

Ny og bedre versjon 2020

MAT110

Statistikk 1

Kompendium 2020, del 1

Per Kristian Rekdal
Bård-Inge Pettersen



Høgskolen i Molde
Vitenskapelig høyskole i logistikk

Dette kompendiet er tillatt å ha med på eksamen.

Innhold

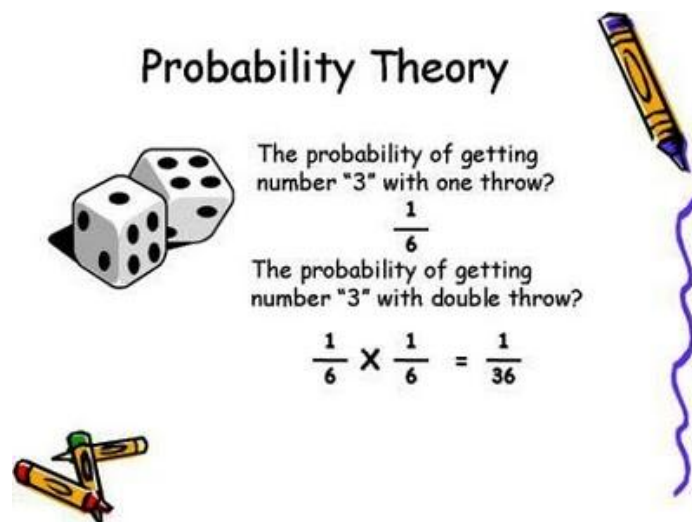
1	Sannsynlighetsteori	7
1.1	Sannsynlighetsmodell	8
1.1.1	Motivasjon	8
1.1.2	Utfallsrommet	11
1.1.3	Mengdelære	13
1.1.4	Begivenheter - delmengder av utfallsrommet	14
1.1.5	Sannsynlighetsmodell	23
1.1.6	Fundamentale setninger i sannsynlighetsteori	24
1.1.7	Diskret sannsynlighetsmodell	40
1.1.8	Uniform sannsynlighetsmodell	45
1.1.9	Kombinatorikk	51
1.2	Betinget sannsynlighet	58
1.2.1	Multiplikasjonssetningen	65
1.2.2	Uavhengighet	67
1.2.3	Bayes lov	72
1.2.4	Sannsynlighetstrær	78
1.2.5	Oppsplitting av Ω	81
2	Stokastiske variabler, forventning og varians	91
2.1	Stokastiske variabler	92
2.1.1	Diskret VS kontinuerlig stokastiske variabler	93
2.2	Forventning og varians	99
2.2.1	Forventning	99
2.2.2	Varians	102
2.2.3	Kovarians	107
2.2.4	Noen regneregler	124
3	Diskrete stokastiske fordelinger	141
3.1	Den binomiske fordelingen	142
3.1.1	Forventningsverdi	158
3.1.2	Varians	160
3.2	Den hypergeometriske fordelingen	166
3.2.1	Forventning og varians	174
3.3	Sammenheng mellom $\text{Hyp}[N, M, n]$ og $\text{Bin}[n, p]$	178
3.3.1	Forventningsverdi	179
3.3.2	Varians	179
3.4	Poissonfordelingen	182

3.4.1	Forventning og varians	188
4	Kontinuerlige stokastiske fordelinger og CLT	193
4.1	Kontinuerlig fordeling	193
4.2	Normalfordelingen (kontinuerlig)	198
4.2.1	Standardisering	203
4.2.2	Standardisering = omskalering	204
4.2.3	Sammenhengen mellom $P(Z \leq z)$ og $G(z)$	207
4.2.4	Diskret vs kontinuerlig fordeling: en viktig forskjell	221
4.2.5	Standardavvik σ og %-vis areal	222
4.3	Oversikt: Bin, Hyp, Poi og N	230
4.4	Sentralgrensesetningen	233
4.5	Diskrete fordelinger $\rightarrow$ normalfordeling	249
4.5.1	Sammenheng: Bin, Hyp, Poi og N	250
4.6	Sum av uavhengige stokastiske variabler	251
I	Statistisk inferens	255
5	Statistisk inferens	257
5.1	Fra sannsynlighetsteori til statistisk inferens	258
5.2	Steg 1: Tilfeldig utvalg	264
5.2.1	Populasjonsvariabler og forsøksvariabler	270
5.3	Steg 2: Gjennomføring av forsøksrekken	273
5.4	Steg 3 : Beskrivende statistikk	275
5.4.1	Lokaliseringsmål	277
5.4.2	Spredingsmål	278
5.5	Statistisk modell	291
6	Estimering og konfidensintervaller	299
6.1	Motivasjon - statistisk inferens	300
6.2	Estimatorer	307
6.3	Konfidensintervaller	314
6.4	Student's t -fordeling og χ_k^2 -fordeling	340
6.4.1	χ_k^2 -fordeling ("kji"-fordeling)	342
6.4.2	Student's t -fordeling	344
6.4.3	Eksakte $(1 - \alpha)$ -konfidensintervaller for μ og σ	347
7	Hypotesetesting	357
7.1	Motivasjon - hypotesetesting	358
7.1.1	Hypotesetest	360
7.1.2	Tilstrekkelig bevis?	361
7.1.3	Type-I feil og type-II feil	361
7.1.4	Analogi til rettsak	363
7.1.5	Test som gir tilstrekkelig bevis	365
7.1.6	Revidert spørsmål	366
7.1.7	En hypotesetest med signifikansnivå $\alpha = 0.05$	367
7.1.8	Hva blir konklusjonen dersom vi bruker ψ_2 ?	370

7.2	Hypotesetesting	371
7.3	To-utvalgs test	375
7.3.1	Statistisk modell for to utvalg	376
7.3.2	Formulering av nullhypotese og alternativ hypotese	378
7.3.3	Konstruksjon av hypotesetest med signifikansnivå $\alpha = 0.05$	379
7.3.4	Realisering og konklusjon	383
8	Regresjonsanalyse	385
8.1	Introduksjon	386
8.2	Statistiske mål (to variabler)	387
8.3	Teoretisk modell vs estimert modell	393
8.4	Residual og <i>sse</i>	394
8.5	Minste kvadraters regresjonslinje	397
8.6	Forklaringsfrakt og <i>sst</i>	405
A	Mengdelære	413
A.1	Venn diagrammer	414
B	Kombinatorikk	419
B.1	Koblinger	420
B.2	4 situasjoner (endelig populasjon)	424
B.3	Binomialkoeffisienten	432
B.4	Kombinatoriske sannsynligheter	433

Kapittel 1

Sannsynlighetsteori



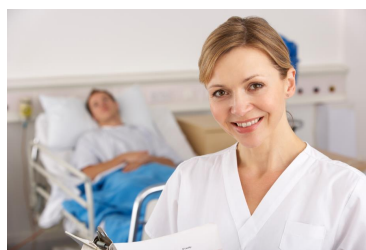
Figur 1.1: Sannsynlighetsteori.

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 sprø tanke. Du kaster bort tiden min! (ennå mer irritert)



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 pasienter.

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 fungerer er 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

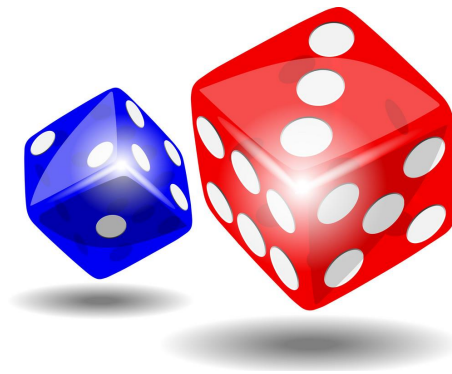
Siden en definerer de mulige utfallene av et eksperiment som en *mengde*, er det viktig med en liten introduksjon til mengdelære.

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* én 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 én 1er *og* at summen av terningene er lik 7?"

S_5 = "hva er sannsynligheten for å få minst én 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 lere?”)

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

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

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 lere 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 :

$$S_2 = \bar{S}_1 = \Omega - S_1 \quad (1.12)$$

Sannsynligheten tilhørende lign.(1.12) er:

$$\underline{P(S_2)} = P(\bar{S}_1) = P(\Omega - S_1) = P(\Omega) - P(S_1) = \overbrace{1 - \frac{11}{36}}^{\text{komplement}} = \underline{\underline{\frac{25}{36}}} \quad (1.13)$$

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

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

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

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

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

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

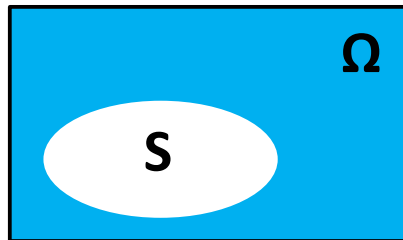
I tillegg må **sannsynlighetsmodellen** oppfylle et sett med fundamentale egenskaper ved beregning av sannsynligheter av delmengder.

■

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 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)

$P(S) \geq 0$	(ikke-negativitet)	(1.27)
$P(S_1 \cup S_2 \cup \dots) = P(S_1) + P(S_2) + \dots$	(addisjon)	(1.28)
$P(\Omega) = 1$	(normalisering)	(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) kan omhandle 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 via *læresetninger* eller bare *setninger*.¹⁴

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

¹⁴På engelsk heter det *theorem*.

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):

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

$\Updownarrow$

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

Siden $P(-) \geq 0$, så er:

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

■

Setning 2: (den spesielle addisjonssetningen) ¹⁵

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

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

■

Bevis:

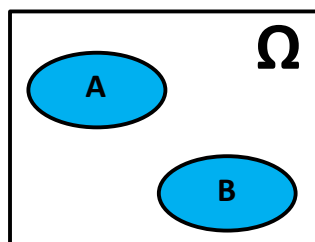
Lag følgen av begivenhetene $S_1 = A$, $S_2 = B$, $S_3 = \emptyset$, $S_4 = \emptyset$, $\dots$

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):

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

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

■



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

¹⁵Den 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:

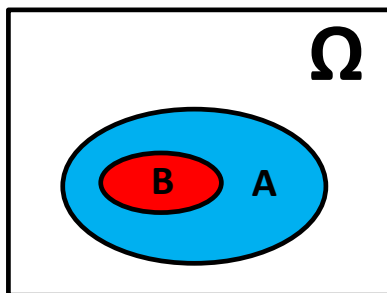
Siden $B \subset A$, er $A - B$ disjunkt med B og $(A - B) \cup B = A$.
Lign. (1.34) gir dermed:

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

$\Updownarrow$

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

■



Figur 1.7: Differansen mellom to begivenheter.

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

For begivenhetene A og B gjelder:

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

Bevis:

For begivenhetene A og B gjelder alltid formelen:

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

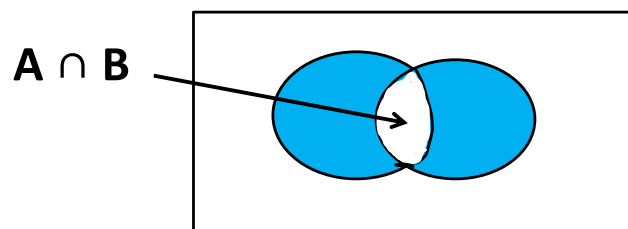
Siden A og $B - A \cap B$ er disjunkte gir den spesielle addisjonssetningen at: ¹⁶

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

Siden $A \cap B \subset B$, gir setningen om differansen mellom to begivenheter videre at: ¹⁷

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

■



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

¹⁶ Lign. (1.34) i setning 2.

¹⁷ Lign. (1.37) i setning 3.

Setning 5: ([komplementær](#)setningen)

La A være en vilkårlig begivenhet, da gjelder alltid:

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

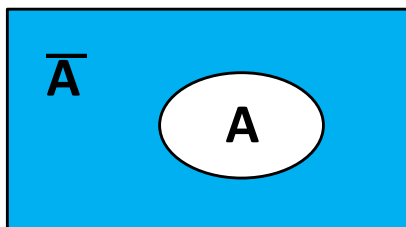
■

Bevis:

Siden en alltid har at $A \subset \Omega$ og $\overline{A} = \Omega - A$, gir setning 3 at: ¹⁸

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

■



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

¹⁸Ved 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:

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

■

Bevis:

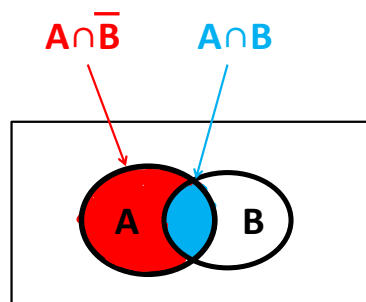
Siden $\Omega = \bar{B} \cup B$ og $A \subset \Omega$ har en alltid at (se figur 1.10):¹⁹

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

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

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

■



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

¹⁹Her 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) ²⁰

For begivenhetene A og B gjelder:

$$P(\overline{A} \overset{\text{og}}{\cap} \overline{B}) = 1 - P(A \overset{\text{eller}}{\cup} B) \quad (1.51)$$

■

Bevis:

For to mengder A og B gjelder alltid:

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

Dermed fås:

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

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

■

²⁰Hvorfor er “tvillingsetningen” et hensiktsmessig navn for lign.(1.51)? Dersom en bytter A med $\overline{A}$ og B med $\overline{B}$ i lign. (1.51), fås ligningen:

$$P(A \overset{\text{og}}{\cap} B) = 1 - P(\overline{A} \overset{\text{eller}}{\cup} \overline{B}) \quad (1.50)$$

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

Eksempel: (gravid - addisjonssetningen)

Eksperiment:

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

Vi gjør følgende eksperiment: trekker en tilfeldig person.

Utfallsrom:

Utfallsrommet består av alle ordnede par:

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

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:

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

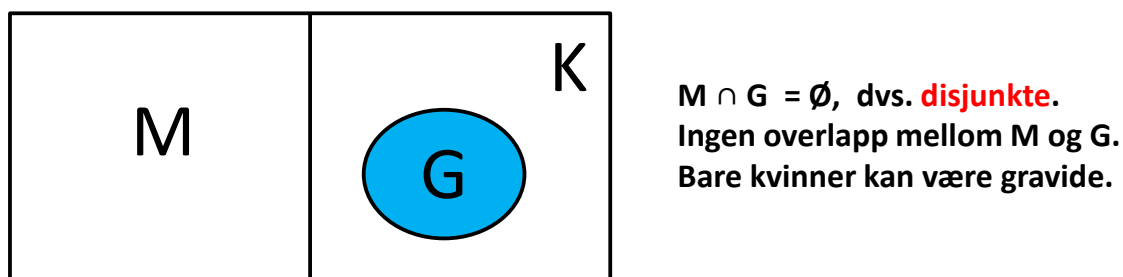
og

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

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

Løsning:

Venn-diagram av situasjonen:



Figur 1.12: Begivenhetene M , K og G .

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

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

dvs. disjunkte ²¹ så gjelder den *spesielle* addisjonssetningen. “Eller”-sannsynligheten finnes da ved å SUMMERE de individuelle sannsynlighetene slik som lign.(1.34) sier:

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

■

²¹Mann 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.

Eksperiment:

Ole Hansen tar første rute 1 og observerer om bussen er ledig eller ikke, deretter tar han rute 2 og observerer om den bussen er ledig. La observasjonene være betegnet ved:

$$l_1 \stackrel{\text{def.}}{=} \text{ledige plasser i bussen på rute 1} \quad (1.62)$$

$$\bar{l}_1 \stackrel{\text{def.}}{=} \text{ingen ledige plasser i bussen på rute 1} \quad (1.63)$$

$$l_2 \stackrel{\text{def.}}{=} \text{ledige plasser i bussen på rute 2} \quad (1.64)$$

$$\bar{l}_2 \stackrel{\text{def.}}{=} \text{ingen ledige plasser i bussen på rute 2} \quad (1.65)$$

Utfallsrom:

Utfallsrommet består av alle ordnede par:²²

$$\Omega = \left\{ (l_1, l_2), (l_1, \bar{l}_2), (\bar{l}_1, l_2), (\bar{l}_1, \bar{l}_2) \right\} \quad (1.66)$$



Figur 1.13: Buss i Kristiansund.

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

Definer begivenhetene:

$$\bar{L}_1 = \left\{ (\bar{l}_1, l_2), (\bar{l}_1, \bar{l}_2) \right\} \quad (\text{ingen ledige plasser i bussen på rute 1}) \quad (1.67)$$

$$\bar{L}_2 = \left\{ (l_1, \bar{l}_2), (\bar{l}_1, \bar{l}_2) \right\} \quad (\text{ingen ledige plasser i bussen på rute 2}) \quad (1.68)$$

$$L_{12} = \left\{ (l_1, l_2) \right\} \quad (\text{begge bussene har ledige sitteplasser}) \quad (1.69)$$

Anta Ole har anslått at:

$$P(\bar{L}_1) = 0.6 \quad (1.70)$$

$$P(\bar{L}_2) = 0.7 \quad (1.71)$$

$$P(L_{12}) = 0.2 \quad (1.72)$$

- a) Definer følgende begivenhet:

$$A = \text{Ole Hansen finner ledige sitteplasser p\aa minst \u00e9n av bussene} \quad (1.73)$$

Uttrykk A ved hjelp av L_1 og L_2 og bruk den generelle addisjonssetningen til \aa regne ut $P(A)$.

- b) Definer følgende begivenhet:

$$B = \text{Ole Hansen finner ledige sitteplasser p\aa rute 1, men ikke p\aa rute 2} \quad (1.74)$$

Uttrykk B ved hjelp av L_1 og L_2 og bruk setningen om total sannsynlighet til \aa regne ut $P(B)$.

- c) Hva er sannsynligheten for at Ole Hansen verken finner ledige sitteplasser p\aa rute 1 eller rute 2?

Løsning:

- a) Begivenheten “*Ole Hansen finner ledige sitteplasser på minst en av bussene*” er lik begivenheten at rute 1 er ledig **eller** rute 2 er ledig.

Uttrykk A ved hjelp av L_1 og L_2 :

$$A = L_1 \cup L_2 \quad (1.75)$$

Generelle addisjonssetningen:

$$P(A) = P(L_1 \cup L_2) \quad (1.76)$$

$$= P(L_1) + P(L_2) - P(L_1 \cap L_2) \quad (1.77)$$

Vi innser at:

$$\underline{L_1 \cap L_2} = \overbrace{\text{Ole Hansen finner ledige sitteplasser på rute 1 og på rute 2}}^{= L_{12}} = \underline{L_{12}} \quad (1.78)$$

Dermed:

$$\underline{P(A)} = P(L_1 \cup L_2) \quad (1.79)$$

$$= P(L_1) + P(L_2) - \underline{P(L_1 \cap L_2)} \quad (1.80)$$

$$= (1 - P(\bar{L}_1)) + (1 - P(\bar{L}_2)) - \underline{P(L_{12})} \quad (1.81)$$

$$= (1 - 0.6) + (1 - 0.7) - 0.2 \quad (1.82)$$

$$= \underline{0.5} \quad (1.83)$$

Sannsynligheten for at Ole Hansen finner ledige sitteplasser på minst en av bussene er dermed 0.5.

- b) Begivenheten “*Ole Hansen finner ledige sitteplasser på rute men ikke på rute 2*” er lik begivenheten at rute 1 er ledig og rute 2 er ikke ledig, dvs:

$$B = L_1 \cap \bar{L}_2 \quad (1.84)$$

Total sannsynlighet:

$$P(E) = P(E \cap F) + P(E \cap \bar{F}) \quad (1.85)$$

med $E = L_1$ og $F = \bar{L}_2$: ($\bar{\bar{L}_2} = L_2$)

$$P(L_1) = \overbrace{P(L_1 \cap \bar{L}_2)}^{=B} + \overbrace{P(L_1 \cap \bar{\bar{L}_2})}^{=P(L_1 \cap L_2) = P(L_{12})} \quad (1.86)$$

$$\Downarrow$$

$$\overbrace{P(L_1 \cap \bar{L}_2)}^{=B} = P(L_1) - P(L_{12}) \quad (1.87)$$

Setter inn tall:

$$\underline{P(B)} = P(L_1) - P(L_{12}) \quad (1.88)$$

$$= (1 - P(\bar{L}_1)) - P(L_{12}) \quad (1.89)$$

$$= (1 - 0.6) - 0.2 \quad (1.90)$$

$$= \underline{0.2} \quad (1.91)$$

Sannsynligheten for at Ole Hansen finner ledige sitteplasser på rute 1, men ikke på rute 2 er 0.2.

c) Begivenheten

$$C = \text{Ole Hansen finner } \underline{\text{ingen ledige}} \text{ sitteplasser p\aa{} hverken rute 1 rute 2} \quad (1.92)$$

er den *motsatte* begivenheten til begivenheten at rute 1 *eller* rute 2 er ledig som vi regnet ut i deloppgave a, dvs: ²³

$$C = \bar{L}_1 \cap \bar{L}_2 = \overline{L_1 \cup L_2} = \bar{A} \quad (1.93)$$

Komplement rsetningen:

$$\underline{P(C)} = P(\bar{A}) \quad (1.94)$$

$$= 1 - P(A) \quad (1.95)$$

$$= 1 - 0.5 \quad (1.96)$$

$$= \underline{0.5} \quad (1.97)$$

Sannsynligheten for at Ole Hansen ikke finner ledige sitteplasser p  hverken rute 1 eller rute 2 er 0.5.

■

²³Tegn opp et Venn-diagram. Da innser du dette lettere.

1.1.7 Diskret sannsynlighetsmodell

Motiverende eksempel: (myntkast)

Eksperiment 1: (myntkast - 1 kast)

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

$$\Omega = \{ K, M \} \quad (1.98)$$

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

$$P(K) = P(M) = \frac{1}{2} \quad (1.99)$$

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

$$S_1 = \emptyset, \quad S_2 = \{ M \}, \quad S_3 = \{ K \}, \quad S_4 = \Omega \quad (1.100)$$

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

$$\underline{\underline{P(S_1)}} = P(\emptyset) = \underline{\underline{0}} \quad (1.101)$$

$$\underline{\underline{P(S_2)}} = P(\{M\}) = \underline{\underline{\frac{1}{2}}} \quad (1.102)$$

$$\underline{\underline{P(S_3)}} = P(\{K\}) = \underline{\underline{\frac{1}{2}}} \quad (1.103)$$

$$\underline{\underline{P(S_4)}} = P(\{M, K\}) = P(\{M\}) + P(\{K\}) = \frac{1}{2} + \frac{1}{2} = \underline{\underline{1}} \quad (1.104)$$

Eksperiment 2: (myntkast - 3 kast)

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

$$\Omega = \left\{ \overbrace{\text{KKK}, \text{KKM}, \text{KMK}, \text{KMM}, \text{MKK}, \text{MKM}, \text{MMK}, \text{MMM}}^{8 \text{ mulige utfall}} \right\} \quad (1.105)$$

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 :

$$P(KKM) = \frac{1}{8}, \quad (1.106)$$

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

$$S_i \subset \Omega, \quad i = 1, 2, 3, \dots, 2^8 \quad (1.107)$$

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:

$$S_{MM} = \left\{ \overbrace{MMK, MKM, KMM}^{3 \text{ mulige utfall}} \right\} = \{ MMK \} \cup \{ MKM \} \cup \{ KMM \} \quad (1.108)$$

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

$$\underline{\underline{P(S_{MM})}} = P(\{ MMK \} \cup \{ MKM \} \cup \{ KMM \}) \quad (1.109)$$

$$= P(MMK) + P(MKM) + P(KMM) \quad (1.110)$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \underline{\underline{\frac{3}{8}}} \quad (1.111)$$

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.:

$$\Omega = \left\{ u_1, u_2, \dots, u_n \right\} \quad (1.112)$$

kalles en **diskret sannsynlighetsmodell**. ■

Anta sannsynlighetene til de n utfallene er gitt ved:

$$P(u_i) = p_i, \quad i = 1, 2, \dots, n \quad (1.113)$$

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

$$\underline{\underline{P(\Omega)}} = P(\{u_1, u_2, \dots, u_n\}) = \sum_{i=1}^n P(u_i) = \sum_{i=1}^n p_i = \underline{\underline{1}} \quad (1.114)$$

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

$$S = \left\{ u_{j_1}, u_{j_2}, \dots, u_{j_k} \right\} \quad (1.115)$$

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

$$P(S) = \sum_{i=1}^k p_{j_i} \quad (1.116)$$

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 :

$$\Omega = \left\{ u_1, u_2, \dots, u_n \right\} \quad (1.117)$$

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

$$P(u_i) = p_i = \frac{1}{n} \quad (1.118)$$

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

■

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

$$S = \left\{ u_{j_1}, u_{j_2}, \dots, u_{j_k} \right\} \quad (1.119)$$

Lign. (1.116) for diskrete sannsynlighetsmodeller gir da at

$$P(S) = \sum_{i=1}^k p_{j_i} = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n} \quad (1.120)$$

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:

$$\Omega = \{ 1, 2, 3, 4, 5, 6 \} \quad (1.121)$$

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**.

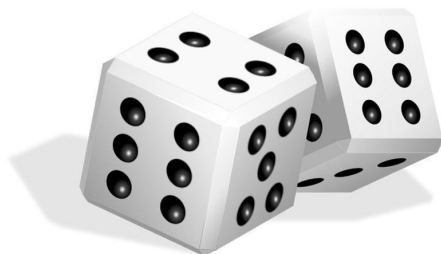
La oss denne gangen se på begivenhetene:

$$A = \text{resultat med odde antall øyne} = \{ 1, 3, 5 \} \quad (1.122)$$

$$B = \text{resultat 4 øyne eller mer} = \{ 4, 5, 6 \} \quad (1.123)$$

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

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



Figur 1.15: Terning.

Løsning:

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

$$\underline{\underline{P(A)}} = \frac{\text{antall } \textcolor{red}{gunstige} \text{ utfall}}{\text{antall } mulige \text{ utfall}} = \frac{\textcolor{red}{3}}{6} = \frac{1}{2} = \underline{\underline{0.5}} \quad (1.124)$$

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**²⁴. Det er 3 stk. gunstige utfall i begivenhetene A og B . Dermed:

$$P(A) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{3}{6} = \frac{1}{2} \quad (1.125)$$

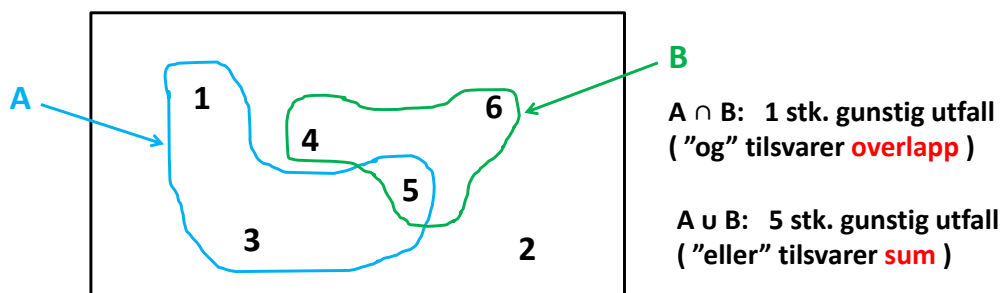
$$P(B) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{3}{6} = \frac{1}{2} \quad (1.126)$$

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

$$P(\overbrace{A \cap B}^{\text{og}}) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{1}{6} \quad (1.127)$$

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

$$\underline{\underline{P(\overbrace{A \cup B}^{\text{eller}})}} = \underbrace{P(A)}_{=1/2} + \underbrace{P(B)}_{=1/2} - \underbrace{P(\overbrace{A \cap B}^{\text{og}})}_{=1/6} = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \underline{\underline{\frac{5}{6}}} \quad (1.128)$$



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

²⁴Vi 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:

$$\underline{\underline{P(A \overset{\text{eller}}{\cup} B)}} = \frac{\text{antall } \textit{gunstige} \text{ utfall}}{\text{antall } \textit{mulige} \text{ utfall}} = \frac{5}{\underline{\underline{6}}} \quad (1.129)$$

som selvfølgelig er det samme svaret som vi fikk via den generelle addisjonssetningen, se lign.(1.128).

■

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: ²⁵

$$\underline{\underline{P(\text{kredittverdig})}} = \frac{\text{antall } \textit{gunstige} \text{ utfall}}{\text{antall } \textit{mulige} \text{ utfall}} = \frac{862}{1000} = \underline{\underline{0.862}} \quad (1.132)$$

■



Figur 1.17: Kredittverdig.

²⁵Senere i dette kapitlet skal vi lære om **komplementsetningen**. Med denne setningen kan vi bruke lign.(1.132) til å finne sannsynligheten for at en tilfeldig valgt kunde **ikke** er kredittverdig:

$$\underline{\underline{P(\text{IKKE kredittverdig})}} \stackrel{\text{kompl.setn.}}{=} 1 - P(\text{kredittverdig}) \quad (1.130)$$

$$= 1 - 0.862 = \underline{\underline{0.138}} \quad (1.131)$$

Kunne man funnet denne sannsynligheten på en annen måte også?

1.1.9 Kombinatorikk

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

Se tillegg B.

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

Eksperiment:

- 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.

Utfallsrom:

Utfallsrommet består av alle tripler:

$$\Omega = \left\{ (P_1, P_2, P_3) \right\} \quad (1.133)$$

hvor P_1 , P_2 og P_3 er forskjellige pasienter.

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

Definer følgende begivenheter:

$$A = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er menn} \right\} \quad (1.134)$$

$$B = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er kvinner} \right\} \quad (1.135)$$

$$C = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er 2 kvinner og en mann} \right\} \quad (1.136)$$

$$D = \left\{ (P_1, P_2, P_3) \mid \text{hvor alle pasientene er 1 kvinne og 2 menn} \right\} \quad (1.137)$$



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 ²⁶
- utfallene ikke påvirker hverandre, dvs. de er **uavhengige** ²⁷

Dermed har vi en tellesituasjon som kan beskrives ved urnemodellen.

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

$$P(A) = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner for } A}{\text{antall } \textit{mulige} \text{ kombinasjoner totalt}} \quad (1.138)$$

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:

$$\text{situasjon 3} \quad (1.139)$$

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

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

$$\underbrace{\# \text{ ikke-ordnede komb. u/tilbakelegging}}_{\text{situasjon 3}} = \frac{N!}{(N-s)! s!} \equiv \underbrace{\binom{N}{s}}_{\text{binomialkoeff.}} \quad (1.140)$$

²⁶Samme sannsynlighet for å trekke en pasient.

²⁷Om 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:

$$\underline{\underline{P(A)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.141)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.142)$$

$$= \frac{\overbrace{\binom{4}{3}}^{=4} \cdot \overbrace{\binom{5}{0}}^{=1}}{\underbrace{\binom{9}{3}}_{=84}} = \frac{4 \cdot 1}{84} = \underline{\underline{0.048}} \quad (= 4.8 \%) \quad (1.143)$$

b) $P(B)$:

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

$$\underline{\underline{P(B)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.144)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.145)$$

$$= \frac{\overbrace{\binom{4}{0}}^{=1} \cdot \overbrace{\binom{5}{3}}^{=10}}{\underbrace{\binom{9}{3}}_{=84}} = \frac{1 \cdot 10}{84} = \underline{\underline{0.119}} \quad (1.146)$$

c) $P(C)$:

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

$$\underline{\underline{P(C)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.147)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.148)$$

$$= \frac{\overbrace{\binom{4}{1}}^{= 4} \cdot \overbrace{\binom{5}{2}}^{= 10}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{4 \cdot 10}{84} = \underline{\underline{0.476}} \quad (1.149)$$

d) $P(D)$:

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

$$\underline{\underline{P(D)}} = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner}}{\text{antall } \textit{mulige} \text{ kombinasjoner}} \quad (1.150)$$

$$= \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (1.151)$$

$$= \frac{\overbrace{\binom{4}{2}}^{= 6} \cdot \overbrace{\binom{5}{1}}^{= 5}}{\underbrace{\binom{9}{3}}_{= 84}} = \frac{6 \cdot 5}{84} = \underline{\underline{0.357}} \quad (1.152)$$

Kommentarer:

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

$$P(\text{dersom 2 kategorier}) = \frac{2 \text{ stk. binomialkoeff.}}{\text{antall mulige kombinasjoner}} \quad (1.153)$$

- 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.(1.149).



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.

trying to find $P(A|B)$ know
read "probability of A given B"

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.

Eksperiment:

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

Utfallsrom:

$$\Omega = \{ K_1, K_2, \dots, K_{66} \} \quad (1.154)$$

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

Definer følgende begivenheter: ²⁸

$$A = \{ K \mid \text{hvor kunden } K \text{ er en god kunde} \} \quad (1.155)$$

$$B = \{ K \mid \text{hvor kunden } K \text{ er en kunde med stor omsetning} \} \quad (1.156)$$

$$A \cap B = \{ K \mid \text{hvor kunden } K \text{ er en god kunde med stor omsetning} \} \quad (1.157)$$

$$A|B = \{ K \mid \text{hvor kunden } K \text{ er en god kunde} \\ \text{dersom vi } \underline{\text{vet}} \text{ at kunden har stor omsetning} \} \quad (1.158)$$



Figur 1.20: Sparebanken Møre.

²⁸Legg merke til at notasjonen $A|B$ er ny og står for betinget sannsynlighet.

	B (stor omsetn.)	$\overline{\mathbf{B}}$ (liten omsetn.)	totalt
A (god kunde)	21	13	34
$\overline{\mathbf{A}}$ (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:

$$\underline{\underline{P(A)}} = \frac{\text{antall } gode \text{ (A) kunder}}{\text{antall totalt}} = \frac{34}{66} \approx \underline{\underline{0.52}} \quad (1.159)$$

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

$$\underline{\underline{P(B)}} = \frac{\text{antall kunder med } stor \text{ (B) omsetning}}{\text{antall totalt}} = \frac{36}{66} \approx \underline{\underline{0.55}} \quad (1.160)$$

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

$$\underline{\underline{P(A \overset{\text{og}}{\cap} B)}} = \frac{\text{antall } gode \text{ (A) kunder med } stor \text{ (B) omsetning}}{\text{antall totalt}} = \frac{21}{66} \approx \underline{\underline{0.32}} \quad (1.161)$$

- 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.

$$\underline{\underline{P(A|B)}} = \frac{\text{antall gode (A) kunder med stor (B) omsetning}}{\text{antall totalt med stor omsetning}} = \frac{21}{36} \quad (1.162)$$

$$= \frac{\frac{21}{66}}{\frac{36}{66}} = \frac{\overbrace{P(A \cap B)}^{= 21/66}}{\underbrace{P(B)}_{= 36/66}} \approx \underline{\underline{0.58}} \quad (1.163)$$

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 kunde dersom vi **vet** at kunden har stor omsetning, $P(A|B) \approx 0.58$.



Definisjon: (betinget sannsynlighet)

Den betingede sannsynligheten er definert ved:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.164)$$

hvor

$$P(A|B) = \text{sannsynligheten for } A \text{ gitt } B \text{ (} B \text{ allerede har inntruffet)} \quad (1.165)$$

(betinget sannsynlighet)

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

$$P(B) = \text{sannsynligheten for } B \neq 0 \quad (1.166)$$

(ubetinget sannsynlighet)

■

Lign.(1.164) er en definisjon og trenger således ikke å bevises.

1.2.1 Multiplikasjonssetningen

Fra definisjonen av betinget sannsynlighet i lign.(1.164) følger trivielt det som kalles multiplikasjonssetningen:

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

For begivenhetene A og B gjelder:

$$P(\overbrace{A \cap B}^{\text{og}}) = P(A|B) P(B) \quad (1.167)$$

■

Bevis:

Følger trivielt fra definisjonen av betinget sannsynlighet:

$$P(A|B) \stackrel{\text{lign. (1.164)}}{=} \frac{P(A \cap B)}{P(B)} \quad \Bigg| \cdot P(B) \quad (1.168)$$

$\Updownarrow$

$$\underline{\underline{P(A \cap B) = P(A|B) P(B)}} \quad (1.169)$$

■

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:

$$\overbrace{P(B \cap A)}^{= P(A \cap B)} = P(B|A) P(A) \quad (1.170)$$

■

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

$$P(B|A) = P(B \mid \underbrace{\text{når vi vet}}_{\text{vet}} A) \quad (1.171)$$

1.2.2 Uavhengighet

Definisjon: (uavhengighet)

To begivenheter A og B er uavhengige dersom

$$P(A|B) = P(A) \quad (1.172)$$

■

Lign.(1.172) 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:

$$P(\overbrace{A \cap B}^{\text{og}}) \stackrel{\text{uavh.}}{=} P(A) P(B) \quad (1.173)$$

■

Bevis:

Multiplikasjonssetningen i lign.(1.167) gir:

$$P(A \cap B) \stackrel{\text{lign.(1.167)}}{=} P(A|B) P(B) \quad (1.174)$$

Men siden A og B er uavhengige så er er

$$P(A|B) = P(A) \quad (1.175)$$

Dermed:

$$\underline{\underline{P(A \cap B) = P(A) P(B)}} \quad (1.176)$$

■

NB:

Ikke bland sammen begrepene “uavhengighet” og “disjunkthet”: ²⁹

$$\text{i) } \overbrace{\text{Uavhengighet}}^{\text{vanskelig \AA visualisere}} : ^{30} \quad P(A \overbrace{\cap}^{\text{og}} B) = P(A) P(B) \quad (\text{ se lign. (1.173) })$$

$$\text{ii) } \underbrace{\text{Disjunkt}}_{\text{ikke overlap}} : ^{31} \quad P(A \overbrace{\cup}^{\text{eller}} B) = P(A) + P(B) \quad (\text{ se lign. (1.34) })$$

²⁹Disse begrepene kan ogs\aa v\are koblet p\aa f\olgende m\aater:

$$\text{disjunkt} \quad \xRightarrow{\text{nesten alltid}} \quad \text{avhengig} \quad (1.177)$$

$$\text{disjunkt} \quad \xleftarrow{\text{ikke n\o}dvendigvis} \quad \text{avhengig} \quad (1.178)$$

³⁰Alts\aa: $P(\text{og}) = \text{PRODUKT}$ av ubetingede sannsynligheter

³¹Alts\aa: $P(\text{eller}) = \text{SUM}$ av ubetingede sannsynligheter

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*.

$$P(A) \cdot P(B) \tag{1.179}$$

$$P(A) + P(B) \tag{1.180}$$

$$P(A \cap B) \tag{1.181}$$

$$P(A \cup B) \tag{1.182}$$

- 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 1.22: 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 $P(B|A) = P(B)$.



1.2.3 Bayes lov

Setning: (*Bayes lov*)

For begivenhetene A og B gjelder:

$$P(A|B) = P(B|A) \frac{P(A)}{P(B)} \quad (1.183)$$

■

Bevis:

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

$$P(A \cap B) = P(B \cap A) \quad (1.184)$$

$\Updownarrow$

$$P(A|B) P(B) = P(B|A) P(A) \quad \left| \cdot \frac{1}{P(B)} \right. \quad (1.185)$$

$$\underline{\underline{P(A|B) = P(B|A) \frac{P(A)}{P(B)}}} \quad (1.186)$$

■

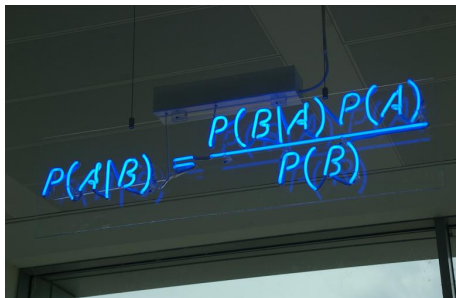
På same måte som multiplikasjonssetningen så kan vi snu på Bayes lov og skrive denne på en alternativ måte:

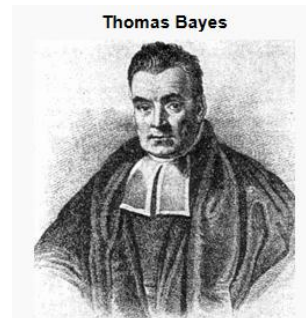
Setning: (*Bayes lov*)

For begivenhetene A og B gjelder:

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)} \quad (1.187)$$

■


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



Figur 1.23: 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 1.24: Relative hyppigheter.

Vi har altså definert begivenhetene:

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



Figur 1.25: Lindorff.

Eksperiment:

Vi gjør følgende eksperiment: trekker en tilfeldig bedrift.

Utfallsrom:

Utfallsrommet består av alle 2-tupler:

$$\Omega = \left\{ (A, B), (A, \overline{B}), (\overline{A}, B), (\overline{A}, \overline{B}) \right\} \quad (1.188)$$

- 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:

$$\underline{\underline{P(A)}} = 0.45 + 0.20 + 0.15 = \underline{\underline{0.80}} \quad (1.189)$$

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

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

$$\underline{\underline{P(B)}} = 0.45 + 0.02 = \underline{\underline{0.47}} \quad (1.190)$$

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.(1.167):

$$\begin{aligned} \underline{\underline{P(A|B)}} &= \frac{\text{hyppighet god likv. (A) og langsikt. (B)}}{\underbrace{\text{hyppighet langsiktige investeringer}}_{\text{utfallsrommet 'reduert' til kun B}}} \\ &\stackrel{\text{lign. (1.167)}}{=} \frac{\overbrace{P(A \cap B)}^{= 0.45}}{\underbrace{P(B)}_{= 0.47}} = \frac{0.45}{0.47} \approx \underline{\underline{0.96}} \end{aligned} \quad (1.191)$$

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.(1.187):

$$\underline{\underline{P(B|A)}} \stackrel{\text{lign. (1.187)}}{=} \overbrace{P(A|B)}^{\approx 0.96} \overbrace{\frac{P(B)}{P(A)}}^{= 0.47} = 0.96 \cdot \frac{0.47}{0.80} = \underline{\underline{0.56}} \quad (1.192)$$

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.(1.191), dvs. multiplikasjonssetningen:

$$\begin{aligned} \underline{\underline{P(B|A)}} &= \frac{\text{hyppighet langsikt. (B) og god (A) likv.}}{\underbrace{\text{hyppighet god likviditet}}_{\text{utfallsrommet 'reduert' til kun A}}} \\ &\stackrel{\text{lign. (1.170)}}{=} \frac{\overbrace{P(B \cap A)}^{= P(A \cap B) = 0.45}}{\underbrace{P(A)}_{= 0.80}} = \frac{0.45}{0.80} \approx \underline{\underline{0.56}} \end{aligned} \quad (1.193)$$

dvs. samme svar som i lign.(1.192), selvfølgelig. ■

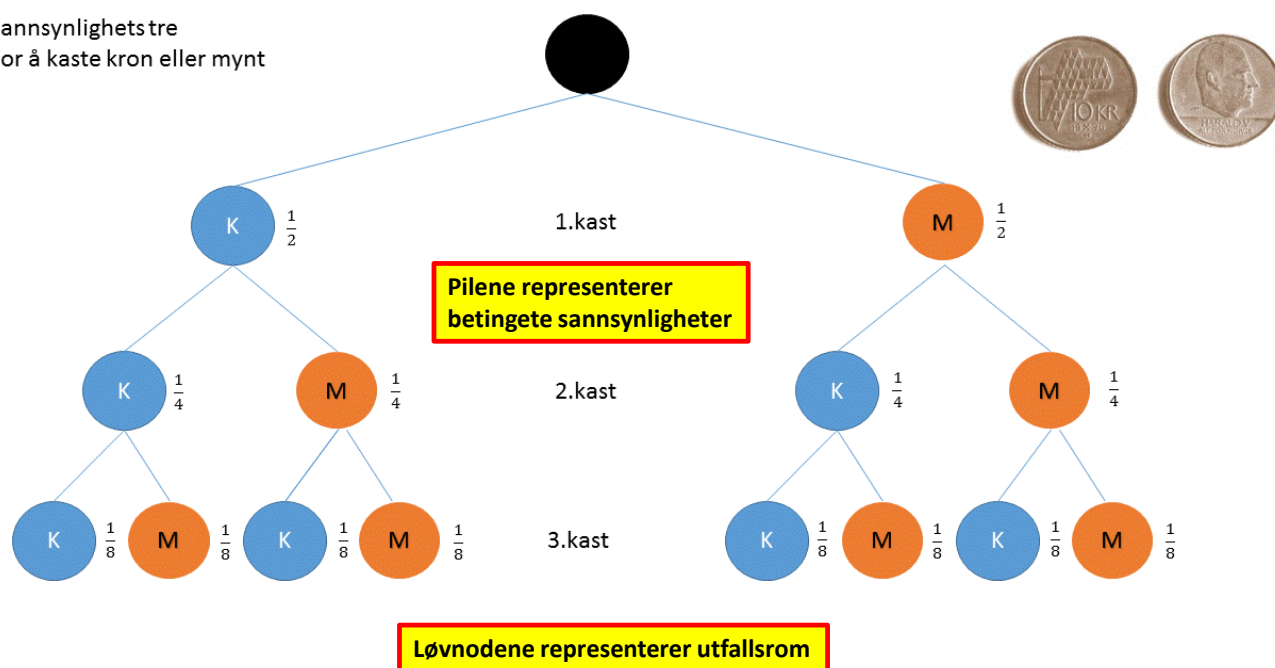
1.2.4 Sannsynlighetstrær

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**.³²

Sannsynlighetstre
For å kaste kron eller mynt



Figur 1.26: Sannsynlighetstre - sekvens av steg.

³²Utfallsrommet 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 investeringer i teknologi?

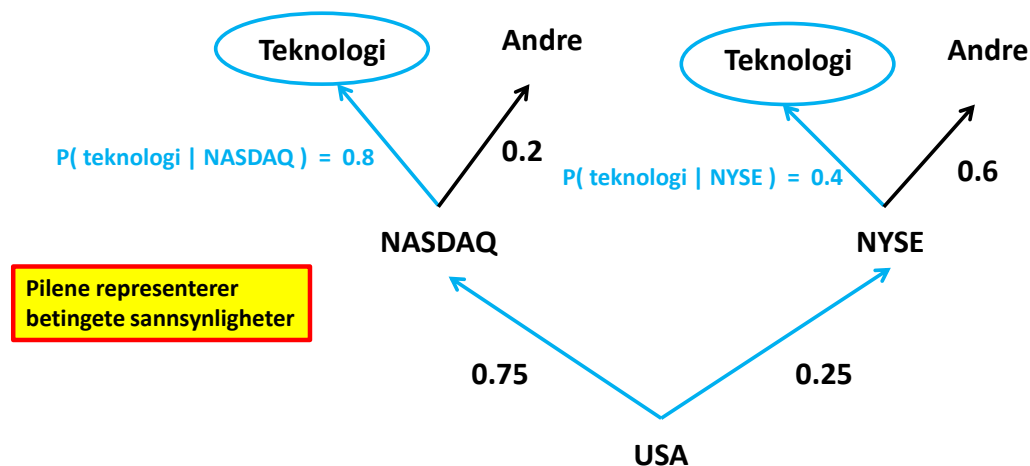


Figur 1.27: NASDAQ og NYSE.

Løsning:

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

Derfor er det her hensiktsmessig å bruke et sannsynlighetstre: ³³



Figur 1.28: **Sannsynlighetstre** for fond i USA.

Sannsynlighetstreet i figur (1.28) er en ^{NB!}visuell fremstilling av opplysningene. Med en slik fremstilling er det lettere å se at: ³⁴

$$\underline{\underline{P(\text{aksjene investert i teknologi})}} = 0.75 \cdot 0.8 + 0.25 \cdot 0.4 = \underline{\underline{0.70}} \quad (1.194)$$

■

³³Utfallsrommet er gitt ved løvnodene. Pilene representerer betingede sannsynligheter.

³⁴Følg de blå linjene i sannsynlighetstreet.

1.2.5 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

Ekspirement:

Trekker et fond som amerikaneren har investert i.

Utfallsrom:

Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ (B_1, A), (B_1, \bar{A}), (B_2, A), (B_2, \bar{A}) \right\} \quad (1.195)$$

Definer følgende begivenheter:

$$B_1 = \left\{ (B_1, A), (B_1, \bar{A}) \right\} \quad (\text{et NASDAQ-fond trekkes}) \quad (1.196)$$

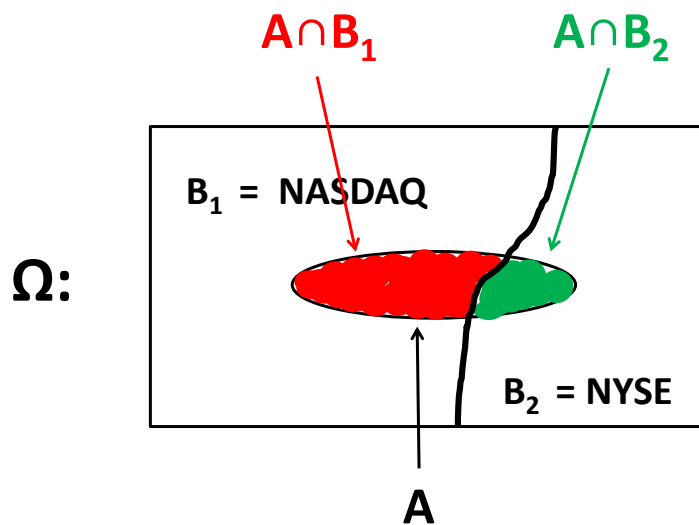
$$B_2 = \left\{ (B_2, A), (B_2, \bar{A}) \right\} \quad (\text{et NYSE-fond trekkes}) \quad (1.197)$$

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

$$\Omega = \left\{ B_1, B_2 \right\} \quad (1.198)$$

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:

$$A = (A \cap B_1) \underbrace{\cup}_{\text{eller}} (A \cap B_2) \quad (1.199)$$



Figur 1.29: **Oppsplitting** av sannsynlighetsrommet i to, $\Omega = B_1 \cup B_2$.

Via lign.(1.199) finner vi sannsynligheten $P(A)$:

$$P(A) = P\left[(A \cap B_1) \underbrace{\cup}_{\text{eller}} (A \cap B_2)\right] \quad (1.200)$$

Av figur (1.29) ser vi at det ikke er noe overlapp mellom $A \cap B_1$ og $A \cap B_2$, dvs. de er disjunkte³⁵. Da gjelder den spesielle addisjonssetningen for “eller”-sannsynligheten i lign.(1.200):³⁶

$$P(A) = P(A \cap B_1) + P(A \cap B_2) \quad (1.201)$$

Denne ligningen inneholder to “og”-sannsynligheter. Dermed kan vi bruke multiplikasjonssetningen³⁷ på disse:

$$\underline{\underline{P(A)}} = \overbrace{P(A|B_1) \cdot P(B_1)}^{= P(A|B_1) \cdot P(B_1)} + \overbrace{P(A|B_2) \cdot P(B_2)}^{= P(A|B_2) \cdot P(B_2)} \quad (1.202)$$

$$\stackrel{\text{lign. (1.167)}}{=} \overbrace{P(A|B_1) \cdot P(B_1)}^{=0.8 \quad =0.75} + \overbrace{P(A|B_2) \cdot P(B_2)}^{=0.4 \quad =0.25} \quad (1.203)$$

$$= 0.8 \cdot 0.75 + 0.4 \cdot 0.25 = \underline{\underline{0.7}} \quad (1.204)$$

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

■

³⁵At det ikke er noe overlapp, dvs. disjunkte, betyr at $(A \cap B_1) \cup (A \cap B_2) = \emptyset$.

³⁶Den spesielle addisjonssetningen i lign.(1.34) sier at: $P(A \cup B) = P(A) + P(B)$.

³⁷Multiplikasjonssetningen i lign.(1.167) 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:

$$P(A) = P(A|B_1) P(B_1) + P(A|B_2) P(B_2) \quad (1.205)$$

■

Bevis:

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

$$A = (A \cap B_1) \overset{\text{eller}}{\cup} (A \cap B_2) \quad (\text{oppsplitting}) \quad (1.206)$$

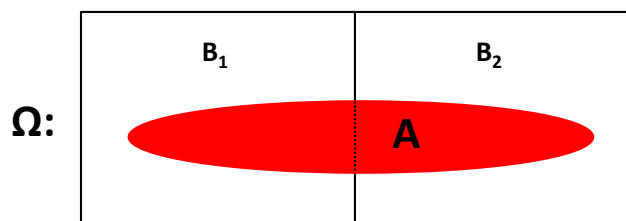
dvs. $B_1 \cap B_2 = \emptyset$. Dermed er tilhørende sannsynlighet:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) \quad (1.207)$$

Via multiplikasjonssetningen i lign.(1.167) får vi dermed:

$$\underline{\underline{P(A) = P(A|B_1) P(B_1) + P(A|B_2) P(B_2)}} \quad (1.208)$$

■



Figur 1.30: **Oppsplitting** av sannsynlighetsrom Ω , jfr. lign.(1.206).

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: ³⁸

$$P(H) = 0.32 \quad (\text{sannsynlighet for høy økonomisk vekst}) \quad (1.210)$$

$$P(M) = 0.51 \quad (\text{sannsynlighet for middels økonomisk vekst}) \quad (1.211)$$

$$P(L) = 0.17 \quad (\text{sannsynlighet for lav økonomisk vekst}) \quad (1.212)$$



Figur 1.31: Økonomisk vekst.

³⁸Legg merke til at lign.(1.210)-(1.212) representerer en gyldig sannsynlighetsfordeling siden:

$$\underline{\underline{P(\text{høy}) + P(\text{middels}) + P(\text{lav}) = 0.32 + 0.51 + 0.17 = \underline{1}}} \quad (1.209)$$

Ovenfor har vi notasjonen:

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

Eksperiment:

Observerer hvordan valutaen styrker seg.

Utfallsrom:

Utfallsrommet består av alle dubletter:

$$\Omega = \left\{ \overbrace{(H, S), (H, \bar{S})}^{= H}, \overbrace{(M, S), (M, \bar{S})}^{= M}, \overbrace{(L, S), (L, \bar{S})}^{= L} \right\} \quad (1.213)$$

$$= \{ H, M, L \} \quad (1.214)$$

hvor vi har definert følgende begivenheter:

$$H = \{ (H, S), (H, \bar{S}) \} \quad (\text{høy økonomisk vekst}) \quad (1.215)$$

$$M = \{ (M, S), (M, \bar{S}) \} \quad (\text{middels økonomisk vekst}) \quad (1.216)$$

$$L = \{ (L, S), (L, \bar{S}) \} \quad (\text{lav økonomisk vekst}) \quad (1.217)$$

$$S = \{ (H, S), (M, S), (L, S) \} \quad (\text{valutaen styrker seg}) \quad (1.218)$$

hvor (se lign.(1.210)-(1.212))

$$\underline{\underline{P(H) + P(M) + P(L) = 0.32 + 0.51 + 0.17 = \underline{\underline{1}}}} \quad (1.219)$$

representerer en gyldig sannsynlighetsfordeling.

a) Hva er $P(S)$? ³⁹

b) Hva er $P(H|S)$? ⁴⁰

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

³⁹Dvs. finn sannsynligheten for at valutaen styrker seg i Østerrike.

⁴⁰Dvs. 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.(1.210) og (1.212), så kan vi bruke setning om *oppsplitting* av Ω i lign.(1.207):

$$\underline{\underline{P(S)}} = \underbrace{P(S|H)}_{=0.7} \underbrace{P(H)}_{=0.32} + \underbrace{P(S|M)}_{=0.5} \underbrace{P(M)}_{=0.51} + \underbrace{P(S|L)}_{=0.2} \underbrace{P(L)}_{=0.17} \quad (1.220)$$

$$= 0.7 \cdot 0.32 + 0.5 \cdot 0.51 + 0.2 \cdot 0.17 \approx \underline{\underline{0.513}} \quad (1.221)$$

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.(1.183):

$$\underline{\underline{P(H|S)}} \stackrel{\text{lign. (1.183)}}{=} \underbrace{P(S|H)}_{=0.7} \cdot \frac{\overbrace{P(H)}^{=0.32}}{\underbrace{P(S)}_{=0.513}} = 0.7 \cdot \frac{0.32}{0.513} \approx \underline{\underline{0.44}} \quad (1.222)$$

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.(1.173).

- i) Siden vi kjenner $P(H|S)$ og $P(S)$, så kan vi finne $P(S \cap H)$ via multiplikasjonssetningen i lign.(1.167):

$$\underline{P(H \cap S)} \stackrel{\text{lign. (1.167)}}{=} \underbrace{P(H|S)}_{\approx 0.44} \underbrace{P(S)}_{= 0.513} \approx \underline{0.22} \quad (1.223)$$

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* (1.170) gir:

$$\underline{P(S \cap H)} \stackrel{\text{lign. (1.170)}}{=} \underbrace{P(S|H)}_{= 0.7} \underbrace{P(H)}_{= 0.32} \approx \underline{0.22} \quad (1.224)$$

dvs. samme svar som i lign.(1.223), selvfølgelig.

- ii) Produktet er lett å regne ut:

$$\underline{P(S) P(H)} = 0.513 \cdot 0.32 \approx \underline{0.16} \quad (1.225)$$

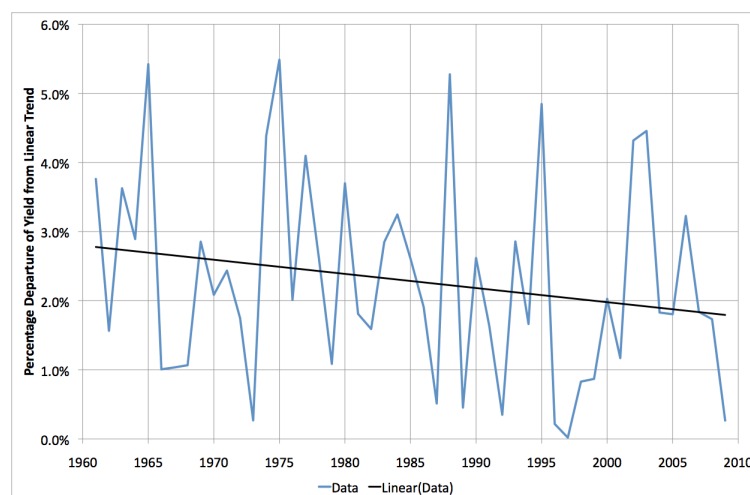
Konklusjon:

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

■

Kapittel 2

Stokastiske variabler, forventning og varians



Figur 2.1: Forventning og varians.

2.1 Stokastiske variabler

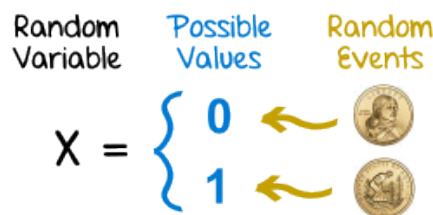
Definisjon: (stokastisk variabel) ¹

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)

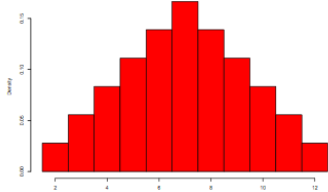
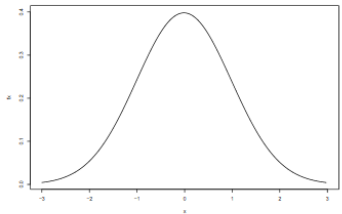
En diskret stokastisk variabel er en størrelse X som kan ha et endelig antall eller et tellbart uendelig antall mulige ulike verdier x med ulike sannsynligheter.



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 $P(X = x)$.

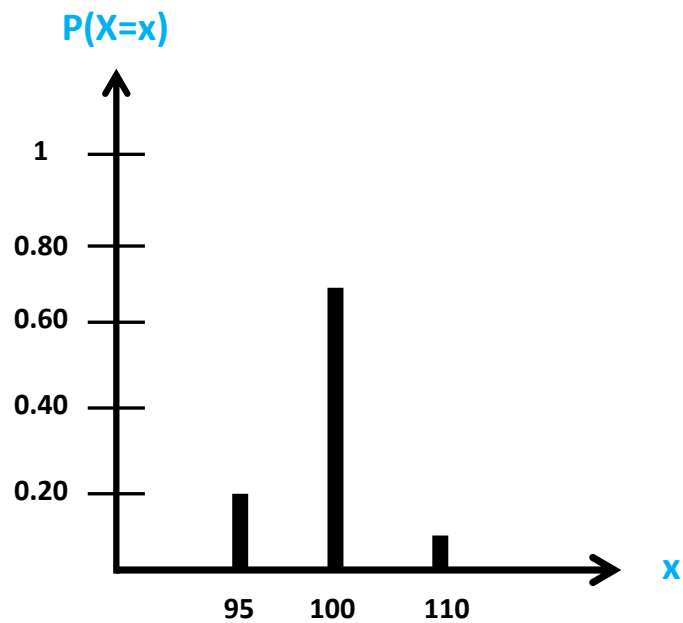


Figur 2.4: National Oilwell Varco, NOV.

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 $\sum_i P(x_i) = 1$.

■

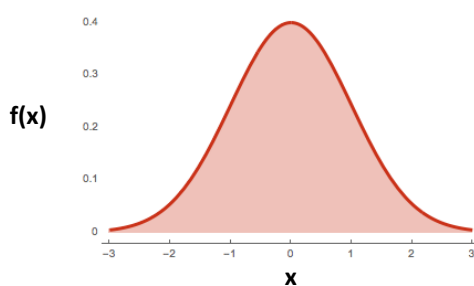
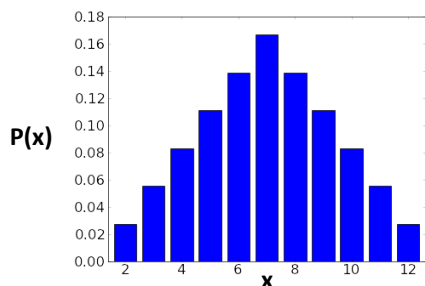
Definisjon: ([sannsynlighetsfordeling](#), kontinuert)

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

■



Figur 2.6: Diskret VS kontinuert.

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:

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

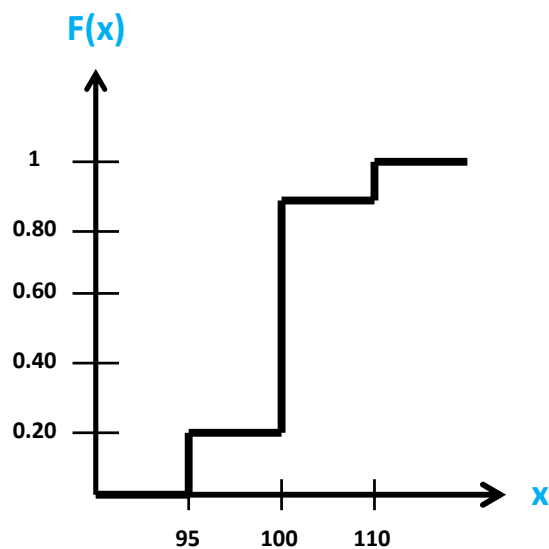
Tilsvarende: sannsynligheten at den tilfeldige variabelen X har verdi *mindre eller lik* 100 er (se lign.(2.2) og (2.3))

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

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

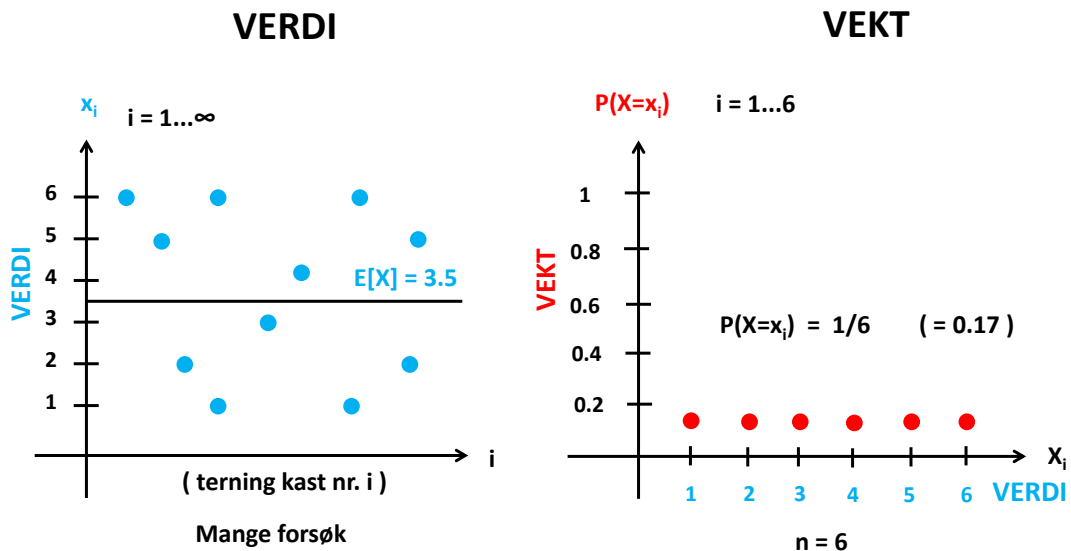
Forventningsverdi:

- 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:

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

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

Varians $Var[X]$ i antall øyne i et teringkast:

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

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

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

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

Varians $Var[X]$ er på formen:

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

■

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:

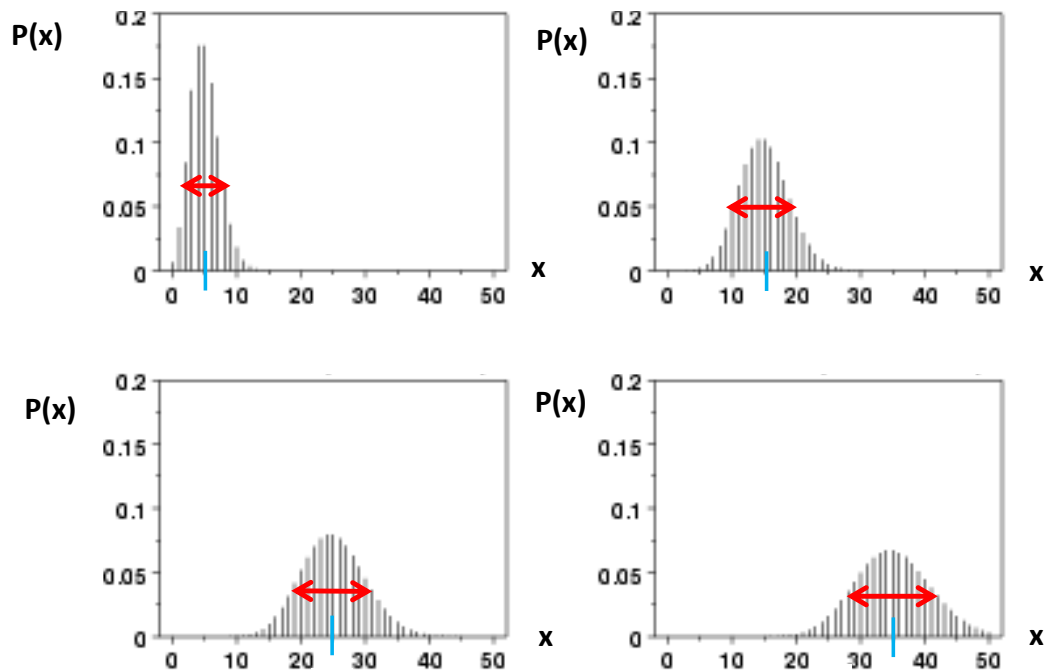
$$Var[X] \stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^n (x_i - E[X])^2 P(X = x_i) \quad (2.29)$$

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:

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

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

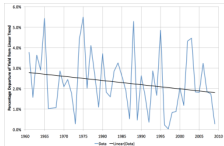
$$= E[X^2] - 2E[X] E[X] + (E[X])^2 \quad (2.34)$$

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

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:

$$\overbrace{\text{Cov}[X, Y]}^{\text{samvariasjon/korrelasjon}} = E[(X - E[X])(Y - E[Y])] \quad (2.36)$$

■

Kovariansen er et mål på grad av samvariasjon/korrelasjon:

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



Figur 2.11: Samvariasjon.

Diskret:

$$Cov[X, Y] = \sum_{i=1}^n \sum_{j=1}^m \left(x_i - E[X] \right) \left(y_j - E[Y] \right) p(x_i, y_j) \quad (2.37)$$

Kontinuerlig:

$$Cov[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - E[X] \right) \left(y - E[Y] \right) p(x, y) \, dx \, dy \quad (2.38)$$

Definisjon: (*simultan*fordeling)

La X og Y være to stokastiske variabler. Med simultanfordeling menes:

$$P(X = x, Y = y) = p(x, y) \quad (2.39)$$

■

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

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

Definisjon: (*marginal*fordeling)

La X og Y være to stokastiske variabler. Med marginalfordeling menes:

$$P(X = x) = \sum_y p(x, y) \quad (2.40)$$

$$P(Y = y) = \sum_x p(x, y) \quad (2.41)$$

■

Marginalfordeling er som en “**individuell**” sannsynlighet:

- den sier noe om hvor stor sannsynlighet det er for at $X = x$, uansett hva verdien av Y er
- den sier noe om hvor stor sannsynlighet det er for at $Y = y$, uansett hva verdien av X er

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 til X og Y .



Figur 2.13: Oslo Børs og Seadrill.

Regn ut kovariansen for de stokastiske variablene X og Y .

Vil store Y høre sammen med små Y ? Eller omvendt?

Løsning:

$$\underline{\underline{Cov[X, Y]}} \stackrel{\text{lin. (2.36)}}{=} E \left[(X - E[X])(Y - E[Y]) \right] \quad (2.42)$$

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

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

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

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

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

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

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

$$\begin{aligned} &= (-30) \cdot (-20\,000) \cdot \frac{1}{12} + (-30) \cdot 20\,000 \cdot \frac{3}{12} \\ &+ 0 + 0 \\ &+ 30 \cdot (-20\,000) \cdot \frac{3}{12} + 30 \cdot 20\,000 \cdot \frac{1}{12} \end{aligned} \quad (2.43)$$

$$= \underline{\underline{-200\,000}} \quad (2.44)$$

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” , den [generelle](#))

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

$$\overbrace{Cov[X, Y]}^{\text{samvariasjon/korrelasjon}} = E[XY] - E[X]E[Y] \quad (2.45)$$

Bevis:

Vi hopper over beviset i dette kompendiet.



Figur 2.14: Samvariasjon.

Definisjon: (uavhengighet) ⁴

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

$$P(X = x, Y = y) = P(X = x) P(Y = y) \quad (2.46)$$

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.(1.173).

- Her, i lign.(2.46), dreier deg seg uavhengighet mellom *stokastiske variabler* X og Y .
- I lign.(1.173), dvs. den spesielle multiplikasjonssetningen $P(A \cap B) = P(A) P(B)$, dreier deg seg uavhengighet mellom *begivenheter* A og B .

⁴Jamfør den *analoge* setn. for uavhengighet mellom begivenheter A og B i lign.(1.173): $P(A \cap B) = P(A) \cdot P(B)$.

Setning: (uavhengighet)

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

$$\boxed{E[XY] = E[X] E[Y]} \quad (2.47)$$

■

Bevis:

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

$$\overbrace{P(X = x_i, Y = y_j)}^{\text{uavhengig}} \stackrel{\text{lin. (2.46)}}{=} \underbrace{P(X = x_i)}_{= p(x_i)} \underbrace{P(Y = y_j)}_{= p(y_j)} \quad (2.48)$$

Ut fra definisjonen av forventning får vi da:

$$\underline{\underline{E[XY]}} \stackrel{\text{lin. (2.17)}}{=} \sum_{i=1}^m \sum_{j=1}^n x_i \cdot y_j \cdot p(x_i, y_j) \quad (2.49)$$

$$\stackrel{\text{uavh.}}{=} \sum_{i=1}^m \sum_{j=1}^n x_i \cdot y_j \cdot p(x_i) \cdot p(y_j) \quad (2.50)$$

$$= \underbrace{\left(\sum_{i=1}^m x_i \cdot p(x_i) \right)}_{= E[X]} \underbrace{\left(\sum_{j=1}^n y_j \cdot p(y_j) \right)}_{= E[Y]} \quad (2.51)$$

$$= \underline{\underline{E[X] E[Y]}} \quad (2.52)$$

■

Setning: (“kovarianssetningen” , den spesielle)

La X og Y være to stokastiske variabler. Dersom X og Y er uavhengige, så er ⁵

$$\underbrace{Cov[X, Y]}_{\text{ukorrelererte}} = 0 \quad (2.53)$$

■

Bevis:

Dersom X og Y er uavhengige så vet vi fra setningen i lign.(2.47) på siden 117 at

$$\overbrace{E[X \cdot Y]}^{X \text{ og } Y \text{ uavhengig}} = E[X] E[Y] \quad (2.54)$$

Dermed:

$$Cov[X, Y] = \overbrace{E[X \cdot Y]}^{\text{uavh.}} - \cancel{E[X] E[Y]} = 0 \quad (2.55)$$

dvs.

$$\underbrace{Cov[X, Y]}_{X \text{ og } Y \text{ ukorrelererte}} = 0 \quad (2.56)$$

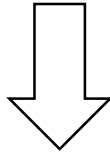
■

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

Konklusjon: Dersom X og Y er uavhengige så er de også ukorrelerte. ⁶

uavhengig

(verdensrekord)



ukorrelert

(norgesrekord)

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



Figur 2.16: Skøyteløper.

⁶Uavhengighet er altså et **st**erkere krav enn ukorrelert.
Dersom en skøyteløper har verdenrekorden på 10 000 meter så har han/hun også norgesrekorden.

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

La oss igjen se på eksemplet fra side 111:

- 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			
		Y = 80 000	Y = 120 000	P(X=x)	
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 2.17: **Simultan**fordelingen $p(x, y)$ og **marginal**fordelingene $P(X = x)$ og $P(Y = y)$.

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



Figur 2.18: Oslo Børs og Seadrill.

Løsning:

Forventet aksjekurs (pris) på Seadrill-aksjene:

$$\begin{aligned}\underline{E[X]} &\stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^3 x_i \cdot P(X = x_i) \\ &= \left(70 \cdot \frac{1}{3} + 100 \cdot \frac{1}{3} + 130 \cdot \frac{1}{3} \right) \text{ NOK} = \underline{100 \text{ NOK}}\end{aligned}\quad (2.57)$$

Forventet antall aksjoner som omsettes:

$$\begin{aligned}\underline{E[Y]} &\stackrel{\text{lign. (2.17)}}{=} \sum_{i=1}^2 y_i \cdot P(Y = y_i) \\ &= 80\,000 \cdot \frac{1}{2} + 120\,000 \cdot \frac{1}{2} = \underline{100\,000}\end{aligned}\quad (2.58)$$

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

$$\begin{aligned}\underline{E[XY]} &= \sum_{i=1}^3 \sum_{j=1}^2 x_i \cdot y_j \cdot p(x_i, y_j) \\ &= \left(70 \cdot 80\,000 \cdot \frac{1}{12} + 100 \cdot 80\,000 \cdot \frac{1}{6} + 130 \cdot 80\,000 \cdot \frac{3}{12} \right. \\ &\quad \left. + 70 \cdot 120\,000 \cdot \frac{3}{12} + 100 \cdot 120\,000 \cdot \frac{1}{6} + 130 \cdot 120\,000 \cdot \frac{1}{12} \right) \text{ NOK} \\ &= \underline{9\,800\,000 \text{ NOK}}\end{aligned}\quad (2.59)$$
$$(2.60)$$

Fra lign.(2.57) og (2.58) ser vi at $E[X] \cdot E[Y] = 10\,000\,000$, dvs.:

$$E[XY] \neq E[X] \cdot E[Y] \quad (2.61)$$

Konklusjon: X og Y er avhengige.

■

Eksempel: (økonomi - kovarians)

La oss atter en gang se på eksemplet fra side 111.
Ved å bruke definisjonen av

$$\begin{aligned}\underline{\underline{Cov[X, Y]}} &\stackrel{\text{lign. (2.36)}}{=} E\left[(X - E[X])(Y - E[Y])\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m (x_i - E[X])(y_j - E[Y]) p(x_i, y_j)\end{aligned}\tag{2.62}$$

så innså vi, se side 113, at det var mye arbeid.

Regn ut kovariansen $Cov[X, Y]$ uten å bruke definisjonen i lign.(2.62).



Figur 2.19: Oslo Børs og Seadrill.

Løsning:

Siden vi kjenner $E[X \cdot Y]$ så vel som $E[X]$ og $E[Y]$, se side 121, så får vi:

$$Cov[X, Y] \stackrel{\text{lign. (2.45)}}{=} \underbrace{E[X \cdot Y]}_{= 9\,800\,000} - \underbrace{E[X]}_{= 100\,000} \cdot \underbrace{E[Y]}_{= 100} = \underline{\underline{-200\,000}}\tag{2.63}$$

Med andre ord:

svært arbeidsbesparende i forhold å bruke definisjonen.

■

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:

$$E[a] = a \quad (2.64)$$

$$E[a + X] = a + E[X] \quad (2.65)$$

$$E[a \cdot X] = a E[X] \quad (2.66)$$

$$E[aX + bY] = aE[X] + bE[Y] \quad (2.67)$$

Regneregler for varians:

$$Var[a] = 0 \quad (2.68)$$

$$Var[a + X] = Var[X] \quad (2.69)$$

$$Var[a X] = a^2 Var[X] \quad (2.70)$$

$$Var[X] = Var[-X] \quad (2.71)$$

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2ab Cov[X, Y] \quad (2.72)$$

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

Regnereglen i lign.(2.72) på side 124 er så viktig at den formuleres ofte i en egen setning:

Setning: (kovarians II)

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

$$\overbrace{\text{Var}[aX + bY]}^{\text{variasjon/(spredning)}} = a^2 \text{Var}[X] + b^2 \text{Var}[Y] + 2ab \overbrace{\text{Cov}[X, Y]}^{\text{samvariasjon}} \quad (2.73)$$

hvor a og b er konstanter.

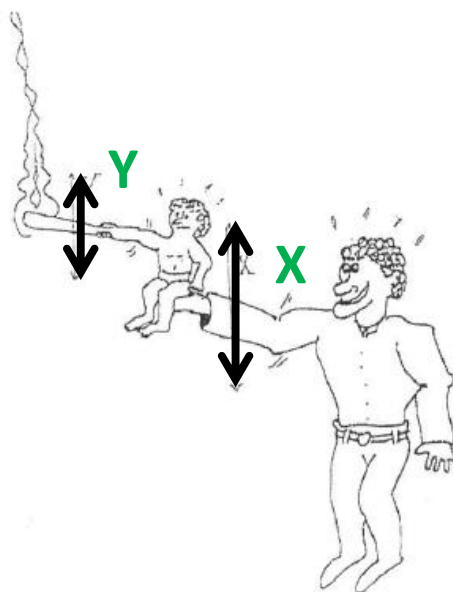
■

Bevis:

Lign.(2.73) bevises ved å bruke første kvadratsetning, (se f.eks. læreboken).

■

Visualisering av variasjon og samvariasjon:



Figur 2.20: **Variasjon** og **samvariasjon**, dvs. $Var[X + Y]$ og $Cov[X, Y]$.

La:

- X = guttens plassering sett fra mannens ståsted, se figur (2.20)
- Y = fakkelens plassering sett fra guttens ståsted, se figur (2.20)
- $X + Y$ = fakkelens plassering i forhold til mannens ståsted (bakken)

Fra lign.(2.73): ($a = 1$ og $b = 1$)

$$Var[1X + 1Y] \stackrel{\text{lign. (2.73)}}{=} 1^2 Var[X] + 1^2 Var[Y] + 2 \cdot 1 \cdot 1 Cov[X, Y] \quad (2.74)$$

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 [ukorrelerete](#), dvs.:

$$\overbrace{Cov[X, Y]}^{X \text{ og } Y \text{ ukorrelerete}} = 0 \quad (2.75)$$

så er:

$$\boxed{Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] \quad (2.76)}$$

■

Eksempel: (økonomi - gler , 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:

$$X = \text{antall mål som KBK scorer i løpet av en tilfeldig valgt kamp} \quad (2.77)$$

Basert på historiske data har de funnet ut at X har fordelingen som vist i figur (2.21): ⁷

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

Figur 2.21: Sannsynlighetsfordeling for $P(X = x_i)$.



Figur 2.22: Kristiansund Ballklubb.

⁷Ut 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:

$$B = cX^2 + d \quad (2.78)$$

hvor $c = 750$ NOK og $d = 1500$ NOK.⁸

- b) Forklar hvorfor B er en stokastisk variabel.
- c) Hva er forventet bonus for en tilfeldig valgt kamp, $E[B]$?

⁸Dette betyr at c og d er **konstanter**, mens X er en stokastisk **variabel**.

Løsning:

- a) Forventet antall mål scoret per kamp:

$$\underline{\underline{E[X]}} = \sum_{i=0}^5 x_i P(X = X_i) \quad (2.79)$$

$$= 0 \cdot 0.20 + 1 \cdot 0.30 + 2 \cdot 0.20 + 3 \cdot 0.15 + 4 \cdot 0.10 + 5 \cdot 0.05 \quad (2.80)$$

$$= \underline{\underline{1.8}} \quad (2.81)$$

- 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:

$$\underline{E[B]} = E[cX^2 + d] \quad (2.82)$$

$$\stackrel{\text{lign. (2.67)}}{=} \underline{cE[X^2] + d} \quad (2.83)$$

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

$$\underline{E[X^2]} = \sum_{i=0}^5 x_i^2 P(X = X_i) \quad (2.84)$$

$$= 0^2 \cdot 0.20 + 1^2 \cdot 0.30 + 2^2 \cdot 0.20 + 3^2 \cdot 0.15 + 4^2 \cdot 0.10 + 5^2 \cdot 0.05 \quad (2.85)$$

$$= \underline{\underline{5.3}} \quad (2.86)$$

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

$$\underline{\underline{E[B]}} = cE[X^2] + d \quad (2.87)$$

$$= \left(750 \cdot 5.3 + 1\,500 \right) \text{ NOK} = \underline{\underline{5\,475 \text{ NOK}}} \quad (2.88)$$

■

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

$$N = 100\,000 \quad (2.89)$$

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 :

$$N \cdot a = \text{antall aksjer som investeres i selskap } A, \text{ Statoil} \quad (2.90)$$

$$N \cdot b = \text{antall aksjer som investeres i selskap } B, \text{ Seadrill} \quad (2.91)$$

$$N \cdot c = \text{antall aksjer som investeres i selskap } C, \text{ Yara} \quad (2.92)$$

hvor

$$a + b + c = 1 \quad (2.93)$$

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



Figur 2.23: 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 én aksje (altså aksjekurs) for selskap A , Statoil (i NOK), **om ett år** (2.94)

Y = pris på én aksje (altså aksjekurs) for selskap B , Seadrill (i NOK), **om ett år** (2.95)

Z = pris på én aksje (altså aksjekurs) for selskap C , Yara (i NOK), **om ett år** (2.96)

Anta videre at forventet pris (altså aksjekurs) **om ett år** er: ⁹

$$E[X] = 90 \quad , \quad E[Y] = 125 \quad , \quad E[Z] = 180 \quad (2.97)$$

med tilhørende varianser: ¹⁰

$$Var[X] = 100 \quad , \quad Var[Y] = 200 \quad , \quad Var[Z] = 600 \quad (2.98)$$



Figur 2.24: Gjelsten Holding.

⁹Forventningene er i NOK, men dropper benevningen her for enkelhets skyld.

¹⁰Variansene er i NOK².

- a) Vis at fortjenesten F om ett år er gitt ved: ¹¹

$$F = Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \quad (2.100)$$

- b) Vis at den forventede fortjenesten $E[F]$ ved et eventuelt salg av aksjene om ett år er: ¹²

$$E[F] = 10N(-a + 2b + 5c) \quad (2.101)$$

- c) Anta nå at aksjene er uavhengige.
 Bruk lign.(2.100) samt lign.(2.98) og vis at variansen $Var[F]$ til fortjenesten ved et eventuelt salg om ett år er:

$$Var[F] = 100N^2(a^2 + 2b^2 + 6c^2) \quad (2.102)$$

¹¹Bruk gjerne at:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (2.99)$$

for hver av de tre selskapene. Bruk også lign.(2.90)-(2.92) samt lign.(2.94)-(2.96).

¹²Bruk lign.(2.100).

- 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.(2.102) er minst mulig når

$$a = 0.6 \quad , \quad b = 0.3 \quad , \quad c = 0.1 \quad (2.103)$$

(Du skal ikke vise lign.(2.103). 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? ¹³

- e) Dersom du bestemmer deg for å maksimere den forventede fortjenetsen $E[F]$ i lign.(2.101) 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?

¹³Bruk lign.(2.90)-(2.92) sammen med lign.(2.103).

Løsning:

a) Fortjenesten F er:

$$F = \text{antall aksjer} \cdot \text{pris per aksje om ett år} - \text{antall aksjer} \cdot \text{pris per aksje i dag} \quad (2.104)$$

for de tre selskapene så får vi:

$$\begin{aligned} \underline{\underline{F}} &= \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ om ett år}}_{= X} \\ &+ \underbrace{\text{antall aksjer i } B}_{= N \cdot b} \cdot \underbrace{\text{pris per aksje for selskap } B \text{ om ett år}}_{= Y} \\ &+ \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ om ett år}}_{= Z} \end{aligned} \quad (2.105)$$

$$- \underbrace{\text{antall aksjer i } A}_{= N \cdot a} \cdot \underbrace{\text{pris per aksje for selskap } A \text{ i dag}}_{= 100} \quad (2.106)$$

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

$$- \underbrace{\text{antall aksjer i } C}_{= N \cdot c} \cdot \underbrace{\text{pris per aksje for selskap } C \text{ i dag}}_{= 130} \quad (2.107)$$

$$= NaX + NbY + NcZ - Na100 + Nb105 + Nc130 \quad (2.108)$$

$$= \underline{\underline{Na(X - 100) + Nb(Y - 105) + Nc(Z - 130)}}, \quad \text{q.e.d.} \quad (2.109)$$

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

$$\underline{E[F]} = E \left[Na(X - 100) + Nb(Y - 105) + Nc(Z - 130) \right] \quad (2.110)$$

$$= Na \left(\underbrace{E[X]}_{=90} - 100 \right) + Nb \left(\underbrace{E[Y]}_{=125} - 105 \right) + Nc \left(\underbrace{E[Z]}_{=180} - 130 \right) \quad (2.111)$$

$$= Na(90 - 100) + Nb(125 - 105) + Nc(180 - 130) \quad (2.112)$$

$$= Na(-10) + Nb 20 + Nc 50 \quad (2.113)$$

$$= \underline{\underline{10N(-a + 2b + 5c)}} \quad , \quad \text{q.e.d.} \quad (2.114)$$

- c) Fra lign.(2.72) vet vi formelen for variansen av en lineær kombinasjon av to stokastiske variabler X og Y , nemlig:

$$\overbrace{Var[aX + bY]}^{\text{variasjon/(spredning)}} = a^2 Var[X] + b^2 Var[Y] + 2ab \overbrace{Cov[X, Y]}^{\text{samvariasjon}} \quad (2.115)$$

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:

$$\underline{\underline{Var[F]}} = Var \left[\textcolor{blue}{N}a(X - 100) + \textcolor{blue}{N}b(Y - 105) + \textcolor{blue}{N}c(Z - 130) \right] \quad (2.116)$$

$$\begin{aligned} &= (Na)^2 \underbrace{Var[X]}_{= 100} + (Nb)^2 \underbrace{Var[Y]}_{= 200} + (Nc)^2 \underbrace{Var[Z]}_{= 600} \\ &- (Na)^2 \underbrace{Var[100]}_{= 0} + (Nb)^2 \underbrace{Var[105]}_{= 0} + (Nc)^2 \underbrace{Var[130]}_{= 0} \end{aligned} \quad (2.117)$$

$$= N^2 (100a^2 + 200b^2 + 600c^2) \quad (2.118)$$

$$= \underline{\underline{100N^2 (a^2 + 2b^2 + 6c^2)}} \quad , \quad \text{q.e.d.} \quad (2.119)$$

- d) Dersom man ønsker minst mulig risiko i investeringen så tilsvare 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:

$$\underline{\underline{Na}} = 100\,000 \cdot 0.6 = \underline{\underline{60\,000}} \quad (2.120)$$

$$\underline{\underline{Nb}} = 100\,000 \cdot 0.3 = \underline{\underline{30\,000}} \quad (2.121)$$

$$\underline{\underline{Nc}} = 100\,000 \cdot 0.1 = \underline{\underline{10\,000}} \quad (2.122)$$

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

$$E[F] = 10N(-a + 2b + 5c) \quad (2.123)$$

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

$$a = 0 \quad , \quad b = 0 \quad , \quad c = 1 \quad (2.124)$$

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

$$\underline{\underline{E[F]}} = 10N \cdot 5c \quad (2.125)$$

$$= 10 \cdot 100\,000 \cdot 5 \cdot 1 \text{ NOK} = \underline{\underline{5\,000\,000 \text{ NOK}}} \quad (2.126)$$

altså 5 mill. NOK.



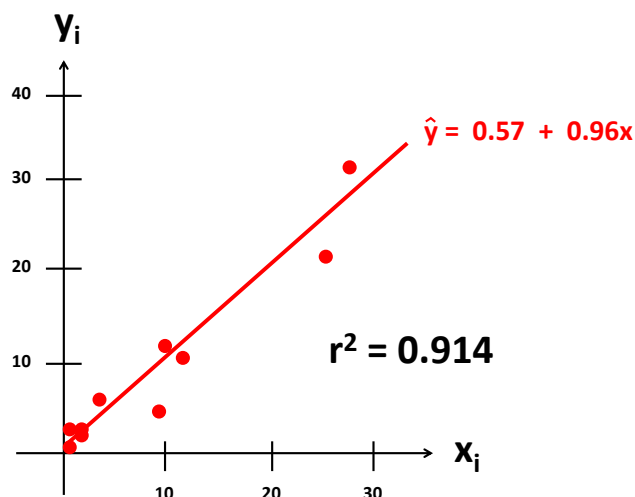
Fra lign.(8.36) vet vi gjennomsnittet $\bar{y} = 9.4$.
Dermed kan regne ut sse fra lign. (8.44):

$$\begin{aligned}
 \underline{sst} &= \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= (3 - 9.4)^2 + (11 - 9.4)^2 \\
 &+ (3 - 9.4)^2 + (1 - 9.4)^2 \\
 &+ (12 - 9.4)^2 + (21 - 9.4)^2 \\
 &+ (6 - 9.4)^2 + (4 - 9.4)^2 \\
 &+ (31 - 9.4)^2 + (2 - 9.4)^2 \\
 &= \underline{858.4}
 \end{aligned} \tag{8.57}$$

Forklaringskraft R^2 er da: (se lign.(8.48)

$$\underline{\underline{r^2}} = 1 - \frac{sse}{sst} = 1 - \frac{74.2}{858.4} = \underline{\underline{0.914}} \tag{8.58}$$

dvs. regresjonslinjen forklarer hele 91.4 % av den totale variasjonen.
Vi sier at modellen har stor forklaringskraft.



Figur 8.13: [Grafisk fremstilling](#) av x og y .

Kommentar:

Som regel gjør man ikke all denne regningen “for hånd”.

Mange dataprogrammer kan hjelpe oss å regne ut statistiske størrelser, f.eks. Excel.

Dersom vi bruker Excel på eksemplet vårt så får vi en utskrift som vist i figuren nedenfor. Da kan vi lese av de størrelsene vi regnet ut “for hånd” direkte fra Excel-utskriften.

A	B	C	D	E	F	G	H	I
SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0,95579703							
R Square	0,91354796		$r^2 = 0.914$					
Adjusted R Square	-1,25							
Standard Error	3,04570245							
Observations	1							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	10	784,1895726	78,41895726	84,5368609	#NUM!			
Residual	8	74,21042743	9,27630343					
Total	18	858,4						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	0,57162987	1,359998689	0,42031649	0,68531665	-2,564532724	3,70779247	-2,56453272	3,70779247
X Variable	0,95960545	0,104368549	9,19439291	1,5833E-05	0,718931143	1,20027975	0,718931143	1,200279754

$\hat{b} = 0.96$

$\hat{a} = 0.57$

$sse = 74.2$

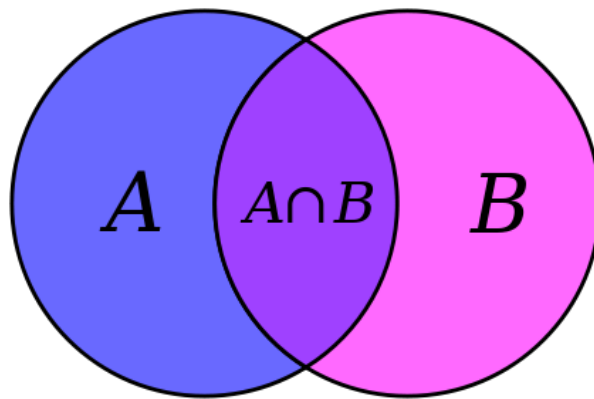
$sst = 858.4$

Figur 8.14: Utskrift fra Excel.



Tillegg A

Mengdelære

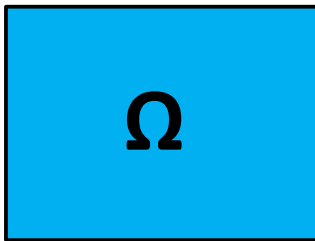


Figur A.1: Mengdelære.

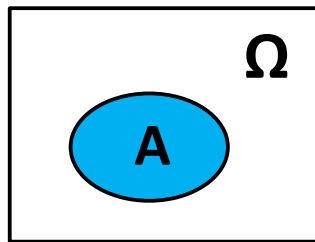
A.1 Venndiagrammer

Venn-diagrammer er illustrasjoner som brukes bl.a. innen mengdelære. De brukes for å visualisere de matematiske forbindelsene mellom ulike grupper av mengder. La oss illustrere dette via de mest velkjente eksemplene.

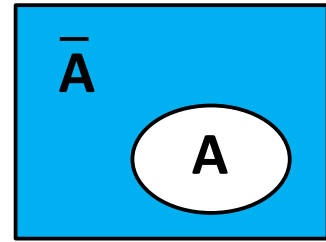
For et eksperiment, la A og B være to begivenheter i utfallsrommet Ω .



Utfallsrommet Ω er hele det blå området.

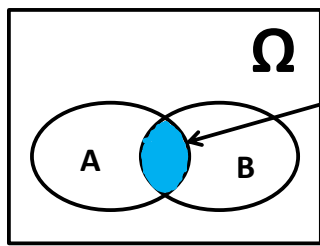


Begivenheten A visualiseres ved det blå området.



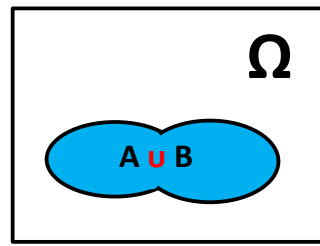
$\bar{A}$ benevnes "ikke" A

Figur A.2: Utfallsrommet Ω , begivenheten A og komplementet $\bar{A}$.



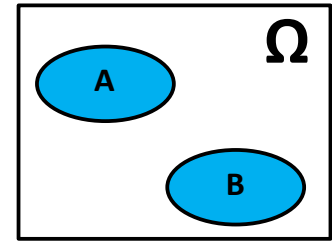
Snitt:

$A \cap B$ betyr A **og** B.
Tilsvare **OVERLAPP**
av mengder.



Union:

$A \cup B$ betyr A **eller** B.
Tilsvare **SUM** av
mengder.



Disjunkt:

$A \cap B = \emptyset$.
A og B inntreffer **ALDRI**
samtidig.
Ingen felles elementer.

Figur A.3: Snitt, union og disjunkt.

Ut fra disse [Venn-diagrammene](#) kan vi merke oss følgende huskereglar:

$\cap$	=	og	, tilsvare OVERLAPP av mengder
$\cup$	=	eller	, tilsvare SUM av mengder

Eksempel: ([terningkast](#))

La oss kaste en terning èn gang. La oss se på begivenhetene:

A = resultat med **1**, **3** eller **4** øyne på terningen (A.1)

B = resultat med **3**, **4** eller **5** øyne på terningen (A.2)

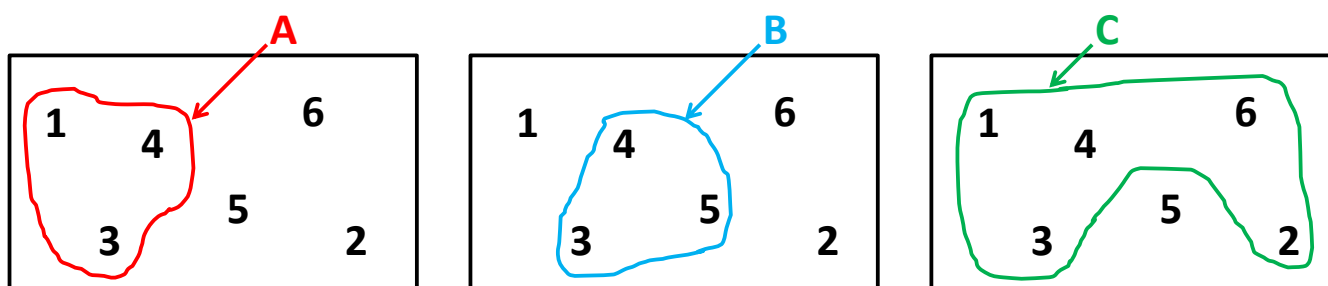
C = resultat hvor 5 ikke fås (A.3)

Matematisk kan disse begivenhetene skrives:

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

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

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



Figur A.4: Begivenhetene A , B og C .



Figur A.5: Terning.

$\cap$ = “og”, tilsvarende OVERLAPP:

$$\begin{array}{lll} \textcolor{red}{A} \overbrace{\cap}^{\text{og}} \textcolor{blue}{B} & = & \{3, 4\} & (\text{ b\aa}de \text{ i } A \text{ og } B) & (A.7) \\ \textcolor{red}{A} \cap \textcolor{green}{C} & = & \{1, 3, 4\} & (\text{ b\aa}de \text{ i } A \text{ og } C) & (A.8) \\ \textcolor{blue}{B} \underbrace{\cap}_{\text{og}} \textcolor{green}{C} & = & \{3, 4\} & (\text{ b\aa}de \text{ i } B \text{ og } C) & (A.9) \end{array}$$

$\cup$ = “eller”, tilsvarende SUM:

$$\textcolor{red}{A} \overbrace{\cup}^{\text{eller}} \textcolor{blue}{B} = \{1, 3, 4, 5\} \quad (A \text{ eller } B \text{ eller begge }) \quad (A.10)$$

$$\textcolor{red}{A} \cup \textcolor{green}{C} = \{1, 2, 3, 4, 6\} \quad (A \text{ eller } C \text{ eller begge }) \quad (A.11)$$

$$\textcolor{blue}{B} \underbrace{\cup}_{\text{eller}} \textcolor{green}{C} = \{1, 2, 3, 4, 5, 6\} \quad (B \text{ eller } C \text{ eller begge }) \quad (A.12)$$

De komplementære begivenhetene er:

$$\overline{A} = \{2, 5, 6\} \quad (\text{ ikke i } A) \quad (A.13)$$

$$\overline{B} = \{1, 2, 6\} \quad (\text{ ikke i } B) \quad (A.14)$$

$$\overline{C} = \{5\} \quad (\text{ ikke i } C) \quad (A.15)$$

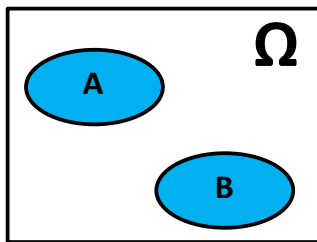
Videre har vi at:

$$A \cap \overline{B} = \{1\} \quad (A \text{ og ikke i } B) \quad (A.16)$$

■

Definisjon: (disjunkte mengder)

To mengder A og B er disjunkte dersom $A \overset{\text{og}}{\cap} B = \emptyset$.



Disjunkt:

$$A \cap B = \emptyset.$$

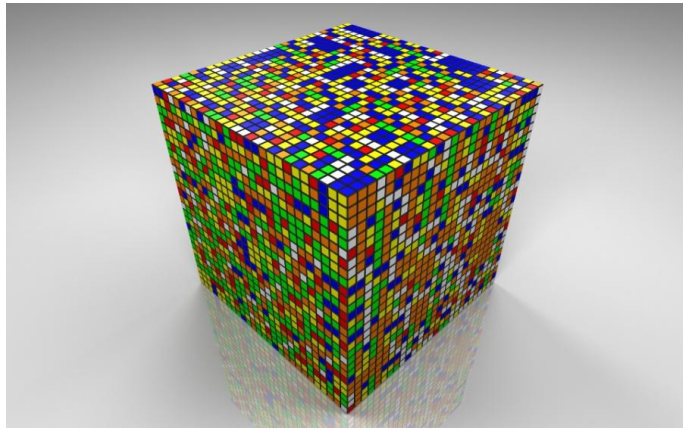
A og B inntreffer ALDRI samtidig.

Ingen felles elementer.

Figur A.6: A og B er disjunkt.

Tillegg B

Kombinatorikk



Figur B.1: Kombinatorikk.

B.1 Koblinger

Definisjon: ([koblinger](#))

Koblinger = forhold som gjør at et bestemt valg kan **påvirke** utfallet av **andre** valg vi skal gjøre.



Grunnprinsipp i kombinatorikk: ([antagelse](#))

Ingen kobling mellom mellom valgmulighetene. (Med koblinger blir det fort vanskelig).

Eksempel: (gutter og jenter)

I en klasse er det 1 gutt og 15 jenter. Vi skal velge et par, en jente og en gutt. Dersom valget av jenta ikke påvirker valget av gutten, så er:

$$\underline{\underline{\text{antall mulige valg}}} = 1 \cdot 15 = \underline{\underline{15}} \quad (\text{B.1})$$

Altså det er 15 mulige forskjellige par i klassen.



Eksempel: (gutter og jenter)

I en klasse er det 2 gutter og 15 jenter. Vi skal velge et par, en jente og en gutt. Dersom valget av jenta ikke påvirker valget av gutten, så er:

$$\underline{\underline{\text{antall mulige valg}}} = 2 \cdot 15 = \underline{\underline{30}} \quad (\text{B.2})$$

Altså det er 30 mulige forskjellige par i klassen.



Figur B.2: Gutter og jenter.

Eksempel: (fotballtipping)

I fotballtipping er det 3 muligheter, H, U og B. Det er 12 kamper på en kupong.

antall mulige valg = $\underbrace{3 \cdot 3 \cdot 3 \cdot 3 \cdot \dots \cdot 3}_{12 \text{ stk.}} \quad = \quad 3^{12} \quad = \quad \underline{\underline{531\,441}}$ (B.3)

(Dette er et eksempel på et **ordnet** utvalg *med* tilbakelegging.
I neste avsnitt kaller vi dette “situasjon 1”.)

■



Figur B.3: Tippekupong.

Eksemplet på forrige side kan vi *generalisere*:

Kombinasjoner: (uten koblinger)

Dersom vi har

$$\begin{aligned} m_1 &= \text{antall muligheter i valg nr. 1} \\ m_2 &= \text{antall muligheter i valg nr. 2} \\ &\cdot \\ &\cdot \\ m_N &= \text{antall muligheter i valg nr. } N \end{aligned} \tag{B.4}$$

da er

$\text{antall } \textcolor{red}{\text{mulige kombinasjoner}} = m_1 \cdot m_2 \cdot m_3 \cdot \dots m_N$	(B.5)
---	-------

B.2 4 situasjoner (endelig populasjon)

Det kan være vanskelig å telle opp antall elementer i utfallsrommet.
Men dersom:

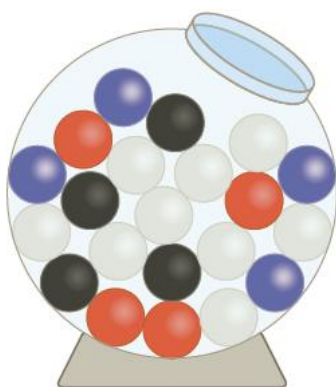
- **uniformt** utfallsrom
- utfallene ikke påvirker hverandre, dvs. de er **uavhengige**

så kan derfor ofte lønne seg å bruke

“**urnemodellen**”

Kommentarer:

- 1) At utfallsrommet er uniformt i forbindelse med urnemodellen betyr at samme sannsynlighet for å trekke kulene i urnen. Altså kulene er **like store**.
- 2) At valgmulighetene er uavhengige i forbindelse med urnemodellen betyr at kulene i urnen er **uavhengige**. Altså kulene er **ikke koplet** på noen måte

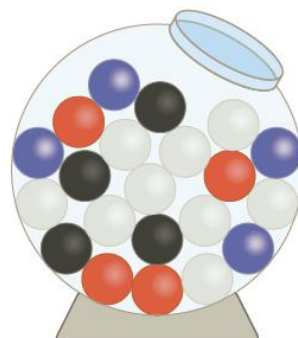


Figur B.4: Urne.

Tenk deg at alle mulige utfall av et eksperiment er representert ved kuler som ligger i en urne. Så trekker vi kuler etter tur.

Da må vi skille mellom:

- er det trekking med eller uten tilbakelegging?
- betyr det noe i hvilken rekkefølge kulene trekkes?



	trekning	
	m/tilbakelegging	u/tilbakelegging
ordnet	situasjon 1	situasjon 2
ikke-ordnet	situasjon 4 (forekommer sjelden)	situasjon 3

Figur B.5: 4 situasjoner for urnemodellen.

1) Ordnet utvalg med tilbakelegging: (situasjon 1)

- rekkefølge har **betydning**: 1, 3 er forskjellig fra 3, 1
- legger tilbake trukket element
- kombinasjonen 1, 1 er da fullt mulig, dvs. et element kan være med flere ganger

2) Ordnet utvalg uten tilbakelegging: (situasjon 2)

- rekkefølge har **betydning**: 1, 3 er forskjellig fra 3, 1
- legger ikke tilbake trukket element
- kombinasjonen 1, 1 er da umulig, dvs. ingen element kan være med flere ganger

3) Ikke-ordnet utvalg uten tilbakelegging: (situasjon 3)

- rekkefølge har ikke **betydning**: 1, 3 er samme som 3, 1
- legger ikke tilbake trukket element
- kombinasjonen 1, 1 er da umulig, dvs. ingen element kan være med flere ganger

4) Ikke-ordnet utvalg med tilbakelegging: (situasjon 4)

- rekkefølge har ikke **betydning**: 1, 3 det samme som 3, 1
- legger tilbake trukket element
- kombinasjonen 1, 1 er da fullt mulig, dvs. element kan være med flere ganger

Eksempel: (situasjon 1) (**ordnet** utvalg *med* tilbakelegging)

Hvor mange måter kan vi velge ut $s = 5$ personer på fra en gruppe på $N = 20$?

Med tilbakelegging.

Anta at alle $N = 20$ personene har et nummer, og at **rekkefølgen betyr noe**.

Løsning:

$$\underline{\underline{\# \text{ ordnede komb. m/tilbakelegging}}} = \underbrace{20 \cdot 20 \cdot 20 \cdot 20 \cdot 20}_{5 \text{ ledd}} = \underbrace{20^5}_{\text{generaliseres lett}} = \underline{\underline{3\,200\,000}} \quad (\text{B.6})$$

■

Eks. i forrige avsnitt, “fotballtiping”, er også et eksempel på situasjon 1 (med $N = 3$ og $s = 12$).

Generelt gjelder: (situasjon 1)

$\# \text{ ordnede komb. m/tilbakelegging} = N \cdot N \cdot \dots \cdot N = N^s \quad (\text{B.7})$
--

hvor

$$N = \underline{\text{totalt}} \text{ antall valgobjekter} \quad (\text{B.8})$$

$$s = \text{antall objekter som } \underline{\text{velges}} \quad (\text{B.9})$$

Eksempel: (situasjon 2) (**ordnet** utvalg *uten* tilbakelegging) ¹

Hvor mange måter kan vi velge ut $s = 5$ personer på fra en gruppe på $N = 20$? *Uten* tilbakelegging. Anta at alle $N = 20$ personene har et nummer, og at **rekkefølgen betyr noe**.

Løsning:

$$\underline{\underline{\# \text{ ordnede komb. u/tilbakelegging}}} = \underbrace{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}_{5 \text{ ledd}} = \underbrace{\frac{20!}{(20-5)!}}_{\text{generaliseres lett}} = \underline{\underline{1\,860\,480}} \quad (\text{B.10})$$

■

Generelt gjelder: (situasjon 2)

$$\# \text{ ordnede komb. u/tilbakelegging} = N \cdot (N-1) \cdot \dots \cdot (N-s+1) = \frac{N!}{(N-s)!} \quad (\text{B.11})$$

hvor

$$N = \underline{\text{totalt}} \text{ antall valgobjekter} \quad (\text{B.12})$$

$$s = \text{antall objekter som } \underline{\text{velges}} \quad (\text{B.13})$$

¹Fakultet: $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$. I tillegg så definerer vi $0! = 1$ og $1! = 1$.

Eksempel: (situasjon 3) (**ikke-ordnet** utvalg *uten* tilbakelegging)

Hvor mange måter kan vi velge ut $s = 5$ personer på fra en gruppe på $N = 20$? *Uten* tilbakelegging. Rekkefølgen betyr *ikke* noe.

Løsning:

$$\# \text{ ordnede komb. u/tilbakelegging} = 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 = \underline{1\,860\,480} \quad \left(= \frac{20!}{(20-5)!} \right) \quad (\text{B.14})$$

Hver gang vi velger ut 5 personer så kan disse personene ordnes på $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ forskjellige måter/("rekkefølge"). Antall **ikke-ordnede** kombinasjoner blir derfor:

$$\begin{aligned} \# \text{ ikke-ordnede komb. u/tilbakelegging} &= \frac{\# \text{ ordnede komb. u/tilbakelegging}}{\# \text{ måter å sortere på}} \\ &= \frac{1\,860\,480}{120} = \frac{20!}{\underbrace{(20-5)! \cdot 5!}_{\text{generaliseres lett}}} = \underline{15\,504} \quad (\text{B.15}) \end{aligned}$$

■

Generelt gjelder: (situasjon 3)

$$\# \text{ ikke-ordnede komb. u/tilbakelegging} = \frac{N!}{(N-s)! s!} \equiv \underbrace{\binom{N}{s}}_{\text{binomialkoeff.}} \quad (\text{B.16})$$

hvor

$$N = \text{totalt antall valgobjekter} \quad (\text{B.17})$$

$$s = \text{antall objekter som velges} \quad (\text{B.18})$$

$$\frac{N!}{(N-s)! s!} \equiv \underbrace{\binom{N}{s}}_{\text{binomialkoeff.}} \quad (\text{"N over s"}) \quad (\text{B.19})$$

Eksempel: (situasjon 4) (ikke-ordnet utvalg *med* tilbakelegging)

Komplisert situasjon. Forekommer sjelden.
Denne situasjonen, dvs. situasjon 4, er ikke et stort tema i dette kurset.



Oppsummering av de 4 siste eksemplene:

	trekning	
	m/tilbakelegging	u/tilbakelegging
ordnet	situasjon 1: 3.2 mill.	situasjon 2: 1.86 mill.
ikke-ordnet	situasjon 4 (forekommer sjelden)	situasjon 3: 15 504

Figur B.6: $N = 20$ og $s = 5$.

trekning		
	m/tilbakelegging	u/tilbakelegging
ordnet	Situasjon 1: # komb. = N^s	Situasjon 2: # komb. = $\frac{N!}{(N-s)!}$
ikke-ordnet	situasjon 4 (forekommer sjelden) (ikke pensum)	Situasjon 3: # komb. = $\binom{N}{s}$

Disse formlene gjelder for alle s : $s \geq 0$

N = antall elementer totalt

s = antall valgte elementer

$s=0$:

Ingen valgte elementer.

$s=1$:

Situasjon 1 = situasjon 2 = situasjon 3 = situasjon 4
 (alle formlene reduserer seg til det samme: # komb. = N)

$s \geq 2$:

Formlene er forskjellige.

Figur B.7: Antall kombinasjoner.

B.3 Binomialkoeffisienten

Egenskaper til binomialkoeffisienten $\binom{N}{s}$: (“N over s”)

$$\underbrace{\binom{N}{s}}_{\text{binomialkoeff.}} = \frac{N!}{(N-s)! s!} \quad (\text{B.20})$$

hvor f.eks.

$$s! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (s-2) \cdot (s-1) \cdot s \quad (\text{B.21})$$

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \quad (\text{B.22})$$

Legg merke til at

$$\binom{N}{0} = \frac{N!}{(N-0)! 0!} = 1 \quad (\text{B.23})$$

$$\binom{N}{1} = \frac{N!}{(N-1)! 1!} = N \quad (\text{B.24})$$

$$\binom{N}{N} = \frac{N!}{(N-N)! N!} = 1 \quad (\text{B.25})$$

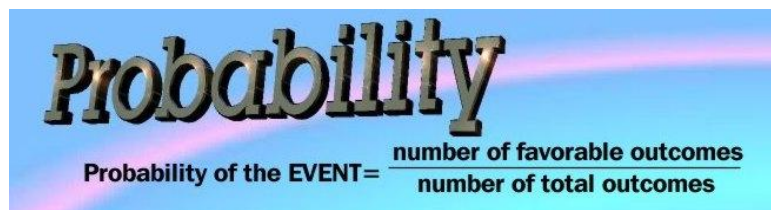
siden $0! = 1$.

B.4 Kombinatoriske sannsynligheter

La oss se på en **tellesituasjon** som kan beskrives med **urnemodellen**. I slike situasjoner er det en sammenheng mellom sannsynlighet og antall kombinasjoner:

La A tilsvare det å trekke kuler av en bestemt type fra urnen. Dette betyr at kulene representerer begivenheten A . Sannsynligheten for å trekke A er da gitt ved: ²

$$P(A) = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner for } A}{\text{antall } \textit{mulige} \text{ kombinasjoner totalt}} \quad (\text{B.26})$$



Figur B.8: Sannsynlighet og telling.

²Dette er helt analogt med **tellesituasjonen** som vi diskuterte i kapittel 1. Se f.eks. eksempel 3 i kapittel 1.

Eksempel: (fond)

Du velger å investere i 2 av i alt 4 ulike aksjefond. Det er ingen spesiell sammenheng mellom fondene. Derfor velger du fondene tilfeldig.

- 4 fond, 2 velges tilfeldig
- a) Hva er sannsynligheten for at du velger det beste fondet, dvs. $P(\text{beste fondet})$?
- b) Hva er sannsynligheten for at du velger de 2 beste fondene, dvs. $P(2 \text{ beste fondene})$?



Figur B.9: Fond.

Løsning:

Siden

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

så kan denne tellesituasjonen beskrives ved “**urnemodellen**”.

Den generelle sammenhengen mellom en slik telling og sannsynlighet er, se lign.(B.26):

$$P(A) = \frac{\text{antall } \textit{gunstige} \text{ kombinasjoner for } A}{\text{antall } \textit{mulige} \text{ kombinasjoner totalt}} \quad (\text{B.27})$$

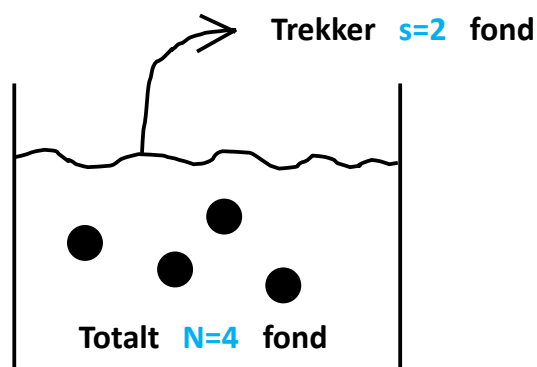
for en gitt begivenhet A .

Er dette situasjon 1, 2, 3 eller 4?

Det er totalt $N = 4$ fond. Du velger å investere i $s = 2$ av disse fondene. Det er ingen avhengighet mellom fondene. Derfor spiller ikke rekkefølgen vi trekker fondene på noen rolle. Når et fond er trukket så kan fondet ikke trekkes på nytt, altså ikke tilbakelegging.

Dette tilsvarer derfor:

situasjon 3



Figur B.10: Totalt 4 fond. Trekker 2 fond.

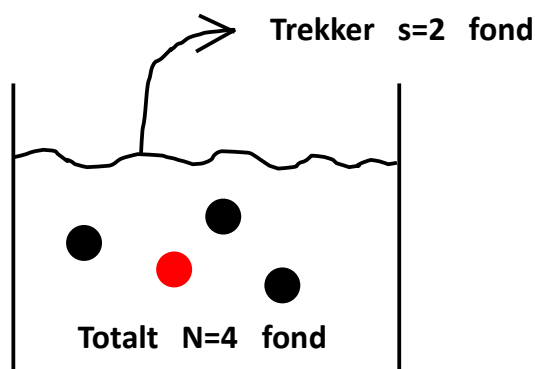
a) $P(\text{beste fondet})$:

Sannsynligheten for å trekke det beste fond er det samme som sannsynligheten for å trekke et *bestemt* fond.

Det er to kategorier: beste fond (rød kule) og ikke beste fond (svarte kuler).
Via lign.(B.27) får vi da:

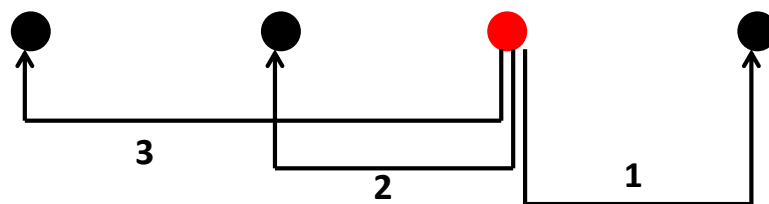
$$\underline{\underline{P(\text{beste fond})}} = \frac{\overbrace{\text{antall } \textit{gunstige} \text{ kombinasjoner}}^{= 3}}{\underbrace{\text{antall } \textit{mulige} \text{ kombinasjoner}}_{= 6}} = \frac{\text{kategori 1} \cdot \text{kategori 2}}{\text{totalt}} \quad (\text{B.28})$$

$$= \frac{\binom{N-1}{s-1} \binom{1}{1}}{\binom{N}{s}} = \frac{\binom{3}{1} \binom{1}{1}}{\binom{4}{2}} = \frac{3 \cdot 1}{6} = \underline{\underline{0.50}} \quad (\text{B.29})$$



Figur B.11: Totalt 4 fond. Trekker 2 fond.

Denne figuren visualiserer at det er 3 måter å trekke kuler på hvor en av dem er den røde slik som telleren i lign.(B.29)



Figur B.12: 3 måter.