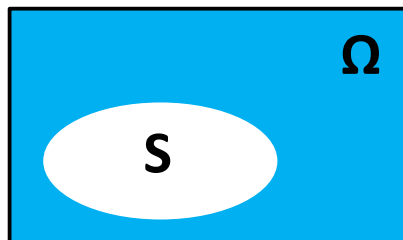
I tillegg må sannsynlighetsmodellen oppfylle et sett med fundamentale egenskaper ved beregning av sannsynligheter av delmengder.

■

Definisjon: ([begivenhet](#)) ¹¹ ¹²

$$\text{begivenhet } S = \text{en mulig delmengde } S \text{ av utfallsrommet } \Omega, \text{ dvs. } S \subseteq \Omega \quad (1.24)$$

■



Figur 1.5: Begivenhet S .

¹¹Noen bruker også “hendelse” istedet for “begivenhet”. I dette kurset brukes samme konvensjon som i læreboken, dvs. “begivenhet”.

¹²Definisjonen av en begivenhet har ordet ”mulig” i seg fordi det viser seg at ved når utfallsrommet f.eks. er alle tall større enn 0 (f.eks. temperatur), finnes det delmengder hvor sannsynligheten ikke er vel-definert. Begivenheter antas derfor alltid å ha en veldefinert sannsynlighet. I det diskrete tilfellet kan sannsynligheten til enhver delmengde evalueres.

1.1.5 Sannsynlighetsmodell

Definisjon: $\overbrace{(\text{sannsynlighetsmodell} - 3 \text{ aksiomer})}^{\text{fundamentale egenskaper}}$

En *sannsynlighetsmodell* for et eksperiment E består av:

$$1) \text{ et utfallsrom } \Omega_E \quad (1.25)$$

$$2) \text{ en sannsynlighetslov } P(-) \quad (1.26)$$

som for hver begivenhet $S \subseteq \Omega$ tilordner en sannsynlighet $P(S) \in [0, 1]$

hvor sannsynlighetsloven P har følgende 3 fundamentale egenskaper: (3 aksiomer)

$$P(S) \geq 0 \quad (\text{ikke-negativitet}) \quad (1.27)$$

$$P(S_1 \cup S_2 \cup \dots) = P(S_1) + P(S_2) + \dots \quad (\text{addisjon}) \quad (1.28)$$

$$P(\Omega) = 1 \quad (\text{normalisering}) \quad (1.29)$$

hvor, lign.(1.28), gjelder for alle parvis disjunkte begivenheter $S_1, S_2, S_3, \dots$



Kommentarer:

- Legg merke til at lign. (1.28) omhandler en *uendelig* følge av begivenheter $S_1, S_2, S_3, \dots$. Dette er viktig når utfallsrommet har uendelig mange utfall, f.eks. alle mulige temperaturer.
- En følge av *parvis disjunkte* begivenheter $S_1, S_2, S_3, \dots$, betyr at *ingen* av begivenhetene overlapper, dvs. $S_i \cap S_j = \emptyset$ for alle $i \neq j$.¹³
- Fra lign. (1.27), (1.28) og (1.29) kan man utlede alle de fundamentale egenskapene som en sannsynlighetsmodell bør ha. Neste avsnitt ramser opp de viktigste.

¹³Brikkene i et puslespill er *parvis disjunkte*, dvs. de overlapper akkurat ikke.

1.1.6 Fundamentale setninger i sannsynlighetsteori

Anta at E er et eksperiment som har en sannsynlighetsmodell med utfallsrom Ω og sannsynlighetslov $P(-)$. Da gjelder følgende setninger:

Setning 1: (sannsynligheten av den tomme begivenheten er null)

Sannsynligheten for den tomme begivenheten $\emptyset$ er lik null, dvs.:

$$P(\emptyset) = 0 \quad (1.30)$$

■

Bevis:

Lag følgen av begivenhetene $S_1 = \Omega$, $S_2 = \emptyset$, $S_3 = \emptyset$, $S_4 = \emptyset$, $\dots$

Begivenhetene i følgen er da parvis disjunkte og siden $S_1 + S_2 + S_3 + \dots = \Omega$, gir lign. (1.28):

$$P(\Omega) = P(S_1) + P(S_2) + P(S_3) + \dots = P(\Omega) + P(\emptyset) + P(\emptyset) + \dots \quad (1.31)$$

$\Updownarrow$

$$0 = P(\emptyset) + P(\emptyset) + \dots \quad (1.32)$$

Siden $P(-) \geq 0$, så er:

$$\underline{\underline{P(\emptyset) = 0}} \quad (1.33)$$

■

Setning 2: (den spesielle addisjonssetningen) ¹⁴

Dersom begivenhetene A og B er disjunkte, dvs. $A \overset{\text{og}}{\cap} B = \emptyset$, så gjelder:

$$P(\overset{\text{eller}}{A \cup B}) = P(A) + P(B) \quad (1.34)$$

■

Bevis:

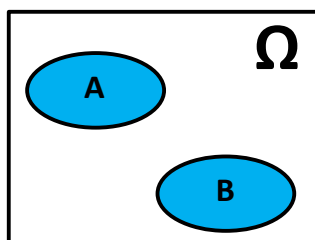
Lag følgen av begivenhetene $S_1 = A$, $S_2 = B$, $S_3 = \emptyset$, $S_4 = \emptyset$, $\dots$

Begivenhetene i følgen er da parvis disjunkte og siden $S_1 \cup S_2 \cup S_3 + \dots = A \cup B$, gir lign. (1.28):

$$\underline{\underline{P(A \cup B)}} = P(S_1) + P(S_2) + P(S_3) + \dots = P(A) + P(B) + \overset{=0}{\overbrace{P(\emptyset)} + \dots} \quad (1.35)$$

$$= \underline{\underline{P(A) + P(B)}} \quad (1.36)$$

■



Figur 1.6: A og B er **disjunkt**.

¹⁴Den spesielle addisjonssetningen kan åpenbart utvides til et endelig antall disjunkte begivenheter.

Setning 3: (differansen mellom to begivenheter)

Anta en har to begivenheter A og B slik at $B \subset A$, da gjelder

$$P(A - B) = P(A) - P(B) \quad (1.37)$$

■

Bevis:

Siden $B \subset A$, er $A - B$ disjunkt med B og $(A - B) \cup B = A$.

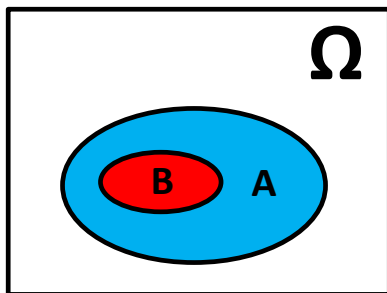
Lign. (1.34) gir dermed:

$$P(A) = P\left((A - B) \cup B\right) \stackrel{\text{disjunkt}}{=} P(A - B) + P(B) \quad (1.38)$$

$\Updownarrow$

$$\underline{\underline{P(A - B) = P(A) - P(B)}} \quad (1.39)$$

■



Figur 1.7: Differansen mellom to begivenheter.

Setning 4: (den [genrelle](#) addisjonssetningen)

For begivenhetene A og B gjelder:

$$P(\overbrace{A \cup B}^{\text{eller}}) = P(A) + P(B) - \underbrace{P(A \cap B)}_{\text{ekstra ledd}} \quad (1.40)$$

Bevis:

For begivenhetene A og B gjelder alltid formelen:

$$A \cup B = A \cup (B - A \cap B) \quad (1.41)$$

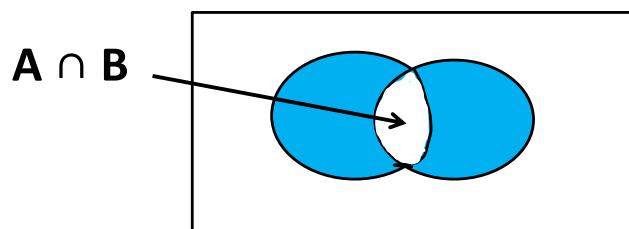
Siden A og $B - A \cap B$ er disjunkte gir den spesielle addisjonssetningen at: ¹⁵

$$P(A \cup B) = P(A \cup (B - A \cap B)) = P(A) + P(B - A \cap B) \quad (1.42)$$

Siden $A \cap B \subset B$, gir setningen om differansen mellom to begivenheter videre at: ¹⁶

$$\begin{aligned} \underline{\underline{P(A \cup B)}} &= P(A) + P(B - A \cap B) = P(A) + (P(B) - P(A \cap B)) \\ &= \underline{\underline{P(A) + P(B) - P(A \cap B)}} \end{aligned} \quad (1.43)$$

■



Figur 1.8: **Overlappen** mellom A og B er $A \cap B$.

¹⁵ Lign. (1.34) i setning 2.

¹⁶ Lign. (1.37) i setning 3.

Setning 5: ([komplementærsetningen](#))

La A være en vilkårlig begivenhet, da gjelder alltid:

$$P(\overline{A}) = 1 - P(A) \quad (1.44)$$

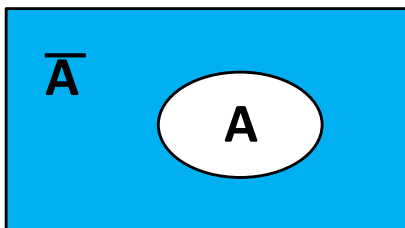
■

Bevis:

Siden en alltid har at $A \subset \Omega$ og $\overline{A} = \Omega - A$, gir setning 3 at: ¹⁷

$$\underline{\underline{P(\overline{A})}} = P(\Omega - A) = P(\Omega) - P(A) = \underline{\underline{1 - P(A)}} \quad (1.45)$$

■



Figur 1.9: Komplementet til A er $\overline{A}$.

¹⁷Ved den siste likheten i lign. (1.45), brukes setning 1 om at $P(\Omega) = 1$.

Setning 6: (total sannsynlighet)

For begivenhetene A og B gjelder:

$$P(A) = P(A \cap \overline{B}) + P(A \cap B) \quad (1.46)$$

■

Bevis:

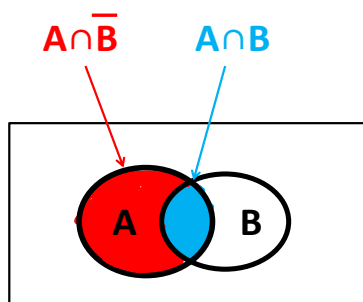
Siden $\Omega = \overline{B} \cup B$ og $A \subset \Omega$ har en alltid at (se figur 1.10): ¹⁸

$$A = A \cap \Omega = A \cap (\overline{B} \cup B) = (A \cap \overline{B}) \cup (A \cap B) \quad (1.48)$$

Siden $A \cap \overline{B}$ er disjunkt med $A \cap B$ gir den spesielle addisjonssetningen ved lign. (1.34):

$$\underline{\underline{P(A)}} = P((A \cap \overline{B}) \cup (A \cap B)) = \underline{\underline{P(A \cap \overline{B}) + P(A \cap B)}} \quad (1.49)$$

■



Figur 1.10: **Oppsplitting** av begivenhenten $A = (A \cap \overline{B}) \cup (A \cap B)$.

¹⁸Her benyttes DeMorgans lov i mengdelære som sier at:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (1.47)$$

for *alle* mengder A , B og C .

Setning 7: ([tvillingsetningen](#)) ¹⁹

For begivenhetene A og B gjelder:

$$P(\overbrace{\bar{A} \cap \bar{B}}^{\text{og}}) = 1 - P(\overbrace{A \cup B}^{\text{eller}}) \quad (1.51)$$

■

Bevis:

For to mengder A og B gjelder alltid:

$$\bar{A} \cap \bar{B} = \overline{A \cup B} \quad (1.52)$$

Dermed fås:

$$\underline{\underline{P(\bar{A} \cap \bar{B})}} = \underline{\underline{P(\overline{A \cup B})}} = \underline{\underline{1 - P(A \cup B)}} \quad (1.53)$$

hvor komplementærsetningen, dvs. lign. (1.44), ble brukt ved siste likhet.

■

¹⁹Hvorfor er “tvillingsetningen” et hensiktsmessig navn for lign.(1.51)? Dersom en bytter A med $\bar{A}$ og B med $\bar{B}$ i lign. (1.51), fås ligningen:

$$P(\overbrace{A \cap B}^{\text{og}}) = 1 - P(\overbrace{\bar{A} \cup \bar{B}}^{\text{eller}}) \quad (1.50)$$

som kan anses som “tvillingsetningen” til lign. (1.51).

Eksempel: (gravid - addisjonssetningen)

Anta at vi har:

- M = mann
- K = kvinne
- G = gravid

Vi gjør følgende eksperiment: trekker en tilfeldig person.

Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ (M, G), (M, \overline{G}), (K, G), (K, \overline{G}) \right\} \quad (1.54)$$

Definer følgende begivenheter:

$$M = \left\{ (M, G), (M, \overline{G}) \right\} \quad (\text{en } \underline{\text{mann}} \text{ trekkes}) \quad (1.55)$$

$$K = \left\{ (K, G), (K, \overline{G}) \right\} \quad (\text{en } \underline{\text{kvinne}} \text{ trekkes}) \quad (1.56)$$

$$G = \left\{ (M, G), (K, G) \right\} \quad (\text{en } \underline{\text{gravid person}} \text{ trekkes}) \quad (1.57)$$



Figur 1.11: Gravid.

Anta at:

$$P(M) = P(K) = 0.5 \quad (1.58)$$

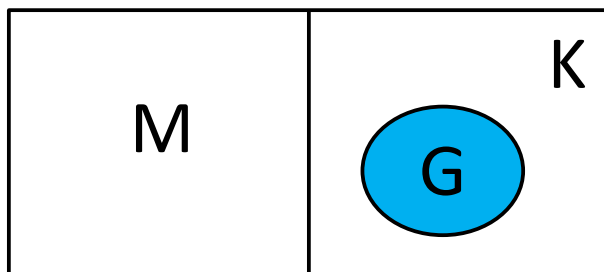
og

$$P(G) = 0.03 \quad (1.59)$$

Hva er “eller”-sannsynligheten $P(M \underbrace{\cup}_{\text{eller}} G)$, dvs. mann eller gravid?

Løsning:

Venn-diagram av situasjonen:



$M \cap G = \emptyset$, dvs. **disjunkte**.
Ingen overlapp mellom M og G.
Bare kvinner kan være gravide.

Figur 1.12: Begivenhetene M , K og G .

Siden “og”-sannsynligheten mann og gravid er 0,

$$M \cap G = \emptyset \quad (1.60)$$

dvs. disjunkte²⁰ så gjelder den *spesielle* addisjonssetningen. “Eller”-sannsynligheten finnes da ved å SUMMERE de individuelle sannsynlighetene slik som lign.(1.34) sier:

$$\underline{\underline{P(M \overset{\text{eller}}{\cup} G)}} = \underbrace{P(M)}_{= 0.5} + \underbrace{P(G)}_{= 0.03} = 0.5 + 0.03 = \underline{\underline{0.53}} \quad (1.61)$$

■

²⁰Mann og gravid har ingen overlapp, se figur (1.12), dvs. de er **disjunkte**).

Eksempel: (logistikk - total sannsynlighet, tvillingsetningene)

En del ansatte ved Vestbase i Kristiansund tar bussen til jobb. En av disse, Ole Hansen, tar bussrute 1, men må skifte til rute 2 underveis. Sannsynligheten for at det er ingen ledige sitteplasser på rute 1 er 0.6. Den tilsvarende sannsynligheten for rute 2 er 0.7. Sannsynligheten for at det er ledige sitteplasser på begge bussene er 0.2.

- a) Hva er sannsynligheten for at Ole Hansen finner ledige sitteplasser på **minst én** av bussene?
- b) Hva er sannsynligheten for at Ole Hansen finner ledige sitteplasser på rute **1**, men **ikke** rute **2**?
- c) Hva er sannsynligheten for at Ole Hansen **verken** finner ledige sitteplasser på rute **1** **eller** rute **2**?



Figur 1.13: Buss i Kristiansund.

Løsning:

Vi må starte med å skaffe oss oversikt over oppgaven/situasjonen. I den sammenheng kan det være lurt å **definere** noen **begivenheter**. I vårt tilfelle kan det være hensiktsmessig å definere følgende begivenheter:

$$L_1 \stackrel{\text{def.}}{=} \text{ ledige plasser i bussen på rute 1 } \quad (1.62)$$

$$L_2 \stackrel{\text{def.}}{=} \text{ ledige plasser i bussen på rute 2 } \quad (1.63)$$

Vi gjør følgende eksperiment: trekker **to** busser ²¹

Utfallsrommet består av alle dubletter: ²²

$$\Omega = \left\{ (L_1, L_2), (L, \overline{L_2}), (\overline{L_1}, L_2), (\overline{L_1}, \overline{L_2}) \right\} \quad (1.64)$$

La oss deretter konvertere opplysningene i oppgaven til et **matematisk språk**.

Det kan man gjøre på følgende måte:

- 1) At sannsynligheten for at det er “*ingen ledige sitteplasser på rute 1 er 0.6*” betyr:

$$P(\overline{L_1}) = 0.6 \quad (1.65)$$

- 2) At sannsynligheten for at der er “*ingen ledige sitteplasser på rute 2 er 0.7*” betyr:

$$P(\overline{L_2}) = 0.7 \quad (1.66)$$

- 3) At sannsynligheten for at det er “*ledige sitteplasser på begge bussene er 0.2*” betyr:

$$P(L_1 \underbrace{\cap}_{\text{og}} L_2) = 0.2 \quad (1.67)$$

²¹En buss rute 1, og en buss rute 2.

²²I prinsipp er dette helt likt med forrige eksempel om graviditet, se lign.(1.54)

- a) Å finne sannsynligheten for at “*Ole Hansen finner ledige sitteplasser på minst én av bussene*” betyr at vi skal finne $P(\underbrace{L_1 \cup L_2}_{\text{eller}})$.

Vi kjenner “og” sannsynligheten $P(\underbrace{L_1 \cap L_2}_{\text{og}})$, og skal finne “eller” sannsynligheten. Setningen som forbinder disse sannsynlighetene er den **generelle addisjonssetningen**, se lign.(1.40):

$$\overbrace{P(L_1 \cup L_2)}^{\text{skal finne}} \stackrel{\text{lign.(1.40)}}{=} P(L_1) + P(L_2) - \overbrace{P(L_1 \cap L_2)}^{= 0.2} \quad (1.68)$$

eller og

hvor sannsynligheten $P(L_1 \cap L_2) = 0.2$ var oppgitt i oppgaven, og hvor sannsynlighetene $P(L_1)$ og $P(L_2)$ finnes via komplementsetningen:

$$\underline{P(L_1)} = 1 - P(\bar{L}_1) = 1 - 0.6 = \underline{0.4} \quad (1.69)$$

$$\underline{P(L_2)} = 1 - P(\bar{L}_2) = 1 - 0.7 = \underline{0.3} \quad (1.70)$$

Dermed har vi alle størrelsene vi trenger i lign.(1.68):

$$\underline{P(L_1 \cup L_2)} = \overbrace{P(L_1)}^{= 0.4} + \overbrace{P(L_2)}^{= 0.3} - \overbrace{P(L_1 \cap L_2)}^{= 0.2} \quad (1.71)$$

$$= 0.4 + 0.3 - 0.2 = \underline{0.5} \quad (1.72)$$

Altså, det er 50 % sannsynlighet for at det er ledige plasser enten på rute 1 eller rute 2 (eller begge).

- b) Å finne sannsynligheten for at “*Ole Hansen finner ledige sitteplasser på rute 1, men ikke rute 2*” betyr at vi skal finne $P(L_1 \underbrace{\cap}_{\text{og}} \bar{L}_2)$.

Sannsynligheten $P(L_1 \cap L_2) = 0.2$ var oppgitt i oppgaven, og vi skal finne $P(L_1 \cap \bar{L}_2)$. Setningen som forbinder disse sannsynlighetene er den **total sannsynlighet**, se lign.(1.46):

$$\underbrace{P(L_1)}_{= 0.4} = \underbrace{P(L_1 \cap L_2)}_{= 0.2} + \overbrace{P(L_1 \underbrace{\cap}_{\text{og}} \bar{L}_2)}^{\text{skal finne}} \quad (1.73)$$

som gir

$$\underline{P(L_1 \cap \bar{L}_2)} = P(L_1) - P(L_1 \cap L_2) \quad (1.74)$$

$$= 0.4 - 0.2 = \underline{0.2} \quad (1.75)$$

Altså, det er 20 % sannsynlighet for at det er ledige plasser på bussen på rute 1, men ikke rute 2.

- c) Å finne sannsynligheten for at “Ole Hansen *verken* finner ledige sitteplasser på rute 1 *eller* rute 2” betyr at vi skal finne $P(\overline{L}_1 \overset{\text{og}}{\cap} \overline{L}_2)$.

Sannsynligheten $P(L_1 \cup L_2) = 0.5$ fant vi i oppgave a, og vi skal finne $P(\overline{L}_1 \cap \overline{L}_2)$. Setningen som forbinder disse sannsynlighetene er en av *tvillingsetningene*, se lign.(1.50):

$$P(\overline{L}_1 \overset{\text{og}}{\cap} \overline{L}_2) = 1 - P(L_1 \overset{\text{eller}}{\cup} L_2) \quad (1.76)$$

som gir

$$\underline{P(\overline{L}_1 \overset{\text{og}}{\cap} \overline{L}_2)} = 1 - \underbrace{P(L_1 \overset{\text{eller}}{\cup} L_2)}_{= 0.5} \quad (1.77)$$

$$= 1 - 0.5 = \underline{0.5} \quad (1.78)$$

Altså, det er 50 % sannsynlighet for at det verken er ledige plasser på bussen på rute 1 eller rute 2.

■

1.1.7 Diskret sannsynlighetsmodell

Motiverende eksempel: (myntkast)

Eksperiment 1: (myntkast - 1 kast)

La oss igjen se på forsøket med myntkast. Utfallsrommet er:

$$\Omega = \{ K, M \} \quad (1.79)$$

hvor K=krone og M=mynt. Det antas at mynten er perfekt balansert, dvs. sannsynligheten for å få mynt/krone er 50%:

$$P(K) = P(M) = \frac{1}{2} \quad (1.80)$$

dvs. like sannsynligheter. Utfallsrommet Ω har totalt 4 mulige begivenheter:

$$S_1 = \emptyset, \quad S_2 = \{ M \}, \quad S_3 = \{ K \}, \quad S_4 = \Omega \quad (1.81)$$

Sannsynlighetsloven $P(-)$ er fullstendig spesifisert ved:

$$\underline{\underline{P(S_1)}} = P(\emptyset) = \underline{\underline{0}} \quad (1.82)$$

$$\underline{\underline{P(S_2)}} = P(\{M\}) = \underline{\underline{\frac{1}{2}}} \quad (1.83)$$

$$\underline{\underline{P(S_3)}} = P(\{K\}) = \underline{\underline{\frac{1}{2}}} \quad (1.84)$$

$$\underline{\underline{P(S_4)}} = P(\{M, K\}) = P(\{M\}) + P(\{K\}) = \frac{1}{2} + \frac{1}{2} = \underline{\underline{1}} \quad (1.85)$$

Eksperiment 2: (myntkast - 3 kast)

Anta mynten kastes 3 ganger hvor resultatet registreres for hvert kast.
Utfallsrommet er tripletter:

$$\Omega = \left\{ \overbrace{\text{KKK} , \text{KKM} , \text{KMK} , \text{KMM} , \text{MKK} , \text{MKM} , \text{MMK} , \text{MMM}}^{8 \text{ mulige utfall}} \right\} \quad (1.86)$$

hvor K=krone og M=mynt. Det antas at mynten er perfekt balansert, dvs. sannsynligheten for å få et av utfallene er $\frac{1}{8}$, ta for eksempel KKM :

$$P(KKM) = \frac{1}{8}, \quad (1.87)$$

De andre utfallene har samme sannsynlighet. Utfallsrommet Ω har totalt 2^8 mulige begivenheter:

$$S_i \subset \Omega, \quad i = 1, 2, 3, \dots, 2^8 \quad (1.88)$$

Sannsynlighetsloven $P(-)$ er fullstendig spesifisert dersom en kan beregne $P(S_i)$ for alle $i = 1, 2, 3, \dots, 2^8$.



Figur 1.14: Myntkast.

Ta for eksempel begivenheten hvor nøyaktig to mynt forekommer:

$$S_{MM} = \left\{ \overbrace{MMK, MKM, KMM}^{3 \text{ mulige utfall}} \right\} = \{ MMK \} \cup \{ MKM \} \cup \{ KMM \} \quad (1.89)$$

Begivenheten S har totalt 3 elementer, og ved den spesielle addisjonssetningen lign. (1.34), fås:

$$\underline{\underline{P(S_{MM})}} = P(\{ MMK \} \cup \{ MKM \} \cup \{ KMM \}) \quad (1.90)$$

$$= P(MMK) + P(MKM) + P(KMM) \quad (1.91)$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \underline{\underline{\frac{3}{8}}} \quad (1.92)$$

Alle de andre begivenhetene kan beregnes på tilsvarende måte ved å [splitte opp begivenheten](#) i de enkelte utfallene og deretter summere sannsynligheten for utfallene (som i dette tilfellet var $\frac{1}{8}$).

Denne fremgangsmåten er felles for *alle* sannsynlighetsmodeller hvor utfallsrommet er *endelig*. Sannsynlighetsmodeller med endelige utfallsrom kalles [diskrete sannsynlighetsmodeller](#).

■

Definisjon: (diskret sannsynlighetsmodell)

En sannsynlighetsmodell $(\Omega, P(-))$ for et eksperiment hvor utfallsrommet er *endelig*, dvs.:

$$\Omega = \{ u_1, u_2, \dots, u_n \} \quad (1.93)$$

kalles en diskret sannsynlighetsmodell.

■

Anta sannsynlighetene til de n utfallene er gitt ved:

$$P(u_i) = p_i, \quad i = 1, 2, \dots, n \quad (1.94)$$

Ved aksiomene om ikke-negativitet og normalisering, dvs. lign. (1.27) og (1.29), følger umiddelbart at $p_i \geq 0$ for alle $i = 1, 2, \dots, n$ og

$$\underline{\underline{P(\Omega)}} = P(\{ u_1, u_2, \dots, u_n \}) = \sum_{i=1}^n P(u_i) = \sum_{i=1}^n p_i = \underline{\underline{1}} \quad (1.95)$$

Siden Ω har n utfall, finnes det totalt 2^n begivenheter. La S være en vilkårlig begivenhet med totalt $k \leq n$ utfall:

$$S = \{ u_{j_1}, u_{j_2}, \dots, u_{j_k} \} \quad (1.96)$$

Fra den spesielle addisjonssetningen (1.34) følger da umiddelbart at:

$$\boxed{P(S) = \sum_{i=1}^k p_{j_i} \quad (1.97)}$$

For en diskret sannsynlighetsmodell er dermed sannsynlighetsloven fullstendig bestemt fra utfalls-sannsynlighetene p_i .

Eksempel: (logistikk - diskret sannsynlighetsmodell)

Se øving 1.



1.1.8 Uniform sannsynlighetsmodell

Definisjon: (uniform sannsynlighetsmodell)

En **diskret** sannsynlighetsmodell $(\Omega, P(-))$ hvor :

$$\Omega = \left\{ u_1, u_2, \dots, u_n \right\} \quad (1.98)$$

kalles **uniform** dersom alle utfallene er **like sannsynlige**, dvs:

$$P(u_i) = p_i = \frac{1}{n} \quad (1.99)$$

hvor $i = 1, 2, \dots, n$

■

La S være en vilkårlig begivenhet med totalt $k \leq n$ utfall:

$$S = \left\{ u_{j_1}, u_{j_2}, \dots, u_{j_k} \right\} \quad (1.100)$$

Lign. (1.97) for diskrete sannsynlighetsmodeller gir da at

$$P(S) = \sum_{i=1}^k p_{j_i} = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n} \quad (1.101)$$

For en *uniform* sannsynlighetsmodell kan en dermed beregne sannsynligheten for enhver begivenhet ved **å telle antall utfall begivenheten består av**. Denne egenskapen til uniforme utfallsrom medfører at beregning av sannsynligheten for en begivenhet blir et *telleproblem*, dvs. fagfeltet **kombinatorikk**.

Eksempel: (**terning** - uniform sannsynlighetsmodell)

La oss atter en gang se på forsøket med en terning. Utfallsrommet er:

$$\Omega = \{ 1, 2, 3, 4, 5, 6 \} \quad (1.102)$$

og er den samme for hvert utfall, $1/6$. Altså **uniformt** utfallsrom. Med et uniformt utfallsrom hvor utfallene ikke påvirker hverandre så er dette en **tellesituasjon**.

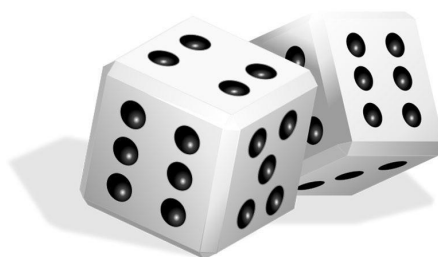
La oss denne gangen se på begivenhetene:

$$A = \text{resultat med odde antall øyne} = \{1, 3, 5\} \quad (1.103)$$

$$B = \text{resultat 4 øyne eller mer} = \{4, 5, 6\} \quad (1.104)$$

a) Hva er sannsynligheten $P(A)$?

b) Hva er “eller”-sannsynligheten $P(\overbrace{A \cup B}^{\text{eller}})$?



Figur 1.15: Terning.

Løsning:

a) Sannsynligheten $P(A)$ finns ved telling:

$$\underline{\underline{P(A)}} = \frac{\text{antall } \textcolor{red}{gunstige} \text{ utfall}}{\text{antall } \textit{mulige} \text{ utfall}} = \frac{\textcolor{red}{3}}{6} = \frac{1}{2} = \underline{\underline{0.5}} \quad (1.105)$$

b) Finner “eller”-sannsynligheten $P(\overbrace{A \cup B}^{\text{eller}})$:

i) Metode 1:

Siden utfallsrommet er uniformt hvor enkeltutfallene ikke påvirker hverandre så er dette en **tellesituasjon**²³. Det er 3 stk. gunstige utfall i begivenhetene A og B . Dermed:

$$P(A) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{3}{6} = \frac{1}{2} \quad (1.106)$$

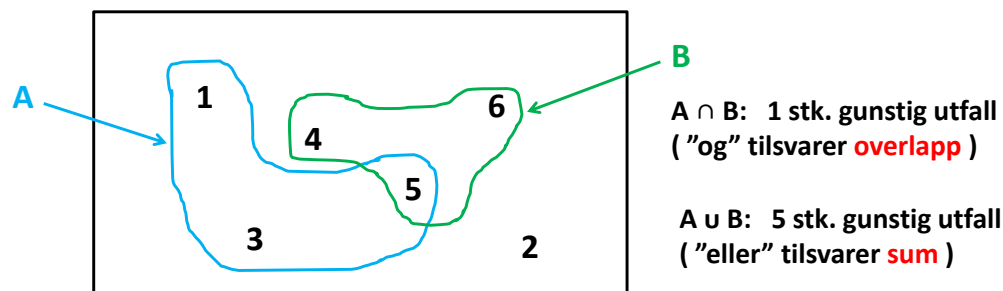
$$P(B) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{3}{6} = \frac{1}{2} \quad (1.107)$$

Av de totalt 6 mulige utfallene er det kun 1 stk. gunstig utfall for $\underbrace{A \cap B}_{\text{og}}$. Dermed:

$$P(\overbrace{A \cap B}^{\text{og}}) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{1}{6} \quad (1.108)$$

Via den **generelle addisjonsetningen**, dvs. lign. Lign.(1.40), får vi da at “eller”-sannsynligheten er:

$$\underline{\underline{P(\overbrace{A \cup B}^{\text{eller}})}} = \underbrace{P(A)}_{=1/2} + \underbrace{P(B)}_{=1/2} - \underbrace{P(\overbrace{A \cap B}^{\text{og}})}_{=1/6} = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \underline{\underline{\frac{5}{6}}} \quad (1.109)$$



Figur 1.16: **Begivenhetene** mellom A og B .

²³Vi kommer mer tilbake til dette med tellesituasjon i kapittel B om kombinatorikk.

ii) Metode 2:

Det er også mulig å finne “eller”-sannsynligheten $P(A \overset{\text{eller}}{\cup} B)$ direkte ut fra figur (1.16). Av de totalt 6 mulige utfallene er det 5 gunstige utfall som er med i A eller B . Dermed:

$$\underline{\underline{P(A \overset{\text{eller}}{\cup} B)}} = \frac{\text{antall } \textit{gunstige} \text{ utfall}}{\text{antall } \textit{mulige} \text{ utfall}} = \frac{5}{\underline{\underline{6}}} \quad (1.110)$$

som selvfølgelig er det samme svaret som vi fikk via den generelle addisjonssetningen, se lign.(1.109).



Eksempel: (kredittverdig, økonomi - uniform sannsynlighetsmodell)

Av 1000 kunder er 862 kredittverdige. Sannsynligheten for å velge en gitt, tilfeldig kunde er $1/1000$, og er den samme for hver kunde. Altså **uniformt** utfallsrom. Utfallene påvirker ikke hverandre, dvs. de er uavhengige. I likhet med eksempel 3 så er dermed også dette en **tellesituasjon**. Sannsynligheten for at en tilfeldig valgt kunde er kredittverdig kan derfor finnes ved telling: ²⁴

$$\underline{\underline{P(\text{kredittverdig})}} = \frac{\text{antall } \textit{gunstige} \text{ utfall}}{\text{antall } \textit{mulige} \text{ utfall}} = \frac{862}{1000} = \underline{\underline{0.862}} \quad (1.113)$$

■



Figur 1.17: Kredittverdig.

²⁴Senere i dette kapitlet skal vi lære om **komplementsetningen**. Med denne setningen kan vi bruke lign.(1.113) til å finne sannsynligheten for at en tilfeldig valgt kunde **ikke** er kredittverdig:

$$\underline{\underline{P(\text{IKKE kredittverdig})}} \stackrel{\text{kompl.setn.}}{=} 1 - P(\text{kredittverdig}) \quad (1.111)$$

$$= 1 - 0.862 = \underline{\underline{0.138}} \quad (1.112)$$

Kunne man funnet denne sannsynligheten på en annen måte også?

1.1.9 Kombinatorikk

For $\overbrace{\text{uniforme}}^{\text{NB!}}$ sannsynlighetsmodeller handler beregning av sannsynligheter for en begivenhet egentlig om *telling*, dvs. fagfeltet *kombinatorikk*

Se tillegg B.

Eksempel: (pasienter , kombinatorikk , telling , uniform sannsynlighetsmodell)

Anta at vi har:

- 9 pasienter
- 4 er menn, M_1 , M_2 , M_3 , M_4
- 5 er kvinner, K_1 , K_2 , K_3 , K_4 , K_5

Vi gjør følgende eksperiment: trekker 3 tilfeldige pasienter fra de 9.

Utfallsrommet består av alle tripler:

$$\Omega = \left\{ (P_1, P_2, P_3) \right\} \quad (1.114)$$

hvor P_1 , P_2 og P_3 er forskjellige pasienter.

Anta at alle utfall har lik sannsynlighet, dvs. uniform sannsynlighetsmodell.

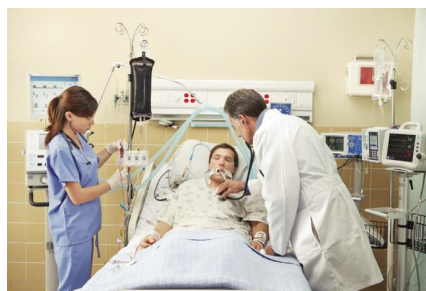
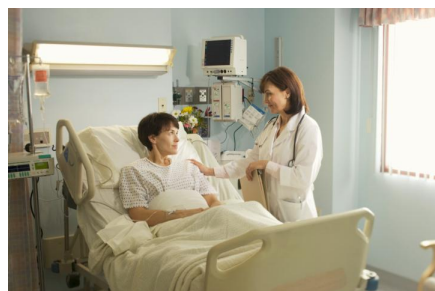
Definer følgende begivenheter:

$$A = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er menn} \right\} \quad (1.115)$$

$$B = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er kvinner} \right\} \quad (1.116)$$

$$C = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er 2 kvinner og en mann} \right\} \quad (1.117)$$

$$D = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er 1 kvinne og 2 menn} \right\} \quad (1.118)$$



Figur 1.18: Kvinnelig (K) og mannlig (M) pasient.

- a) Sannsynligheten for å trekke 3 menn, dvs. $P(A)$?
- b) Sannsynligheten for å trekke 3 kvinner, dvs. $P(B)$?
- c) Sannsynligheten for å trekke 2 kvinner og 1 mann, dvs. $P(C)$?
- d) Sannsynligheten for å trekke 1 kvinne og 2 menn, dvs. $P(D)$?

Løsning:

I dette eksemplet med pasienter som skal trekkes tilfeldig så har vi:

- **uniformt** utfallsrom ²⁵
- utfallene ikke påvirker hverandre, dvs. de er **uavhengige** ²⁶

Dermed har vi en tellesituasjon som kan beskrives ved urnemodellen.

Den generelle sammenhengen mellom telling og sannsynlighet er da gitt ved lign.(B.26):

$$P(A) = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner for } A}{\text{antall } \textit{mulige} \text{ kombinasjoner totalt}} \quad (1.119)$$

Når man trekker en mann eller kvinne så spiller rekkefølgen ingen rolle.

Dessuten, når en mann eller kvinne er trukket så kan han/hun ikke trekkes på nytt.

Dette tilsvarer:

$$\text{situasjon 3} \quad (1.120)$$

dvs. **ikke-ordnet** utvalg **uten** tilbakelegging.

Antall kombinasjoner er da gitt ved lign.(B.16), dvs.:

$$\underbrace{\# \text{ ikke-ordnede komb. u/tilbakelegging } = \frac{N!}{(N-s)! s!} \equiv \underbrace{\binom{N}{s}}_{\text{binomialkoeff.}}}_{\text{situasjon 3}} \quad (1.121)$$

²⁵Samme sannsynlighet for å trekke en pasient.

²⁶Om vi trekker en pasient så har det ikke påvirkning på hva slags pasient vi trekker neste gang.

a) $P(A)$:

Det er to kategorier, 4 menn og 5 kvinner. Dermed:

$$\underline{\underline{P(A)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.122)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.123)$$

$$= \frac{\overbrace{\binom{4}{3}}^{= 4} \cdot \overbrace{\binom{5}{0}}^{= 1}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{4 \cdot 1}{84} = \underline{\underline{0.048}} \quad (= 4.8 \%) \quad (1.124)$$

b) $P(B)$:

Det er to kategorier, 4 menn og 5 kvinner. Dermed:

$$\underline{\underline{P(B)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.125)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.126)$$

$$= \frac{\overbrace{\binom{4}{0}}^{= 1} \cdot \overbrace{\binom{5}{3}}^{= 10}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{1 \cdot 10}{84} = \underline{\underline{0.119}} \quad (1.127)$$

c) $P(C)$:

Det er to kategorier, 4 menn og 5 kvinner. Dermed:

$$\underline{\underline{P(C)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.128)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.129)$$

$$= \frac{\overbrace{\binom{4}{1}}^{= 4} \cdot \overbrace{\binom{5}{2}}^{= 10}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{4 \cdot 10}{84} = \underline{\underline{0.476}} \quad (1.130)$$

d) $P(D)$:

Det er to kategorier, 4 menn og 5 kvinner. Dermed:

$$\underline{\underline{P(D)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.131)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.132)$$

$$= \frac{\overbrace{\binom{4}{2}}^{= 6} \cdot \overbrace{\binom{5}{1}}^{= 5}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{6 \cdot 5}{84} = \underline{\underline{0.357}} \quad (1.133)$$

Kommentarer:

- 1) Legg merke til følgende: Det er 2 kategorier, menn og kvinner. Sannsynligheten blir da:

$$P(\text{dersom 2 kategorier}) = \frac{2 \text{ stk. binomialkoeff.}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.134)$$

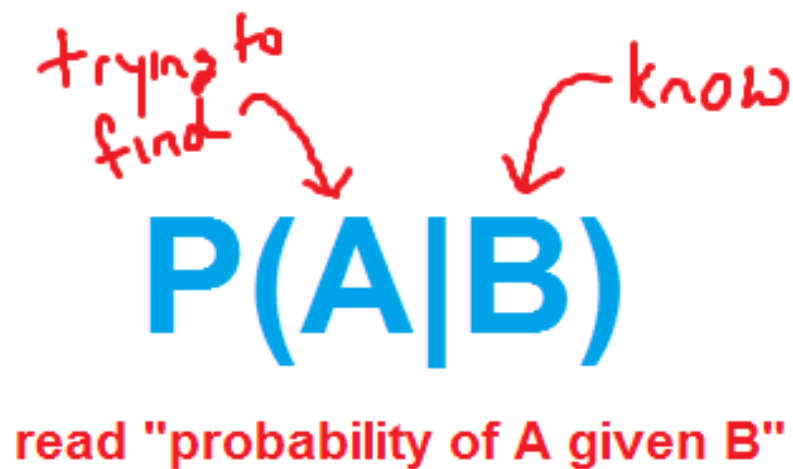
- 2) Antall *kombinasjoner* som finnes av kvinner og menn har ikke noe med hverandre å gjøre. De er uavhengige. Derfor kan vi multiplisere antall kombinasjoner i teller i lign.(1.130).

■

1.2 Betinget sannsynlighet

Noen ganger har vi $\overbrace{\text{tilleggsinformasjon om noe som har skjedd.}}^{\text{betinget sannsynlighet}}$.

Dette kan påvirke beregningen av sannsynlighetene.



Figur 1.19: Betinget sannsynlighet.

Eksempel: (**bankkunder** - betinget sannsynlighet)

Bedriftdivisjonen til Sparebanken Møre har totalt 66 kunder, $K_1, K_2, \dots, K_{66}$. Banken klassifiserer kundene sine hht. figur 1.21.

Banken ønsker å gjøre følgende eksperiment: trekke en vilkårlig kunde.

Utfallsrom:

$$\Omega = \{ K_1, K_2, \dots, K_{66} \} \quad (1.135)$$

dvs. 66 utfall hvor vi antar at alle utfall har lik sannsynlighet, dvs. uniform sannsynlighetsmodell.

Definer følgende begivenheter: ²⁷

$$A = \{ K \mid \text{hvor kunden } K \text{ er en god kunde} \} \quad (1.136)$$

$$B = \{ K \mid \text{hvor kunden } K \text{ er en kunde med stor omsetning} \} \quad (1.137)$$

$$A \cap B = \{ K \mid \text{hvor kunden } K \text{ er en god kunde med stor omsetning} \} \quad (1.138)$$

$$A|B = \{ K \mid \text{hvor kunden } K \text{ er en god kunde} \\ \text{dersom vi } \underline{\text{vet}} \text{ at kunden har stor omsetning} \} \quad (1.139)$$



Figur 1.20: Sparebanken Møre.

²⁷Legg merke til at notasjonen $A|B$ er ny og står for betinget sannsynlighet.

	B (stor omsetn.)	$\overline{\text{B}}$ (liten omsetn.)	totalt
A (god kunde)	21	13	34
$\overline{\text{A}}$ (dårlig kunde)	15	17	32
totalt	36	30	66

Figur 1.21: Bedriftkundemassen til Sparebanken Møre.

Eksperiment: En tilfeldig bedrift velges fra kundemassen.

- a) Hva er sannsynligheten for å trekke en **god** kunde, dvs. $P(A)$?
- b) Hva er sannsynligheten for å trekke en kunde med **stor** omsetning, dvs. $P(B)$?
- c) Hva er sannsynligheten for å trekke en **god** kunde med **stor** omsetning, dvs. $P(A \cap B)$?
- d) Hva er sannsynligheten for å trekke en **god** kunde dersom vi **vet** at kunden har stor omsetning, dvs. $P(A|B)$?

Løsning:

- a) Sannsynligheten for å trekke en **god** (A) kunde:

$$\underline{\underline{P(A)}} = \frac{\text{antall } gode \text{ (A) kunder}}{\text{antall totalt}} = \frac{34}{66} \approx \underline{\underline{0.52}} \quad (1.140)$$

- b) Sannsynligheten for å trekke en kunde med **stor** (B) omsetning:

$$\underline{\underline{P(B)}} = \frac{\text{antall kunder med } stor \text{ (B) omsetning}}{\text{antall totalt}} = \frac{36}{66} \approx \underline{\underline{0.55}} \quad (1.141)$$

- c) Sannsynligheten for å trekke en **god** (A) kunde med **stor** (B) omsetning:

$$\underline{\underline{P(A \overset{\text{og}}{\cap} B)}} = \frac{\text{antall } gode \text{ (A) kunder med } stor \text{ (B) omsetning}}{\text{antall totalt}} = \frac{21}{66} \approx \underline{\underline{0.32}} \quad (1.142)$$

- d) Sannsynligheten for å trekke en **god** kunde dersom vi **vet** at kunden har stor omsetning:

Siden vi vet at kunden har stor (B) omsetning, så kan man si at B allerede har inntruffet. Det betyr at alle kundene med lav omsetning ikke kan bli trukket, mao. utfallsrommet har blitt mindre.

$$\underline{\underline{P(A|B)}} = \frac{\text{antall gode (A) kunder med stor (B) omsetning}}{\text{antall totalt med stor omsetning}} = \frac{21}{36} \quad (1.143)$$

$$= \frac{\frac{21}{66}}{\frac{36}{66}} = \frac{\overbrace{P(A \cap B)}^{= 21/66}}{\underbrace{P(B)}_{= 36/66}} \approx \underline{\underline{0.58}} \quad (1.144)$$

Konklusjon:

Sannsynligheten for å trekke en god kunde, $P(A) \approx 0.52$, er altså **ikke** den samme som sannsynligheten for å trekke en god god kunde dersom vi **vet** at kunden har stor omsetning, $P(A|B) \approx 0.58$.



Definisjon: (*betingset* sannsynlighet)

Den betingsede sannsynligheten er definert ved:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.145)$$

hvor

$$P(A|B) = \text{sannsynligheten for } A \text{ gitt } B \text{ (} B \text{ allerede har inntruffet)} \quad (1.146)$$

(*betingset* sannsynlighet)

$$P(A \cap B) = \text{“og”-sannsynligheten for } A \text{ og } B$$

$$P(B) = \text{sannsynligheten for } B \neq 0 \quad (1.147)$$

(*ubetingset* sannsynlighet)

■

Lign.(1.145) er en definisjon og trenger således ikke å bevises.

1.2.1 Multiplikasjonssetningen

Fra definisjonen av betinget sannsynlighet i lign.(1.145) følger trivielt det som kalles multiplikasjonssetningen:

Setning: (*multiplikasjons*setningen, *generelle*)

For begivenhetene A og B gjelder:

$$P(A \overset{\text{og}}{\cap} B) = P(A|B) P(B) \quad (1.148)$$

■

Bevis:

Følger trivielt fra definisjonen av betinget sannsynlighet:

$$P(A|B) \stackrel{\text{lign. (1.145)}}{=} \frac{P(A \cap B)}{P(B)} \quad \Bigg| \cdot P(B) \quad (1.149)$$

$\Updownarrow$

$$\underline{\underline{P(A \cap B) = P(A|B) P(B)}} \quad (1.150)$$

■

Siden $P(B \cap A) = P(A \cap B)$ følger det umiddelbart at multiplikasjonssetningen alternativt kan skrives:

Setning: (*multiplikasjons*setningen, *generelle*)

For begivenhetene A og B gjelder:

$$\overbrace{P(B \cap A)}^{= P(A \cap B)} = P(B|A) P(A) \quad (1.151)$$

■

For en betinget sannsynlighet står det vi **vet** til høyre for “|”:

$$P(B|A) = P(B \mid \underbrace{\text{når vi vet}}_{\text{vet}} A) \quad (1.152)$$

1.2.2 Bayes lov

Setning: (*Bayes lov*)

For begivenhetene A og B gjelder:

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)} \quad (1.153)$$

■

Bevis:

Bayes lov følger direkte fra [multiplikasjons](#)setningen og det faktum at $P(A \cap B) = P(B \cap A)$:

$$P(A \cap B) = P(B \cap A) \quad (1.154)$$

$\Updownarrow$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \quad \left| \cdot \frac{1}{P(B)} \right. \quad (1.155)$$

$$\underline{\underline{P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}}} \quad (1.156)$$

■

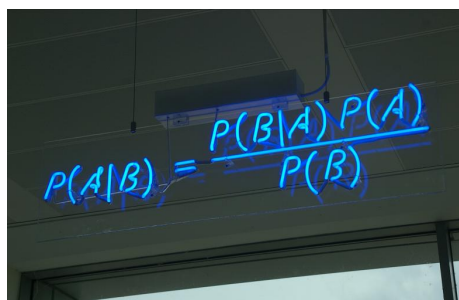
På same måte som multiplikasjonssetningen så kan vi snu på Bayes lov og skrive denne på en alternativ måte:

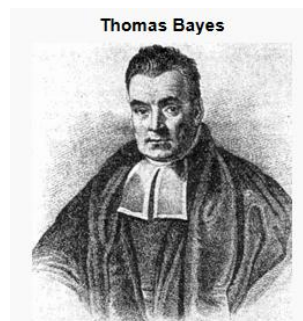
Setning: (*Bayes lov*)

For begivenhetene A og B gjelder:

$$P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)} \quad (1.157)$$

■


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



Figur 1.22: Bayes lov. Thomas Bayes.

Eksempel: (likviditet og økonomi - betinget sannsynlighet)

Lindorff Group AB har gjort en likviditetsvurdering av alle store og mellomstore bedrifter i Møre & Romsdal. De deler bedriftene inn i grupper: god likviditet (A) og dårlig likviditet ($\bar{A}$). Videre deler de bedriftene inn i de som har gjennomført investeringer som forventer avkastning i et langsiktig (B) tidsperspektiv, og de som kun har gjort investeringer med tanke på kortsiktige ($\bar{B}$) gevinster eller av helt nødvendige grunner ($\bar{B}$):

	B	$\bar{B}$		totalt
	langsiktige investeringer	kortsiktige investeringer	nødvendige investeringer	
A (god likviditet)	0.45	0.20	0.15	0.80
$\bar{A}$ (dårlig likviditet)	0.02	0.08	0.10	0.20
totalt	0.47	0.28	0.25	1

Figur 1.23: Relative hyppigheter.

Vi har altså definert begivenhetene:

- A = bedrift har god likviditet
- B = bedrift har gjennomført langsiktige investeringer



Figur 1.24: Lindorff.

Vi gjør følgende eksperiment: trekker en tilfeldig bedrift.
Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ (A, B), (A, \overline{B}), (\overline{A}, B), (\overline{A}, \overline{B}) \right\} \quad (1.158)$$

- a) Hva er sannsynligheten for at en tilfeldig valgt bedrift har **god** likviditet?
- b) Hva er sannsynligheten for at en tilfeldig valgt bedrift har gjennomført **langsiktige** investeringer?
- c) Hva er sannsynligheten at en tilfeldig valgt bedrift har **god** likviditet dersom vi **vet** at bedriften har gjort **langsiktige** investeringer?
- d) Hva er sannsynligheten at en tilfeldig valgt bedrift har gjort **langsiktige** investeringer dersom vi **vet** at bedriften har **god** likviditet?

Løsning:

a) **God** (A) likviditet:

$$\underline{\underline{P(A)}} = 0.45 + 0.20 + 0.15 = \underline{\underline{0.80}} \quad (1.159)$$

som er den ubetingede sannsynligheten for at en tilfeldig valgt bedrift har *god* likviditet.

b) **Langsiktige** (B) investeringer:

$$\underline{\underline{P(B)}} = 0.45 + 0.02 = \underline{\underline{0.47}} \quad (1.160)$$

som er den ubetingede sannsynligheten for at en tilfeldig valgt bedrift har gjennomført *langsiktige* investeringer.

c) **God** (A) likviditet dersom vi **vet** at bedriften har gjort **langsiktige** (B) investeringer:

Siden vi kjenner sannsynlighetene $P(B)$ og $P(A \overset{\text{og}}{\cap} B)$ så kan vi bruke multiplikasjonssetningen i lign.(1.148):

$$\begin{aligned} \underline{\underline{P(A|B)}} &= \frac{\text{hyppighet god likv. (A) og } \textit{langsikt. (B)}}{\underbrace{\text{hyppighet } \textit{langsiktige} \text{ investeringer}}_{\text{utfallsrommet 'reduisert' til kun B}}} \\ \text{lign. (1.148)} \quad \underline{\underline{\frac{P(A \cap B)}{P(B)}}} &= \frac{\overset{= 0.45}{P(A \cap B)}}{\underset{= 0.47}{P(B)}} = \frac{0.45}{0.47} \approx \underline{\underline{0.96}} \quad (1.161) \end{aligned}$$

som er den betingede sannsynligheten for at en tilfeldig bedrift *blant de som har gjort langsiktige* (B) investeringer, har god likviditet (A).

- d) **Langsiktige** investeringer (B) dersom vi **vet** at bedriften har **god** (A) likviditet:

Siden vi kjenner sannsynlighetene $P(A)$, $P(B)$ og $P(A|B)$ så kan vi bruke **Bayes lov** i lign.(1.157):

$$\underline{\underline{P(B|A)}} \stackrel{\text{lign. (1.157)}}{=} \overbrace{P(A|B)}^{\approx 0.96} \cdot \overbrace{\frac{P(B)}{P(A)}}^{= 0.47} = 0.96 \cdot \frac{0.47}{0.80} = \underline{\underline{0.56}} \quad (1.162)$$

som er den betingede sannsynligheten for at en tilfeldig valgt bedrift *blant de som har god likviditet* (A), har gjort langsiktige (B) investeringer.

Alternativt kan vi finne $P(B|A)$ via direkte avlesning av tabellen, analogt med lign.(1.161), dvs. multiplikasjonssetningen:

$$\begin{aligned} \underline{\underline{P(B|A)}} &= \frac{\text{hyppighet langsikt. (B) og god (A) likv.}}{\underbrace{\text{hyppighet god likviditet}}_{\text{utfallsrommet 'reduert' til kun A}}} \\ &\stackrel{\text{lign. (1.151)}}{=} \frac{\overbrace{P(B \cap A)}^{= P(A \cap B) = 0.45}}{\underbrace{P(A)}_{= 0.80}} = \frac{0.45}{0.80} \approx \underline{\underline{0.56}} \end{aligned} \quad (1.163)$$

dvs. samme svar som i lign.(1.162), selvfølgelig.

■

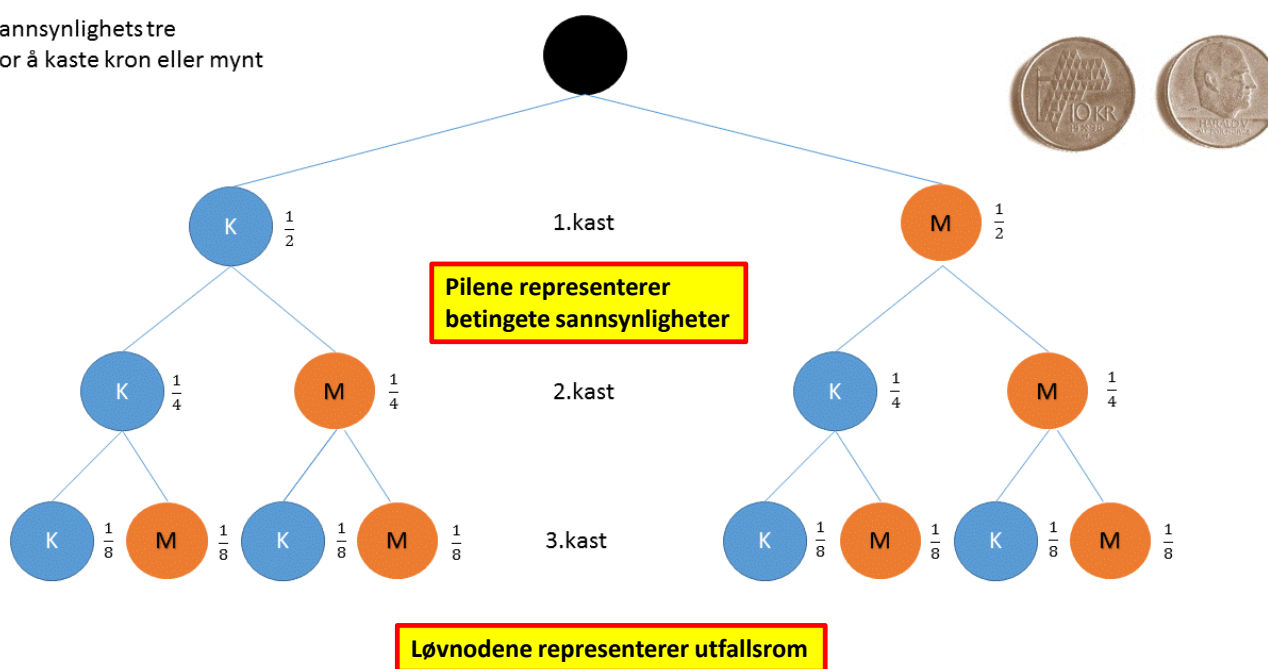
1.2.3 Sannsynlighetstrær

I mange tilfeller har man et eksperiment som gjennomføres som en sekvens av steg, dvs. en bestemt rekkefølge av steg.

I slike tilfeller kan det være hensiktsmessig å tegne opp et tre som viser de ulike mulighetene.

Slike trær kalles **sannsynlighetstrær**.²⁸

Sannsynlighetstre
For å kaste kron eller mynt



Figur 1.25: Sannsynlighetstre - sekvens av steg.

²⁸Utfallsrommet er gitt ved løvnodene. Pilene representerer betingede sannsynligheter.

Eksempel: (fond, økonomi - sannsynlighetstrær)

En amerikaner har bestemt seg for å investere i fond.

Anta at sannsynligheten for at han investerer på NASDAQ-børsen er 75 % og sannsynligheten for å investere på NYSE-børsen er 25 %.

Anta videre at det er 80 % sannsynlighet for at disse NASDAQ-fondene blir innvesret i teknologi.

Anta også at det er 40 % sannsynlighet for at NYSE-fondene blir innvesret i teknologi.

- 75 % NASDAQ
 - av disse blir 80 % investert i teknologi
- 25 % NYSE
 - av disse blir 40 % investert i teknologi

Hva er sannsynligheten for at amerikaneren gjør inovesringer i teknologi?

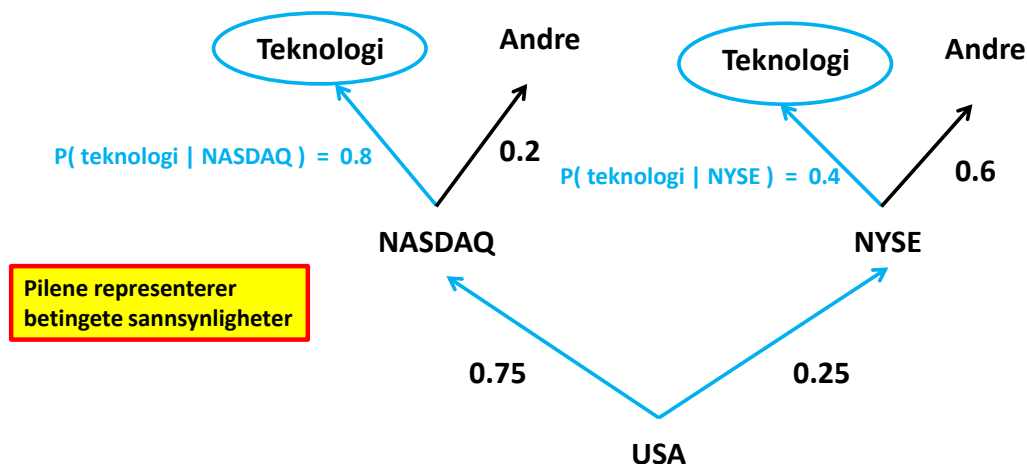


Figur 1.26: NASDAQ og NYSE.

Løsning:

Legg merke til at denne situasjonen med investeringer kan ses på som en sekvens av steg: Først velges børs, deretter om man skal gjøre investeringer i teknologi eller noe annet.

Derfor er det her hensiktsmessig å bruke et sannsynlighetstre: ²⁹



Figur 1.27: **Sannsynlighetstre** for fond i USA.

Sannsynlighetstreet i figur (1.27) er en visuell fremstilling av opplysningene. Med en slik fremstilling er det lettere å se at: ³⁰

$$\underline{\underline{P(\text{aksjene investert i teknologi})}} = 0.75 \cdot 0.8 + 0.25 \cdot 0.4 = \underline{\underline{0.70}} \quad (1.164)$$

■

²⁹Utfallsrommet er gitt ved løvnodene. Pilene representerer betingede sannsynligheter.

³⁰Følg de blå linjene i sannsynlighetstreet.

1.2.4 Oppsplitting av Ω

Eksempel: (fond, økonomi - sannsynlighetstrær)

La oss igjen se på eksempelet i forrige avsnitt. La oss løse oppgaven ved regning.

Anta at vi har:

- A = teknologi
- B_1 = NASDAQ
- B_2 = NYSE

Vi gjør følgende eksperiment: trekker et fond som amerikaneren har investert i. Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ (B_1, A), (B_1, \overline{A}), (B_2, A), (B_2, \overline{A}) \right\} \quad (1.165)$$

Definer følgende begivenheter:

$$B_1 = \left\{ (B_1, A), (B_1, \overline{A}) \right\} \quad (\text{et NASDAQ-fond trekkes}) \quad (1.166)$$

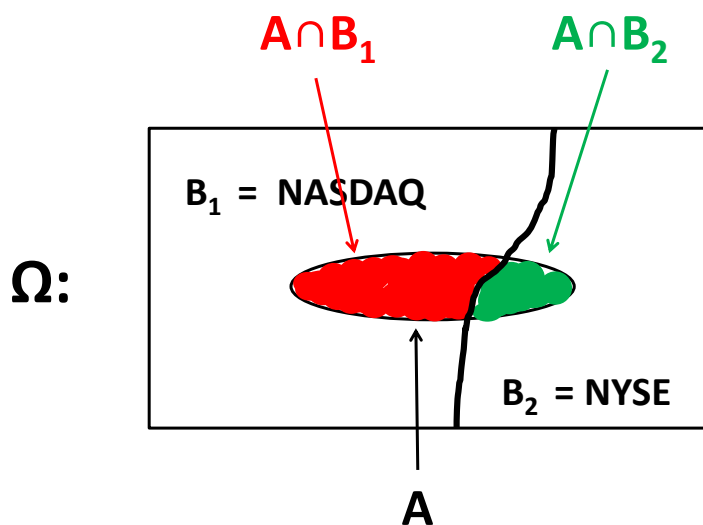
$$B_2 = \left\{ (B_2, A), (B_2, \overline{A}) \right\} \quad (\text{et NYSE-fond trekkes}) \quad (1.167)$$

Vi har altså splittet utfallsrommet $\Omega = B_1 \underbrace{\cup}_{\text{eller}} B_2$ i to:

$$\Omega = \left\{ B_1, B_2 \right\} \quad (1.168)$$

Noe av NASDAQ-fondene investeres i teknologi, $A \cap B_1$. Tilsvarende, noe av NYSE-fondene investeres også i teknologi, $A \cap B_2$. Den totale mengden fond som investeres i teknologi A er summen av disse:

$$A = (A \cap B_1) \underbrace{\cup}_{\text{eller}} (A \cap B_2) \quad (1.169)$$



Figur 1.28: **Oppsplitting** av sannsynlighetsrommet i to, $\Omega = B_1 \cup B_2$.

Via lign.(1.169) finner vi sannsynligheten $P(A)$:

$$P(A) = P\left[\underbrace{(A \cap B_1) \cup (A \cap B_2)}_{\text{eller}}\right] \quad (1.170)$$

Av figur (1.28) ser vi at det ikke er noe overlapp mellom $A \cap B_1$ og $A \cap B_2$, dvs. de er disjunkte³¹. Da gjelder den spesielle addisjonssetningen for “eller”-sannsynligheten i lign.(1.170):³²

$$P(A) = P(A \cap B_1) + P(A \cap B_2) \quad (1.171)$$

Denne ligningen inneholder to “og”-sannsynligheter. Dermed kan vi bruke multiplikasjonssetningen³³ på disse:

$$\underline{\underline{P(A)}} = \underbrace{P(A|B_1) \cdot P(B_1)}_{= P(A \cap B_1)} + \underbrace{P(A|B_2) \cdot P(B_2)}_{= P(A \cap B_2)} \quad (1.172)$$

$$\stackrel{\text{lign. (1.148)}}{=} \overbrace{P(A|B_1)}^{= 0.8} \cdot \overbrace{P(B_1)}^{= 0.75} + \overbrace{P(A|B_2)}^{= 0.4} \cdot \overbrace{P(B_2)}^{= 0.25} \quad (1.173)$$

$$= 0.8 \cdot 0.75 + 0.4 \cdot 0.25 = \underline{\underline{0.7}} \quad (1.174)$$

dvs. samme svar som ved bruk av sannsynlighetstre, selvfølgelig.

■

³¹At det ikke er noe overlapp, dvs. disjunkte, betyr at $(A \cap B_1) \cup (A \cap B_2) = \emptyset$.

³²Den spesielle addisjonssetningen i lign.(1.34) sier at: $P(A \cup B) = P(A) + P(B)$.

³³Multiplikasjonssetningen i lign.(1.148) sier at: $P(A \cap B) = P(A|B) \cdot P(B)$.

Setning: (oppsplitting av Ω i 2)

For betingelsene A og B gjelder at:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) \quad (1.175)$$

■

Bevis:

Anta at utfallsrommet Ω **splittes** i to delrom som er disjunkte:

$$A = (A \cap B_1) \overset{\text{eller}}{\cup} (A \cap B_2) \quad (\text{oppsplitting}) \quad (1.176)$$

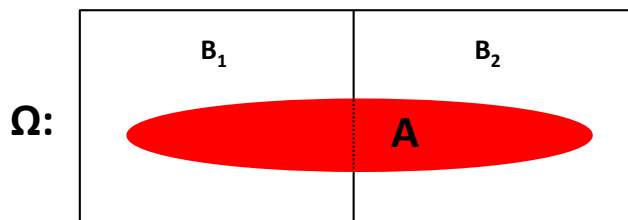
dvs. $B_1 \cap B_2 = \emptyset$. Dermed er tilhørende sannsynlighet:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) \quad (1.177)$$

Via multiplikasjonssetningen i lign.(1.148) får vi dermed:

$$\underline{\underline{P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)}} \quad (1.178)$$

■



Figur 1.29: **Oppsplitting** av sannsynlighetsrom Ω , jfr. lign.(1.176).

1.2.5 Uavhengighet

Definisjon: (uavhengighet)

To begivenheter A og B er uavhengige dersom

$$P(A|B) = P(A) \quad (1.179)$$

■

Lign.(1.179) er en definisjon og trenger således ikke å bevises.

Dette betyr at selv om B har inntruffet så påvirker ikke det sannsynligheten for A .
Med andre ord, A og B er uavhengige.

Setning: (multiplikasjonssetningen, spesielle)

Dersom begivenhetene A og B er uavhengige, så gjelder:

$$P(\overbrace{A \cap B}^{\text{og}}) \stackrel{\text{uavh.}}{=} P(A) \cdot P(B) \quad (1.180)$$

■

Bevis:

Multiplikasjonssetningen i lign.(1.148) gir:

$$P(A \cap B) \stackrel{\text{lign.(1.148)}}{=} P(A|B) P(B) \quad (1.181)$$

Men siden A og B er uavhengige så er er

$$P(A|B) = P(A) \quad (1.182)$$

Dermed:

$$\underline{\underline{P(A \cap B) = P(A) \cdot P(B)}} \quad (1.183)$$

■

NB:

Ikke bland sammen begrepene “uavhengighet” og “disjunkthet”: ³⁴

- i) $\overbrace{\text{Uavhengighet}}^{\text{vanskelig \AA visualisere}} : ^{35} \quad P(A \overset{\text{og}}{\cap} B) = P(A) \cdot P(B) \quad (\text{ se lign. (1.180) })$
- ii) $\underbrace{\text{Disjunkt}}_{\text{ikke overlapp}} : ^{36} \quad P(A \overset{\text{eller}}{\cup} B) = P(A) + P(B) \quad (\text{ se lign. (1.34) })$

³⁴Disse begrepene kan ogs\AA v\AA re koblet p\AA f\AA lgende m\AA ter:

$$\text{disjunkt} \quad \overset{\text{nesten alltid}}{\implies} \quad \text{avhengig} \quad (1.184)$$

$$\text{disjunkt} \quad \overset{\text{ikke n\AA} \text{dvendigvis}}{\impliedby} \quad \text{avhengig} \quad (1.185)$$

³⁵Alts\AA : $P(\text{og}) = \text{PRODUKT}$ av ubetingede sannsynligheter

³⁶Alts\AA : $P(\text{eller}) = \text{SUM}$ av ubetingede sannsynligheter

Eksempel: (økonomisk vekst)

Det er ofte en klar sammenheng mellom økonomisk vekst i et land og styrkingen av valutaen. Grovt sett så følger disse størrelsene et **generelt** mønster:

Ved $\overbrace{\text{høy vekst}}^{=H}$ er det 70 % sannsynlig at $\overbrace{\text{valutaen styrker seg}}^{=S}$, $P(S|H) = 0.7$

Ved $\overbrace{\text{middels vekst}}^{=M}$ er det 50 % sannsynlig at $\overbrace{\text{valutaen styrker seg}}^{=S}$, $P(S|M) = 0.5$

Ved $\overbrace{\text{lav vekst}}^{=L}$ er det 20 % sannsynlig at $\overbrace{\text{valutaen styrker seg}}^{=S}$, $P(S|L) = 0.2$

Anta at sannsynligheten for økonomisk vekst i Østerrike er: ³⁷

$$P(H) = 0.32 \quad (\text{sannsynlighet for høy økonomisk vekst}) \quad (1.187)$$

$$P(M) = 0.51 \quad (\text{sannsynlighet for middels økonomisk vekst}) \quad (1.188)$$

$$P(L) = 0.17 \quad (\text{sannsynlighet for lav økonomisk vekst}) \quad (1.189)$$



Figur 1.30: Økonomisk vekst.

³⁷Legg merke til at lign.(1.187)-(1.189) representerer en gyldig sannsynlighetsfordeling siden:

$$\underline{\underline{P(\text{høy}) + P(\text{middels}) + P(\text{lav})}} = 0.32 + 0.51 + 0.17 = \underline{\underline{1}} \quad (1.186)$$

Ovenfor har vi notasjonen:

- $H = \text{høy}$ økonomisk vekst
- $M =$ middels økonomisk vekst
- $L = \text{lav}$ økonomisk vekst
- $S =$ valutaen styrker seg

Vi gjør følgende eksperiment: observerer hvordan valutaen styrker seg.
Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ \overbrace{(H, S), (H, \bar{S})}^{= H}, \overbrace{(M, S), (M, \bar{S})}^{= M}, \overbrace{(L, S), (L, \bar{S})}^{= L} \right\} \quad (1.190)$$

$$= \{ H, M, L \} \quad (1.191)$$

hvor vi har definert følgende begivenheter:

$$H = \{ (H, S), (H, \bar{S}) \} \quad (\text{høy økonomisk vekst}) \quad (1.192)$$

$$M = \{ (M, S), (M, \bar{S}) \} \quad (\text{middels økonomisk vekst}) \quad (1.193)$$

$$L = \{ (L, S), (L, \bar{S}) \} \quad (\text{lav økonomisk vekst}) \quad (1.194)$$

$$S = \{ (H, S), (M, S), (L, S) \} \quad (\text{valutaen styrker seg}) \quad (1.195)$$

hvor (se lign.(1.187)-(1.189))

$$\underline{\underline{P(H) + P(M) + P(L)}} = 0.32 + 0.51 + 0.17 = \underline{\underline{1}} \quad (1.196)$$

representerer en gyldig sannsynlighetsfordeling.

a) Hva er $P(S)$? ³⁸

b) Hva er $P(H|S)$? ³⁹

c) Er begivenhetene om økonomisk vekst og styrkingen av valuta uavhengige for Østerrike sit tilfelle?

³⁸Dvs. finn sannsynligheten for at valutaen styrker seg i Østerrike.

³⁹Dvs. finn sannsynligheten for at det er høy vekst i Østerrike dersom vi vet at valutaen styrker.

Løsning:

- a) Siden vi kjenner $P(S|H)$ og $P(S|L)$ i tillegg til de ubetingede sannsynlighetene i lign.(1.187) og (1.189), så kan vi bruke setning om *oppsplitting* av Ω i lign.(1.177):

$$\underline{\underline{P(S)}} = \underbrace{P(S|H)}_{=0.7} \underbrace{P(H)}_{=0.32} + \underbrace{P(S|M)}_{=0.5} \underbrace{P(M)}_{=0.51} + \underbrace{P(S|L)}_{=0.2} \underbrace{P(L)}_{=0.17} \quad (1.197)$$

$$= 0.7 \cdot 0.32 + 0.5 \cdot 0.51 + 0.2 \cdot 0.17 \approx \underline{\underline{0.513}} \quad (1.198)$$

Konklusjon:

Det er 51.3 % sannsynlighet for at valutaen styrker seg i Østerrike.

- b) Sannsynlighetene $P(H|S)$ og $P(S|H)$ er *ikke* like. Det er imidlertid en sammenheng mellom dem. Denne sammenhengen finnes via *Bayes lov* i lign.(1.153):

$$\underline{\underline{P(H|S)}} \stackrel{\text{lign.(1.153)}}{=} \underbrace{P(S|H)}_{=0.7} \cdot \frac{\overbrace{P(H)}^{=0.32}}{\underbrace{P(S)}_{=0.513}} = 0.7 \cdot \frac{0.32}{0.513} \approx \underline{\underline{0.44}} \quad (1.199)$$

Konklusjon:

Dersom valutaen styrker seg, så er det 44 % sannsynlighet for at Østerrike har høy økonomisk vekst.

- c) Uavhengighet testes ved å regne ut $P(S \cap H)$ og produktet $P(H) \cdot P(S)$. Dersom disse to er like, så er begivenhetene H og S uavhengige, jfr. den *spesielle* multiplikasjonssetningen i lign.(1.180).

- i) Siden vi kjenner $P(H|S)$ og $P(S)$, så kan vi finne $P(S \cap H)$ via multiplikasjonssetningen i lign.(1.148):

$$\underline{P(H \cap S)} \stackrel{\text{lign. (1.148)}}{=} \underbrace{P(H|S)}_{\approx 0.44} \cdot \underbrace{P(S)}_{= 0.513} \approx \underline{0.22} \quad (1.200)$$

Så et svært viktig poeng: Siden vi også kjenner $P(S|H)$ og $P(H)$, så kan vi alternativt finne $P(S \cap H)$ via disse størrelsene. *Multiplikasjonssetningen* (1.151) gir:

$$\underline{P(S \cap H)} \stackrel{\text{lign. (1.151)}}{=} \underbrace{P(S|H)}_{= 0.7} \cdot \underbrace{P(H)}_{= 0.32} \approx \underline{0.22} \quad (1.201)$$

dvs. samme svar som i lign.(1.200), selvfølgelig.

- ii) Produktet er lett å regne ut:

$$\underline{P(S) \cdot P(H)} = 0.513 \cdot 0.32 \approx \underline{0.16} \quad (1.202)$$

Konklusjon:

Siden $\underbrace{P(S \cap H)}_{\approx 0.22} \neq \underbrace{P(S) \cdot P(H)}_{\approx 0.16}$ så er begivenhetene avhengige.

■

Eksempel: (eksamen 1. juni 2012, [sannsynlighetsregning](#))

La oss se på begivenhetene A og B . Nedenfor ser du fire matematiske uttrykk for disse begivenhetene. To og to av disse hører sammen og danner to forskjellige *setninger*.

$$P(A) \cdot P(B) \tag{1.203}$$

$$P(A) + P(B) \tag{1.204}$$

$$P(A \cap B) \tag{1.205}$$

$$P(A \cup B) \tag{1.206}$$

- a) Sett sammen to og to av disse uttrykkene som hører sammen slik at de danner to setninger.
- b) Hvilke navn har disse to setningene?
- c) For de to setningene i oppgavene foran, hva er forutsetningen for at de skal gjelde?



Figur 1.31: Statistikk.

Løsning:

a) $\underline{P(A \cup B) = P(A) + P(B)}$, $\underline{P(A \cap B) = P(A) \cdot P(B)}$

b) $\underline{\text{Den spesielle addisjonssetningen}}$, $\underline{\text{Den spesielle multiplikasjonssetningen}}$

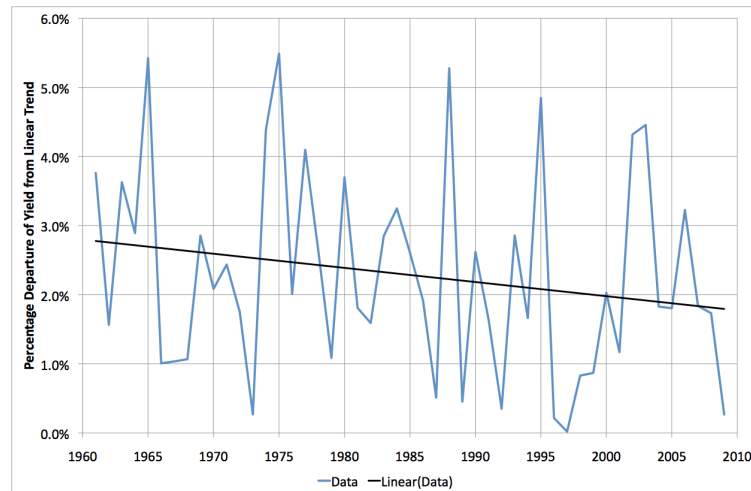
c) Den *spesielle* addisjonssetningen forutsetter:
Begivenhetene A og B er disjunkte, dvs. $A \cap B = \emptyset$, dvs. A og B inntreffer aldri samtidig.

Den *spesielle* multiplikasjonssetningen forutsetter:
Begivenhetene A og B er uavhengige, dvs. $P(A|B) = P(A)$ eller $P(B|A) = P(B)$.



Kapittel 2

Stokastiske variabler, forventning og varians



Figur 2.1: Forventning og varians.

2.1 Stokastiske variabler

Definisjon: (stokastisk variabel) ¹

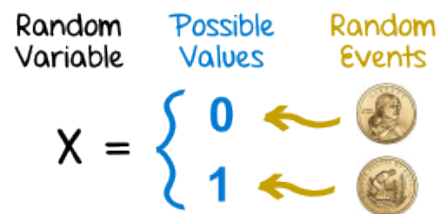
En stokastisk variabel er en størrelse X som kan anta ulike verdier x med ulike sannsynligheter.

tilfeldig

■

Noen kommentarer:

- Vi skiller mellom ^{"hakkete"} diskrete og ^{"glatte"} kontinuerlige stokastiske variabler
- stor X = stokastisk variabel
- liten x = en mulig verdi til variabelen X
- På de neste sidene skal vise at enhver stokastisk variabel har en tilhørende sannsynlighetsfordeling $P(X = x)$.



Figur 2.2: Diskret stokastisk variabel.

¹En mer teknisk (matematisk) versjon av definisjonen av en stokastisk variabel er: Med en tilfeldig variabel mener vi en funksjon X som til ethvert mulig utfall definerer et bestemt reelt tall.

2.1.1 Diskret VS kontinuerlig stokastiske variabler

Definisjon: (^{”hakkete”}
diskret stokastisk variabel)

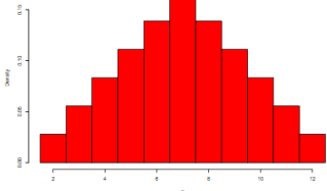
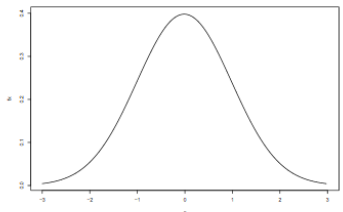
En diskret stokastisk variabel er en størrelse X som kan ha et endelig antall eller et tellbart uendelig antall mulige ulike verdier x med ulike sannsynligheter.

■

Definisjon: (^{”glatt”}
kontinuerlig stokastisk variabel)

En kontinuerlig stokastisk variabel er en størrelse X som kan ha alle verdier i et intervall tallinjen eller hele tallinjen.

■

Diskret stokastisk variabel	Kontinuerlig stokastisk variabel
X	X
Mulige verdier x : Endelig eller tellbart mange	Mulige verdier x : Intervall eller hele R
Eksempel: $\{0, 1, \dots, n\}$	Eksempel: $[0, 1]$ eller $[0, \infty)$
Sannsynlighetsfordeling: $f(x) = P(X = x)$ for alle mulige x	Sannsynlighetsfordeling: $f(x)$ definert for alle reelle x ved $P(a < X < b) = \int_a^b f(x)dx$
	

Figur 2.3: Diskret VS kontinuerlig stokastisk variabel.

Eksempel: (økonomi - utfallsrom , diskret stokastisk variabel , fordeling)

Anta at aksjekursen til National Oilwell Varco, NOV, kun kan ha verdiene 95 NOK, 100 NOK eller 110 NOK. Aksjekursen er med andre ord et *reelt tall*. Derfor kan aksjekursen representeres ved en *tilveldig variabel* X . Utfallsrommet til denne tilveldige variabelen er:

$$\Omega = \{ 95 , 100 , 110 \} \quad (2.1)$$

Avhengig av mange faktorer som f.eks. driftsresultat, omsetning og forventet optimisme i markedet så har NOV vurdert at aksjeprisen for neste kalenderår vil være:

$$P(X = 95) = 0.20 \quad (= P(x_1)) \quad (2.2)$$

$$P(X = 100) = 0.70 \quad (= P(x_2)) \quad (2.3)$$

$$\underbrace{P(X = 110)}_{\text{stor } X \text{ for selve variabelen}} = 0.10 \quad (= \underbrace{P(x_3)}_{\text{liten } x \text{ for verdien}}) \quad (2.4)$$

som selvsagt oppfyller $\sum_{i=1}^3 P(x_i) = 1$, (jfr. lign.(1.29))².

Lag et stolpediagram av $P(X = x)$.

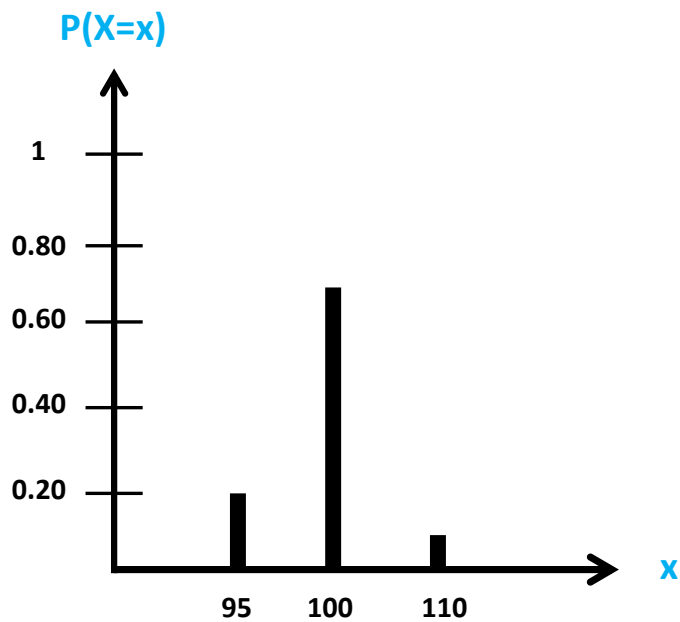


Figur 2.4: National Oilwell Varco, NOV.

² $x_1 = 95, x_2 = 100$ og $x_3 = 110$

Løsning:

Sannsynlighetene i lign.(2.2)-(2.4) kan plottes i et stolpediagram:



Figur 2.5: Sannsynlighetsfordeling $P(X = x)$.

Kommentarer:

- Fordelingen i figur (2.5) er *asymmetrisk* om $X = 100$, dvs. “skjev”.
- En sannsynlighet er som en “**vekt**”. Verdien $x_1 = 95$ har “**vekten**” $P(X = 95) = 0.20$.

■

Definisjon: ([sannsynlighetsfordeling](#), diskret)

En [sannsynlighetsfordeling](#) $P(x_i)$ til en diskret stokastisk variabel X er en funksjon definert ved

$$\underbrace{P(x_i)}_{\text{liten } x \text{ for } \underline{\text{verdier}}} = \underbrace{P(X = x_i)}_{\text{stor } X \text{ for selve } \underline{\text{variabelen}}} \quad (2.5)$$

hvor

$$i = 1, 2, 3, \dots \quad (2.6)$$

$$P(x_i) = \text{sannsynligheten for at } X \text{ har verdien } x_i \quad (2.7)$$

slik at $\sum_i P(x_i) = 1$.

■

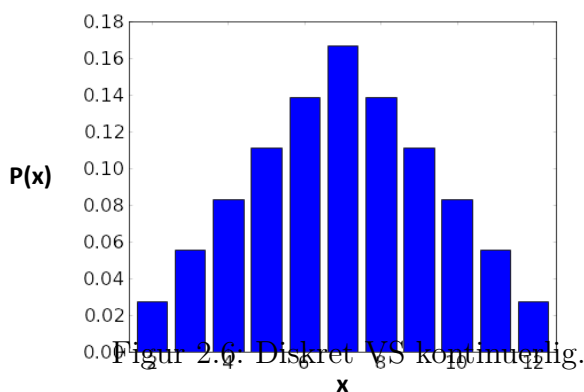
Definisjon: ([sannsynlighetsfordeling](#), kontinuert)

En [sannsynlighetstetthet](#) $f(x)$ til en kontinuert stokastisk variabel X er en funksjon

$$\underbrace{f(x)}_{\text{liten } x \text{ for } \underline{\text{verdier}}} \quad (2.8)$$

slik at $\int_{-\infty}^{\infty} f(x) dx = 1$.

■



Eksempel: (økonomi - utfallsrom , diskret stokastisk variabel , fordeling)

La oss nå forsette med forrige eksempel.

La oss se på sannsynligheten for at den tilfeldige variabelen X har verdi *mindre eller lik* 95.

Fra lign.(2.2) ser vi at:

$$\underline{\underline{P(X \leq 95)}} = P(X = 95) = \underline{\underline{0.20}} \quad (2.9)$$

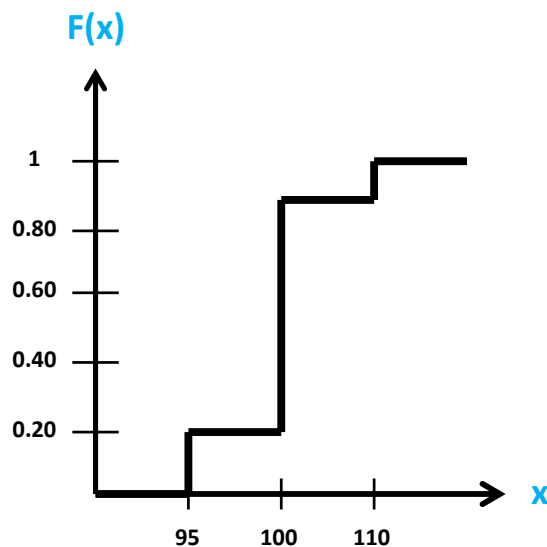
Tilsvarende: sannsynligheten at den tilfeldige variabelen X har verdi *mindre eller lik* 100 er (se lign.(2.2) og (2.3))

$$\underline{\underline{P(X \leq 100)}} = P(X = 95) + P(X = 100) = 0.20 + 0.70 = \underline{\underline{0.90}} \quad (2.10)$$

Og til slutt, sannsynligheten at den tilfeldige variabelen X har verdi *mindre eller lik* 110 er

$$\begin{aligned} \underline{\underline{P(X \leq 110)}} &= P(X = 95) + P(X = 100) + P(X = 110) \\ &= 0.20 + 0.70 + 0.10 = \underline{\underline{1}} \end{aligned} \quad (2.11)$$

$P(X \leq x)$ kalles **kumulativ** sannsynlighetsfordeling. Den kan illustreres slik:



Figur 2.7: **Kumulativ** sannsynlighetsfordeling $F(x) \equiv P(X \leq x)$.

■

Definisjon: ([kumulativ](#) sannsynlighetsfordeling)

Den *kumulative* sannsynlighetsfordeling F til en diskret stokastisk variabel X er definert ved

$$F(x) = P(X \leq x) \quad (2.12)$$

■

Dersom fordelingsfunksjonen $P(X = x)$ er kjent, slik som i eksemplet ovenfor, så er

$$F(x) = \sum_{x_i \leq x} P(X = x_i) \quad (2.13)$$

hvor $P(X = x_i)$ er en diskret sannsynlighetsfordeling.

2.2 Forventning og varians

2.2.1 Forventning

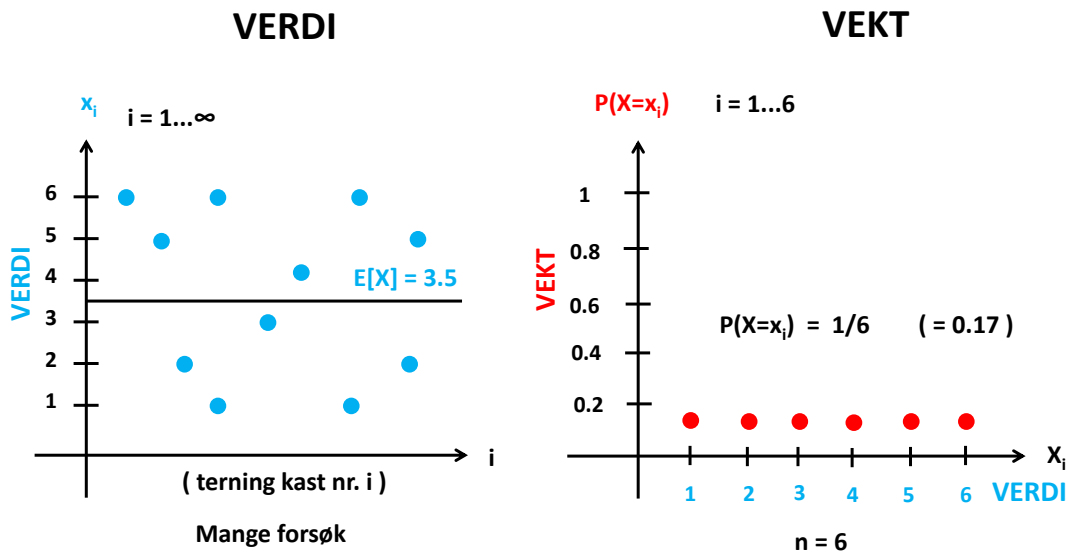
Forventningsverdi:

- et mål for “**tyngdepunktet**”
- “gjennomsnittet” i det lange løp
- må ikke nødvendigvis være lik en mulig verdi av den stokastiske variabelen
- forventning er *ikke* en sannsynlighet
- forventningsverdi er heller nødvendigvis *ikke* den verdien som er mest sannsynlig

Eksempel: (**terningkast** - diskret stokastisk variabel , verdi , vekt , forventningsverdi)

Stokastisk variabel: $(\Omega = \overbrace{\{ 1, 2, 3, 4, 5, 6 \}}^{\text{verdi}})$

$$X = \text{antall øyne i terningen} \quad (2.14)$$



Figur 2.8: Verdi og vekt.

Forventet antall øyne i et teringkast:

$$\underline{\underline{E[X]}} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \underline{\underline{3.5}} \quad (2.15)$$

Forventningen $E[X]$ er på formen:

$$E[X] = \sum_i \underbrace{\text{verdi}_i}_{x_i} \cdot \underbrace{\text{vekt}_i}_{P(X=x_i)} \quad (2.16)$$

■

Definisjon: (forventningsverdi, diskret)

For en *diskret* tilfeldig variabel X med de mulige verdiene $x_1, x_2, \dots, x_n$ er forventningsverdien:

$$E[X] \stackrel{\text{def.}}{=} \sum_{i=1}^n x_i P(X = x_i) \quad (2.17)$$

hvor ³

$$n = \text{antall utfall i utfallsrommet} \quad (2.18)$$

$$P(X = x_i) = \text{punktsannsynligheten} \quad (2.19)$$

■

Definisjon: (forventningsverdi, kontinuert)

For en *kontinuert* tilfeldig variabel X har forventningsverdien:

$$E[X] \stackrel{\text{def.}}{=} \int_{-\infty}^{\infty} x f(x) dx \quad (2.20)$$

hvor

$$f(x) = \text{sannsynlighetstetthet} \quad (2.21)$$

■

³For en terning er $n = 6$.

2.2.2 Varians

Varians:

- et mål for “spredning”
- et mål for [avviket](#) fra forventningsverdien
- gjennomsnittlig [avvik](#) fra forventningsverdien

Eksempel: ([terningkast](#) - diskret stokastisk variabel , verdi , vekt , varians , standardavvik)

Stokastisk variabel:

$$X = \text{antall \u00f8yne i terningen} \quad (2.22)$$

dvs. utfallsrommet er: $\Omega = \{ 1 , 2 , 3 , 4 , 5 , 6 \}$.

Varians $Var[X]$ i antall \u00f8yne i et teringkast:

$$\begin{aligned} \underline{\underline{Var[X]}} &= (1 - 3.5)^2 \cdot \frac{1}{6} + (2 - 3.5)^2 \cdot \frac{1}{6} + (3 - 3.5)^2 \cdot \frac{1}{6} \\ &\quad + (4 - 3.5)^2 \cdot \frac{1}{6} + (5 - 3.5)^2 \cdot \frac{1}{6} + (6 - 3.5)^2 \cdot \frac{1}{6} \end{aligned} \quad (2.23)$$

$$= \underline{\underline{2.92}} \quad (\text{vanskelig \u00e5 tolke}) \quad (2.24)$$

Standardavvik $\sigma[X]$ i antall \u00f8yne i et teringkast:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{2.92} = \underline{\underline{1.71}} \quad (\text{lett \u00e5 tolke}) \quad (2.25)$$

Varians $Var[X]$ er på formen:

$$Var[X] = \sum_i \underbrace{avvik_i^2}_{x_i} \cdot \underbrace{vekt_i}_{P(X=x_i)} \quad (2.26)$$

■

Definisjon: ([varians](#))

For en stokastisk variabel X er variansen

$$Var[X] \stackrel{\text{def.}}{=} E[(X - E[X])^2] \quad (2.27)$$

altså

$$Var[X] = \text{forventet avvik} \quad (2.28)$$

■

Variansen er altså [forventningsverdien](#) til avviket $(X - E[X])^2$.

Dermed kan vi bruke definisjonen av forventningsverdien til å skrive ut lign.(2.27) eksplisitt:

Diskret:

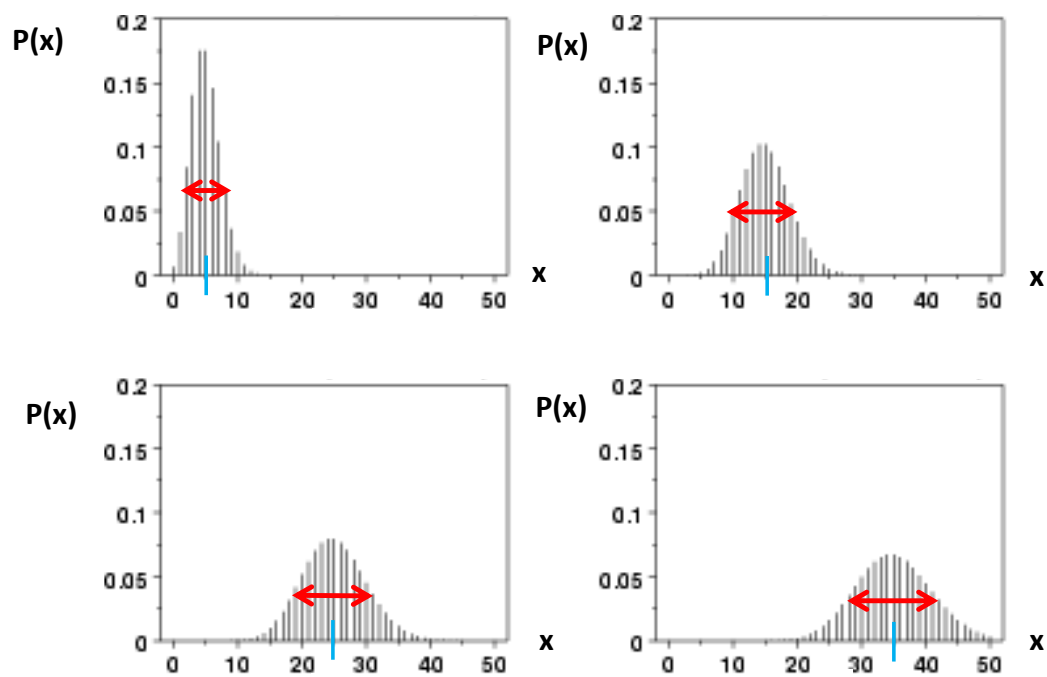
$$Var[X] \stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^n (x_i - E[X])^2 P(X = x_i) \quad (2.29)$$

Kontinuerlig:

$$Var[X] \stackrel{\text{lign. (2.20)}}{=} \int_{-\infty}^{\infty} (x_i - E[X])^2 f(x) \quad (2.30)$$

Diskrete stokastisk variabler med sannsynlighetsfordeling $P(x)$ som har forskjellige forventninger $E[X]$ og forskjellige varianser $Var[X]$ er vist i figur (2.9):

"tyngdepunkt"
"spredning"



Figur 2.9: Diskret $p(x) = P(X = x)$. Forskjellig $E[X]$ og forskjellig $Var[X]$.

Setning: ([varians](#) , “varianssetningen”)

La X være en stokastisk variabel. For variansen gjelder da:

$$\boxed{Var[X] = E[X^2] - E[X]^2} \quad (2.31)$$

■

Bevis:

$$\underline{\underline{Var[X]}} \stackrel{\text{linj. (2.27)}}{=} E\left[(X - E[X])^2\right] \quad (2.32)$$

$$= E\left[X^2 - 2X E[X] + (E[X])^2\right] \quad (2.33)$$

$$= E[X^2] - 2E[X] E[X] + (E[X])^2 \quad (2.34)$$

$$= \underline{\underline{E[X^2] - (E[X])^2}} \quad (2.35)$$

Diskret:

$$Var[X] = \sum_{i=1}^m x_i^2 P(X = x_i) - \left(\sum_{i=1}^n x_i P(X = x_i) \right)^2 \quad (2.36)$$

Kontinuerlig:

$$Var[X] = \int_{-\infty}^{\infty} x_i^2 f(x) - \left(\int_{-\infty}^{\infty} x_i f(x) \right)^2 \quad (2.37)$$

Definisjon: ([standardavvik](#), diskret)

For en stokastisk variabel X er standardavviket

$$\sigma[X] \stackrel{\text{def.}}{=} \sqrt{\text{Var}[X]} \quad (2.38)$$

■

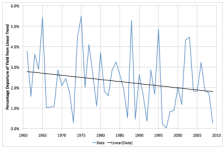
Hvorfor standardavvik?

Standardavvik er [lettere å tolke](#) enn varians, se eksempel side 106.

2.2.3 Kovarians

Kovarians:

- et mål for “samvariasjon / korrelasjon”
- et mål for **avviket** mellom X og $E[X]$ og Y og $E[Y]$
- gjennomsnittlig **avvik** fra forventningsverdien til forskjellige variabler

Var[X]	Cov[X,Y]
Definisjon: $\text{Var}[X] = E[(X - E[X])^2]$	Definisjon: $\text{Cov}[X,Y] = E[(X - E[X])(Y - E[Y])]$
Mål på: • spredning • avvik mellom (X og $E[X]$)	Mål på: • samvariasjon / korrelasjon • avvik mellom (X og $E[X]$) og (Y og $E[Y]$)
Var[X] utskrevet: $\text{Var}[X] = \sum_{i=1}^n (x_i - E[X])^2 P(X=x_i)$	Cov[X,Y] utskrevet: $\text{Cov}[X,Y] = \sum_{i=1}^n \sum_{j=1}^m (x_i - E[X]) (y_j - E[Y]) p(x_i, y_j)$
	

Figur 2.10: $\text{Var}[X]$ VS $\text{Cov}[X, Y]$.

Definisjon: (kovarians)

La X og Y være to stokastiske variabler. Med $\overbrace{\text{kovariansen}}^{\text{samvariasjon/korrelasjon}}$ mellom disse mener vi:

$$\overbrace{Cov[X, Y]}^{\text{samvariasjon/korrelasjon}} = E[(X - E[X])(Y - E[Y])] \quad (2.39)$$

■

Kovariansen er et mål på grad av samvariasjon/korrelasjon:

- $\overbrace{Cov[X, Y] < 0}^{\text{negativ}}$: betyr at store X gir små Y , $\overbrace{X \text{ og } Y \text{ varierer i "motsatt"}}^{\text{"motfase"}}$
- $\overbrace{Cov[X, Y] > 0}^{\text{positiv}}$: betyr at store X gir store Y , $\overbrace{X \text{ og } Y \text{ varierer i "motsatt"}}^{\text{"fase"}}$



Figur 2.11: Samvariasjon.

Diskret:

$$Cov[X, Y] = \sum_{i=1}^n \sum_{j=1}^m \left(x_i - E[X] \right) \left(y_j - E[Y] \right) p(x_i, y_j) \quad (2.40)$$

Kontinuerlig:

$$Cov[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - E[X] \right) \left(y - E[Y] \right) p(x, y) \, dx \, dy \quad (2.41)$$

Definisjon: ([simultanfordeling](#))

La X og Y være to stokastiske variabler. Med [simultan](#)fordeling menes:

$$p(x, y) = P(X = x \text{ og } Y = y) \quad (2.42)$$

■

Simultanfordeling er som en “**og**”-sannsynlighet:

den sier noe om hvor stor sannsynlighet det er for at $X = x$ og **samtidig** $Y = y$.

Eksempel: (økonomi - simultan , marginal , forventning , uavhengighet)

Du jobber som aksjemegler ved Oslo Børs. Du og dine kollegaer ønsker å se nærmere på sammenhengen mellom aksjekurs (pris) og antall aksjer som omsettes per dag.

Siden du har spesialisert deg på offshorenæringen så er aksjekursen til Seadrill interessant for deg. Du bestemmer deg for å definere de stokastiske variablene X og Y :

- X = kursen (prisen) på en Seadrill-aksje en gitt dag
- Y = antall Seadrill-aksjer som omsettes per dag

Basert på erfaring har du funnet følgende $\overbrace{\text{simultanfordelingen}}^{\text{og-sannsynligheter}}$ for X og Y :

		# aksjer omsatt per dag	
		Y = 80 000	Y = 120 000
Kurs (NOK)	X = 70	1/12	3/12
	X = 100	1/6	1/6
	X = 130	3/12	1/12

Figur 2.12: Simultanfordelling (og-sannsynligheter) til X og Y .



Figur 2.13: Oslo Børs og Seadrill.

Regn ut kovariansen for de stokastiske variablene X og Y .

Vil store Y h re sammen med sm  Y ?
Eller omvendt?

Løsning:

$$\underline{\underline{Cov[X, Y]}} \stackrel{\text{lin. (2.39)}}{=} E \left[(X - E[X])(Y - E[Y]) \right] \quad (2.43)$$

$$= \sum_{i=1}^3 \sum_{j=1}^2 \left(x_i - E[X] \right) \left(y_j - E[Y] \right) p(x_i, y_j)$$

$$= \overbrace{(x_1 - E[X])}^{=70-100} \cdot \overbrace{(y_1 - E[Y])}^{=80\,000-100\,000} \cdot \overbrace{p(x_1, y_1)}^{=1/12}$$

$$+ \underbrace{(x_1 - E[X])}_{=70-100} \cdot \underbrace{(y_2 - E[Y])}_{=120\,000-100\,000} \cdot \underbrace{p(x_1, y_2)}_{=3/12}$$

$$+ \overbrace{(x_2 - E[X])}^{=100-100} \cdot \overbrace{(y_1 - E[Y])}^{=80\,000-100\,000} \cdot \overbrace{p(x_2, y_1)}^{=1/6}$$

$$+ \underbrace{(x_2 - E[X])}_{=100-100} \cdot \underbrace{(y_2 - E[Y])}_{=120\,000-100\,000} \cdot \underbrace{p(x_2, y_2)}_{=1/6}$$

$$+ \overbrace{(x_3 - E[X])}^{=130-100} \cdot \overbrace{(y_1 - E[Y])}^{=80\,000-100\,000} \cdot \overbrace{p(x_3, y_1)}^{=3/12}$$

$$+ \underbrace{(x_3 - E[Y])}_{=130-100} \cdot \underbrace{(y_2 - E[Y])}_{=120\,000-100\,000} \cdot \underbrace{p(x_3, y_2)}_{=1/12}$$

$$\begin{aligned} &= (-30) \cdot (-20\,000) \cdot \frac{1}{12} + (-30) \cdot 20\,000 \cdot \frac{3}{12} \\ &+ 0 + 0 \\ &+ 30 \cdot (-20\,000) \cdot \frac{3}{12} + 30 \cdot 20\,000 \cdot \frac{1}{12} \end{aligned} \quad (2.44)$$

$$= \underline{\underline{-200\,000}} \quad (2.45)$$

Konklusjon:

$\overbrace{Cov[X, Y]}^{\text{samvariasjon}} < 0$: , kovariansen er **negativ** og dermed:

- X og Y varierer **motsatt** av hverandre
- store X gir små Y
- lite antall aksjer omsatt er forbundet med høy kurs

■

Som vi ser:

Å bruke definisjonen for regne ut $Cov[X, Y]$ kan være **omstendelig**. Men det finnes en setning som forenkler regningen i en del tilfeller. Setningen er som følger:

Setning: (“kovarianssetningen”)

La X og Y være to stokastiske variabler. For $\overbrace{\text{kovariansen}}^{\text{samvariasjon/korrelasjon}}$ gjelder da:

$$\overbrace{Cov[X, Y]}^{\text{samvariasjon/korrelasjon}} = E[XY] - E[X]E[Y] \quad (2.46)$$

■

Bevis:

Vi hopper over beviset i dette kompendiet.

■



Figur 2.14: Samvariasjon.

Fra lign.(2.46) ser vi at dersom X og Y er uavhengige, dvs.:

$$\overbrace{E[X \cdot Y]}^{X \text{ og } Y \text{ uavhengig}} = E[X] E[Y] \quad (2.47)$$

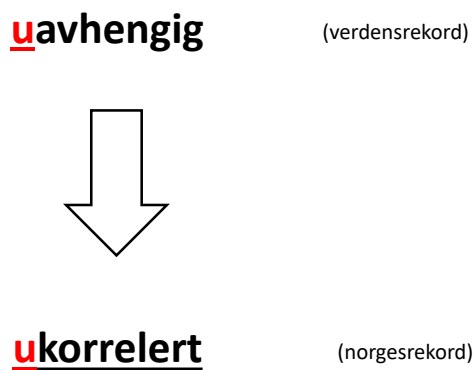
så er

$$\textcolor{red}{Cov}[X, Y] = \overbrace{E[X \cdot Y] - E[X]E[Y]}{= 0 \text{ dersom } X \text{ og } Y \text{ er uavh.}} \quad (2.48)$$

dvs.

$$\underbrace{\textcolor{red}{Cov}[X, Y] = 0}_{X \text{ og } Y \text{ u}korrelerte} \quad (2.49)$$

Konklusjon: Dersom X og Y er uavhengige så er de også ukorrelerte. ⁴



Figur 2.15: Uavhengighet er et **st**erkere krav enn ukorrelert.

⁴Uavhengighet er altså et **st**erkere krav enn ukorrelert.
Dersom en skøyteløper har verdenrekorden på 10 000 meter så har han/hun også norgesrekorden.

Setning ([spesialtilfelle](#) , [kovarians](#))

La X og Y være to stokastiske variabler. Dersom X og Y er uavhengige, dvs.:

$$\underbrace{E[X \cdot Y]}_{\text{uavhengig}} = E[X]E[Y] \quad (2.50)$$

så er ⁵

$$\underbrace{Cov[X, Y]}_{\text{ukorrelerte}} = 0 \quad (2.51)$$

■



Figur 2.16: Skøyteløper.

⁵Uavhengighet er altså et **sterkere** krav enn ukorrelert. Dersom en skøyteløper har verdenrekorden på 10 000 meter så har han/hun også norgesrekorden.

2.2.4 Noen regneregler

La a og b være konstanter. La videre X og Y være to stokastiske variabler. Da gjelder:

Regneregler for forventning:

$$E[a] = a \quad (2.52)$$

$$E[a + X] = a + E[X] \quad (2.53)$$

$$E[a \cdot X] = a E[X] \quad (2.54)$$

$$E[aX + bY] = aE[X] + bE[Y] \quad (2.55)$$

Regneregler for varians:

$$Var[a] = 0 \quad (2.56)$$

$$Var[a + X] = Var[X] \quad (2.57)$$

$$Var[a X] = a^2 Var[X] \quad (2.58)$$

$$Var[X] = Var[-X] \quad (2.59)$$

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2ab Cov[X, Y] \quad (2.60)$$

Dessuten er alltid $Var[X] \geq 0$, dvs. en varians kan aldri være negativ.

Eksempel: (økonomi - regneregler , forventning , diskret stokastisk variabel)

Spillerne i Kristiansund Ballklubb (KBK) ønsker å se nærmere på antall mål de scorer per kamp. I den sammenheng defineres den stokastiske variabelen:

$$X = \text{antall mål som KBK scorer i løpet av en tilfeldig valgt kamp} \quad (2.61)$$

Basert på historiske data har de funnet ut at X har fordelingen som vist i figur (2.17): ⁶

$X=x_i$	0	1	2	3	4	5
$P(X=x_i)$	0.20	0.30	0.20	0.15	0.10	0.05

Figur 2.17: Sannsynlighetsfordeling for $P(X = x_i)$.



Figur 2.18: Kristiansund Ballklubb.

⁶Ut fra denne sannsynlighetsfordelingen ser vi at KBK maksimalt scorer 5 mål per kamp. Denne begrensningen er kun introdusert for å unngå for mye repeterende regning.

- a) Finn forventet antall mål som scores av KBK i løpet av en tilfeldig valgt kamp, $E[X]$.

SpareBank 1 Nordvest er hovedsponsor for KBK. Sponsoravtalen er prestasjonsrettet på en slik måte at de får **bonus** B bestemt av antall mål scoret i løpet av en gitt kamp:

$$B = cX^2 + d \quad (2.62)$$

hvor $c = 750$ NOK og $d = 1500$ NOK.⁷

- b) Forklar hvorfor B er en stokastisk variabel.

- c) Hva er forventet bonus for en tilfeldig valgt kamp, $E[B]$?

⁷Dette betyr at c og d er **konstanter**, mens X er en stokastisk **variabel**.

Løsning:

a) Forventet antall mål scoret per kamp:

$$\underline{\underline{E[X]}} = \sum_{i=0}^5 x_i P(X = X_i) \quad (2.63)$$

$$= 0 \cdot 0.20 + 1 \cdot 0.30 + 2 \cdot 0.20 + 3 \cdot 0.15 + 4 \cdot 0.10 + 5 \cdot 0.05 \quad (2.64)$$

$$= \underline{\underline{1.8}} \quad (2.65)$$

b) En funksjon av en tilfeldig variabel er bare en ny tilfeldig variabel. Derfor:
 $B = cX^2 + d$ er en stokastisk variabel fordi den er en funksjon av X , hvor X er en stokastisk variabel.

c) Forventet bonus per kamp:

$$\underline{\underline{E[B]}} = E[cX^2 + d] \quad (2.66)$$

$$\stackrel{\text{lign. (2.55)}}{=} \underline{\underline{cE[X^2] + d}} \quad (2.67)$$

Vi trenger altså $E[X^2]$ for å finne bonusen. Denne må vi regne ut:

$$\underline{\underline{E[X^2]}} = \sum_{i=0}^5 x_i^2 P(X = X_i) \quad (2.68)$$

$$= 0^2 \cdot 0.20 + 1^2 \cdot 0.30 + 2^2 \cdot 0.20 + 3^2 \cdot 0.15 + 4^2 \cdot 0.10 + 5^2 \cdot 0.05 \quad (2.69)$$

$$= \underline{\underline{5.3}} \quad (2.70)$$

Forventet bonus per kamp $E[B]$ blir derfor:

$$\underline{\underline{E[B]}} = cE[X^2] + d \quad (2.71)$$

$$= \left(750 \cdot 5.3 + 1\,500 \right) \text{ NOK} = \underline{\underline{5\,475 \text{ NOK}}} \quad (2.72)$$

■

Eksempel: (økonomi - forventning , varians, usikkerhet , regneregler)

Anta at du jobber i investeringsselskapet Gjelsten Holding som skal investere i tre ulike selskaper:

Statoil , Seadrill og Yara .
selskap A selskap B selskap C

Prisen på aksjene i dag er **100** NOK for selskap A , **105** NOK for selskap B og **130** NOK for selskap C . Du bestemmer deg for å kjøpe

$$N = 100\,000 \quad (2.73)$$

antall aksjer, men du er usikker på hvor mange aksjer du skal kjøpe i de respektive selskapene. Derfor definerer du konstantene a , b og c :

$$N \cdot a = \text{antall aksjer som investeres i selskap } A, \text{ Statoil} \quad (2.74)$$

$$N \cdot b = \text{antall aksjer som investeres i selskap } B, \text{ Seadrill} \quad (2.75)$$

$$N \cdot c = \text{antall aksjer som investeres i selskap } C, \text{ Yara} \quad (2.76)$$

hvor

$$a + b + c = 1 \quad (2.77)$$

dvs. a er den brøkdelen av de $N = 100\,000$ aksjene som investeres i selskap A . Tilsvarende for b og c .



Figur 2.19: Selskap A , B og C .

For å avgjøre hvor mange aksjer du skal kjøpe i de respektive selskapene så ønsker du å finne ut mer om den potensielle fortjenesten av aksjene ved et eventuelt salg **om ett år**.

Derfor defineres følgende stokastiske variabler:

X = pris på en aksje (altså aksjekurs) for selskap A , Statoil (i NOK), **om ett år** (2.78)

Y = pris på en aksje (altså aksjekurs) for selskap B , Seadrill (i NOK), **om ett år** (2.79)

Z = pris på en aksje (altså aksjekurs) for selskap C , Yara (i NOK), **om ett år** (2.80)

Anta videre at forventet pris (altså aksjekurs) **om ett år** er: ⁸

$$E[X] = 90 \quad , \quad E[Y] = 125 \quad , \quad E[Z] = 180 \quad (2.81)$$

med tilhørende varianser: ⁹

$$Var[X] = 100 \quad , \quad Var[Y] = 200 \quad , \quad Var[Z] = 600 \quad (2.82)$$



Figur 2.20: Gjelsten Holding.

⁸Forventningene er i NOK, men dropper benevnningen her for enkelhets skyld.

⁹Variansene er i NOK².

- a) Vis at fortjenesten F om ett år er gitt ved: ¹⁰

$$F = Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \quad (2.84)$$

- b) Vis at den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er: ¹¹

$$E[F] = 10N(-a + 2b + 5c) \quad (2.85)$$

- c) Anta nå at aksjene er uavhengige.
 Bruk lign.(2.84) samt lign.(2.82) og vis at variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år er:

$$Var[F] = 100N^2(a^2 + 2b^2 + 6c^2) \quad (2.86)$$

¹⁰Bruk gjerne at:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (2.83)$$

for hver av de tre selskapene. Bruk også lign.(2.74)-(2.76) samt lign.(2.78)-(2.80).

¹¹Bruk lign.(2.84).

- d) Du ønsker å minimere risikoen i din investering, dvs. du ønsker å minimere variansen $Var[F]$.
Man kan vise at variansen $Var[F]$ i lign.(2.86) er minst mulig når

$$a = 0.6 \quad , \quad b = 0.3 \quad , \quad c = 0.1 \quad (2.87)$$

(Du skal ikke vise lign.(2.87). Bare ta dette faktum for gitt.)

Hva blir fordelingen av antall aksjer for de tre selskapene dersom du ønsker å minimere risikoen i din investering? ¹²

- e) Dersom du bestemmer deg for å maksimere den forventede fortjenetsen $E[F]$ i lign.(2.85) istedet for å minimere risikoen $Var[F]$, hva slags selskap bør du investere i da?

Hva er verdien på den forventede fortjenesten $E[F]$ i det tilfellet?

¹²Bruk lign.(2.74)-(2.76) sammen med lign.(2.87).

Løsning:

a) Fortjenesten F er:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (2.88)$$

for de tre selskapene så får vi:

$$\begin{aligned} \underline{\underline{F}} &= \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ om ett år}}_{= X} \\ &+ \underbrace{\text{antall aksjer i } B}_{= N \cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ om ett år}}_{= Y} \\ &+ \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ om ett år}}_{= Z} \end{aligned} \quad (2.89)$$

$$- \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ i dag}}_{= 100} \quad (2.90)$$

$$- \underbrace{\text{antall aksjer i } B}_{= N \cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ i dag}}_{= 105}$$

$$- \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ i dag}}_{= 130} \quad (2.91)$$

$$= NaX + NbY + NcZ - Na100 + Nb105 + Nc130 \quad (2.92)$$

$$= \underline{\underline{Na(X - 100) + Nb(Y - 105) + Nc(Z - 130)}} \quad , \quad \text{q.e.d.} \quad (2.93)$$

b) Den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er:

$$\underline{E[F]} = E \left[Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \right] \quad (2.94)$$

$$= Na \left(\underbrace{E[X]}_{=90} - 100 \right) + Nb \left(\underbrace{E[Y]}_{=125} - 105 \right) + Nc \left(\underbrace{E[Z]}_{=180} - 130 \right) \quad (2.95)$$

$$= Na(90 - 100) + Nb(125 - 105) + Nc(180 - 130) \quad (2.96)$$

$$= Na(-10) + Nb 20 + Nc 50 \quad (2.97)$$

$$= \underline{\underline{10N(-a + 2b + 5c)}} \quad , \quad \text{q.e.d.} \quad (2.98)$$

c) Fra lign.(2.60) vet vi formelen for variansen av en lineær kombinasjon av to stokastiske variabler X og Y , nemlig:

$$\overbrace{Var[aX + bY]}^{\text{variasjon/(spredning)}} = a^2 Var[X] + b^2 Var[Y] + 2ab \overbrace{Cov[X, Y]}^{\text{samvariasjon}} \quad (2.99)$$

Dersom X og Y er uavhengige så er $Cov[X, Y] = 0$.

For vårt tilfelle er de stokastiske variablene uavhengige. Da er også $Cov[X, Y, Z] = 0$. Variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år blir da:

$$\underline{\underline{Var[F]}} = Var \left[\textcolor{blue}{N}a(X - 100) + \textcolor{blue}{N}b(Y - 105) + \textcolor{blue}{N}c(Z - 130) \right] \quad (2.100)$$

$$\begin{aligned} &= (Na)^2 \underbrace{Var[X]}_{= 100} + (Nb)^2 \underbrace{Var[Y]}_{= 200} + (Nc)^2 \underbrace{Var[Z]}_{= 600} \\ &- (Na)^2 \underbrace{Var[100]}_{= 0} + (Nb)^2 \underbrace{Var[105]}_{= 0} + (Nc)^2 \underbrace{Var[130]}_{= 0} \end{aligned} \quad (2.101)$$

$$= N^2 (100a^2 + 200b^2 + 600c^2) \quad (2.102)$$

$$= \underline{\underline{100N^2 (a^2 + 2b^2 + 6c^2)}} \quad , \quad \text{q.e.d.} \quad (2.103)$$

- d)** Dersom man ønsker minst mulig risiko i investeringen så tilsvarer det at variansen $Var[F]$ er minst mulig. I oppgaven er det oppgitt at denne variansen er minst når $a = 0.6$, $b = 0.3$ og $c = 0.1$.

Antall aksjer som må kjøpes i de forskjellige selskapene for å oppnå dette er da:

$$\underline{\underline{Na}} = 100\,000 \cdot 0.6 = \underline{\underline{60\,000}} \quad (2.104)$$

$$\underline{\underline{Nb}} = 100\,000 \cdot 0.3 = \underline{\underline{30\,000}} \quad (2.105)$$

$$\underline{\underline{Nc}} = 100\,000 \cdot 0.1 = \underline{\underline{10\,000}} \quad (2.106)$$

- e) Fra oppgave **b** vet vi at den forventede fortjenesten $E[F]$ er

$$E[F] = 10N(-a + 2b + 5c) \quad (2.107)$$

Fra denne ligningen ser vi umiddelbart at $E[F]$ blir størst når vi kun investerer kun investerer i selskap C , dvs.

$$a = 0 \quad , \quad b = 0 \quad , \quad c = 1 \quad (2.108)$$

Verdien på den forventede fortjenesten $E[F]$ blir da:

$$\underline{\underline{E[F]}} = 10N \cdot 5c \quad (2.109)$$

$$= 10 \cdot 100\,000 \cdot 5 \cdot 1 \text{ NOK} = \underline{\underline{5\,000\,000 \text{ NOK}}} \quad (2.110)$$

altså 5 mill. NOK.



Definisjon: (uavhengighet) ¹³

La X og Y være to stokastiske variabler. Disse er uavhengige dersom:

$$P(X = x \text{ og } Y = y) = P(X = x) \cdot P(Y = y) \quad (2.111)$$

for alle $X = x$ og $Y = y$.



NB:

Denne definisjonen er helt *analog* med spesielle multiplikasjonssetningen i forbindelse med kapitlet om betinget sannsynlighet, se lign.(1.180).

- Her, i lign.(2.111), dreier deg seg uavhengighet mellom *stokastiske variabler* X og Y .
- I lign.(1.180), dvs. den spesielle multiplikasjonssetningen $P(A \cap B) = P(A) \cdot P(B)$, dreier deg seg uavhengighet mellom *begivenheter* A og B .

¹³Jamfør den *analoge* setn. for uavhengighet mellom begivenheter A og B i lign.(1.180): $P(A \cap B) = P(A) \cdot P(B)$.

Setning: (uavhengighet)

Dersom X og Y er to uavhengige stokastiske variabler så gjelder:

$$\boxed{E[X \cdot Y] = E[X] E[Y]} \quad (2.112)$$

■

Bevis:

Dersom de stokastiske variablene X og Y er uavhengige så er:

$$\overbrace{P(X = x_i \text{ og } Y = y_j)}^{\text{uavhengig}} \stackrel{\text{lign. (2.111)}}{=} \underbrace{P(X = x_i)}_{= p(x_i)} \cdot \underbrace{P(Y = y_j)}_{= p(y_j)} \quad (2.113)$$

Ut fra definisjonen av forventning får vi da:

$$\underline{\underline{E[X \cdot Y]}} \stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^m \sum_{j=1}^n x_i \cdot y_i \cdot \textcolor{red}{p(x_i, y_j)} \quad (2.114)$$

$$\stackrel{\text{uavh.}}{=} \sum_{i=1}^m \sum_{j=1}^n x_i \cdot y_i \cdot \textcolor{red}{p(x_i)} \cdot \textcolor{red}{p(y_j)} \quad (2.115)$$

$$= \underbrace{\left(\sum_{i=1}^m x_i \cdot p(x_i) \right)}_{= E[X]} \underbrace{\left(\sum_{j=1}^n y_j \cdot p(y_j) \right)}_{= E[Y]} \quad (2.116)$$

$$= \underline{\underline{E[X] E[Y]}} \quad (2.117)$$

■

Eksempel: (økonomi - simultan , marginal , forventning , uavhengighet)

La oss igjen se på eksemplet fra side 117:

- X = kursen (prisen) på en Seadrill-aksje en gitt dag
- Y = antall Seadrill-aksjer som omsettes per dag

Basert på erfaring har du funnet følgende **simultan**fordelingen for X og Y :

		# aksjer omsatt per dag		P(X=x)	
		Y = 80 000	Y = 120 000		
Kurs (NOK)	X = 70	1/12	3/12	1/3	P(X=70)
	X = 100	1/6	1/6	1/3	P(X=100)
	X = 130	3/12	1/12	1/3	P(X=130)
P(Y=y)		1/2	1/2		
		P(Y=80 000)	P(Y=120 000)		

Figur 2.21: **Simultan**fordelingen $p(x, y)$ og **marginmal**fordelingene $P(X = x)$ og $P(Y = y)$.

Er aksjekurs (pris) og antall aksjer som omsettes per dag uavhengig?



Figur 2.22: Oslo Børs og Seadrill.

Løsning:

Forventet aksjekurs (pris) på Seadrill-aksjene:

$$\begin{aligned}\underline{E[X]} &\stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^3 x_i \cdot P(X = x_i) \\ &= \left(70 \cdot \frac{1}{3} + 100 \cdot \frac{1}{3} + 130 \cdot \frac{1}{3} \right) \text{ NOK} = \underline{100 \text{ NOK}}\end{aligned}\quad (2.118)$$

Forventet antall aksjoner som omsettes:

$$\begin{aligned}\underline{E[Y]} &\stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^2 y_i \cdot P(Y = y_i) \\ &= \textcolor{red}{80\,000} \cdot \frac{1}{2} + \textcolor{blue}{120\,000} \cdot \frac{1}{2} = \underline{100\,000}\end{aligned}\quad (2.119)$$

Siden omsetning = pris $\times$ antall aksjer omsatt = $X \cdot Y$, så er forventet omsetning:

$$\underline{E[X \cdot Y]} = \sum_{i=1}^3 \sum_{j=1}^2 x_i \cdot y_j \cdot p(x_i, y_j) \quad (2.120)$$

$$\begin{aligned}&= \left(\textcolor{red}{70} \cdot 80\,000 \cdot \frac{\textcolor{red}{1}}{\textcolor{red}{12}} + \textcolor{blue}{100} \cdot 80\,000 \cdot \frac{\textcolor{blue}{1}}{\textcolor{blue}{6}} + \textcolor{green}{130} \cdot 80\,000 \cdot \frac{\textcolor{green}{3}}{\textcolor{green}{12}} \right. \\ &+ \left. \textcolor{red}{70} \cdot 120\,000 \cdot \frac{\textcolor{red}{3}}{\textcolor{red}{12}} + \textcolor{blue}{100} \cdot 120\,000 \cdot \frac{\textcolor{blue}{1}}{\textcolor{blue}{6}} + \textcolor{green}{130} \cdot 120\,000 \cdot \frac{\textcolor{green}{1}}{\textcolor{green}{12}} \right) \text{ NOK} \\ &= \underline{9\,800\,000 \text{ NOK}}\end{aligned}\quad (2.121)$$

Fra lign.(2.118) og (2.119) ser vi at $E[X] \cdot E[Y] = 10\,000\,000$, dvs.:

$$E[X \cdot Y] \neq E[X] \cdot E[Y] \quad (2.122)$$

Konklusjon: X og Y er avhengige.



Eksempel: (økonomi - kovarians)

La oss atter en gang se på eksemplet fra side 117.
Ved å bruke definisjonen av

$$\begin{aligned}\underline{\underline{Cov[X, Y]}} &\stackrel{\text{lign. (2.39)}}{=} E\left[(X - E[X])(Y - E[Y])\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m (x_i - E[X])(y_j - E[Y]) p(x_i, y_j)\end{aligned}\tag{2.123}$$

så innså vi, se side 119, at det var mye arbeid.

Regn ut kovariansen $Cov[X, Y]$ **uten** å bruke definisjonen i lign.(2.123).



Figur 2.23: Oslo Børs og Seadrill.

Løsning:

Siden vi kjenner $E[X \cdot Y]$ så vel som $E[X]$ og $E[Y]$, se side 140, så får vi:

$$Cov[X, Y] \stackrel{\text{lign. (2.46)}}{=} \underbrace{E[X \cdot Y]}_{= 9\,800\,000} - \underbrace{E[X]}_{= 100\,000} \cdot \underbrace{E[Y]}_{= 100} = \underline{\underline{-200\,000}}\tag{2.124}$$

Med andre ord:

svært arbeidsbesparende i forhold å bruke definisjonen.



Regnereglen i lign.(2.60) på side 124 er så viktig at vi tar nærmere titt på den her.

Setning: (kovarians II)

La X og Y være to stokastiske variabler. Generelt gjelder da følgende sammenhengen mellom variansen og kovariansen:

$$\overbrace{Var[aX + bY]}^{\text{variasjon/(spredning)}} = a^2 Var[X] + b^2 Var[Y] + 2ab \cdot \overbrace{Cov[X, Y]}^{\text{samvariasjon}} \quad (2.125)$$

hvor a og b er konstanter.

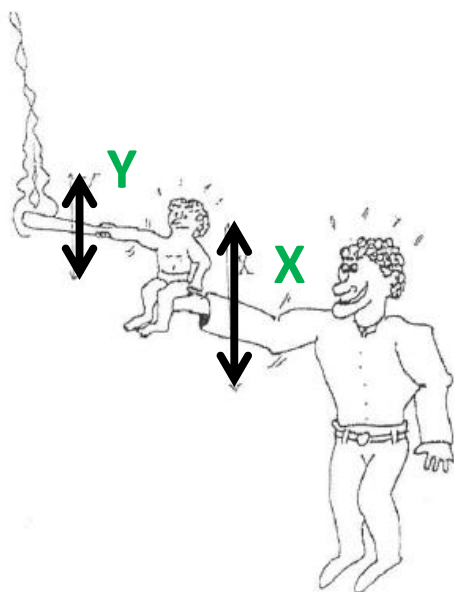
■

Bevis:

Lign.(2.125) bevises ved å bruke første kvadratsetning, (se f.eks. læreboken).

■

Visualisering av variasjon og samvariasjon:



Figur 2.24: **V**ariasjon og **s**am**v**ariasjon, dvs. $Var[X + Y]$ og $Cov[X, Y]$.

La:

- X = guttens plassering sett fra mannens ståsted, se figur (2.24)
- Y = fakkelens plassering sett fra guttens ståsted, se figur (2.24)
- $X + Y$ = fakkelens plassering i forhold til mannens ståsted (bakken)

Fra lign.(2.125): ($a = 1$ og $b = 1$)

$$Var[1X + 1Y] \stackrel{\text{lign. (2.125)}}{=} 1^2 Var[X] + 1^2 Var[Y] + 2 \cdot 1 \cdot 1 Cov[X, Y] \quad (2.126)$$

Da har vi:

- $Cov[X, Y] > 0$: armene i bevegges i **samme** retning $\Rightarrow$ *stor* samvariasjon
- $Cov[X, Y] < 0$: armene i bevegges i **motsatt** retning $\Rightarrow$ *liten* samvariasjon
- $Cov[X, Y] = 0$: **ingen sammenheng** mellom guttens og mannens armbevegelser

Setning ([spesialtilfelle](#) , [kovarians II](#))

La X og Y være to stokastiske variabler. Dersom X og Y er [ukorrelererte](#), dvs.:

$$\overbrace{Cov[X, Y]}^{X \text{ og } Y \text{ ukorrelererte}} = 0 \quad (2.127)$$

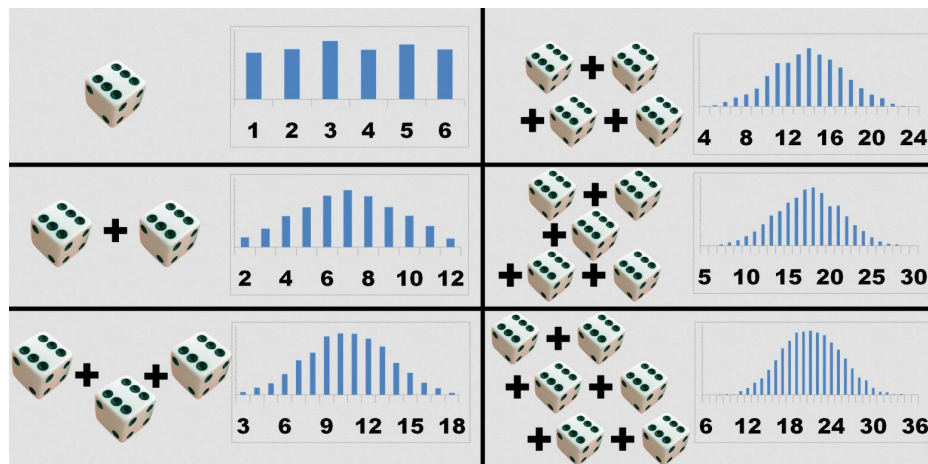
så er:

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] \quad (2.128)$$

■

Kapittel 3

Diskrete stokastiske fordelinger

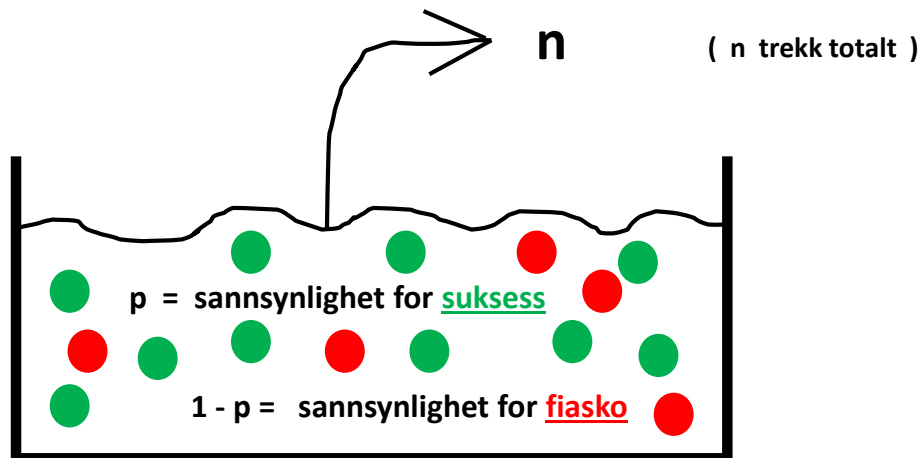


Figur 3.1: Sentralgrensesetningen (“CLT”).

3.1 Den binomiske fordelingen

Binomisk forsøksserie:

1. Hvert forsøk skal ha 2 mulige utfall, s (suksess) eller f (fiasko).
2. Det skal være samme sannsynlighet p for suksess i alle n forsøkene.
3. Alle forsøk er uavhengige.
4. Vi gjennomfører et bestemt antall forsøk, n .



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for **suksess**: p

Figur 3.2: Binomisk forsøksserie.

Utfallet er enten $\overbrace{\text{suksess}}^s$ eller $\overbrace{\text{fiasko}}^f$, dvs.

$$\boxed{P(s) + P(f) = 1} \quad (3.1)$$

Siden

$$P(s) = p \quad (s = \text{suksess}) \quad (3.2)$$

så følger umiddelbart at

$$P(f) = 1 - p \quad (f = \text{fiasko}) \quad (3.3)$$

Eksempel: (binomisk forsøksserie , *gitt* sekvens)

La oss se på en forsøksserie som består av $n = 5$ forsøk.

La oss videre anta at disse forsøkene også oppfyller de 4 kriteriene til **binomisk** forsøksserie.

Hva er sannsynligheten $P(ssf sf)$ for 3 suksesser av totalt $n = 5$ forsøk?

Løsning:

Siden alle $n = 5$ forsøkene er per antagelse uavhengige, så kan vi benytte oss av den spesielle multiplikasjonssetningen, se lign.(1.180).

Dermed:

$$\underline{\underline{P(ssf sf)}} = P(s) \cdot P(s) \cdot P(f) \cdot P(s) \cdot P(f) \quad (3.4)$$

$$= p \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) = \underline{\underline{p^3 \cdot (1 - p)^2}} \quad (3.5)$$

som er sannsynligheten for at den *gitte* sekvensen $ssf sf$ skal inntreffe.



Merk:

Sannsynligheten i lign.(3.5) er sannsynligheten for en *gitt* sekvens.

Ut fra dette eksemplet er det lett å generalisere: For en *gitt sekvens* med et gitt antall suksesser n_{suksess} så blir dermed sannsynligheten:

$$P(n_{\text{suksess}} \text{ suksesser av totalt } n \text{ forsøk for } \textit{gitt sekvens}) = p^{n_{\text{suksess}}} \cdot (1 - p)^{n - n_{\text{suksess}}} \quad (3.6)$$

hvor

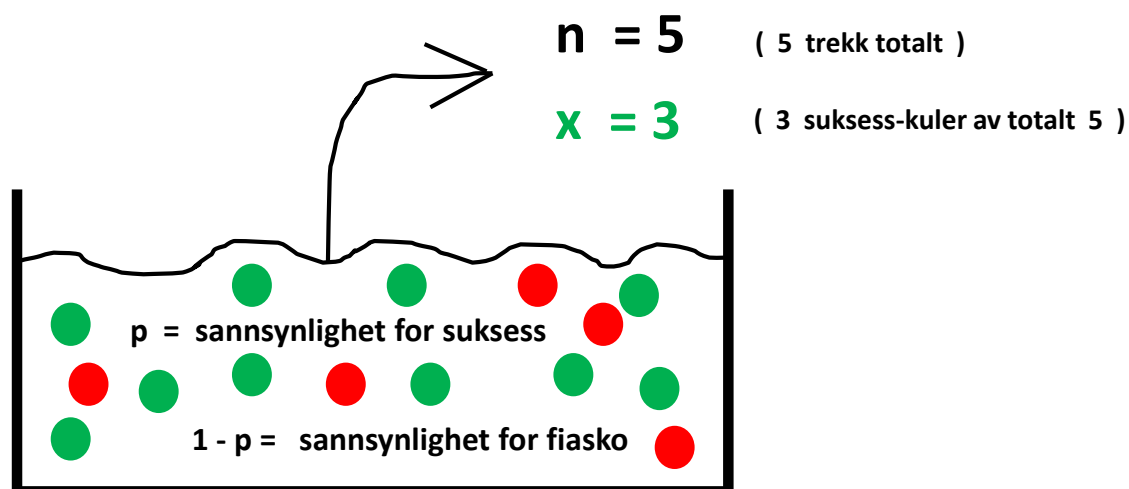
$$n_{\text{suksess}} = \text{antall suksesser} \quad (3.7)$$

$$p = \text{sannsynligheten for suksess for ett forsøk} \quad (3.8)$$

Eksempel: (^{diskret} binomisk forsøksserie , 3 suksesser)

Vi gjør igjen $n = 5$ forsøk. Hva er sannsynligheten for $x = 3$ suksesser, dvs. hva er

$$P(3 \text{ suksesser av totalt } 5 \text{ forsøk}) = ? \quad (3.9)$$



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: p

Figur 3.3: $n = 5$ binomiske forsøk.

Løsning:

Sannsynligheten for 3 suksesser for en (hvilken som helst) gitt sekvens er gitt ved lign. (3.5);

$$P(\text{gitt mulig sekvens}) = p^3 \cdot (1 - p)^2 \quad (3.10)$$

Men 3 suksesser kan fås på flere forskjellige måter:

$$\underbrace{\text{sssf}, \text{ssfs}, \text{ssff}, \text{sfssf}, \text{sfsfs}, \text{sffss}, \text{fsssf}, \text{fssfs}, \text{fsfss}, \text{ffss}}_{= \text{totalt 10 mulige måter/sekvenser å få 3 suksesser på}} \quad (3.11)$$

Alle disse sekvensene har **samme sannsynlighet** . Dermed kan vi bruke “**urnemodellen**” for å telle opp antall mulige måter å få 3 suksesser på (av totalt 5 forsøk). ¹

Egenskapene til forsøksserien i lign.(3.11) er:

- siden sannsynligheten p er uendret etter trekking så tilsvarende et valg **uten** tilbakelegging
- rekkefølge har **ikke betydning**, dvs. *ikke*-ordnet utvalg

Dette er derfor et *ikke*-ordnet utvalg *uten* tilbakelegging, dvs. **situasjon 3** i kap. (B.2). Dermed gjelder lign.(B.16):

$$\# \text{ ikke-ordnede komb. u/tilbakelegging} = \frac{5!}{(5-3)! 3!} \equiv \underbrace{\binom{5}{3}}_{\text{binomialkoeff.}} = 10 \quad (3.12)$$

¹ “**Urnemodellen**” ble introdusert i vedlegg B.4.

hvor

$$n = 5 \quad (\text{antall valgobjekter/trekk}) \quad (3.13)$$

$$x = 3 \quad (\text{antall objekter som velges}) \quad (3.14)$$

Dermed, siden hver sekvens svarer til enkeltutfall ², så gjelder den **spesielle addisjonssetningen** fra lign.(1.34):

$$\underline{\underline{P(3 suksesser av totalt 5 forsøk)}} \stackrel{\text{lign. (1.34)}}{=} P(\text{sssff} \cup \text{ssfsf} \cup \dots \cup \text{ffss}) \quad (3.15)$$

$$\stackrel{\text{spes. add.}}{=} P(\text{sssff}) + P(\text{ssfsf}) + \dots + P(\text{ffss}) \quad (3.16)$$

$$= p^3 \cdot (1-p)^2 + p^3 \cdot (1-p)^2 + \dots + p^3 \cdot (1-p)^2 \quad (3.17)$$

$$= 10 \cdot p^3 \cdot (1-p)^2 \quad (3.18)$$

$$= \begin{aligned} &\text{antall måter å få en gitt mulig sekvens på} \\ &\times \\ &P(\text{gitt mulig sekvens}) \end{aligned} \quad (3.19)$$

$$= \underline{\underline{\binom{5}{3} \cdot p^3 \cdot (1-p)^2}} \quad (3.20)$$

■

Dette eksemplet kan vi lett generalisere:

²Dvs. begivenhetene/utfallene er **disjunkte** (ingen felles elementer)

Definisjon: ($\overbrace{\text{binomisk}}^{\text{diskret}}$ fordeling , $X \sim \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}}$) ³

Punktsannsynlighetene for en binomisk fordeling er:

$$P(X = x) \stackrel{\text{def.}}{=} \binom{n}{x} p^x (1 - p)^{n-x} \quad (3.21)$$

hvor

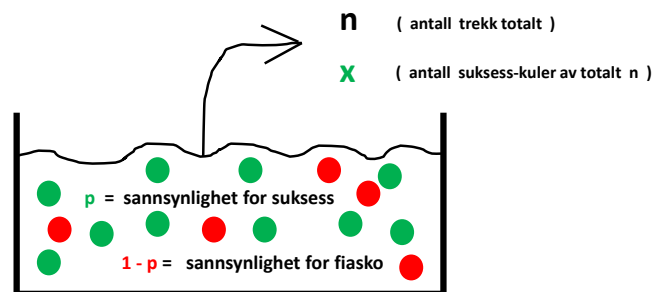
$$\begin{aligned} X &= \text{stokastisk variabel} \\ &= \text{antall "suksesser" i en binomisk forsøksserie på totalt } n \text{ forsøk} \end{aligned} \quad (3.22)$$

$$p = \text{sannsynlighet for "suksess"} \quad (3.23)$$

$$n = \text{totalt antall forsøk} \quad (3.24)$$

$$\underbrace{\binom{n}{x}}_{\text{binomialkoeff.}} = \frac{n!}{(n-x)! x!} \quad ("n \text{ over } x") \quad (3.25)$$

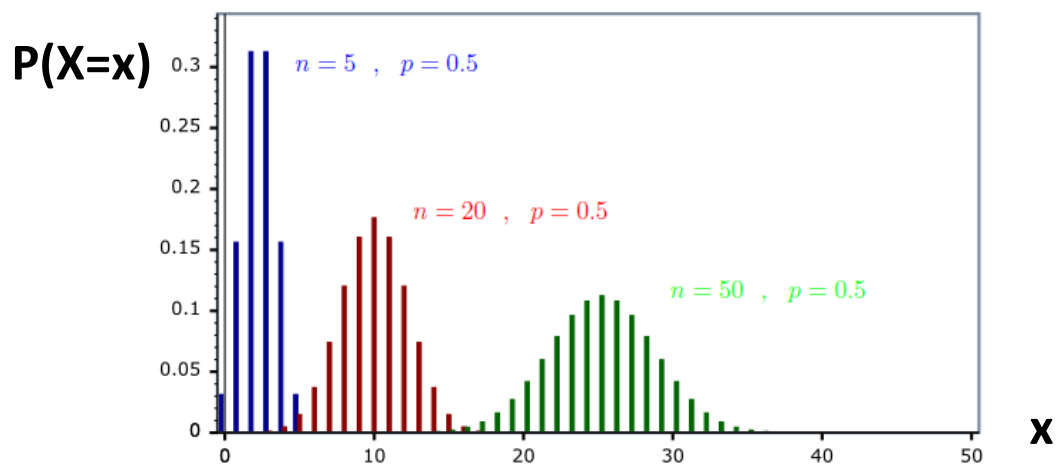
■



- 1) To typer kuler: suksess-kuler
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: p

Figur 3.4: Binomisk forsøksserie.

³Skrivemåten $X \sim \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}}$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er binomisk fordelt.



Figur 3.5: **Binomiske** sannsynlighetsfordelinger $P(X = x)$.

Eksempel: ($\overbrace{\text{binomisk}}^{\text{diskret}}$ forsøksserie , Sport Management)

La oss se på en fotballspiller som prøver å score mål i løpet av en fotballkamp. Spilleren har en målscoreingssannsynlighet på 25 % for hvert skuddforsøk mot mål. Sannsynligheten antas altså å være den samme for hvert skudd, uansett om det foregående skuddet gikk i mål eller ikke. Alle skuddforsøkene er med andre ord uavhengige. Vi antar en spilleren får 4 slike skuddsjanser i løpet av en kamp. Vi definerer den stokastiske variabelen

$$X = \text{antall mål som scores av spilleren i løpet av en kamp} \\ \text{dersom hun får 4 skuddsjanser i løpet av kampen} \quad (3.26)$$

- a) Hva er sannsynligheten for at han scorer 0 mål i løpet av en $\overbrace{4\text{-sjanserskamp}}^{\text{!}}$, dvs. $P(X = 0)$?
- b) Hva er $P(X = 1)$?
- c) Hva er $P(X = 2)$?
- d) Hva er $P(X = 3)$?
- e) Hva er $P(X = 4)$?



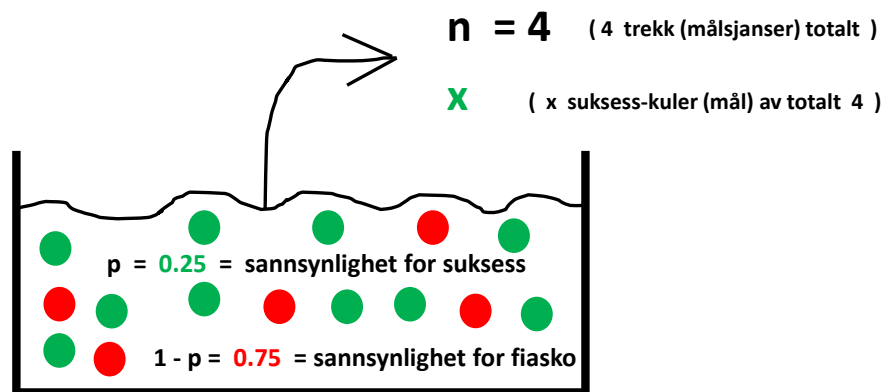
Figur 3.6: Skudd på mål.

Løsning:

Dette forsøket med skudd på mål har følgende egenskaper:

1. Hvert forsøk har bare 2 mulige utfall,
 - $p = 0.25$, (et gitt skudd går i mål, “suksess”)
 - $1 - p = 1 - 0.25 = 0.75$, (et gitt skudd går *ikke* i mål, “fiasko”).
2. Det er **samme sannsynlighet** $p = 0.25$ for mål (“suksess”) i alle $n = 4$ forsøkene.
3. Alle $n = 4$ forsøkene er **uavhengige**.
4. Vi gjennomfører et bestemt antall forsøk, $n = 4$.

Dermed er dette en **binomiske** fordeling, dvs. vi kan bruke formel i lign.(3.21).



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: **$p = 0.25$**

Figur 3.7: $n = 4$ binomiske forsøk.

- a) Sannsynlighet for at spilleren scorer 0 mål i løpet av en 4-sjanserskamp:

$$\underline{\underline{P(X=0)}} = \binom{n}{0} p^0 (1-p)^{n-0} = \binom{4}{0} 0.25^0 (1-0.25)^{4-0} = \underline{\underline{0.32}} \quad (3.27)$$

- b) Sannsynlighet for at spilleren scorer 1 mål i løpet av en 4-sjanserskamp:

$$\underline{\underline{P(X=1)}} = \binom{n}{1} p^1 (1-p)^{n-1} = \binom{4}{1} 0.25^1 (1-0.25)^{4-1} = \underline{\underline{0.42}} \quad (3.28)$$

- c) Sannsynlighet for at spilleren scorer 2 mål i løpet av en 4-sjanserskamp:

$$\underline{\underline{P(X=2)}} = \binom{n}{2} p^2 (1-p)^{n-2} = \binom{4}{2} 0.25^2 (1-0.25)^{4-2} = \underline{\underline{0.21}} \quad (3.29)$$

- d) Sannsynlighet for at spilleren scorer 3 mål i løpet av en 4-sjanserskamp:

$$\underline{\underline{P(X=3)}} = \binom{n}{3} p^3 (1-p)^{n-3} = \binom{4}{3} 0.25^3 (1-0.25)^{4-3} = \underline{\underline{0.05}} \quad (3.30)$$

- e) Sannsynlighet for at spilleren scorer 4 mål i løpet av en 4-sjanserskamp:

$$\underline{\underline{P(X=4)}} = \binom{n}{4} p^4 (1-p)^{n-4} = \binom{4}{4} 0.25^4 (1-0.25)^{4-4} = \underline{\underline{0.004}} \quad (3.31)$$

■

Eksempel: (revisjon I , ^{diskret} binomisk forsøksserie)

Vi gjør et tilfeldig utvalg med 150 bilag fra en revisjon. Dette tilfeldige utvalget på 150 er slik at sannsynligheten for at et *gitt* bilag er feil, er 2 %, dvs. $p = 0.02$. La videre

$$X = \text{antall bilag med feil i utvalget på 150} \quad (3.32)$$

- a) Hva er sannsynligheten for at **ingen** av de 150 bilagene har noen feil ⁴, dvs. $P(X = 0)$?
- b) Hva er sannsynligheten for at kun **ett** av de 150 bilagene inneholder feil, dvs. $P(X = 1)$?
- c) Hva er sannsynligheten for at **mindre enn 2** av de 150 bilagene har feil, dvs. $P(X \leq 1)$?
- d) Hva er sannsynligheten for at **minst 2** av de 150 bilagene har feil, dvs. $P(X \geq 2)$?



Figur 3.8: Revisjon I.

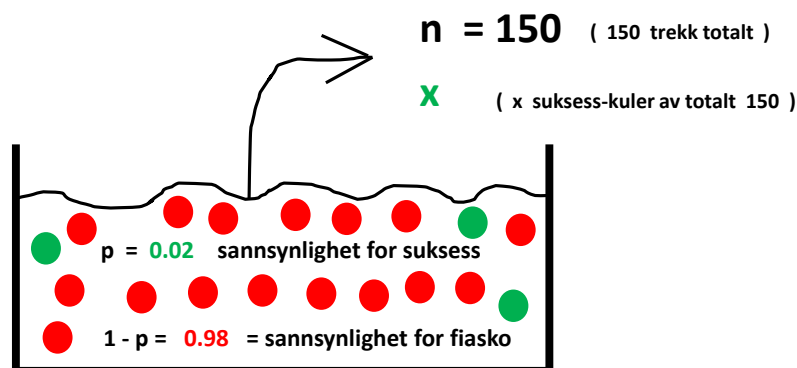
⁴Alternativt og ekvivalent spørsmål: Hva er sannsynligheten for at alle 150 bilagene feilfrie, dvs. $P(X = 0)$?

Løsning:

Dette forsøket med et tilfeldig utvalg med 150 bilag fra en revisjon har følgende egenskaper:

1. Hvert forsøk har bare 2 mulige utfall,
 - $p = 0.02$, (feil i bilag, “suksess”) ⁵
 - $1 - p = 1 - 0.02 = 0.98$, (ingen feil i bilag, “fiasko”).
2. Det er samme sannsynlighet $p = 0.02$ for feil i bilag (“suksess”) i alle $n = 150$ forsøkene.
3. Alle $n = 150$ forsøkene er uavhengige.
4. Vi gjennomfører et bestemt antall forsøk, $n = 150$.

Dermed er dette en binomisk fordeling, dvs. vi kan bruke formel i lign.(3.21).



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: $p = 0.02$

Figur 3.9: $n = 150$ binomiske forsøk.

⁵“Suksess” og “fiasko” er bare **betegnelser/merkelapper**. Disse merkelappene kan vi velge som vi vil. Det kan virke litt rart at “suksess” assosieres med feil i bilag, men det er altså bare merkelapper. “Suksess” behøver ikke nødvendigvis å være noe positivt.

a) Sannsynlighet for at **ingen** av de **150** bilagene har noen feil, dvs. alle er feilfrie:

$$\underline{\underline{P(X=0)}} = \binom{n}{0} p^0 (1-p)^{n-0} = \binom{150}{0} 0.02^0 (1-0.02)^{150} = \underline{\underline{0.0483}} \quad (3.33)$$

b) Sannsynlighet for at kun **ett** bilag av de **150** har feil:

$$\underline{\underline{P(X=1)}} = \binom{n}{1} p^1 (1-p)^{n-1} = \binom{150}{1} 0.02^1 (1-0.02)^{150-1} = \underline{\underline{0.1478}} \quad (3.34)$$

c) Sannsynlighet for at **mindre enn 2** bilag av de **150** har feil: ⁶

$$\underline{\underline{P(X \leq 1)}} = \underbrace{P(X=0) + P(X=1)}_{\text{mindre enn 2}} = 0.0483 + 0.1478 = \underline{\underline{0.1961}} \quad (3.35)$$

d) Sannsynlighet for at **minst 2** bilag av de **150** har feil:

$$\underline{\underline{P(X \geq 2)}} \stackrel{\text{kompl. setn.}}{=} 1 - P(X \leq 1) = 1 - 0.1961 = \underline{\underline{0.8039}} \quad (3.36)$$

■

⁶Siden alle forsøk er uavhengige så kan vi addere sannsynlighetene, jfr. den **spesielle** **addisjonssetningen** i lign.(1.34).

Kommentar:

Binomiske forsøk benyttes i omtrent alle sammenhenger der vi ^{binomisk} teller opp antall “suksesser” i en forsøksrekke. Her er noen eksempler:

- antall suksesser i n uavhengige forsøk
- antall “ja” i en spørreundersøkelse
- antall bilag med feil i en revisjon
- antall defekte enheter i et vareparti
- antall oppmøtte studenter til en forelesning
- antall mål som scores i en fotballkamp

3.1.1 Forventingsverdi

Definerer følgende stokastiske variabel:

$$X = \text{antall observerte utfall av typen "suksess"} \quad (3.37)$$

dvs. X = antall suksesser. La oss nå introdusere en stokastisk variabel med kun to utfall:^{7 8}

$$Y_i = \begin{cases} 1 & , \text{ dersom "suksess" for forsøk nr. } i \\ 0 & , \text{ dersom "fiasko" for forsøk nr. } i \end{cases} \quad (3.38)$$

Denne binære variabelen har altså kun to mulige utfall, 0 eller 1. Dermed innser vi at

$$X = Y_1 + Y_2 + \dots + Y_n \quad (3.39)$$

Forventingen $E[Y_i]$ er dermed⁹

$$\underline{E[Y_i]} = \overbrace{p(\text{"sukcess"}) \cdot \text{verdi for "sukcesss"} + p(\text{"fiasko"}) \cdot \text{verdi for "fiasko"}}^{\text{forventning}} \quad (3.40)$$

$$= p \cdot 1 + (1 - p) \cdot 0 = \underline{p} \quad (3.41)$$

Generelt for stokastiske variabler så gjelder setningen i lign.(2.55), dvs. $E[X + Y] = E[X] + E[Y]$. Dette gjelder uansett om X og Y er uavhengige eller ikke. Forventingen av variabelen X er dermed:

$$\underline{\underline{E[X]}} = E[Y_1 + Y_2 + \dots + Y_n] \stackrel{\text{lign. (2.55)}}{=} \overbrace{\underbrace{E[Y_1]}_{=p} + \underbrace{E[Y_2]}_{=p} + \dots + \underbrace{E[Y_n]}_{=p}}^{= n \text{ stk. ledd}} = \underline{\underline{n \cdot p}} \quad (3.42)$$

Dette er et viktig resultat. La oss derfor formulere dette i en egen setning.

⁷«Ja-nei»-variabel, også kalt en stokastisk binær variabel.

⁸For de av dere som har emnet SCM200 så kjenner dere igjen denne type variabler som «ja-nei»-beslutninger.

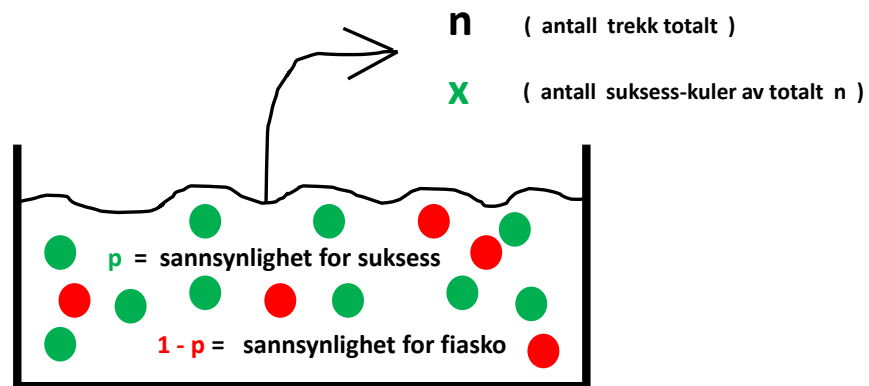
⁹Fra kap.(2) og lign.(2.17) vet vi at forventning er definert ved: $E[X] = \sum_{i=1}^m x_i \cdot P(X = x_i)$.

Setning: (forventing av $X \sim \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}}$)

La X være en binomisk variabel, dvs. $X \sim \text{Bin}[n, p]$. Da gjelder:

$$E[X] = n \cdot p \quad (3.43)$$

■



- 1) To typer kuler: suksess-kuler
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: p

Figur 3.10: Binomisk forsøksserie.

3.1.2 Varians

Tilsvarende som i forrige avsnitt så regne ut variansen for en gitt **hjelpevariabel** H_i :

$$Var[H_i] = E[H_i^2] - E[H_i]^2 \quad (\text{jfr. lign. (2.31)}) \quad (3.44)$$

Dermed ser vi at vi behøver

$$\underline{E[H_i^2]} = \overbrace{p(\text{“sukess”}) \cdot (\text{verdi for “sukess”})^2 + p(\text{“fiasko”}) \cdot (\text{verdi for “fiasko”})^2}^{\text{forventning}} \quad (3.45)$$

$$= p \cdot 1^2 + (1-p) \cdot 0^2 = \underline{p} \quad (3.46)$$

Videre behøver vi $E[H_i]^2$. Men vi regnet ut $E[H_i] = p$ i forrige avsnitt. Dermed har vi alt vi trenger for å regne ut variansen i lign. (3.44):

$$\underline{Var[H_i]} = E[H_i^2] - E[H_i]^2 = p - p^2 = \underline{p(1-p)} \quad (3.47)$$

Nå kommer et viktig poeng:

I en binomisk fordeling er alle forsøkene **uavhengige**. Dermed gjelder setningen i lign. (2.128), dvs. **$Var[Y+Z] = Var[Y] + Var[Z]$** . Her **må** Y og Z være uavhengige for at dette skal gjelde. Variansen av variabelen X er dermed:

$$\underline{\underline{Var[X]}} = Var[H_1 + H_2 + \dots + H_n] = \overbrace{Var[H_1] + Var[H_2] + \dots + Var[H_n]}^{= n \text{ stk. ledd}} \quad (3.48)$$

$$= \underline{\underline{n \cdot p(1-p)}} \quad (3.49)$$

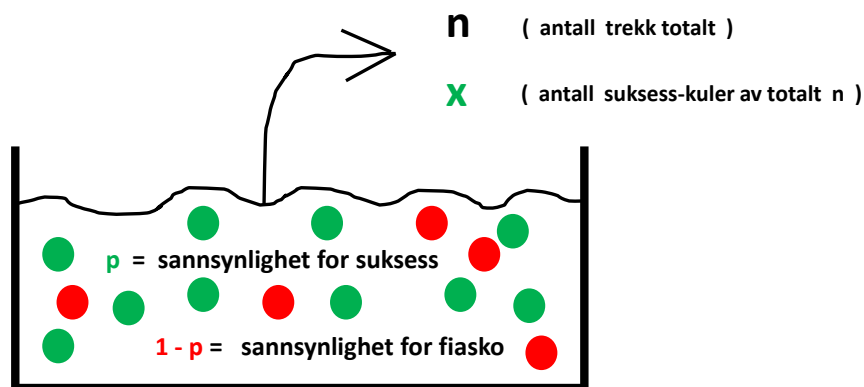
Dette er også et viktig resultat. Vi formulerer derfor dette i en egen setning.

Setning: (varians av $X \sim \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}}$ ¹⁰)

La X være en binomisk variabel, dvs. $X \sim \text{Bin}[n, p]$. Da gjelder:

$$\text{Var}[X] = n \cdot p(1 - p) \quad (3.50)$$

■



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling,
men vi kjenner sannsynligheten for suksess: **p**

Figur 3.11: Binomisk forsøksserie.

¹⁰Skrivemåten $X \sim \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}}$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er **binomisk** fordelt.

Kommentar:

Husk sammenhengen mellom standardavvik og varians:

$$\sigma[X] \stackrel{\text{def.}}{=} \sqrt{\text{Var}[X]} \quad (3.51)$$

Dermed gjelder, for en binomisk variabel $X \sim \text{Bin}[n, p]$, at $\sigma[X] \stackrel{\text{Bin.}}{=} \sqrt{n \cdot p(1 - p)}$.

■

Eksempel: (revisjon II , ^{diskret}binomisk forsøksserie)

Anta at vi har $n = 7600$ regnskap med bilag. La

$$X = \text{antall bilag med feil i disse 7600 regnskapene} \quad (3.52)$$

Anta videre at sannsynligheten for at et regnskap har et bilag med feil er $p = 5\%$.

- a) Hva slags fordeling har X ?
- b) Hva er forventningen $E[X]$?
- c) Hva er standardavviket $\sigma[X]$?



Figur 3.12: Revisjon II.

Løsning:

- a) La oss betegne s = suksess = feil i bilag, og f = fiasko = ingen feil i bilag ¹¹. Dette er en situasjon hvor vi **teller opp antall “suksesser”**. Vår serie med forsøk **oppfyller alle 4 punktene** i kapittel 3.1. Dermed er dette en

binomisk forsøksserie

- b) Siden X er binomisk fordelt, dvs. $X \sim \text{Bin}[n, p]$, så er forventningen gitt ved lign.(3.43):

$$\underline{\underline{E[X]}} \stackrel{\text{Bin.}}{=} n \cdot p = 7600 \cdot 0.05 = \underline{\underline{380}} \quad (3.53)$$

- c) Siden X er binomisk fordelt, dvs. $X \sim \text{Bin}[n, p]$, så er standardavviket gitt ved kvadratroten av lign.(3.50):

$$\underline{\underline{\sigma[X]}} \stackrel{\text{def.}}{=} \sqrt{\text{Var}[X]} \stackrel{\text{Bin.}}{=} \sqrt{n \cdot p (1 - p)} \quad (3.54)$$

$$= \sqrt{7600 \cdot 0.05 \cdot (1 - 0.05)} = \sqrt{361} = \underline{\underline{19}} \quad (3.55)$$

■

¹¹ “Suksess” og “fiasko” er bare **betegnelser/merkelapper**. Disse merkelappene kan vi velge som vi vil. Det kan virke litt rart at “suksess” assosieres med feil i bilag, men det er altså bare merkelapper. “Suksess” behøver ikke nødvendigvis å være noe positivt.

Merk:

Noen ganger oppgis **forventning** og **standardavvik** på følgende måte:

$$\underline{\underline{E[X] \pm \sigma[X]}} = \text{tyngdepunkt} \pm \text{spredning} = \underline{\underline{380 \pm 19}} \quad (3.56)$$

■

3.2 Den hypergeometriske fordelingen

Vi skiller mellom to situasjoner: ukjent og kjent fordeling mellom kulene i urnen.

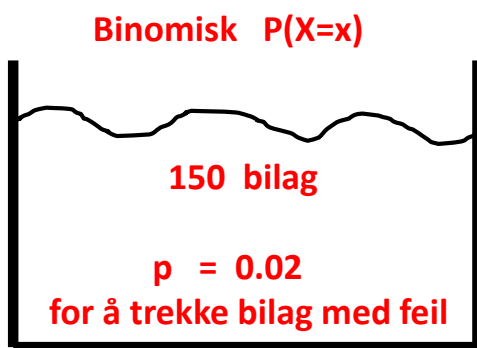
- 1) $\overbrace{\text{Ukjent fordeling mellom kulene i urnen:}}^{\text{Binomisk } P(X=x)}$

La oss igjen se på eksempel 4 på side 160.

- Vi vet **ikke** den eksakte fordelingen mellom antall bilag med feil og antall bilag uten feil
- Vi kjenner derimot sannsynligheten for at et tilfeldig trukket bilag inneholder feil, $p = 0.02$

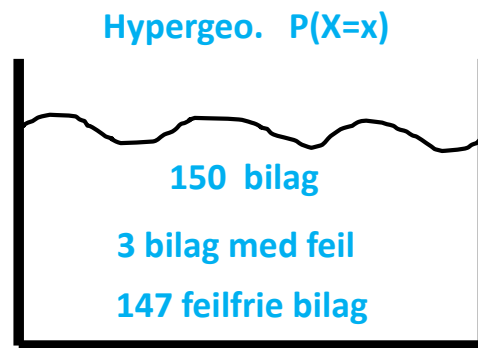
- 2) $\overbrace{\text{Kjent fordeling mellom kulene i urnen:}}^{\text{Hypergeo. } P(X=x)}$

- Vi **vet** den eksakte fordelingen mellom antall bilag med feil og antall bilag uten feil
 - 3 bilag med feil
 - 147 bilag som er feilfrie



Ukjent fordeling på kulene

(samme sannsynlighet "p")

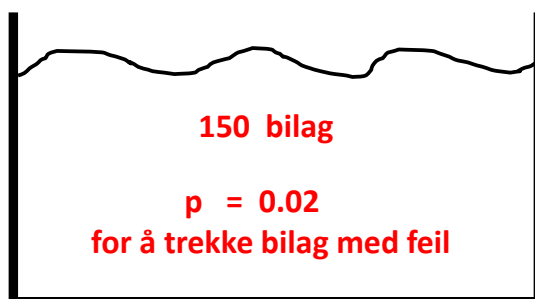


Kjent fordeling på kulene

(forskjellige sannsynligheter)

Figur 3.13: Ukjent og kjent fordeling mellom kulene i urnen.

- 1) $\overbrace{\text{Ukjent fordeling mellom kulene i urnen:}}^{\text{Binomisk } P(X=x)} \quad (p_1 = p_2 = p_3 \dots \equiv p)$



Figur 3.14: “Urne” med 150 bilag.

Vi skal nå argumentere for at enhver trekning har **alltid samme sannsynlighet** $p = 0.02$ til å inneholde et bilag med feil. Uansett resultatet av foregående trekninger:

Dersom vi trekker et tilfeldig bilag fra den **røde** urnen så er sannsynligheten

$$p_1 = 0.02 \quad (3.57)$$

for at dette tilfeldig trukne bilaget inneholder feil. Dersom vi nå gjør en ny trekning, dvs. trekning nr. 2 fra den **røde** urnen, så er sannsynligheten fortsatt

$$p_2 = 0.02 \quad (3.58)$$

for at det 2. tilfeldig trukne bilaget inneholder feil. Uansett hva utfallet av 1. trekning ble. Dersom vi trekker en 3. gang, så er også sannsynligheten fortsatt

$$p_3 = 0.02 \quad (3.59)$$

for at det 3. tilfeldig trukne bilaget inneholder feil. Uansett hva utfallet av 1. trekning og 2. trekning ble. Med andre ord, binomial-betingelsen

$$\text{“det skal være } \underline{\text{samme}} \text{ sannsynlighet } p \text{ for suksess i alle } n \text{ forsøkene”} \quad (3.60)$$

er oppfylt. Dette er altså den situasjonen vi så på i eksempel 4 side 160 i dette kapitlet, dvs. **binomisk** forsøksserie.

2) Kjent fordeling: $(p_1 \leq p_2 \leq p_3 \dots)$



Figur 3.15: Urne med 150 bilag.

La oss igjen se på 150 bilag fra en revisjon. Anta denne gangen at vi *vet* at det er 3 bilag som inneholder feil. Disse bilagene samles i en **blå** urne, se figur (3.15). Det betyr at dersom vi trekker et tilfeldig bilag fra den **blå** urnen så er sannsynligheten

$$p_1 = \frac{3}{150} \quad (= 0.02) \quad (3.61)$$

for at dette tilfeldig trukne bilaget inneholder feil. Altså samme sannsynlighet som for den røde urnen. La oss nå gjør en ny trekning, dvs. trekning nr. 2 fra den **blå** urnen. Vi legger ikke bilaget vårt tilbake i urnen (!). For trekning nr. 2 så gjelder følgende sannsynlighet:

$$p_2 = \begin{cases} \frac{2}{149} & , \text{ dersom bilaget i 1. trekning inneholder feil} \\ \frac{3}{149} & , \text{ dersom bilaget i 1. trekning ikke inneholder feil} \end{cases} \quad (3.62)$$

dvs. sannsynligheten for trekning nr. 2 er **avhengig** av utfallet i 1. trekning. Forutsendingene for en binomisk forsøksserie er dermed ikke oppfylt.

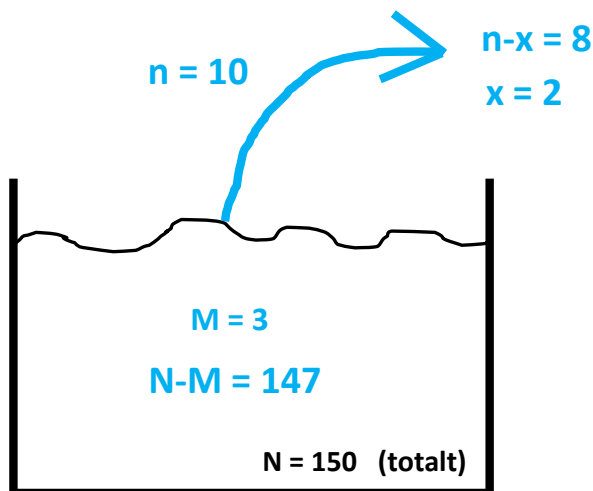
Denne situasjonen illustrerer forutsetningene for en **hypergeometrisk** forsøksserie.

Eksempel: (revisjon III , $\overbrace{\text{hypergeometrisk}}^{\text{diskret}}$ forsøksserie)

La oss atter en gang se på et tilfeldig utvalg med 150 bilag fra en revisjon¹².

Av de $N = 150$ bilagene *vet* vi at $M = 3$ bilag inneholder feil, se figur (3.16). La oss anta at vi gjør en forsøksserie med $n = 10$ trekninger fra vår urne. La oss definere den stokastiske variabelen

$$X = \text{antall bilag av de } n = 10 \text{ trukne bilagene som inneholder feil} \quad (3.63)$$



Figur 3.16: Trekker $n = 10$ bilag.

Hva er sannsynligheten for at $x = 2$ av disse $n = 10$ trukne bilagene inneholder feil, dvs. hva er $P(X = 2)$?

¹²La oss se på situasjonen [ii](#)) fra side 173 og 174.

Løsning:

Sannsynligheten er den samme for å trekke de forskjellige bilagene. Dette er derfor en tellesituasjon som kan beskrives med **urnemodellen**. Fra vedlegg B.4 vet vi da at det er en sammenheng mellom sannsynlighet og antall kombinasjoner:

$$P(X = 2) = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner for å oppnå } x = 2}{\text{antall } \textit{mulige} \text{ kombinasjoner totalt}} \quad (3.64)$$

Vi legger ikke bilagene/kulene tilbake i urnen. Rekkefølgen betyr heller ikke noe. Derfor derier dette seg om:

situasjon 3

Det betyr at vi skal bruke binomialkoeffisienten, se vedlegg B.4, for å finne # gunstige og # mulige. Dermed:

$$\underline{\underline{P(X = 2)}} = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}} = \frac{\overbrace{\binom{3}{2}}^{\text{2 med feil}} \cdot \overbrace{\binom{147}{8}}^{\text{8 feilfrie}}}{\binom{150}{10}} = \frac{63}{\underline{\underline{5513}}} \quad (\approx 1.14\%) \quad (3.65)$$

■

Dette eksemplet kan vi lett generalisere:

Definisjon: ($\overbrace{\text{hypergeometrisk}}^{\text{diskret}}$ fordeling , $X \sim \overbrace{\text{Hyp}[N, M, n]}^{3 \text{ param.}}$ ¹³)

Punktsannsynlighetene for en hypergeometrisk fordeling er

$$P(X = x) \stackrel{\text{def.}}{=} \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad (3.66)$$

hvor

$$\begin{aligned} X &= \text{stokastisk variabel} \\ &= \text{antall “spesielle” elementer i det tilfeldige utvalget p\aa} \\ &\quad n \text{ trukne elementer} \end{aligned} \quad (3.67)$$

$$N = \text{antall elementer i grunnmengden} \quad (3.68)$$

$$M = \text{antall “spesielle” elementer} \quad (3.69)$$

$$N - M = \text{antall “vanlige” elementer} \quad (3.70)$$

$$n = \text{antall trukne elementer} \quad (3.71)$$

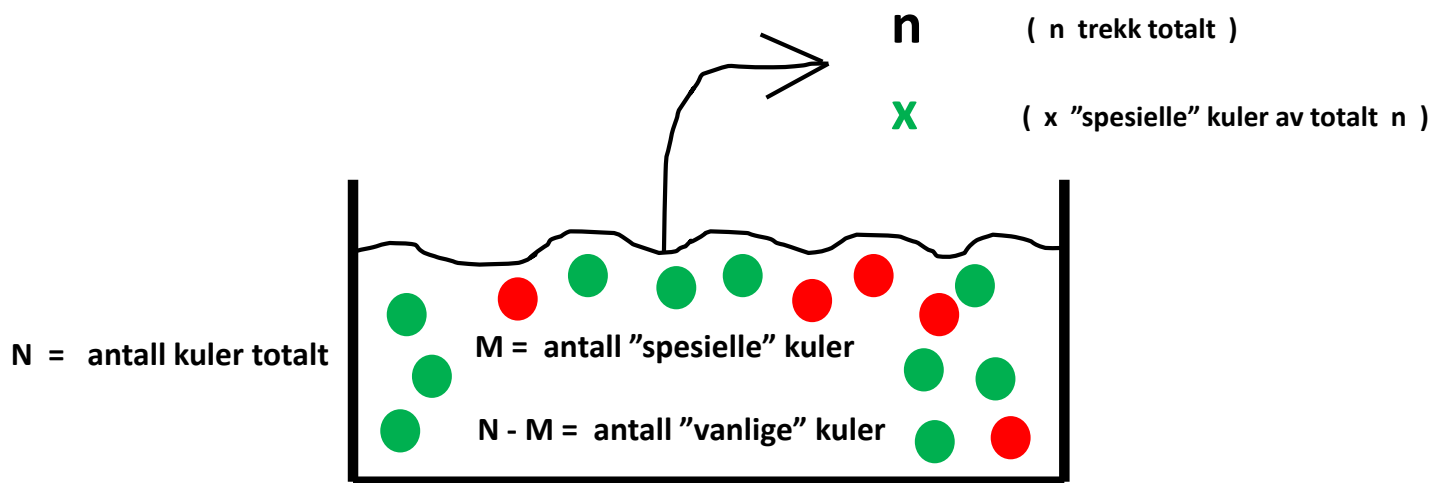
$$\underbrace{\binom{N}{n}}_{\text{binomialkoeff.}} = \frac{N!}{(N-n)! n!} \quad (“N \text{ over } n”) \quad (3.72)$$

■

Merk:

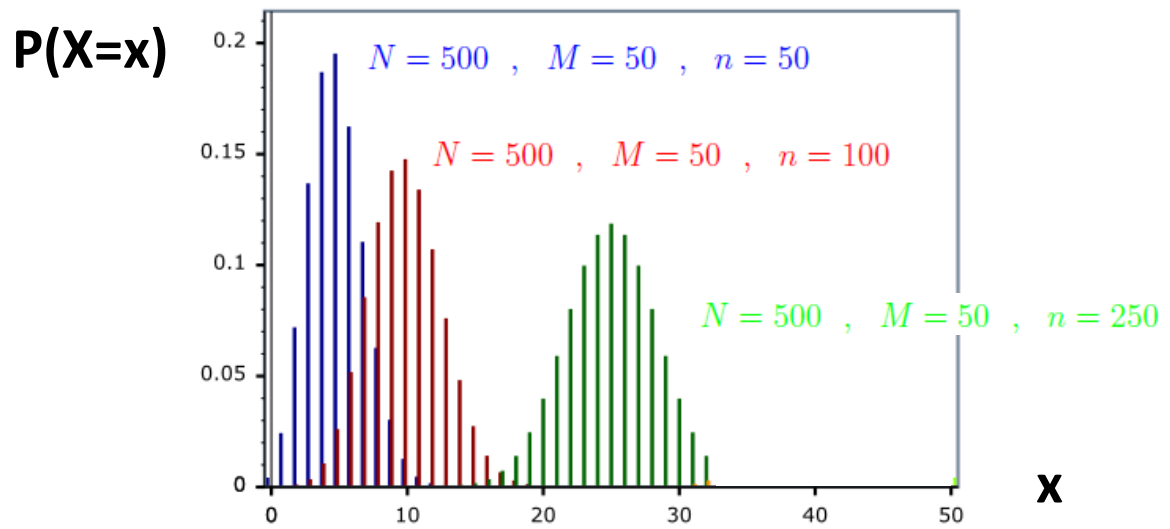
En hypergeometrisk forsøksserie tilsvarer en situasjon med trekning *uten* tilbakelegging.

¹³Skrivem\aa ten $X \sim \text{Hyp}[N, M, n]$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er hypergeometrisk fordelt.



- 1) To typer kuler: "spesielle" kuler
"vanlige" kuler
- 2) Kjent fordeling mellom kulene

Figur 3.17: Hypergeometrisk forsøk.



Figur 3.18: **Hypergeometriske** sannsynlighetsfordelinger $P(X = x)$.

3.2.1 Forventning og varians

Ut fra fordelingsfunksjonen $P(X = x)$ til en hypergeometrisk forsøksserie så kan vi utlede forventning og varians. Utregningen av dette er matematisk og ganske teknisk krevende. Derfor utelater vi det her. Vi presenterer imidlertid resultatene:

Setning: (forventing av $X \sim \overbrace{\text{Hyp}[N, M, n]}^{3 \text{ param.}}$ ¹⁴)

La X være en hypergeometrisk variabel, dvs. $X \sim \text{Hyp}[N, M, n]$. Da gjelder:

$$E[X] = n \cdot \frac{M}{N} \quad (3.73)$$

■

Setning: (varians av $X \sim \overbrace{\text{Hyp}[N, M, n]}^{3 \text{ param.}}$)

La X være en hypergeometrisk variabel, dvs. $X \sim \text{Hyp}[N, M, n]$. Da gjelder:

$$\text{Var}[X] = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \quad (3.74)$$

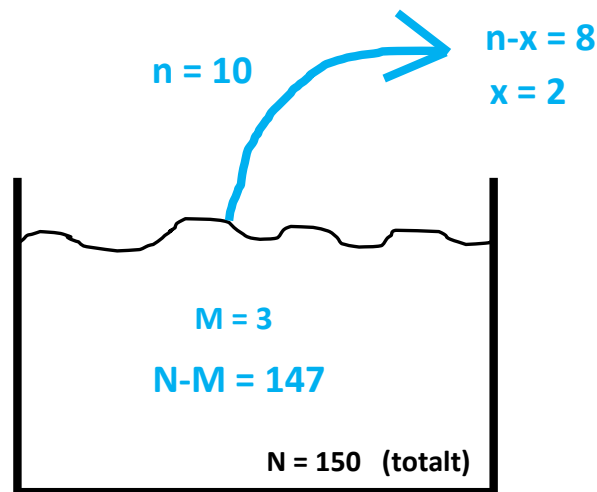
■

¹⁴Skrivemåten $X \sim \overbrace{\text{Hyp}[N, M, n]}^{3 \text{ param.}}$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er hypergeometrisk fordelt.

Eksempel: (revisjon IV , ^{diskret} hypergeometrisk forsøksserie)

La oss igjen en gang se på et tilfeldig utvalg med 150 bilag fra en revisjon fra eksempel 5 i dette avsnittet (se side 175). La X være den stokastiske variabelen definert ved

$$X = \text{antall bilag av de } n = 10 \text{ trukne bilagene som inneholder feil} \quad (3.75)$$



Figur 3.19: Trekker $n = 10$ bilag.

- a) i) Hva er $E[X]$?
- ii) Tolk resultatet $E[X]$.

- b) ii) Hva er $\sigma[X]$?
- ii) Tolk resultatet $\sigma[X]$.

Løsning:

Situasjonen i figur (3.19) beskriver en **hypergeometrisk** forsøksserie. Dermed kan bruke formelene for $E[X]$ og $\sigma[X]$

En **hypergeometrisk** stokastisk variabel $X \sim \text{Hyp}[N, M, n]$ bestemmes av 3 størrelser:

$$N = 150 \quad (\text{antall elementer i grunnmengden}) \quad (3.76)$$

$$M = 3 \quad (\text{antall “spesielle” elementer}) \quad (3.77)$$

$$n = 10 \quad (\text{trekte elementer}) \quad (3.78)$$

a) i) Forventningen til X :

$$\underline{\underline{E[X]}} \stackrel{\text{Hyp}}{=} n \cdot \frac{M}{N} = 10 \cdot \frac{3}{150} = \underline{\underline{0.2}} \quad (3.79)$$

ii) $E[X] =$ forventet verdi for den antall bilag med feil
i den **hypergeometriske** forsøksserien, er 0.2

Kommentar:

Selv om antall bilag med feil er heltall så behøver ikke forventningen av den stokastiske variabelen X å være et heltall.

b) ii) Siden standardavviket til X er

$$\sigma[X] = \sqrt{Var[X]} \quad (3.80)$$

så må vi først finne variansen. Variansen er gitt ved lign.(3.74):

$$\underline{Var[X]} \stackrel{\text{Hyp.}}{=} \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \quad (3.81)$$

$$= \frac{150-10}{150-1} \cdot 10 \cdot \frac{3}{150} \cdot \left(1 - \frac{3}{15}\right) = \frac{\overbrace{686}^{\approx 0.1842}}{3725} \quad (3.82)$$

Dermed er standardavviket

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} = \sqrt{\frac{686}{3725}} \approx \underline{\underline{0.4291}} \quad (3.83)$$

ii) $\sigma[X]$ = standardavviket (som er et mål for spredning) av antall bilag med feil i den **hypergeometrisk** forsøksserien, er 0.4291

Kommentar:

Legg merke til at

$$E[X] < \sigma[X] \quad (3.84)$$

betyr at den stokastiske variabelen X har stor spredning.

■

3.3 Sammenheng mellom $\text{Hyp}[N, M, n]$ og $\text{Bin}[n, p]$

Anta at grunnmengden N i en hypergeo. forsøkserie er mye større enn utvalget n , dvs. anta at:

$$\boxed{N \gg n} \quad (3.85)$$

Dersom dette vilkåret er oppfylt så vil første leddet i variansen til $X \sim \text{Hyp}[N, M, n]$ være nesten 1, se lign. (3.74):

$$\frac{N - n}{N - 1} \stackrel{N \gg n}{\approx} 1 \quad (3.86)$$

I dette tilfellet blir **ikke** sannsynligheten **nevneverdig endret** for hver gang vi trekker. Den er nesten konstant. Denne nesten konstante sannsynligheten er

$$P(X = x) \stackrel[\text{Hyp.}]{\text{lign. (3.66)}} \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\underbrace{\binom{N}{n}}_{\approx \text{konstant}}} \approx \frac{M}{N} \quad (3.87)$$

hvor vi kan innføre notasjonen

$$p \approx \frac{M}{N} \quad (3.88)$$

3.3.1 Forventningsverdi

Forventningsverdien til $X \sim \text{Hyp}[N, M, n]$ **reduserer** seg da til

$$\underline{\underline{E[X]}} \stackrel[\text{Hyp.}]{\text{lin. (3.73)}} n \cdot \overbrace{\frac{M}{N}}^{=p} = \underline{\underline{n \cdot p}} \quad (3.89)$$

dvs. den samme som forventningsverdien til $X \sim \text{Bin}[n, p]$.

3.3.2 Varians

Tilsvarende, variansen til $X \sim \text{Hyp}[N, M, n]$, se lin. (3.74):

$$\underline{\underline{Var[X]}} \stackrel[\text{Hyp.}]{\text{lin. (3.74)}} \overbrace{\frac{N-n}{N-1}}^{\approx 1} \cdot n \cdot \overbrace{\frac{M}{N}}^{=p} \cdot \left(1 - \overbrace{\frac{M}{N}}^{=p}\right) \stackrel[N \gg n]{\approx} \underline{\underline{n \cdot p \cdot (1-p)}} \quad (3.90)$$

dvs. den samme som variansen til $X \sim \text{Bin}[n, p]$.

Setning: $(\overbrace{\text{Hyp}[N, M, n]}^{3 \text{ param.}} \longrightarrow \overbrace{\text{Bin}[n, p]}^{2 \text{ param.}})$

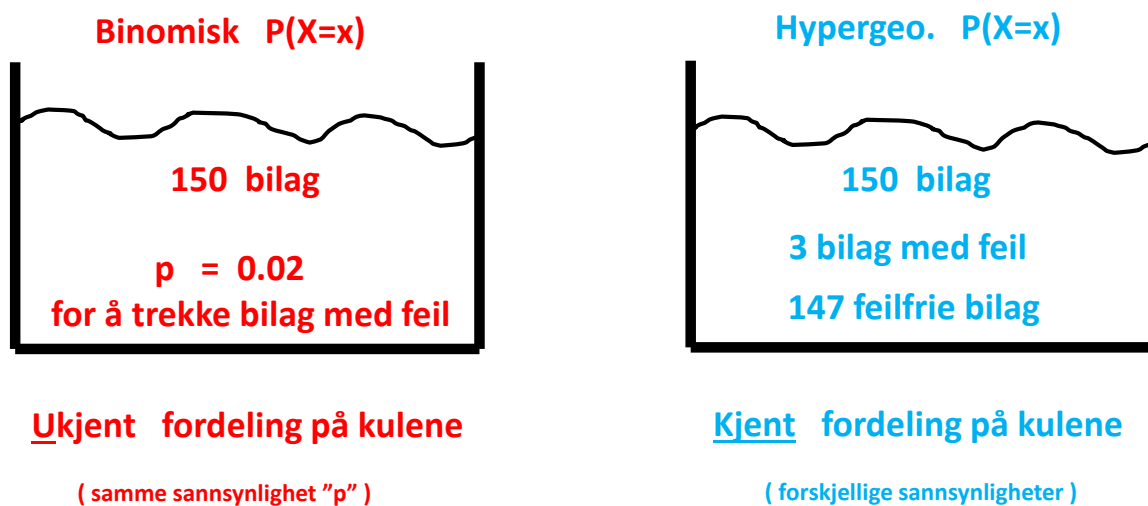
Dersom grunnmengden N i en hypergeometrisk forsøkserie er mye større enn utvalget n , typisk

$$N \gtrsim 20 \cdot n \quad (3.91)$$

så er

$$X \sim \text{Hyp}[N, M, n] \stackrel{N \gtrsim 20 \cdot n}{\approx} X \sim \text{Bin}[n, p] \quad (3.92)$$

■



Figur 3.20: Ukjent og kjent fordeling mellom kulene i urnen.

Noen kommentarer:

1. Alternativt til lign.(3.92) kan vi skrive:

$$\underbrace{\text{hypergeometrisk forsøksserie}}_{\text{u / tilbakelegging}} \stackrel{N \gtrsim 20 \cdot n}{\approx} \underbrace{\text{binomisk forsøksserie}}_{\text{m / tilbakelegging}} \quad (3.93)$$

2. På “godt norsk” betyr lign.(3.91) og (3.92):

Dersom grunnmengden N er stor i forhold til antall forsøk n
så betyr det ikke så mye om det er **med** eller **uten** tilbakelegging. (3.94)

3. hypergeometrisk forsøksserie:

kan beskrives med urnemodellen **uten** tilbakelegging ¹⁵

4. binomisk forsøksserie:

kan beskrives med urnemodellen **med** tilbakelegging ¹⁶

¹⁵Uten tilbakelegging er det generelt ikke samme sannsynlighet hver gang vi trekker. Men dersom $N \gtrsim 20 \cdot n$ så er det **nesten samme sannsynlighet** p hver gang vi trekker.

¹⁶Med tilbakelegging er det lik sannsynlighet p hver gang vi trekker.

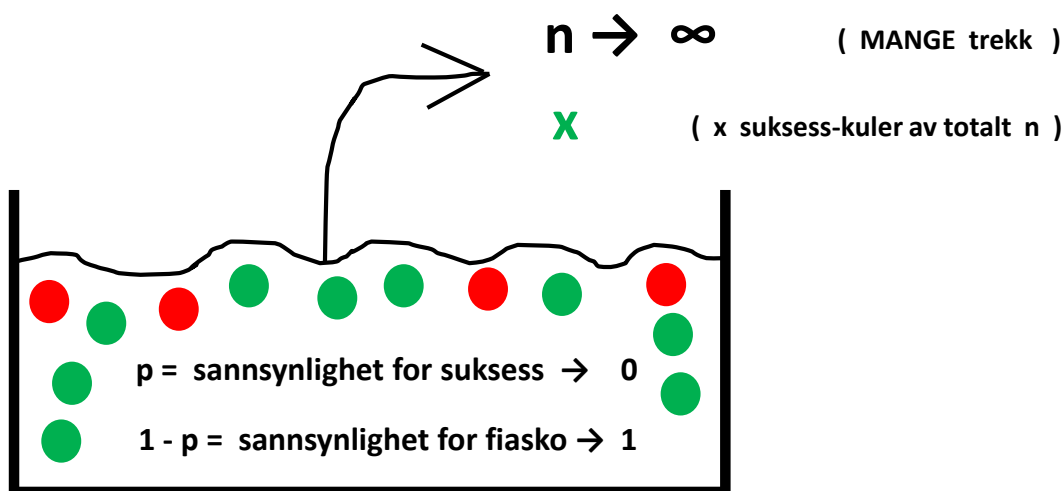
3.4 Poissonfordelingen

Den binomiske fordelingsfunksjonen $P(X = x)$ fra lign. (3.21) er vanskelig å regne ut i de tilfellene hvor

$$p \text{ er liten} \quad \text{og} \quad n \text{ er stor} \quad (3.95)$$

Matematisk betyr dette

$$p \longrightarrow 0^+ \quad \text{og} \quad n \longrightarrow \infty \quad (3.96)$$



- 1) To typer kuler: **suksess-kuler**
fiasko-kuler
- 2) Ukjent fordeling mellom kulene,
men vi kjenner sannsynligheten for suksess: $p \rightarrow 0$

Figur 3.21: Poisson fordeling = binomisk fordeling når $p \longrightarrow 0^+$ og $n \longrightarrow \infty$.

I slike “ekstremtilfeller” av den binomiske fordelingsfunksjonen så er binomialkoeffisienten så stor at verken en kalkulator eller PC kan håndtere det (!). Men, dersom

$$p \cdot n = \underbrace{\text{konstant}}_{\text{RATE}} = \lambda \quad (3.97)$$

så representerer denne konstanten en RATE, dvs.

$$\lambda = \underbrace{\text{konstant}}_{\text{RATE}} = \text{antall begivenheter innenfor et en gitt tid eller gitt rom}$$

Eksempler på slike konstante rater kan være: ¹⁷

- antall fly som ankommer en flyplass per time
- antall bakteriekolonier på en glassplate per cm²
- antall bilulykker på en bestemt strekning
- antall kunder som venter på å bli ekspedert i en bank i løpet av èn time

I slike tilfeller er det en ny fordeling, **Poisson**fordelingen, som beskriver situasjonen. Denne nye sannsynlighetsfordelingen er altså bare et spesialtilfelle av den binomiske sannsynlighetsfordelingen, nemlig når lign.(3.96) inntreffer. Legg også merke til at **biominal**fordelingen har **to parametre**, n og p , mens **Poisson**fordelingen kun har **èn**, nemlig λ .

¹⁷Poissonfordelingen benyttes i såkalte **telleforsøk**.

Definisjon: ($\overbrace{\text{Poissonfordeling}}^{\text{diskret}}$, $X \sim \overbrace{\text{Poi}[\lambda]}^{1 \text{ param.}}$) ^{18 19})

Punktsannsynlighetene for en Poissonfordeling er

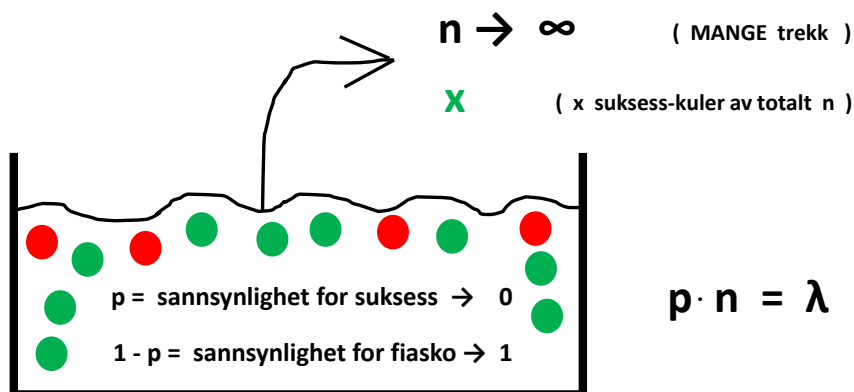
$$P(X = x) \stackrel{\text{def.}}{=} \frac{\lambda^x}{x!} e^{-\lambda} \quad (3.98)$$

hvor

X = stokastisk variabel
 = **antall begivenheter** som inntreffer innenfor en gitt tid eller gitt rom (3.99)

λ = **gjennomsnittlig** antall begivenheter innenfor en gitt tid eller gitt rom (3.100)

■



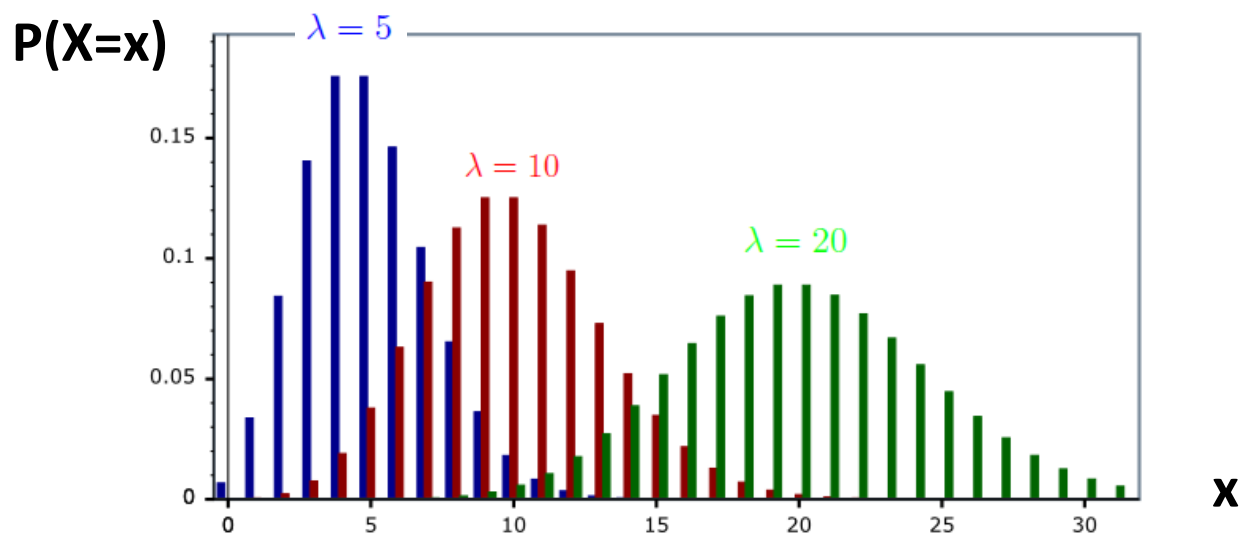
1) To typer kuler: **suksess-kuler**
fiasko-kuler

2) Ukjent fordeling mellom kulene,
 men vi kjenner sannsynligheten for suksess: $p \rightarrow 0$

Figur 3.22: Poisson fordeling = binomisk fordeling når $p \rightarrow 0^+$ og $n \rightarrow \infty$.

¹⁸Kalles ofte “*loven om sjeldne begivenheter*”.

¹⁹Skrivemåten $X \sim \text{Poi}[\lambda]$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er Poissonfordelt.



Figur 3.23: **Poisson**fordelinger $P(X = x)$.

Setning: $(\overbrace{\text{Bin}[n, p]}^{2 \text{ param.}} \longrightarrow \overbrace{\text{Poi}[\lambda]}^{1 \text{ param.}})$

I en **binomisk** sannsynlighetsfordeling, dersom sannsynligheten for “suksess” p er svært liten og antall forsøk n er svært stor, typisk

$$n \gtrsim 50 \quad \text{og} \quad p \lesssim 0.05 \quad (3.101)$$

så er

$$X \sim \text{Bin}[n, p] \quad \overset{n \gtrsim 50 \text{ og } p \lesssim 0.05}{\approx} \quad X \sim \text{Poi}[\lambda] \quad (3.102)$$

■

Noen kommentarer:

1. Binomialfordelingen har **to** parametre, n og p , mens **Poisson**fordelingen kun har **èn**, λ .
2. **Poisson**fordelingen benyttes i såkalte **telleforsøk**, dersom vi kjenner raten λ .
3. Raten ifm. **Poisson**fordelingen er:

$$\lambda = \underbrace{\text{konstant}}_{\text{RATE}} = \text{antall begivenheter innenfor et en gitt } \underline{\text{tid}} \text{ eller gitt } \underline{\text{rom}}$$

4. Poissonfordelingen kalles ofte “*loven om sjeldne begivenheter*”.



3.4.1 Forventning og varians

Ut fra fordelingsfunksjonen $P(X = x)$ til en Poissonfordelingen så kan vi utlede forventning og varians. Utregningen av dette er matematisk og ganske teknisk krevende. Derfor utelater vi det her. Vi presenterer imidlertid resultatene:

Setning: (forventning av $X \sim \text{Poi}[\lambda]$ ²⁰)

La X være en Poissonfordelt variabel, dvs. $X \sim \text{Poi}[\lambda]$. Da gjelder:

$$E[X] = \lambda \quad (3.103)$$

■

Setning: (variens av $X \sim \text{Poi}[\lambda]$)

La X være en Poissonfordelt variabel, dvs. $X \sim \text{Poi}[\lambda]$. Da gjelder:

$$\text{Var}[X] = \lambda \quad (3.104)$$

■

²⁰Skrivemåten $X \sim \text{Poi}[\lambda]$ betyr at den stokastiske variabelen X har en sannsynlighetsfordeling $P(X = x)$ som er Poissonfordelt.

Eksempel: ($\overbrace{\text{Poissonfordeling}}^{\text{diskret}}$, logistikk)

Det viser seg at det kommer 30 kunder innom Shell-stasjonen på Rensvik i Kristiansund hver time for å fylle diesel. En god antagelse i denne sammenheng er at disse kundene kommer uavhengige av hverandre. Anta videre at hver kunde bruker 5 minutter på å fylle diesel. Det er 2 dieselpumper på stasjonen. Anta videre at dersom alle pumpene er opptatt så kjører de potensielle kundene videre til nærmeste konkurrent. Dermed er salget tapt for vår Shell-stasjonen på Rensvik.

La oss definere den stokastiske variabelen:

$$X = \begin{array}{l} \text{antall dieselkunder som kommer inneom Shell-stasjonen} \\ \text{i en 5 minutters periode} \end{array} \quad (3.105)$$

- a) Hva slags sannsynlighetsfordeling har den stokastiske variabelen X ?
- b) Hvor mange kunder svinger innem innom Shell-stasjonen på Rensvik hver 5. minutt?
- c) Hva er sannsynligheten for at Shell-stasjonen på Rensvik mister en kunde?
- d) Hva er forventet antall kunder innenfor en 5. minutters periode?
- e) Hva er standardavviket for X ?



Figur 3.24: Shell-stasjonen på Rensvik i Kristiansund.

Løsning:

- a) Her er ikke “*suksess*”-sannsynligheten p og antall forsøk n oppgitt. Så en binomisk sannsynlighetsfordeling er ikke hensiktsmessig for situasjonen som beskrevet i oppgaven. Av tilsvarende grunn er ei heller hypergeometrisk aktuell. Men siden denne oppgaven dreier seg om antall begivenheter **innenfor et gitt tidsintervall** (5 minutter), så vil den stokastiske variabelen X beskrives av en

$$\underline{\underline{\text{Poissonfordeling}}} \quad (3.106)$$

- b) I oppgaveteksten opplyses det at det i gjennomsnitt kommer 30 kunder per time, 30 kunder per 60 minutter. Antall kunder som i gjennomsnitt kommer hvert 5. minutt er derfor:

$$\underline{\underline{\lambda}} = 30 \cdot \frac{5 \text{ minutter}}{60 \text{ minutter}} = \underline{\underline{2.5}} \quad (3.107)$$

- c) Shell-stasjonen på Rensvik mister en kunde dersom det kommer flere kunder enn det er pumper, dvs. dersom:

$$\underline{\underline{P(\text{miste kunde})}} = P(X > 2) \quad (3.108)$$

$$= 1 - P(X \leq 2) \quad (3.109)$$

$$= 1 - \sum_{x=0}^2 \frac{\lambda^x}{x!} e^{-\lambda} \quad (3.110)$$

$$= 1 - \left(\frac{2.5^0}{0!} e^{-2.5} + \frac{2.5^1}{1!} e^{-2.5} + \frac{2.5^2}{2!} e^{-2.5} \right) \quad (3.111)$$

$$\stackrel{\text{kalkis}}{\approx} \underline{\underline{0.46}} \quad (3.112)$$

- d) Forventet antall kunder innenfor en 5. minutters periode er:

$$\underline{\underline{E[X]}} \stackrel{\text{Poi.}}{=} \lambda = \underline{\underline{2.5}} \quad (3.113)$$

- e) Standardavviket for X :

$$\underline{\underline{\sigma[X]}} \stackrel{\text{def.}}{=} \sqrt{\text{Var}[X]} \stackrel{\text{Poi.}}{=} \sqrt{\lambda} = \sqrt{2.5} \approx \underline{\underline{1.58}} \quad (3.114)$$

■



Figur 3.25: Fylle diesel.