



MAT110

Statistikk 1

Kompendium 2020

Per Kristian Rekdal

Bård-Inge Pettersen



Innhold

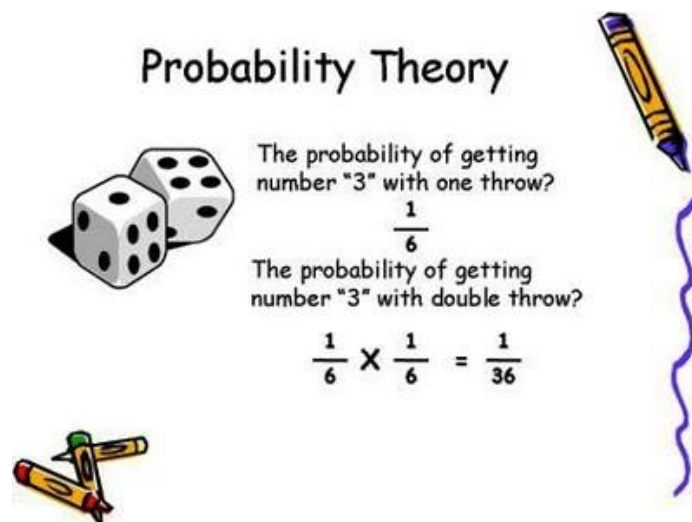
1	Sannsynlighetsteori	7
1.1	Sannsynlighetsmodell	8
1.1.1	Motivasjon	8
1.1.2	Utfallsrommet	11
1.1.3	Mengdelære	13
1.1.4	Begivenheter - delmengder av utfallsrommet	14
1.1.5	Sannsynlighetsmodell	23
1.1.6	Fundamentale setninger i sannsynlighetsteori	24
1.1.7	Diskret sannsynlighetsmodell	40
1.1.8	Uniform 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	79
1.2.5	Oppsplitting av Ω	82
2	Stokastiske variabler, forventning og varians	92
2.1	Stokastiske variabler	93
2.1.1	Diskret VS kontinuerlig stokastiske variabler	94
2.2	Forventning og varians	100
2.2.1	Forventning	100
2.2.2	Varians	103
2.2.3	Kovarians	108
2.2.4	Noen regneregler	125
3	Diskrete stokastiske fordelinger	142
3.1	Den binomiske fordelingen	143
3.1.1	Forventningsverdi	159
3.1.2	Varians	161
3.2	Den hypergeometriske fordelingen	167
3.2.1	Forventning og varians	175
3.3	Sammenheng mellom $\text{Hyp}[N, M, n]$ og $\text{Bin}[n, p]$	179
3.3.1	Forventningsverdi	180
3.3.2	Varians	180
3.4	Poissonfordelingen	183

3.4.1	Forventning og varians	189
4	Kontinuerlige stokastiske fordelinger og CLT	194
4.1	Kontinuerlig fordeling	194
4.2	Normalfordelingen (kontinuerlig)	199
4.2.1	Standardisering	204
4.2.2	Standardisering = omskalering	205
4.2.3	Sammenhengen mellom $P(Z \leq z)$ og $G(z)$	208
4.2.4	Diskret vs kontinuerlig fordeling: en viktig forskjell	222
4.2.5	Standardavvik σ og %-vis areal	223
4.3	Oversikt: Bin, Hyp, Poi og N	231
4.4	Sentralgrensesetningen	234
4.5	Diskrete fordelinger $\rightarrow$ normalfordeling	250
4.5.1	Sammenheng: Bin, Hyp, Poi og N	251
4.6	Sum av uavhengige stokastiske variabler	252
5	Statistisk inferens	256
5.1	Fra sannsynlighetsteori til statistisk inferens	257
5.2	Steg 1: Tilfeldig utvalg	263
5.2.1	Populasjonsvariabler og forsøksvariabler	269
5.3	Steg 2: Gjennomføring av forsøksrekken	272
5.4	Steg 3 : Beskrivende statistikk	274
5.4.1	Lokaliseringsmål	276
5.4.2	Spredingsmål	277
5.5	Statistisk modell	290
6	Estimering og konfidensintervaller	298
6.1	Motivasjon - statistisk inferens	299
6.2	Estimatorer	306
6.3	Konfidensintervaller	313
6.4	Student's t -fordeling og χ_k^2 -fordeling	339
6.4.1	χ_k^2 -fordeling ("kji"-fordeling)	341
6.4.2	Student's t -fordeling	343
6.4.3	Eksakte $(1 - \alpha)$ -konfidensintervaller for μ og σ	346
7	Hypotesetesting	356
7.1	Motivasjon - hypotesetesting	357
7.1.1	Hypotesetest	359
7.1.2	Tilstrekkelig bevis?	360
7.1.3	Type-I feil og type-II feil	360
7.1.4	Analogi til rettsak	362
7.1.5	Test som gir tilstrekkelig bevis	364
7.1.6	Revidert spørsmål	365
7.1.7	En hypotesetest med signifikansnivå $\alpha = 0.05$	366
7.1.8	Hva blir konklusjonen dersom vi bruker ψ_2 ?	369
7.2	Hypotesetesting	370
7.3	To-utvalgs test	374
7.3.1	Statistisk modell for to utvalg	375

7.3.2	Formulering av nullhypotese og alternativ hypotese	377
7.3.3	Konstruksjon av hypotesetest med signifikansnivå $\alpha = 0.05$	378
7.3.4	Realisering og konklusjon	382
8	Regresjonsanalyse	384
8.1	Introduksjon	385
8.2	Statistiske mål (to variabler)	386
8.3	Teoretisk modell vs estimert modell	392
8.4	Residual og <i>sse</i>	393
8.5	Minste kvadraters regresjonslinje	396
8.6	Forklaringskraft og <i>sst</i>	404
A	Mengdelære	412
A.1	Vennndiagrammer	413
B	Kombinatorikk	418
B.1	Koblinger	419
B.2	4 situasjoner (endelig populasjon)	423
B.3	Binomialkoeffisienten	431
B.4	Kombinatoriske sannsynligheter	432

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

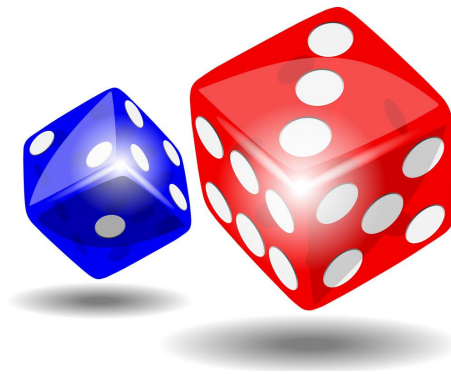
Siden en definerer de mulige utfallene av et eksperiment som en *mengde*, er det viktig med en liten introduksjon til mengdelære.

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 order *probability* som betyr *sannsynlighet* på norsk.
 $P(\Omega)$ kan leses som: ”sannsynligheten for at et av utfallene fra eksperimentet er med i mengden Ω , er lik 1”.
⁶Denne intusjonen er korrekt.

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

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

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

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

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

⁷Legg merke til at:

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

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

Sannsynligheten tilhørende lign.(1.12) er:

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

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

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

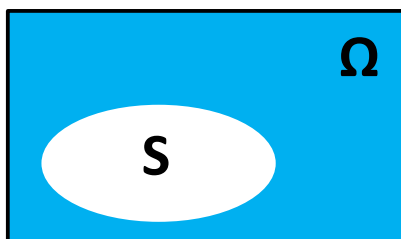
I tillegg må **sannsynlighetsmodellen** oppfylle et sett med fundamentale egenskaper ved beregning av sannsynligheter av delmengder.

■

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:

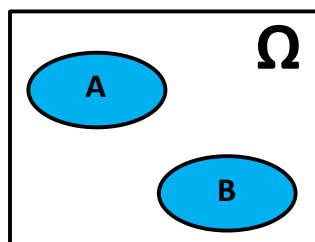
Lag følgen av begivenhetene $S_1 = A$, $S_2 = B$, $S_3 = \emptyset$, $S_4 = \emptyset$, $\dots$

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:

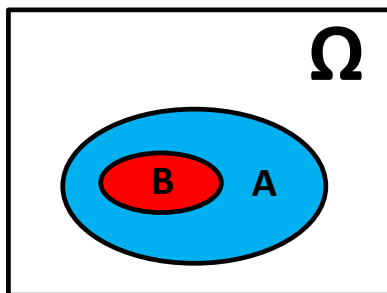
Siden $B \subset A$, er $A - B$ disjunkt med B og $(A - B) \cup B = A$.
Lign. (1.34) gir dermed:

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

$\Updownarrow$

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

■



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:

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

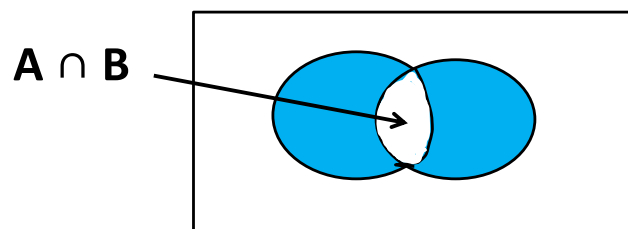
Siden A og $B - A \cap B$ er disjunkte gir den spesielle addisjonssetningen at: ¹⁶

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

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 $A \cap B$.

¹⁶Lign. (1.34) i setning 2.

¹⁷Lign. (1.37) i setning 3.

Setning 5: ([komplementærsetningen](#))

La A være en vilkårlig begivenhet, da gjelder alltid:

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

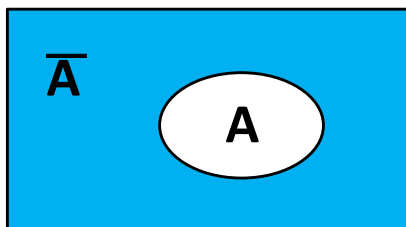
■

Bevis:

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

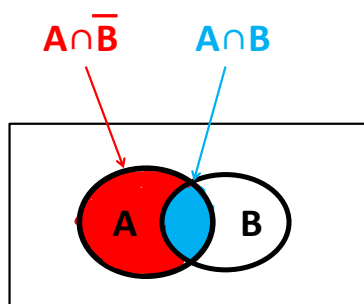
Siden $\Omega = \bar{B} \cup B$ og $A \subset \Omega$ har en alltid at (se figur 1.10):¹⁹

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

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:

La oss trekke en tilfeldig person blant kvinner og menn.

Utfallsrom:

Utfallsrommet består av alle ordnede par:

$$\Omega = \left\{ (m, g), (m, \bar{g}), (k, g), (k, \bar{g}) \right\} \quad (1.54)$$

med $\overbrace{\text{notasjonen}}^{\text{små bokstaver}}$

- m = mann
- k = kvinne
- g = gravid

Definer følgende $\overbrace{\text{begivenheter}}^{\text{store bokstaver}}$:

$$M = \left\{ (m, g), (m, \bar{g}) \right\} \quad (\text{en } \underline{\text{mann}} \text{ trekkes}) \quad (1.55)$$

$$K = \left\{ (k, g), (k, \bar{g}) \right\} \quad (\text{en } \underline{\text{kvinne}} \text{ trekkes}) \quad (1.56)$$

$$G = \left\{ (m, g), (k, g) \right\} \quad (\text{en } \underline{\text{gravid person}} \text{ trekkes}) \quad (1.57)$$



Figur 1.11: Gravid.

Anta at:

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

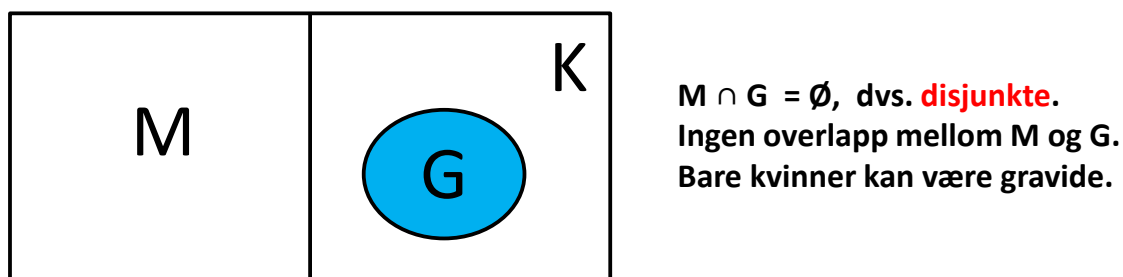
og

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

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:

(notasjon av mulige utfall)
små bokstaver

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

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

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

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

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:

$$A = \text{Ole Hansen finner ledige sitteplasser på minst \textcolor{blue}{\text{en}} av bussene} \quad (1.73)$$

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

- b) Definer følgende begivenhet:

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

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

- c) Hva er sannsynligheten for at Ole Hansen **verken** finner ledige sitteplasser på 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

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

Komplementærsetningen:

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

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

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

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

Sannsynligheten for at Ole Hansen ikke finner ledige sitteplasser p\aa 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 $\overbrace{k=\text{krone og } m=\text{mynt}}^{\text{små bokstaver}}$. Det antas at mynten er perfekt balansert, dvs. sannsynligheten for å få mynt/krone er 50%: ²⁴

$$P(k) = P(m) = \frac{1}{2} \quad (1.99)$$

dvs. like sannsynligheter. Utfallsrommet Ω har totalt 4 mulige $\overbrace{\text{begivenheter}}^{\text{store bokstaver}}$:

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

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

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

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

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

$$\underline{\underline{P(S_4)}} = P(\{k, m\}) = P(\{k\} \cup \{m\}) = P(m) + P(k) = \frac{1}{2} + \frac{1}{2} = \underline{\underline{1}} \quad (1.104)$$

²⁴Vanlig konvesjon: $P(\{m\}) = P(m)$.

Eksperiment 2: (myntkast - 3 kast)

Anta mynten kastes 3 ganger hvor resultatet registreres for hvert kast.

Utfallsrommet består av 3-tuppler:

$$\Omega = \left\{ \overbrace{(k, k, k), (k, k, m), (k, m, k), (k, m, m), (m, k, k), (m, k, m), (m, m, k), (m, m, m)}^{8 \text{ mulige utfall}} \right\} \quad (1.105)$$

hvor $\overbrace{k, m}^{\text{små bokstaver}}$ = 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 :

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

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 $i = 1, 2, 3, \dots, 2^8$.

Ta for eksempel begivenheten hvor nøyaktig to mynt forekommer:

$$S_{mm} = \left\{ \overbrace{(m, m, k), (m, k, m), (k, m, m)}^{3 \text{ mulige utfall}} \right\} = \{(m, m, k)\} \cup \{(m, k, m)\} \cup \{(k, m, m)\} \quad (1.108)$$

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

$$\underline{\underline{P(S_{mm})}} = P\left(\{(m, m, k)\} \cup \{(m, k, m)\} \cup \{(k, m, m)\}\right) \quad (1.109)$$

$$= P(\{(m, m, k)\}) + P(\{(m, k, m)\}) + P(\{(k, m, m)\}) \quad (1.110)$$

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

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](#).

■



Figur 1.14: Myntkast.

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**.

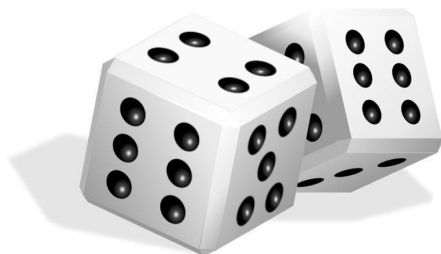
La oss denne gangen se på begivenhetene:

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

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

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**²⁵. Det er 3 stk. gunstige utfall i begivenhetene A og B . Dermed:

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

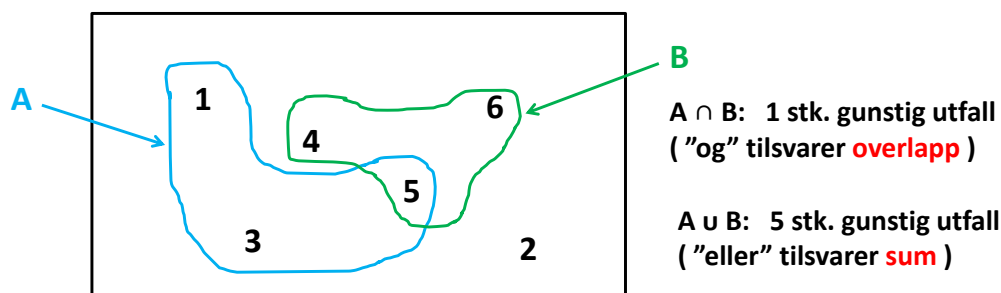
$$P(B) = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}} = \frac{3}{6} = \frac{1}{2} \quad (1.126)$$

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 .

²⁵Vi 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: ²⁶

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

■



Figur 1.17: Kredittverdig.

²⁶Senere 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: ($\overbrace{\text{notasjon av mulige utfall}}^{\text{små bokstaver}}$)

- 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 3-tuppler:

$$\Omega = \left\{ (m_1, m_2, m_3), (m_1, k_1, m_2), \dots, (k_1, k_2, k_3) \right\} \quad (1.133)$$

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

Definer følgende $\overbrace{\text{begivenheter}}^{\text{store bokstaver}}$: (q_1 , q_2 og q_3 er forskjellige pasienter)

$$A = \left\{ (q_1, q_2, q_3) \mid \text{hvor alle pasientene er menn} \right\} \quad (1.134)$$

$$B = \left\{ (q_1, q_2, q_3) \mid \text{hvor alle pasientene er kvinner} \right\} \quad (1.135)$$

$$C = \left\{ (q_1, q_2, q_3) \mid \text{hvor alle pasientene er 2 kvinner og en mann} \right\} \quad (1.136)$$

$$D = \left\{ (q_1, q_2, q_3) \mid \text{hvor alle pasientene er 1 kvinne og 2 menn} \right\} \quad (1.137)$$



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 ²⁷
- utfallene ikke påvirker hverandre, dvs. de er **uavhengige** ²⁸

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.:

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

²⁷Samme sannsynlighet for å trekke en pasient.

²⁸Om 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

know

$P(A|B)$

read "probability of A given B"

Figur 1.19: Betinget sannsynlighet.

Eksempel: (nytt produkt - betinget sannsynlighet)

Cocal Cola i Norge planlegger lanseringen av et nytt produkt de kaller Lady Cola. De er svært spent på hvilken respons markedet gir. De velger derfor å gjennomføre en markedsundersøkelse hvor de spør tilfeldige valgte personer hvorvidt de er villig til å kjøpe Lady Cola eller ikke.

Spesielt er de interessert i sannsynligheten for at Lady Cola kjøpes, gitt at personen er en kvinne.

De klassifiserer kundene sine etter følgende kriterier:

Kjøper/Kjøper ikke:

- b = kunden er villig til å kjøpe Lady Cola
- $\bar{b}$ = kunden er *ikke* villig til å kjøpe Lady Cola

Kjønn:

- k = den spurte personen er er kvinne
- m = den spurte personen er en mann



Figur 1.20: Coca Cola.

Coca Cola definerer følgende sannsynlighetsmodell:

Eksperiment:

Coca Cola trekker en tilfeldig person og registrer hvilket kjønn personen har og spør vedkommende om han/hun er villig til å kjøpe Lady Cola eller ikke.

Utfallsrom:

Utfallsrommet består av alle mulige 2-tuppler:

$$\Omega = \{ (m, b), (m, \bar{b}), (k, b), (k, \bar{b}) \} \quad (1.154)$$

Data:

Coca Cola gjennomførte eksperimentet på 66 tilfeldige personer, og fikk følgende svar:

	b (vil kjøpe)	$\bar{b}$ (vil ikke kjøpe)	totalt
k (kvinne)	21	13	34
m (mann)	15	17	32
totalt	36	30	66

Figur 1.21: Respons fra markedsundersøkelsen.

Begivenheter:

Definerer begivenhetene: ²⁹

$$K = \left\{ (k, b), (k, \bar{b}) \right\} \quad (\text{den spurte personen er en kvinne}) \quad (1.155)$$

$$B = \left\{ (m, b), (k, b) \right\} \quad (\text{den spurte personen er villig til å kjøpe Lady Cola}) \quad (1.156)$$

$$K \cap B = \left\{ (k, b) \right\} \quad (\text{den spurte personen er villig til å kjøpe Lady Cola} \quad (1.157)$$

$$\text{og er en kvinne.}) \quad (1.158)$$

$$B|K = \quad (\text{den spurte personen er villig til å kjøpe Lady Cola,} \quad (1.159)$$

gitt at en kvinne spørres.

- a) Hva er sannsynligheten for at en **kvinne** spørres, dvs. $P(K)$?
- b) Hva er sannsynligheten for at personen er **villig** til å kjøpe Lady Cola, dvs. $P(B)$?
- c) Hva er sannsynligheten for at personen er villig til å kjøpe Lady Cola **og** at personen er en kvinne, dvs. $P(B \cap K)$?
- d) Hva er sannsynligheten for at personen er villig til å kjøpe Lady Cola, dersom vi *vet* at vi spør en kvinne, dvs. $P(B|K)$?

²⁹Legg merke til at notasjonen $A|B$ er ny og står for betinget sannsynlighet.

Løsning:

- a) Sannsynligheten for at en **kvinne** spørres (K):

$$\underline{\underline{P(K)}} = \frac{\text{antall } \textit{kvinner}}{\text{antall totalt}} = \frac{34}{66} \approx \underline{\underline{0.52}} \quad (1.160)$$

- b) Sannsynligheten for at personen er **villig** til å kjøpe Lady Cola (B):

$$\underline{\underline{P(B)}} = \frac{\text{antall kunder } \textit{villige} \text{ til å kjøpe Lady Cola}}{\text{antall totalt}} = \frac{36}{66} \approx \underline{\underline{0.55}} \quad (1.161)$$

- c) Sannsynligheten for at personen er villig til å kjøpe Lady Cola **og** personen er en kvinne:

$$\underline{\underline{P(B \overset{\text{og}}{\cap} K)}} = \frac{\text{antall kvinner som er villige til å kjøpe Lady Cola}}{\text{antall totalt}} = \frac{21}{66} \approx \underline{\underline{0.32}}$$

- d) Sannsynligheten for at personen er villig til å kjøpe Lady Cola, dersom vi *vet* at det er en kvinne:

Siden vi vet at vi spør en kvinne (K), så kan vi si at K allerede har inntruffet. Det betyr at vi restrikerer oss kun til kvinner mao. utfallsrommet har blitt mindre.

$$\underline{\underline{P(B|K)}} = \frac{\text{antall som er villige til å kjøpe Lady Cola}}{\text{antall **kvinner** totalt}} = \frac{21}{36} \quad (1.162)$$

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

Konklusjon:

Sannsynligheten for å spørre en kvinne, $P(K) \approx 0.52$, er altså **ikke** den samme som sannsynligheten for at personen er villig til å kjøpe Lady Cola dersom *vet* at vi spør en kvinne, $P(B|K) \approx 0.58$.

■

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)

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

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

(*betinget* 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”: ³⁰

$$\text{i) } \overbrace{\text{Uavhengighet}}^{\text{vanskelig \AA visualisere}} : ^{31} \quad P(A \overset{\text{og}}{\cap} B) = P(A) P(B) \quad (\text{ se lign. (1.173) })$$

$$\text{ii) } \underbrace{\text{Disjunkt}}_{\text{ikke overlap}} : ^{32} \quad P(A \overset{\text{eller}}{\cup} B) = P(A) + P(B) \quad (\text{ se lign. (1.34) })$$

³⁰Disse begrepene kan ogs\AA v\AA re koblet p\AA f\AA lgende m\AA ter:

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

$$\text{disjunkt} \quad \overset{\text{ikke n\AA dvendigvis}}{\impliedby} \quad \text{avhengig} \quad (1.178)$$

³¹Alts\AA : $P(\text{og}) = \text{PRODUKT}$ av ubetingede sannsynligheter

³²Alts\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*.

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

$$P(A) + P(B) \quad (1.180)$$

$$P(A \cap B) \quad (1.181)$$

$$P(A \cup B) \quad (1.182)$$

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

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 $P(A \cap B) = P(B \cap A)$:

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

$\Updownarrow$

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

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

■

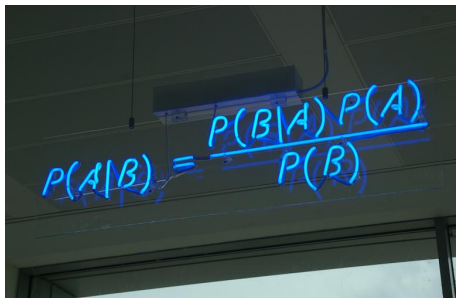
På same måte som multiplikasjonssetningen så kan vi snu på Bayes lov og skrive denne på en alternativ måte:

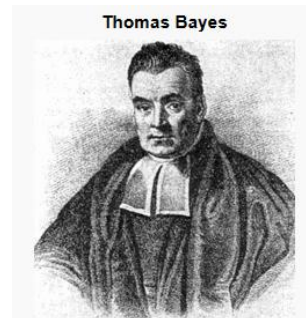
Setning: (*Bayes lov*)

For begivenhetene A og B gjelder:

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

■





Figur 1.23: Bayes lov. Thomas Bayes.

Eksempel: (mislighold og økonomi - betinget sannsynlighet)

Lindorff Group AB har blitt leid inn av DnB til å gjøre en statistisk analyse sammenhengen mellom mislighold av huslån og årsinntekt.

Spesielt er de interessert i sannsynligheten for at en husholdning med årsinntekt over 600 000 NOK misligholder huslånet sitt.

De klassifiserer kundene sine etter følgende kriterier:

Mislighold/Ikke mislighold:

- m = kunden misligholder huslånet
- $\overline{m}$ = kunden misligholder *ikke* huslånet sitt

Årsinntekt:

- i_1 = kunden har årsinntekt mellom 0 og 300 000 NOK
- i_2 = kunden har årsinntekt mellom 300 000 NOK og 600 000 NOK
- i_3 = kunden har årsinntekt mer enn 600 000 NOK



Figur 1.24: Lindorff.

Lindorff definerer følgende sannsynlighetsmodell:

Eksperiment:

Lindorff trekker en tilfeldig kunde og registrerer om kunden har misligholdt huslånet eller ikke og i hvilken årsinntekt kunden har.

Utfallsrom:

Utfallsrommet består av alle mulige 2-tuppler:

$$\Omega = \left\{ (m, i_1), (m, i_2), (m, i_3), (\bar{m}, i_1), (\bar{m}, i_2), (\bar{m}, i_3) \right\} \quad (1.188)$$

Data:

Lindorff gjennomførte eksperimentet på 200 kunder, og fikk følgende estimat på sannsynlighetsloven $P(-)$:

P(m, i_k)	i₁ [0, 300 000]	i₂ <300 000, 600 000>	i₃ [600,->]	totalt
m (mislighold)	0.15	0.20	0.02	0.37
m̄ (ikke mislighold)	0.5	0.13	0.45	0.63
totalt	0.20	0.33	0.47	1

Figur 1.25: **Estimerte utfallssannsynligheter.**

Begivenheter:

Definerer begivenhetene:³³

$$M = \left\{ (m, i_1), (m, i_2), (m, i_3) \right\} \quad (\text{en kunde misligholder huslånet}) \quad (1.189)$$

$$I_3 = \left\{ (m, i_3), (\bar{m}, i_3) \right\} \quad (\text{en kunde har inntekt over 600 000 NOK}) \quad (1.190)$$

³³Legg merke til at vi bruker store bokstaver for begivenheter og små bokstaver for utfall.

- a) Hva er sannsynligheten for at en tilfeldig valgt kunde har **misligholder** huslånet?
- b) Hva er sannsynligheten for at en kunde har inntekt over 600 000 NOK?
- c) Hva er sannsynligheten for at en tilfeldig valgt kunde **misligholder** huslånet dersom vi **vet** at kunden har **årsinntekt** over 600 000?
- d) Hva er sannsynligheten for at en tilfeldig valgt kunde har **årsinntekt** over 600 000 NOK dersom vi **vet** at kunden har **misligholdt** huslånet?

Løsning:

a) **Mislighold** (A):

$$\underline{\underline{P(M)}} = P(\{ (m, i_1) , (m, i_2) , (m, i_3) \}) \quad (1.191)$$

$$= P(m, i_1) + P(m, i_2) + P(m, i_3) \quad (1.192)$$

$$= 0.15 + 0.20 + 0.02 = \underline{\underline{0.37}} \quad (1.193)$$

som er den ubetingede sannsynligheten for at en tilfeldig valgt kunde misligholder huslånet.

b) **Årsinntekt over 600 000 NOK** (B):

$$\underline{\underline{P(I_3)}} = P(\{ (m, i_3) , (\overline{m}, i_3) \}) \quad (1.194)$$

$$= P(m, i_3) + P(\overline{m}, i_3) \quad (1.195)$$

$$= 0.02 + 0.45 = \underline{\underline{0.47}} \quad (1.196)$$

som er den ubetingede sannsynligheten for at en tilfeldig valgt kunde har årsinntekt over 600 000 NOK.

c) **Mislighold** (M) dersom vi **vet** at kunden har en **årsinntekt** over 600 000 NOK(I_3):

$$\underline{\underline{P(M|I_3)}} = \frac{P(M \cap I_3)}{P(I_3)} \quad / M \cap I_3 = \{(m, i_3)\} \quad (1.197)$$

$$= \frac{P(m, i_3)}{P(I_3)} \quad (1.198)$$

$$= \frac{0.02}{0.47} \quad (1.199)$$

$$= \underline{\underline{0.0425}} \quad (1.200)$$

som er den *betingede* sannsynligheten for at en tilfeldig kunde *blant de som har årsinntekt over 600 000 NOK* (I_3), misligholder huslånet (M).

d) **Årsinntekt over 600 000 NOK** (I_3) dersom vi **vet** at kunden har *misligholdt* (M) huslånet:

Siden vi kjenner sannsynlighetene $P(M)$, $P(I_3)$ og $P(M|I_3)$ så kan vi bruke **Bayes lov** i lign.(1.187):

$$\underline{\underline{P(I_3|M)}} = \frac{P(M|I_3)P(I_3)}{P(M)} \quad (1.201)$$

$$= \frac{0.0425 \cdot 0.47}{0.37} \quad (1.202)$$

$$= \underline{\underline{0.054}} \quad (1.203)$$

som er den *betingede* sannsynligheten for at en tilfeldig valgt kunde *blant de som ikke har misligholdt huslånet* (M), har årsinntekt over 600 000 NOK.

■

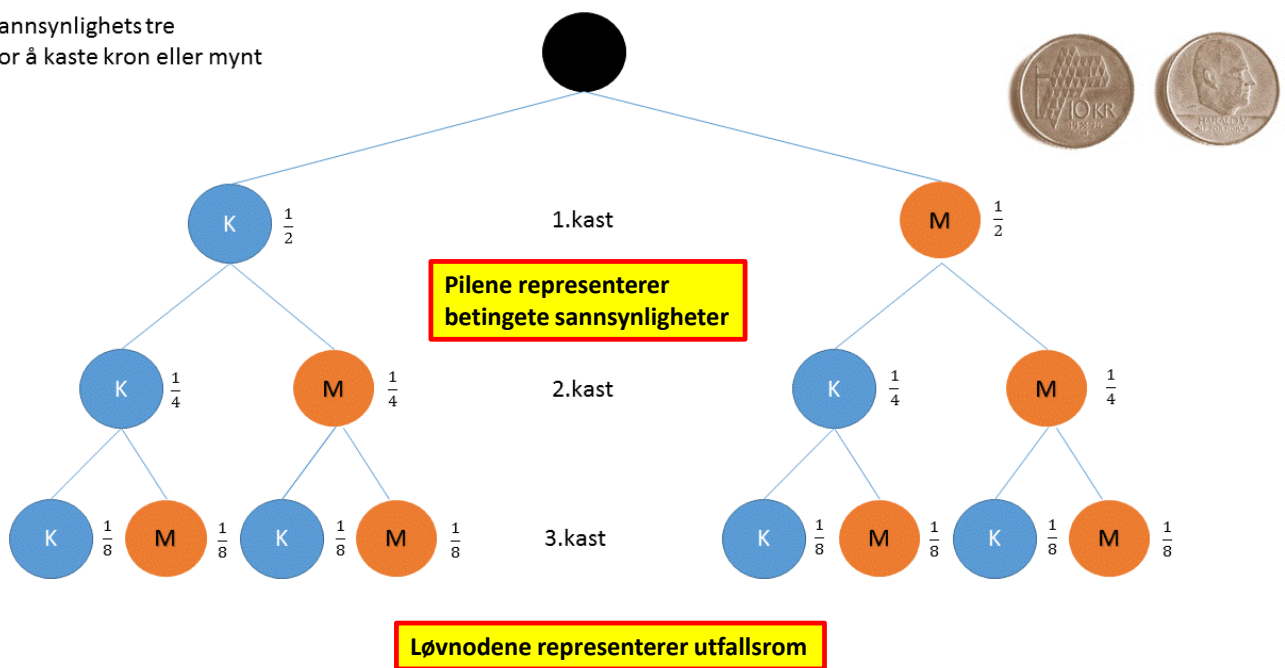
1.2.4 Sannsynlighetstrær

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**.³⁴

Sannsynlighets tre
For å kaste kron eller mynt



Figur 1.26: Sannsynlighetstre - sekvens av steg.

³⁴Utfallsrommet 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?

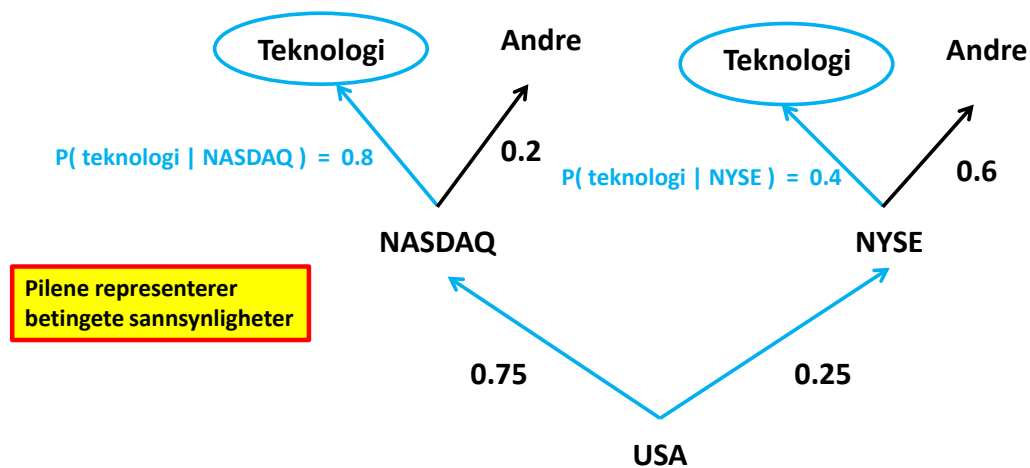


Figur 1.27: NASDAQ og NYSE.

Løsning:

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

Derfor er det her hensiktsmessig å bruke et sannsynlighetstre: ³⁵



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: ³⁶

$$\underline{\underline{P(\text{aksjene investert i teknologi})}} = 0.75 \cdot 0.8 + 0.25 \cdot 0.4 = \underline{\underline{0.70}} \quad (1.204)$$

³⁵Utfallsrommet er gitt ved løvnodene. Pilene representerer betingede sannsynligheter.

³⁶Følg 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.

Eksperiment:

Trekker et fond som amerikaneren har investert i.

Utfallsrom:

Utfallsrommet består av alle mulige 2-tuppler:

$$\Omega = \left\{ (t, b_1), (\bar{t}, b_1), (t, b_2), (\bar{t}, b_2) \right\} \quad (1.205)$$

hvor vi har introdusert $\overbrace{\text{notasjonen}}^{\text{små bokstaver}}$:

- t = utfallet teknologi
- b_1 = utfallet er børs 1, dvs. NASDAQ
- b_2 = utfallet er børs 2, dvs. NYSE

Definer følgende begivenheter:

$$T = \left\{ (t, b_1), (t, b_2) \right\} \quad (\text{teknologi}) \quad (1.206)$$

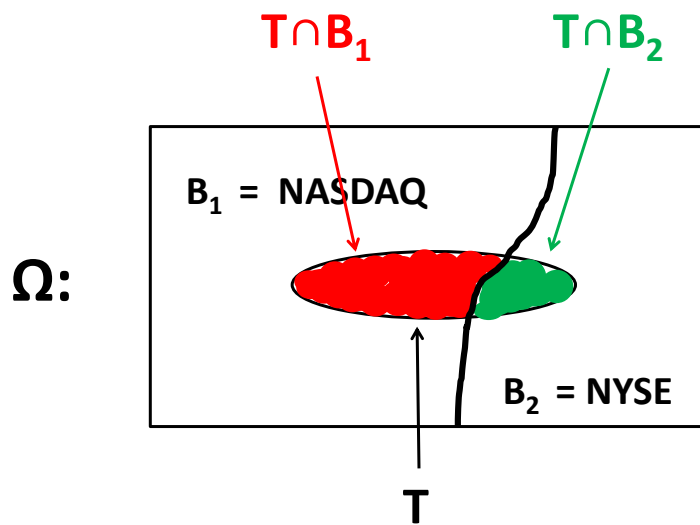
$$B_1 = \left\{ (t, b_1), (\bar{t}, b_1) \right\} \quad (\text{et NASDAQ-fond trekkes}) \quad (1.207)$$

$$B_2 = \left\{ (t, b_2), (\bar{t}, b_2) \right\} \quad (\text{et NYSE-fond trekkes}) \quad (1.208)$$

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

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

$$T = (T \cap B_1) \underbrace{\cup}_{\text{eller}} (T \cap B_2) \quad (1.209)$$



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

Via lign.(1.209) finner vi sannsynligheten $P(A)$:

$$P(T) = P\left[\underbrace{(T \cap B_1) \cup (T \cap B_2)}_{\text{eller}}\right] \quad (1.210)$$

Av figur (1.29) ser vi at det ikke er noe overlapp mellom $T \cap B_1$ og $T \cap B_2$, dvs. de er disjunkte³⁷. Da gjelder den spesielle addisjonssetningen for “eller”-sannsynligheten i lign.(1.210):³⁸

$$P(T) = P(T \cap B_1) + P(T \cap B_2) \quad (1.211)$$

Denne ligningen inneholder to “og”-sannsynligheter. Dermed kan vi bruke multiplikasjonssetningen³⁹ på disse:

$$\underline{\underline{P(T)}} = \overbrace{P(T \cap B_1)}^{= P(T|B_1) \cdot P(B_1)} + \overbrace{P(T \cap B_2)}^{= P(T|B_2) \cdot P(B_2)} \quad (1.212)$$

$$\stackrel{\text{lign. (1.167)}}{=} \overbrace{P(T|B_1)}^{= 0.8} \cdot \overbrace{P(B_1)}^{= 0.75} + \overbrace{P(T|B_2)}^{= 0.4} \cdot \overbrace{P(B_2)}^{= 0.25} \quad (1.213)$$

$$= 0.8 \cdot 0.75 + 0.4 \cdot 0.25 = \underline{\underline{0.7}} \quad (1.214)$$

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

■

³⁷At det ikke er noe overlapp, dvs. disjunkte, betyr at $(T \cap B_1) \cup (T \cap B_2) = \emptyset$.

³⁸Den spesielle addisjonssetningen i lign.(1.34) sier at: $P(A \cup B) = P(A) + P(B)$.

³⁹Multiplikasjonssetningen i lign.(1.167) sier at: $P(A \cap B) = P(A|B) \cdot P(B)$.

Setning: (oppsplitting av Ω i 2)

For begivenhetene A og B gjelder at:

$$P(A) = P(A|B_1) P(B_1) + P(A|B_2) P(B_2) \quad (1.215)$$

■

Bevis:

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

$$A = (A \cap B_1) \overset{\text{eller}}{\cup} (A \cap B_2) \quad (\text{oppsplitting}) \quad (1.216)$$

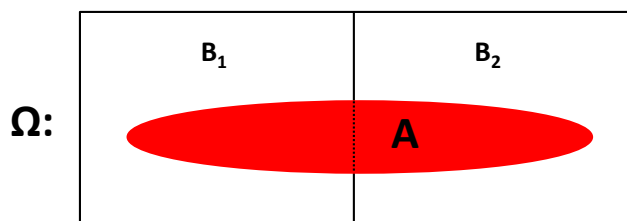
dvs. $B_1 \cap B_2 = \emptyset$. Dermed er tilhørende sannsynlighet:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) \quad (1.217)$$

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

$$\underline{\underline{P(A) = P(A|B_1) P(B_1) + P(A|B_2) P(B_2)}} \quad (1.218)$$

■



Figur 1.30: **Oppsplitting** av sannsynlighetsrom Ω , jfr. lign.(1.216).

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 høy vekst er det 70 % sannsynlig at valutaen styrker seg

Ved middels vekst er det 50 % sannsynlig at valutaen styrker seg

Ved lav vekst er det 20 % sannsynlig at valutaen styrker seg

Eksperiment:

Observerer hvordan valutaen styrker seg.

Utfallsrom:

Utfallsrommet består av alle 2-tuppler:

$$\Omega = \left\{ (h, s), (h, \bar{s}), (m, s), (m, \bar{s}), (l, s), (l, \bar{s}) \right\} \quad (1.219)$$

hvor vi har introdusert $\overbrace{\text{små bokstaver}}^{\text{små bokstaver}}$ notasjon av mulige utfall:

- h = utfallet med høy økonomisk vekst
- m = utfallet med middels økonomisk vekst
- l = utfallet med lav økonomisk vekst
- s = utfallet at valutaen styrker seg



Figur 1.31: Økonomisk vekst.

store bokstaver
Begivenheter:

$$H = \{ (h, s), (h, \bar{s}) \} \quad (\text{høy økonomisk vekst}) \quad (1.220)$$

$$M = \{ (m, s), (m, \bar{s}) \} \quad (\text{middels økonomisk vekst}) \quad (1.221)$$

$$L = \{ (l, s), (l, \bar{s}) \} \quad (\text{lav økonomisk vekst}) \quad (1.222)$$

$$S = \{ (h, s), (m, s), (l, s) \} \quad (\text{valutaen styrker seg}) \quad (1.223)$$

Anta: ⁴⁰

$$P(H) = 0.32 \quad (\text{sannsynlighet for høy økonomisk vekst}) \quad (1.225)$$

$$P(M) = 0.51 \quad (\text{sannsynlighet for middels økonomisk vekst}) \quad (1.226)$$

$$P(L) = 0.17 \quad (\text{sannsynlighet for lav økonomisk vekst}) \quad (1.227)$$

Med begivennene ovenfor så kan sannsynlighetene som som oppgitt tidligere i oppgaven, skrives:

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

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

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

⁴⁰Legg merke til at lign.(1.225)-(1.227) representerer en gyldig sannsynlighetsfordeling siden:

$$P(H) + P(M) + P(L) = 0.32 + 0.51 + 0.17 = 1 \quad (1.224)$$

a) Hva er $P(S)$? ⁴¹

b) Hva er $P(H|S)$? ⁴²

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

⁴¹Dvs. finn sannsynligheten for at valutaen styrker seg i Østerrike.

⁴²Dvs. 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.(1.225) og (1.227), så kan vi bruke setning om *oppsplitting* av Ω i lign.(1.217):

$$\underline{\underline{P(S)}} = \underbrace{P(S|H)}_{=0.7} \underbrace{P(H)}_{=0.32} + \underbrace{P(S|M)}_{=0.5} \underbrace{P(M)}_{=0.51} + \underbrace{P(S|L)}_{=0.2} \underbrace{P(L)}_{=0.17} \quad (1.228)$$

$$= 0.7 \cdot 0.32 + 0.5 \cdot 0.51 + 0.2 \cdot 0.17 \approx \underline{\underline{0.513}} \quad (1.229)$$

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):

$$\underline{\underline{P(H|S)}} \stackrel{\text{lign. (1.183)}}{=} \underbrace{P(S|H)}_{=0.7} \cdot \frac{\overbrace{P(H)}^{=0.32}}{\underbrace{P(S)}_{=0.513}} = 0.7 \cdot \frac{0.32}{0.513} \approx \underline{\underline{0.44}} \quad (1.230)$$

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 $P(H) P(S)$. 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):

$$\underline{P(H \cap S)} \stackrel{\text{lign. (1.167)}}{=} \underbrace{P(H|S)}_{\approx 0.44} \underbrace{P(S)}_{= 0.513} \approx \underline{0.22} \quad (1.231)$$

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:

$$\underline{P(S \cap H)} \stackrel{\text{lign. (1.170)}}{=} \underbrace{P(S|H)}_{= 0.7} \underbrace{P(H)}_{= 0.32} \approx \underline{0.22} \quad (1.232)$$

dvs. samme svar som i lign.(1.231), selvfølgelig.

- ii) Produktet er lett å regne ut:

$$\underline{P(S) P(H)} = 0.513 \cdot 0.32 \approx \underline{0.16} \quad (1.233)$$

Konklusjon:

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

■