4. Sannsynlighet.

4.1. Noen grunnbegreper.

For å komme videre med statistikken, må du vite litt om *sannsynlighet*. Og ikke minst må du *forstå* noen grunnleggende trekk ved sannsynlighetsregningen.

Mange synes at sannsynlighetsregning er både ullent og vanskelig. For å unngå noen av de vanligste fallgruvene, bør du snarest mulig bli fortrolig med noen grunnleggende begreper.

- Et ***stokastisk eksperiment*** er en aktivitet der vi ikke kan forutsi resultatet.
  Eksempel: Slå en terning, trekke kort fra en kortstokk, skyte på blink.

- Et ***utfall*** er et resultat fra et stokastisk eksperiment som ikke kan detaljeres mer (splittes opp i flere komponenter).

Vi har stor nytte av *mengdelære* når vi jobber med sannsynlighet. Da sier vi at

Mengden av alle mulige utfall fra et stokastisk eksperiment utgjør ***utfallsrommet***.

Eksempel 4.1: Hvilke utfallsrom får vi
a) Når vi slår en terning
b) Når vi trekker et kort fra en kortstokk.

Løsning:
a) Når vi slår en terning, kan vi få 6 ulike utfall. Utfallsrommet blir $\{1,2,3,4,5,6\}$.
b) Når vi trekker ett kort fra en kortstokk, kan vi få 52 ulike utfall. Utfallsrommet består av spar ess, spar konge, …, kløver 2.

Noen ganger er det hensiktsmessig å slå sammen flere ulike utfall til en ***hendelse***. Eller:

En ***hendelse*** er en delmengde av utfallsrommet.

Det er helt opp til deg å bestemme hvilke elementer som skal inngå i en hendelse. Når du trekker ett kort fra en kortstokk, kan en *hendelse* være at du trekker en konge (består av 4 elementer fra utfallsrommet), eller at du trekker en spar (består av 13 elementer fra utfallsrommet), eller en annen kombinasjon av utfall som du finner hensiktsmessig.

Eksempel 4.2: Du slår to terninger samtidig, og ser at summen av ”øynene” på terningene blir 7. Er dette et ”utfall” eller en ”hendelse”?

Løsning: Vi kan tenke oss at vi slår de to terningene etter tur. Da kan det første terningkastet få 6 ulike utfall, og det siste kan også få 6 ulike utfall. Til sammen kan du da få $6 \cdot 6 = 36$ ulike utfall. Utfallsrommet vil da bestå av de 36 tallparene

$$\{(1,1),(1,2),\dots,(2,1),\dots,(6,6)\}.$$

At summen av ”øynene” blir 7, er da en ”hendelse” som består av utfallene

$(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$ og $(6,1)$.

Terminologien ovenfor brukes ikke alltid konsekvent. Det kan ha sammenheng med at på engelsk brukes begrepet *event* både om hendelse og utfall. Om nødvendig sier vi *simple event* når vi mener *utfall*, og bare *event* når vi mener *hendelse*. På norsk ser vi noen ganger at ordet *enkeltutfall* brukes, mens *utfall* da blir en generell betegnelse som både dekker *utfall* og *hendelse* slik jeg har definert disse begrepene.

Vi bruker ofte store bokstaver *A*, *B*, osv. for å betegne hendelser (eller delmengder av utfalls-rommet). Da kan vi også benytte symboler fra mengdelæra, for eksempel slik:

Mengden $A \cup B$ består av utfall som inngår i *enten A eller B eller begge*.
Mengden $A \cap B$ består av utfall som inngår i *både A og B*.
Mengden $\overline{A}$ består av alle elementer i utfallsrommet som *ikke* inngår i *A*.

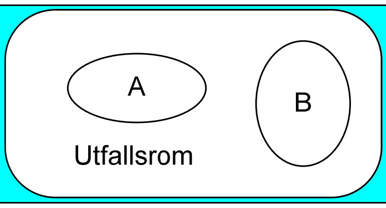
Eksempel 4.3: Du trekker kort fra en kortstokk. La *E* være mengden av Ess, mens *S* er mengden av Spar. Hvilke elementer inngår da i disse mengdene:

a) $E \cup S$.
b) $E \cap \overline{S}$.
c) $E \cap \overline{S}$.

Løsning:

a) Mengden $E \cup S$ består av alle de 13 sparene pluss hjerter ess, ruter ess og kløver ess, til sammen 16 elementer.
b) Mengden $E \cap S$ består av kun ett element, nemlig spar ess.
c) Mengden $E \cap \overline{S}$ består av alle ess som *ikke* er spar, d.v.s. hjerter ess, ruter ess og kløver ess.

Vi trenger noen flere definisjoner:



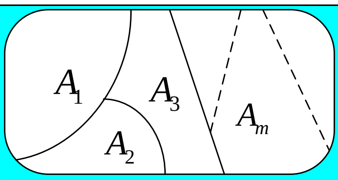
Vi sier at to hendelser *A* og *B* er ***disjunkte*** eller ***innbyrdes ekskluderende*** dersom de ikke kan forekomme samtidig. Tenker vi utfallsrom, blir

$$A \cap B = \varnothing.$$

Situasjonen er illustrert til venstre.

Denne definisjonen kan greit utvides til flere hendelser: Vi sier at *m* hendelser A_1, A_2, …, A_m er innbyrdes disjunkte dersom $A_i \cap A_j = \varnothing$ for alle $i \neq j$.

La oss videre anta at disse *m* hendelsene dekker *hele* utfallsrommet. Da definerer vi:



Dersom *hele* utfallsrommet kan deles opp i *m* disjunkte hendelser $A_1,\ A_2,\ \ldots,\ A_m$, sier vi at disse hendelsene er ***komplementære***.

Situasjonen er illustrert til venstre.

Merk spesielt at to hendelser *A* og $\overline{A}$ er komplementære.

4.2. Hva er "sannsynlighet"?

Du har sikkert en intuitiv forståelse av hva vi mener med "sannsynlighet". Men kan du gi en klar, entydig definisjon av begrepet? Så vidt jeg vet, har ingen greid det hittil. Før vi går i gang med ulike definisjoner, skal vi klargjøre notasjonen.

- Vi bruker bokstaven *p* (eller *P*) for sannsynlighet (**P**robability).
- Aktuelle utfall eller hendelser kalles *A*, *B*, *C*, …
- Sannsynligheten for at hendelse (utfall) *A* skal forekomme skriver vi $P(A)$.

Vi kan definere *sannsynlighet* på flere måter, bl.a. disse tre:

1. ***Eksperimentelt:*** Dersom vi utfører samme stokastiske eksperiment på identisk måte *N* ganger, og hendelse *A* forekommer n_A av disse gangene, er brøken
$$\frac{n_A}{N}$$
et estimat for $P(A)$. Jo større *N* er, jo bedre blir dette estimatet. Vi kan da sette at
$$P(A) = \lim_{N\to\infty} \frac{n_A}{N}.$$

Eksempel 4.4: Hvordan vil du gå fram for å finne sannsynligheten for at et tilfeldig valgt nyfødt barn skal være en gutt?

Løsning: Til daglig antar vi at guttefødsel og jentefødsel er like sannsynlig. I virkeligheten er guttefødsel litt mer sannsynlig enn jentefødsel. Dersom vi tar for oss *N* tilfeldig valgte fødsler, og n_G av disse er guttefødsler, så er
$$\frac{n_G}{N}$$
et estimat for sannsynligheten for guttefødsel. Jo større *N* er, jo bedre blir estimatet. Altså kan vi si at sannsynligheten for guttefødsel er
$$P(\text{Gutt}) = \lim_{N\to\infty} \frac{n_G}{N}.$$

2. ***Med utgangspunkt i utfallsrom.*** Anta at det er *N* mulige utfall i utfallsrommet, og at det er *lik sjanse for hvert av disse utfallene*. Dersom en hendelse *A* består av n_A slike utfall, er sannsynligheten for at hendelse *A* skal forekomme definert som

$$P(A) = \frac{n_A}{N}.$$

Eksempel 4.5: Hva er sannsynligheten for at summen av to terninger skal bli lik 5?

Løsning: Dersom vi slår to terninger, vil de 4 utfallene (1,4), (2,3), (3,2) og (4,1) gi summen 5. Siden utfallsrommet består av $6 \cdot 6 = 36$ elementer, blir sannsynligheten for å få summen 5:

$$P(\text{sum } 5) = \tfrac{4}{36} = \underline{\underline{\tfrac{1}{9}}}.$$

Her forutsetter vi at det er samme sjanse for alle de 36 utfallene, d.v.s. at begge terningene er "ærlige".

3. ***Basert på erfaring.*** Vi bruker vår erfaring til å *anslå* verdien av $P(A)$.

Eksempel 4.3: Når du fyller ut en tippekupong, benytter du subjektive sannsynligheter for utfallet av hver kamp. Du bruker også subjektive sannsynligheter for å anslå sannsynligheten for at det skal bli regn i morgen, eller for å anslå sannsynligheten for at du skal stå til eksamen.

I praksis er det definisjon 2 ovenfor som er mest nyttig. Flere av de regnereglene som vi skal behandle nedenfor, kan utledes eller begrunnes med utgangpunkt i denne definisjonen. Reglene gjelder imidlertid uansett hvilken definisjon vi bruker.

4.3. Regneregler for sannsynlighet.

Skrivemåten $P(A)$ betyr altså "sannsynligheten for at hendelse *A* forekommer". Ved å bruke notasjon fra mengdelæra, får vi:

- $P(\overline{A})$ betyr "sannsynligheten for at hendelse *A* *ikke* forekommer".
- $P(A \cap B)$ betyr "sannsynligheten for at både hendelse *A* og hendelse B forekommer".
- $P(A \cup B)$ betyr "sannsynligheten for at enten hendelse *A* eller hendelse B eller begge forekommer".

Det er også vanlig å bruke utsagns-notasjon. Da skriver vi $P(A \wedge B)$ istedenfor $P(A \cap B)$, og $P(A \vee B)$ istedenfor $P(A \cup B)$.

La meg starte med å slå fast noen egenskaper som bør være helt opplagt:

$0 \le P(A) \le 1$.

Dersom $P(A) = 0$, vil *A* aldri forekomme.

Dersom $P(A) = 1$, vil *A* alltid forekomme.

4.3.1. Addisjonsregelen.

Du husker sikker at to hendelser er ***disjunkte*** dersom de ikke kan forekomme samtidig. Da gjelder:

> Hvis og bare hvis *A* og *B* er disjunkte hendelser, er
> $$P(A \cup B) = P(A) + P(B).$$

Setningen bevises ved å telle elementer i mengdene. Lar vi antall elementer i utfallsrommet være *N*, mens det er henholdsvis n_A og n_B elementer i *A* og *B*, får vi at

$$P(A \cup B) = \frac{n_{A\cup B}}{N} = \frac{n_A + n_B}{N} = \frac{n_A}{N} + \frac{n_B}{N} = \underline{P(A) + P(B)}.$$

Merk at regelen $P(A \cup B) = P(A) + P(B)$ *kun* gjelder når *A* og *B* er disjunkte hendelser. Men hva skjer dersom *A* og *B* *ikke* er disjunkte? Da må vi bruke den generelle addisjons-regelen nedenfor:

> La *A* og *B* være to hendelser. Da er
> $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Utledningen bygger på ”telleregelen” for antall elementer i unionen av to mengder, som ser slik ut:

$$n_{A\cup B} = n_A + n_B - n_{A\cap B}.$$

Det siste leddet korrigerer for at elementene i $A \cap B$ vil telles to ganger. Da får vi:

$$P(A \cup B) = \frac{n_{A\cup B}}{N} = \frac{n_A + n_B - n_{A\cap B}}{N} = \frac{n_A}{N} + \frac{n_B}{N} - \frac{n_{A\cap B}}{N} = \underline{P(A) + P(B) - P(A \cap B)}.$$

I praksis er det lett å glemme $P(A \cap B)$-leddet. Et lite eksempel: La *A* være hendelsen ”Rosenborg blir seriemester” mens hendelsen *B* er ”Rosenborg blir cupmester”. En optimistisk trønder antar at $P(A) = 0.8$ mens $P(B) = 0.5$. Sannsynligheten for at Rosenborg skal vinne et mesterskap er da $P(A \cup B)$. Dersom du glemmer $P(A \cap B)$-leddet, får du at

$$P(A \cup B) = P(A) + P(B) = 0.8 + 0.5 = 1.3$$

som åpenbart må være galt.

For å få rett resultat, må du ta hensyn til at *A* og *B* *ikke* er innbyrdes ekskluderende hendelser. Du må trekke fra sannsynligheten for at Rosenborg skal bli *både* cup- og seriemester, d.v.s. at du må trekke fra $P(A \cap B)$-leddet. Hvis vi anslår at $P(A \cap B) = 0.4$, får vi at

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.5 - 0.4 = \underline{\underline{0.9}}$$

som selv den mest ihuga Rosenborg-patriot bør være fornøyd med.

Hvis du er med på resonnementene ovenfor, innser du sikkert at resultatene kan generaliseres:

Anta at *m* hendelser $A_1,\ A_2,\ \ldots,\ A_m$ er innbyrdes disjunkte. Da har vi at

$$P(A_1 \cup A_2 \cup \cdots \cup A_m) = P(A_1) + P(A_2) + \cdots + P(A_m).$$

La oss videre anta at disse *m* hendelsene er komplementære, d.v.s. at de dekker *hele* utfalls-rommet. Da gjelder:

Dersom *m* hendelser $A_1,\ A_2,\ \ldots,\ A_m$ er komplementære, blir

$$P(A_1) + P(A_2) + \cdots + P(A_m) = 1.$$

Setningen bevises slik: Dersom det er henholdsvis $n_1,\ n_2,\ \ldots,\ n_m$ utfall i hendelsene $A_1,\ A_2,\ \ldots,\ A_m$, og *N* utfall i alt, blir

$$n_1 + n_2 + \cdots + n_m = N \quad \Leftrightarrow \quad \frac{n_1 + n_2 + \cdots + n_m}{N} = 1$$

$$\Leftrightarrow \quad \frac{n_1}{N} + \frac{n_2}{N} + \cdots + \frac{n_m}{N} = 1 \quad \Leftrightarrow \quad \underline{P(A_1) + P(A_2) + \cdots + P(A_m) = 1}$$

Vi har allerede nevnt at *A* og $\bar{A}$ er komplementære hendelser. Da har vi følgende viktige spesialtilfelle av regelen over:

$$P(A) + P(\bar{A}) = 1 \quad \Leftrightarrow \quad P(\bar{A}) = 1 - P(A)$$

Eksempel 4.4: Anta at karakterfordelingen ved en eksamen med 75 kandidater var slik:

Karakter	A	B	C	D	E	F
Antall	5	10	15	20	15	10

Dersom karakterene A, B, C, D og E gir ”bestått” mens F er ”ikke bestått”, hva er da sannsynligheten for at en tilfeldig valgt besvarelse skal være ”bestått”?

Løsning: Vi kan benytte at

$$P(\text{bestått}) = P(A \vee B \vee C \vee D \vee E) = P(A) + P(B) + P(C) + P(D) + P(E)$$

$$= \tfrac{5}{75} + \tfrac{10}{75} + \tfrac{15}{75} + \tfrac{20}{75} + \tfrac{15}{75} = \tfrac{65}{75} = \underline{\underline{\tfrac{13}{15}}}$$

Eller vi kan gjøre det enklere:

$$P(\text{bestått}) = 1 - P(\text{ikke bestått}) = 1 - \tfrac{10}{75} = \underline{\underline{\tfrac{13}{15}}}.$$

Her er det viktig å merke seg at kandidatene får en og kun en karakter til eksamen, slik at vi snakker om innbyrdes ekskluderende hendelser. Videre er ”bestått” og ”ikke bestått” komplementære hendelser.

4.3.2. Multiplikasjonsregelen.

Vi skal nå se nærmere på sannsynligheten for at *både* hendelse *A og* hendelse *B* skal forekomme. Denne sannsynligheten skriver vi $P(A \cap B)$. Intuitivt kan vi resonnere slik: For at *både A* og *B* skal forekomme, kan vi tenke oss at *A* forekommer først og *B* etterpå. Sannsynligheten for at *A* skal forekomme, er $P(A)$. I en andel $P(B)$ av de tilfellene der *A* forekommer, vil også *B* forekomme. Sannsynligheten for at *både A og B* skal forekomme blir da

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

Det er imidlertid en alvorlig svakhet ved dette resonnementet. Det kan jo hende at $P(B)$ avhenger av om *A* har forekommet eller ikke! I slike tilfeller får vi bruk for ***betinget sannsynlighet***, som vi skriver slik:

Skrivemåten

$$P(B \mid A)$$

betyr ”sannsynligheten for at *B* skal forekomme når *A* har forekommet”. Vi leser dette som ”sannsynligheten for *B* gitt *A*”.

Eksempel 4.5: Du trekker to kort fra en kortstokk. La hendelse *A* være at ”det første kortet er et ess”, mens hendelse *B* er at ”neste kort er et ess”. Hva er $P(B \mid A)$ og $P(B \mid \overline{A})$?

Løsning: Dersom det første kortet var et ess, er det 3 ess igjen blant de 51 gjenværende kortene, slik at

$$P(B \mid A) = \underline{\underline{\tfrac{3}{51}}}.$$

Men dersom første kort *ikke* var et ess, er det fremdeles 4 ess igjen, slik at

$$P(B \mid \overline{A}) = \underline{\underline{\tfrac{4}{51}}}.$$

Nå er vi klar til å gi en fullgod multiplikasjonsregel:

La *A* og *B* være to hendelser. Da er

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

der $P(B \mid A)$ er sannsynligheten for at *B* skal forekomme når *A* har forekommet.

Noen lærebøker bruker denne regelen til å definere betinget sannsynlighet. De sier at:

$$P(B \mid A) \overset{\text{def.}}{=} \frac{P(A \cap B)}{P(A)}.$$

Nå er tiden inne til å se på et viktig begrep. Vi skal definere hva vi mener med at to hendelser er ***uavhengige***:

Hendelse *B* er ***uavhengig*** av hendelse *A* hvis og bare hvis $P(B \mid A) = P(B)$.

Hvis vi setter dette uttrykket for $P(B \mid A)$ inn i multiplikasjonsregelen, får vi at

$$P(A \cap B) = P(A) \cdot P(B \mid A) = P(A) \cdot P(B),$$

Men multiplikasjonsregelen må også kunne skrives

$$P(A \cap B) = P(B) \cdot P(A \mid B).$$

Settes høyresidene lik hverandre, får vi at

$$P(B) \cdot P(A \mid B) = P(A) \cdot P(B) \quad \Leftrightarrow \quad P(A) = P(A \mid B).$$

Dette betyr at *A* er uavhengig av *B*. Vi sier at *A* og *B* er ***uavhengige hendelser***:

Hvis og bare hvis to hendelser *A* og *B* er ***uavhengige***, er:

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

$$P(A) = P(A \mid B)$$

$$P(B) = P(B \mid A)$$

Vi skal nå kombinere multiplikasjonsregelen med addisjonsregelen for komplementære hendelser. Da får vi regelen nedenfor. Vi skal ikke føre noe formelt bevis for regelen. Men den virker rimelig – dersom du først skjønner hva som står der:

Dersom A_1, A_2, …, A_m er ***komplementære hendelser***, blir sannsynligheten for at hendelse *B* skal inntreffe:

$$P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + \cdots + P(A_m) \cdot P(B \mid A_m).$$

Spesielt har vi at:

$$P(B) = P(A) \cdot P(B \mid A) + P(\bar{A}) \cdot P(B \mid \bar{A}).$$

Vi skal illustrere regelen med eksemplet nedenfor:

Eksempel 4.6: Gå tilbake til eksempel 4.5 der du trekker to kort fra en kortstokk, og regn ut sannsynligheten for at det *andre* kortet du trekker skal være et ess, d.v.s. du skal finne $P(B)$.

Løsning: Det første kortet kan være et ess, etterfulgt av et nytt ess. Eller det første kortet er *ikke* et ess, men det andre kortet er et ess. Samler du disse mulighetene, får du:

$$P(B) = P(A)\cdot P(B\mid A) + P(\bar{A})\cdot P(B\mid \bar{A})$$

$$= \frac{4}{52}\cdot\frac{3}{51} + \frac{48}{52}\cdot\frac{4}{51} = \frac{12+192}{52\cdot 51} = \frac{204}{52\cdot 51} = \underline{\frac{4}{52}}$$

Var dette overraskende?

Eksempel 4.7: Vi undersøker røykevanene til en gruppe på 50 studenter (22 gutter og 28 jenter). Resultatet oppsummeres i denne tabellen:

	Gutt	Jente	Sum
Røyker	10	20	30
Røyker ikke	12	8	20
Sum	22	28	50

a) Bestem disse sannsynlighetene:
$P(\text{røyker|gutt})$, $P(\text{røyker|jente})$, $P(\text{røyker})$.

b) Er røyking uavhengig av kjønn?

Løsning:

a) Av tabellen ser vi at av de 22 guttene røyker 10. Sannsynligheten for at en tilfeldig valgt gutt skal røyke er da

$$P(\text{røyker|gutt}) = \underline{\underline{\tfrac{10}{22}}}.$$

Siden 20 av de 28 jentene røyker, er sannsynligheten for at en tilfeldig valgt jente skal røyke

$$P(\text{røyker|jente}) = \underline{\underline{\tfrac{20}{28}}}.$$

Av tabellen ser vi at 30 av de 50 studentene røyker. Da får vi at

$$P(\text{røyker}) = \tfrac{30}{50} = \underline{\underline{\tfrac{3}{5}}}.$$

Det siste resultatet kan vi også finne ved å bruke addisjonsregelen for komplementære hendelser. Siden 22 av de 50 studentene er gutter, blir

$$P(\text{gutt}) = \tfrac{22}{50}.$$

På samme måte blir

$$P(\text{jente}) = \tfrac{28}{50}$$

eller:

$$P(\text{jente}) = 1 - P(\text{gutt}) = 1 - \tfrac{22}{50} = \tfrac{28}{50}$$

fordi gutt og jente er komplementære hendelser.

Sannsynligheten for at en tilfeldig valgt student skal røyke, blir da

$$P(\text{røyker}) = P(\text{gutt})\cdot P(\text{røyker}|\text{gutt}) + P(\text{jente})\cdot P(\text{røyker}|\text{jente})$$

$$= \tfrac{22}{50}\cdot\tfrac{10}{22} + \tfrac{28}{50}\cdot\tfrac{20}{28} = \tfrac{10}{50} + \tfrac{20}{50} = \tfrac{30}{50} = \underline{\underline{\tfrac{3}{5}}}$$

b) Er røyking uavhengig av kjønn? Dersom det er uavhengighet, må vi ha at

$$P(\text{røyker}) = P(\text{røyker}|\text{gutt}) = P(\text{røyker}|\text{jente}).$$

Vi ser at dette langt fra er tilfelle i vårt talleksempel. Men dersom våre tall er et tilfeldig utvalg fra en større populasjon, kan vi da påstå at det sannsynligvis er sammenheng mellom røyking og kjønn i *populasjonen*? Selv om det *ikke* er sammenheng mellom røyking og kjønn i populasjonen, må vi forvente at vi kan finne slike sammenhenger i et *tilfeldig* utvalg. Er utslagene så store at de ikke bare kan skyldes tilfeldigheter? Slike problem skal vi ta for oss senere.

4.3.3. Bayes regel.

La som før A og B være to hendelser. Da har vi satt opp at

$$P(A\cap B) = P(A)\cdot P(B\,|\,A).$$

Men vi kan like godt sette opp at

$$P(A\cap B) = P(B\cap A) = P(B)\cdot P(A\,|\,B).$$

Altså er

$$P(A)\cdot P(B\,|\,A) = P(B)\cdot P(A\,|\,B) \quad\Leftrightarrow\quad \underline{P(A\,|\,B) = \frac{P(A)\cdot P(B\,|\,A)}{P(B)}}.$$

Denne formelen kalles ***Bayes regel***:

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

Eksempel 4.8: Bruk resultatene fra eksempel 4.7 og Bayes regel til å finne $P(\text{gutt}|\text{røyker})$.

Løsning: Fra eksempel 4.7 vet vi at:

$$P(\text{røyker}|\text{gutt}) = \tfrac{10}{22}, \qquad P(\text{røyker}) = \tfrac{3}{5}, \qquad P(\text{gutt}) = \tfrac{22}{50}.$$

Av Bayes formel får vi da

$$P(\text{gutt}|\text{røyker}) = \frac{P(\text{gutt})\cdot P(\text{røyker}|\text{gutt})}{P(\text{røyker})} = \frac{\tfrac{22}{50}\cdot\tfrac{10}{22}}{\tfrac{30}{50}} = \underline{\underline{\frac{1}{3}}}.$$

Av tabellen ser vi at 10 av de 30 røykerne er gutter, noe som stemmer med at

$$P(\text{gutt}|\text{røyker}) = \frac{10}{30} = \frac{1}{3}.$$

Bayes regel kan gi en del overraskende resultater, som eksemplet nedenfor viser.

Eksempel 4.9: Vi skal nå se nærmere på en dopingtest. Enhver medisinsk test, også doping-tester, er beheftet med en viss sannsynlighet for feil. To typer feil kan forekomme:

- Testen gir utslag selv om testpersonen ikke er dopet.
- Testen gir ikke utslag selv om testpersonen er dopet.

Vi antar at laboratoriet har erfaring for at (oppdiktede tall, heldigvis):

1. Det er 0.1% sannsynlighet for at testen gir utslag selv om testpersonen ikke er dopet.
2. 5% av de dopede slipper gjennom uten at testen gir utslag.
3. 1% av testpersonene er dopet.

Hva er sannsynligheten for at testpersonen ikke er dopet når testen gir utslag?

Løsning: Først må vi ”oversette” de gitte opplysningene til vår terminologi. Vi starter med å definere disse hendelsene:

- A: Testpersonen er dopet.
- B: Testen gir utslag.

De gitte opplysningene blir da:

1. $P\left(B \mid \bar{A}\right) = 0.001$.
2. $P\left(\bar{B} \mid A\right) = 0.05$
3. $P\left(A\right) = 0.01$

Vi skal altså finne $P\left(\bar{A} \mid B\right)$. Bayes regel gir

$$P\left(\bar{A} \mid B\right) = \frac{P\left(\bar{A}\right) \cdot P\left(B \mid \bar{A}\right)}{P\left(B\right)}.$$

Her kjenner vi $P\left(B \mid \bar{A}\right)$, men mangler de andre to faktorene. Vi må da foreta et par mellom-regninger. Først finner vi $P\left(\bar{A}\right)$:

$$P\left(\bar{A}\right) = 1 - P\left(A\right) = 1 - 0.01 = \underline{0.99}$$ (99% av testpersonene er ikke dopet).

Så skal vi finne $P\left(B\right)$. Da må vi bruke addisjonsregelen for komplementære hendelser:

$$P\left(B\right) = P\left(A\right) \cdot P\left(B \mid A\right) + P\left(\bar{A}\right) \cdot P\left(B \mid \bar{A}\right).$$

Her kjenner vi alt unntatt $P\left(B \mid A\right)$. Men denne sannsynligheten finner vi raskt:

$$P\left(B \mid A\right) = 1 - P\left(\bar{B} \mid A\right) = 1 - 0.05 = \underline{0.95}$$ (95% av de dopede avsløres av testen).

Nå nøster vi oss tilbake:

$$\begin{aligned} P\left(B\right) &= P\left(A\right) \cdot P\left(B \mid A\right) + P\left(\bar{A}\right) \cdot P\left(B \mid \bar{A}\right) \\ &= 0.01 \cdot 0.95 + 0.99 \cdot 0.001 = \underline{0.01049} \end{aligned}$$

(1.049% av alle testene gir utslag).

Da gir Bayes regel:

$$P\left(\bar{A} \mid B\right) = \frac{P\left(\bar{A}\right) \cdot P\left(B \mid \bar{A}\right)}{P\left(B\right)} = \frac{0.99 \cdot 0.001}{0.01049} = \underline{\underline{0.0944}}.$$

Dette betyr at i 9.44% av alle tilfellene der testen gir utslag, er utøveren *ikke* dopet!

Tilsvarende analyser kan foretas for eksempel for brannalarmer, der det er en viss sannsynlighet for at det går alarm uten at det er brann (falsk alarm), og en viss sannsynlighet for at det ikke går alarm hvis det brenner. Hva er da sannsynligheten for at det virkelig brenner når alarmen går?

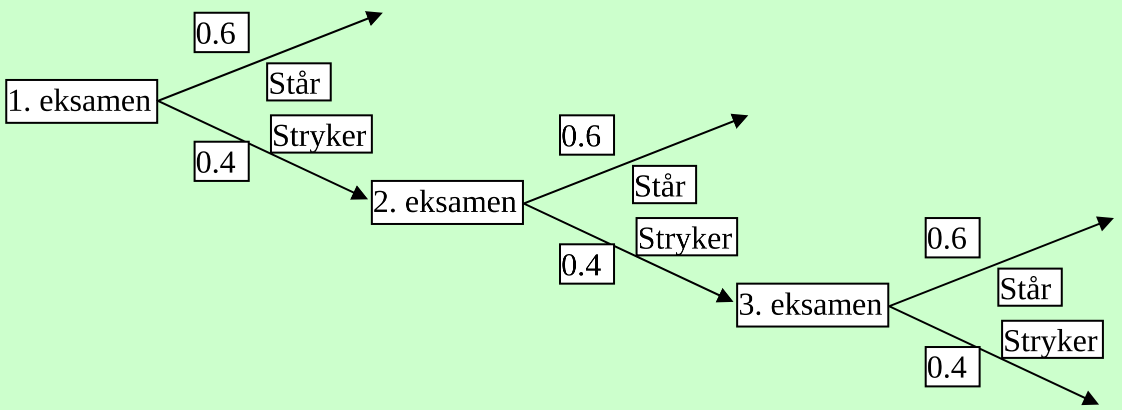
4.3.4. Hendelsestrær.

Det er ofte nyttig å kunne beregne sannsynligheter for sekvenser av hendelser. Et formelt oppsett av utfallsrom kan bli tungvint. Da kan det være lettere å illustrere hendelsene i et *tre-diagram*.

Jeg vil illustrere framgangsmåten med eksemplet nedenfor.

Eksempel 4.10: Dersom en student stryker til eksamen, kan han gå opp på nytt. Han har maksimalt tre forsøk. Anta at sannsynligheten for at en tilfeldig valgt student skal stå til eksamen er 0.6, og at denne sannsynligheten er den samme for hver eksamen. Hva er sannsynligheten for at studenten står i løpet av de tre forsøkene?

Løsning: Vi illustrerer situasjonen med å tegne et hendelsestre, der vi tegner inn situasjonene (eksamen nr. 1, 2 og 3), og føyer til mulige utfall med tilhørende sannsynligheter. Treet blir slik:



Sannsynligheten for at denne kandidaten skal stå i løpet av de tre forsøkene blir da:

$$0.6 + 0.4 \cdot 0.6 + 0.4 \cdot 0.4 \cdot 0.6 = \underline{\underline{0.936}}.$$

Dette resultatet kan vi finne på en annen måte. Sannsynligheten for å stryke alle tre gangene er $0.4^3 = \underline{0.064}$. Da blir sannsynligheten for *ikke* å stryke alle tre gangene (d.v.s. stå før eller senere) $1 - 0.064 = \underline{\underline{0.936}}$.

Slike hendelsestrær brukes mye innenfor risikoanalyse, der man setter opp oversikter over hvilke kombinasjoner av hendelser som kan føre til bestemte uønskede situasjoner. Man kan da beregne sannsynligheter for at slike situasjoner kan oppstå, og man kan også analysere hvordan man best (og billigst) kan redusere sannsynligheten for at uønskede situasjoner skal oppstå.

4.4. Kombinatorikk og sannsynlighet.

Det kan ofte være vanskelig å telle opp antall elementer i utfallsrommet og antall utfall som gir en bestemt hendelse. De grunnleggende tellereglene er gjennomgått i et eget notat om kombinatorikk, men vi skal repetere hovedtrekkene og benytte dem i sannsynlighets-regningen.

Det kan ofte lønne seg å benytte "urnemodellen": Tenk deg at alle mulige utfall av et eksperiment er kuler som ligger i en urne. Så trekker du kuler etter tur. Da må du skille mellom:

- Er det trekking med eller uten tilbakelegging?
- Betyr det noe i hvilken rekkefølge kulene trekkes?

Den grunnleggende telleregelen er:

Dersom en hendelse *A* kan forekomme på n_A forskjellige måter, og en annen hendelse *B* kan forekomme på n_B forskjellige måter, så kan kombinasjonen av *A* og *B* forekomme på $n_A \cdot n_B$ forskjellige måter.

Eksempel 4.11: Hvor mange ulike tipperekker fins det når det er 12 kamper på kupongen?

Løsning: Hver av kampene kan få 3 ulike utfall. Med 12 kamper kan vi da få

$$3^{12} = \underline{\underline{531441}}$$

forskjellige tipperekker.

Eksemplet ovenfor illustrerer en generell regel som gjelder for "trekking med tilbakelegging":

Dersom er eksperiment som kan få *r* ulike utfall gjentas *n* ganger, kan vi få r^n ulike kombinasjoner.

Vi får oftere bruk for "trekking uten tilbakelegging". La oss starte med et eksempel:

Eksempel 4.12: Du trekker to kort fra en kortstokk. Hvor mange forskjellige utfall kan du få i disse tilfellene:

a) Trekkingsrekkefølgen har betydning?
b) Trekkingsrekkefølgen er uten betydning?

Løsning:

a) Første trekking kan ha 52 ulike utfall. Andre trekking kan ha 51 ulike utfall. Til sammen blir det $52 \cdot 51 = \underline{\underline{2652}}$ ulike utfall dersom trekkingsrekkefølgen har betydning.

b) Dersom trekkingsrekkefølgen *ikke* har betydning, må du dele tallet ovenfor på antall måter å stokke de to kortene som er trukket ut. Dette kan gjøres på to måter. Du kan derfor få

$$\frac{52 \cdot 51}{2} = \underline{\underline{1326}}$$

ulike utfall dersom trekkingsrekkefølgen ikke har betydning.

Eksemplet ovenfor kan danne bakgrunn for reglene nedenfor, som vi setter opp uten flere forklaringer:

Når vi har *n* forskjellige objekter, og alle disse trekkes uten tilbakelegging, kan vi få $n!$ forskjellige kombinasjoner.

Vi sier også at *n* forskjellige objekter kan ***permuteres*** på $n!$ måter.

I mange tilfeller trekker vi ikke alle de *n* objektene vi har til disposisjon. Da gjelder disse reglene:

Dersom vi trekker *r* av de *n* objektene, og trekningsrekkefølgen er av betydning, kan vi få

$$n(n-1)\cdot\ldots\cdot(n-(r-1)) = \frac{n!}{(n-r)!}$$

forskjellige kombinasjoner. Denne formelen kalles ofte ***permutasjonsformelen***.

Dersom trekningsrekkefølgen *ikke* er av betydning, kan vi få

$$\frac{n!}{(n-r)! \cdot r!} = \binom{n}{r}$$

forskjellige kombinasjoner. Denne formelen kalles ofte ***kombinasjonsformelen***.

Eksempel 4.13: En pokerhand består av 5 kort. Hvor mange forskjellige pokerhender fins det?

Løsning: Du er kun interessert i *hvilke* kort du har fått, uten hensyn til trekkingsrekkefølgen. Da fins det

$$\frac{52!}{(52-5)! \cdot 5!} = \underline{\underline{2598960}}$$

ulike pokerhender.

Kombinasjonsformelen i ramma ovenfor kan tolkes slik: De *n* objektene (kulene i urna) deles inn i to grupper, en gruppe på *r* objekter og en annen gruppe på $(n-r)$ objekter. Når den innbyrdes rekkefølgen i disse gruppene er likegyldig, kan våre objekter ordnes på

$$\frac{n!}{r!\cdot(n-r)!}$$

ulike måter.

Dette resonnementet kan generaliseres slik:

Anta at *n* objekter kan deles inn i *k* grupper som inneholder $n_1,\ n_2,\ \ldots,\ n_k$ objekter der innbyrdes rekkefølge i hver gruppe er likegyldig, og rangeringen av grupper er likegyldig. Da kan de *n* objektene ordnes på

$$\frac{n!}{n_1!\cdot n_2!\cdot\cdots\cdot n_k!}$$

ulike måter.

Eksempel 4.14: Hvor mange *forskjellige* ord kan du danne ved å stokke om på bokstavene i ordet MISSISSIPPI ?

Løsning: Ordet MISSISSIPPI inneholder 4 S'er, 4 I'er2 P'er og 1 M. Ved å stokke om på bokstavene i ordet, kan du få i alt

$$\frac{11!}{4!\cdot 4!\cdot 2!\cdot 1!} = \underline{\underline{34650}}$$

ulike ord, fordi innbyrdes rekkefølge av de 4 I-ene og av de 4 S-ene og av de 2 P-ene (og den ene M-en) er likegyldig.

Nå som vi har lært å telle, er vi klar til å benytte disse kunnskapene til å beregne sannsynligheter. Vi støtter oss da til en av definisjonene på sannsynlighet:

Dersom et stokastisk eksperiment kan få *n* ulike utfall som alle er like sannsynlige, og n_g av disse er "gunstige", er sannsynligheten for "gunstig" utfall

$$\frac{n_g}{n}.$$

Dessuten får vi ofte bruk for den grunnleggende telleregelen: Dersom en hendelse *A* kan forekomme på n_A forskjellige måter, og en annen hendelse *B* kan forekomme på n_B forskjellige måter, så kan kombinasjonen av *A* og *B* forekomme på $n_A\cdot n_B$ forskjellige måter.

Eksempel 4.15: Du trekker 5 kort fra en kortstokk. Hva er sannsynligheten for å trekke
a) Ingen ess
b) Nøyaktig ett ess.
c) Nøyaktig to ess.

Løsning: Vi skal løse alle disse problemene på to måter, ved å telle antall kombinasjoner og ved å benytte hendelsestrær.

a) Du vet allerede at når du trekker 5 kort fra en kortstokk med 52 kort, og trekkingsrekkefølgen ikke har betydning, kan dette gjøres på

$$\frac{52!}{(52-5)!\cdot 5!} = \frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5!}$$

forskjellige måter.

Av de 52 kortene er det 48 kort som ikke er ess. Disse kortene kan permuteres på

$$\frac{48!}{(48-5)!\cdot 5!} = \frac{48\cdot 47\cdot 46\cdot 45\cdot 44}{5!}$$

forskjellige måter når trekkingsrekkefølgen ikke har betydning. Sannsynligheten for ikke å få noen ess er derfor

$$\frac{\frac{48\cdot 47\cdot 46\cdot 45\cdot 44}{5!}}{\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5!}} = \frac{48\cdot 47\cdot 46\cdot 45\cdot 44}{52\cdot 51\cdot 50\cdot 49\cdot 48} \approx \underline{\underline{0.659}}\,.$$

Vi kan også resonnere slik: Sannsynligheten for at det første kortet ikke er ess er $\frac{48}{52}$. Nå er det 51 kort igjen i stokken hvorav 47 ikke er ess. Sannsynligheten for at også det andre kortet er noe annet enn et ess er derfor $\frac{47}{51}$. Slik kan vi fortsette med alle de 5 kortene.

Sannsynligheten for at du ikke skal få noen ess er derfor

$$\frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50}\cdot\frac{45}{49}\cdot\frac{44}{48}$$

som er samme uttrykk som du fikk i sted.

b) Når du trekker ett ess, kan vi tenke at de 5 kortene deles i to grupper: En gruppe med ett ess og en gruppe med 4 ”ikke-ess”. Antall kombinasjoner av de 4 ”ikke-essene” er

$$\frac{48!}{(48-4)!\cdot 4!} = \frac{48\cdot 47\cdot 46\cdot 45}{4!}\,.$$

Hver av disse kombinasjonene kan du igjen kombinere med ett av de 4 essene.
Sannsynligheten for å få ett ess blant de 5 kortene er derfor

$$\frac{\frac{48\cdot 47\cdot 46\cdot 45}{4!}\cdot 4}{\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5!}} = \frac{48\cdot 47\cdot 46\cdot 45\cdot 4}{52\cdot 51\cdot 50\cdot 49\cdot 48}\cdot\frac{5!}{4!} = \frac{48\cdot 47\cdot 46\cdot 45\cdot 4}{52\cdot 51\cdot 50\cdot 49\cdot 48}\cdot 5 \approx \underline{\underline{0.299}}\,.$$

Vi kan også resonnere slik: Anta at de fire første kortene du trekker er ”ikke-ess” og at det femte kortet er et ess. Sannsynligheten for dette er

$$\frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50}\cdot\frac{45}{49}\cdot\frac{4}{48}\,.$$

Men du kan like gjerne få esset som første kort, eller som andre kort, eller som tredje kort, eller som fjerde kort. Vi må derfor multiplisere sannsynligheten ovenfor med 5, slik at sannsynligheten for å få nøyaktig ett ess blir

$$\frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50}\cdot\frac{45}{49}\cdot\frac{4}{48}\cdot 5$$

som er samme uttrykk som før.

c) Når du trekker to ess, kan vi tenke at de 5 kortene deles i to grupper: En gruppe med to ess og en gruppe med 3 ”ikke-ess”. Antall kombinasjoner av de 3 ”ikke-essene” er

$$\frac{48!}{(48-3)!\cdot 3!} = \frac{48\cdot 47\cdot 46}{3!}\,,$$

mens de to essene kan kombineres på

$$\frac{4!}{(4-2)!\cdot 2!}=\frac{4\cdot 3}{2!}$$

forskjellige måter. Sannsynligheten for å få nøyaktig to ess blant de 5 kortene er derfor

$$\frac{\frac{48\cdot 47\cdot 46}{3!}\cdot\frac{4\cdot 3}{2!}}{\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5!}}=\frac{48\cdot 47\cdot 46\cdot 4\cdot 3}{52\cdot 51\cdot 50\cdot 49\cdot 48}\cdot\frac{5!}{3!\cdot 2!}=\frac{48\cdot 47\cdot 46\cdot 4\cdot 3}{52\cdot 51\cdot 50\cdot 49\cdot 48}\cdot\frac{5\cdot 4}{2}\approx \underline{\underline{0.040}}.$$

Også dette resultatet kan du komme fram til med et annet resonnement: Anta at de tre første kortene du trekker er ”ikke-ess” og at de to siste kortene er ess. Sannsynligheten for dette er

$$\tfrac{48}{52}\cdot\tfrac{47}{51}\cdot\tfrac{46}{50}\cdot\tfrac{4}{49}\cdot\tfrac{3}{48}.$$

Men du kan like gjerne få essene som første og andre kort, eller som første og tredje kort, eller som en av de andre $\tfrac{5!}{(5-2)!\cdot 2!}$ mulige plasseringene. Sannsynligheten for å få nøyaktig to ess blir derfor

$$\tfrac{48}{52}\cdot\tfrac{47}{51}\cdot\tfrac{46}{50}\cdot\tfrac{4}{49}\cdot\tfrac{3}{48}\cdot\tfrac{5!}{3!\cdot 2!}$$

som er samme uttrykk som før.

Nå kan du innvende at det er ikke så viktig å bestemme sannsynlighetene for ulike utfall når du trekker fort fra en kortstokk. Men resonnementene ovenfor går igjen i en lang rekke praktiske situasjoner, som vist i neste eksempel:

Eksempel 4.16: En forening har 17 medlemmer, 10 menn og 7 kvinner. Det skal nedsettes en komite med 5 medlemmer, der det skal være minst to representanter av hvert kjønn. Hva er sannsynligheten for at dette kravet er oppfylt dersom komiteen nedsettes ved at medlemmer trekkes ut tilfeldig?

Løsning: Kravet ”to av hvert kjønn” innebærer to menn og tre kvinner, eller tre menn og to kvinner. Vi kan trekke ut 5 medlemmer av 17 på $\tfrac{17!}{(17-5)!\cdot 5!}=\tfrac{17!}{12!\cdot 5!}$ ulike måter. På samme måte kan vi trekke to menn av 10 på $\tfrac{10!}{(10-2)!\cdot 2!}=\tfrac{10!}{8!\cdot 2!}$ måter, og tre kvinner av 7 på $\tfrac{7!}{(7-3)!\cdot 3!}=\tfrac{7!}{4!\cdot 3!}$ ulike måter. En komite med to menn og tre kvinner kan da settes sammen på $\tfrac{10!}{8!\cdot 2!}\cdot\tfrac{7!}{4!\cdot 3!}$ ulike måter. Et tilsvarende resonnement fører til at en komite med tre menn og to kvinner kan settes sammen på $\tfrac{10!}{7!\cdot 3!}\cdot\tfrac{7!}{5!\cdot 2!}$ ulike måter. Sannsynligheten for at en tilfeldig uttrukket komite skal ha to av hvert kjønn blir da

$$\frac{\tfrac{10!}{8!\cdot 2!}\cdot\tfrac{7!}{4!\cdot 3!}+\tfrac{10!}{7!\cdot 3!}\cdot\tfrac{7!}{5!\cdot 2!}}{\tfrac{17!}{12!\cdot 5!}}=\underline{\underline{\frac{45}{68}}}\approx \underline{\underline{0.66}}.$$

Sannsynlighetsregningen oppstod i sin tid fordi gamblere hadde behov for å beregne sannsynligheter ved pengespill. Fremdeles ligger det mye fin sannsynlighetsregning i slike anvendelser (se for eksempel et lite notat om sannsynlighetsregning i poker). I dag brukes sannsynlighetsregning på en rekke områder, for eksempel innen forsikringsbransjen, som verktøy i forskning, og i forbindelse med risikoanalyser. Men uansett anvendelse er det de samme grunnleggende prinsippene som ligger til grunn.