

Sannsynlighet

Per G. Østerlie

20. april 2023

1	Sannsynlighetsregning	4
1.1	Hva og hvorfor?	4
1.2	Hvordan kan vi finne sannsynlighet?	4
1.3	Et stokastisk forsøk	5
1.4	Representasjon av sannsynlighet	6
1.5	Noen ord og uttrykk	6
1.6	Symboler	8
1.7	Sannsynlighetsmodeller	8
2	Illustrasjon av utfallsrom	12
2.1	Vennndiagram	12
2.2	Union og snitt	12
2.3	Krysstabell	14
3	Empirisk sannsynlighet	16
3.1	Introduksjon	16
3.2	De store talls lov	16
3.3	Simuleringer	17
3.4	Simulering med Python	18
4	Teoretisk sannsynlighet	21
4.1	Introduksjon	21
4.2	Enten den ene eller den andre	21
4.3	Både og	26
4.4	Bayes teorem	30
5	Kombinatorikk	36
5.1	Kombinatorikk	36
5.2	Multiplikasjonsprinsippet	36
5.3	Vi trekker kuler fra ei urne	39
5.4	Ordna utvalg med tilbakelegging	39
5.5	Ordna utvalg uten tilbakelegging	41
5.6	Uordna utvalg uten tilbakelegging	45
5.7	Uordna utvalg med tilbakelegging	47
5.8	En oppsummering	49
5.9	Oppgaver	50
6	Binomisk forsøk	54
6.1	Hva er et binomisk forsøk?	54
6.2	Et ektepar bestemmer seg for å få 6 barn	54
6.3	Tegnestifter	55
6.4	Et generelt binomisk forsøk	56
6.5	Andre eksempler på binomiske forøksserier	57
6.6	Eksempel: Straffespark	57
6.7	Newton-Pepys problem	60

7	Hypergeometrisk forsøk	62
7.1	Hypergeometrisk forsøk	62
7.2	Et klassisk eksempel	62
7.3	Et generelt utvalg	63
7.4	Et eksempel	64
	7.4.1 Simulering med Python	67
7.5	Lotto	67
7.6	Poker	69

Her er litt om sannsynlighetsregning.

Teksten er skrevet med $\text{\LaTeX}$.



Creative Commons Navngivelse-IkkeKommersiell-DelPåSammeVilkår 4.0 Internasjonal Offentlige Lisens

1 Sannsynlighetsregning

1.1 Hva og hvorfor?

Hva?

Vi får høre at det er sannsynlig at et eller annet kommer til å skje. Sannsynligheten for å bli truffet av en meteoritt skal være 1 til 1700 000 000. Sannsynligheten for å dø av røyking blant røykere er 50 %. Statistisk sett er sjansen for å bli truffet av lynet $\frac{1}{23000}$.

Hva er så sannsynlighet? Hva forteller opplysningene over? I sannsynlighetsregninga skal vi finne en verdi for hvor sannsynlig noe er. Det kan vi oppgi som «1 til 1700 000 000» eller «50%» eller « $\frac{1}{23000}$ ». Alt dette er verdier som forteller hvor sannsynlig det er at noe skjer. I sannsynlighetsregninga skal denne verdien være mellom 0 og 1. Alle verdiene over kan vi skrive om til en slik verdi. Ellers i dagliglivet omgås vi sannsynlighet uten å direkte tallfeste. Tenk bare på alle gangene du benytter ord som «umulig», «mulig», «helt sikkert», «fifty-fifty» osv. I nesten hver setning hvor disse inngår er det sannsynlighet et tema som ligger i bakgrunnen.

Hvorfor?

Hvorfor skulle vi være interesserte i å finne en sannsynlighet? Det er flere grunner til det. Her er noen eksempler.

Forretningsvirksomhet For den som driver butikk er det viktig å vite noe om sannsynligheten for å få solgt noe eller sannsynligheten for hva som vil skje i et marked. Tenk bare på vurderingene en aksjemegler må ta.

Forsikring All forsikring baserer seg på å vite noe om sannsynligheten for at et eller annet skjer. Ut fra det kan forsikringspremien settes

Investering Investeringsanalyse er sannsynlighetsregning

Spill Spiller du Lotto? Tipper? Sannsynligheten vil avgjøre om du vinner noe. Sannsynligvis gjør du ikke det.

Risikoanalyse Når risiko vurderes er det stort sett basert på sannsynlighet.

Værmelding Bare se på yr.no eller liknende tjenester. Der oppgir de til og med sannsynligheten for at været blir som de melder.

1.2 Hvordan kan vi finne sannsynlighet?

Det er i hovedsak to typer sannsynlighet: empirisk og teoretisk. Empiri er basert på erfaring. Tenk at vi kaster et pappbeger og at vi observerer hvordan det lander. Enten vil det lande pent med bunnen ned, på sida eller opp-ned. Før vi kaster er det vanskelig å tallfeste sannsynlighetene. Kaster vi dette pappbegeret tusen ganger og ser på forholdene mellom antall ganger det lander på de forskjellige måtene, kan vi begynne å uttale oss om sannsynligheten. Vi har

gjort en erfaring om kan tallfeste sannsynligheten. Da har vi funnet en empirisk sannsynlighet gjennom det som kalles de store talls lov.

Bytter vi ut pappbegeret med en mynt har vi et annet utgangspunkt. Ut fra de fysiske forutsetningene til mynten kan vi si noe om sannsynligheten for at den skal lande med mynt eller kron opp. Hvis mynten ikke skiller seg for mye ut fra vanlige mynter vil vi kunne si at det er like stor sannsynlighet for begge og at det ikke er mulig at den blir stående på høykant. Da kan vi tallfeste sannsynligheten til å være 0.5. Vi har funnet en teoretisk sannsynlighet. Det utelukker ikke at vi også kan finne denne sannsynligheten empirisk. Det er myntens symmetri, eller geometri, som avgjør konklusjonen vår. Av den grunn sier vi at vi benytter en geometrisk eller symmetrisk modell for å finne sannsynligheten.

Hva om vi trekker et tilfeldig kort fra en kortstokk og skal finne sannsynligheten for å trekke en kløver? Da er det rimelig å anta at sannsynligheten for det vil være $\frac{1}{4}$. Vi kan begrunne det med at en firedel av kortstokken inneholder kløver-kort. Her er det ikke geometrien, eller symmetrien som avgjør, men forholdet mellom de gunstige, kløver, og resten av kortene. Vi benytter en uniform modell.

I tillegg har vi også den subjektive sannsynlighet – en sannsynlighet basert på oppfatning. At det etter min oppfatning 73 prosent sannsynlighet for at Rosenborg vil vinne neste kamp er et eksempel på det.

I sannsynlighetsregning holder vi oss til empirisk og teoretisk sannsynlighet.

1.3 Et stokastisk forsøk

Kast med pappbeger eller mynt kaller vi stokastiske forsøk. Vi vet ikke helt hva det vil ende med: Det er tilfeldig.

I matematikken har du møtt matematiske modeller som funksjoner. De beskriver sammenhengen mellom to størrelser. Et eksempel er at sammenhengen mellom volumet av ei eske, V , og sida, x , i grunnflata er gitt ved:

$$V(x) = x^2(120 - 2x)$$

Her er det ikke rom for tvil. Hvis sida er 40 cm, vil volumet være

$$V(40) = 40^2(120 - 2 \cdot 40) = 64000$$

Vi kaller en slik modell for *deterministisk*. Det er bare ett eneste resultat når vi vet en innverdi. Slik er det ikke i *stokastiske* forsøk som beskrives av stokastiske modeller. Vi aner ikke hva som kommer til å skje når vi kaster en terning eller trekker et tilfeldig kort fra en kortstokk. Alle er eksempler på stokastiske forsøk. Sjøl om vi ikke kan si akkurat hva som vil skje, kan vi si noe om hva som er mulig kan skje. Vi kan også gjenta forsøkene flere ganger.

Definisjon 1 Stokastisk forsøk

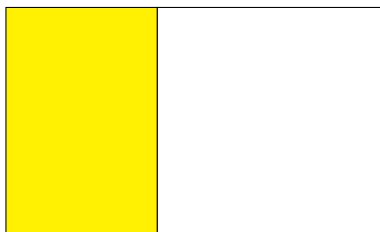
I et stokastisk forsøk kan vi

1. ikke forutsi hva resultatet av forsøket blir
2. gjenta forsøket flere ganger under samme forhold
3. beskrive de mulige utfallene av forsøket

1.4 Representasjon av sannsynlighet

Som et areal

Vi kan framstille sannsynlighet som et areal. Bruker vi bokstaven A som en erstatning for at noe skal skje, at du skal få en meteoritt i hodet eller vinne i Lotto, kan det gule arealet under vise hvor sannsynlig det er.



Tenk deg at dette er et åpent område og at det regner jevnt fordelt. Hvor stor del av de totale regndråpene vil falle på den gule delen? Svaret blir det samme som sannsynligheten for at A skal skje.

1.5 Noen ord og uttrykk

Utfall

Et utfall er et resultat av et stokastisk (tilfeldig) forsøk. Kaster vi tegnestift har vi utfallene «spiss opp» eller «spiss ned». Kaster vi en terning er eksempler på utfall: «firer», «ener», osv.

Definisjon 2 Utfall

Et utfall er ett av de enkelte resultatene et stokastisk forsøk kan få

Utfallsrom

Et utfallsrom er alle mulige utfall et forsøk kan ha. Ser vi på det tilfeldige forsøket å kaste en ideell terning med seks sider er utfallsrommet at terningen ender med en, to, tre, fire, fem eller seks øyne opp. Slik:

Med symboler kan vi skrive:

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

Inne i krøllparentesene står elementene, alle de enkelte delene, som utgjør mengden. Kaster vi

en tegnestift er utfallsrommet

$$U = \{\text{«spiss opp»}, \text{«spiss ned»}\}$$

Legg merke til at vi må ha med alle utfallene i utfallsrommet.

Definisjon 3 Utfallsrom

Utfallsrommet er mengden av alle utfallene ved et stokastisk forsøk

Hendinger

Det benyttes flere ord som betyr det samme. Mens noen kaller det en begivenhet, bruker andre hendelse, eller hending. Alle ordene betyr det samme. Uttrykk som «minst fire» og «to mynt på rad» kaller vi hendinger. Vi bruker store bokstaver som symboler for hendinger. Bokstavene kan vi velge sjøl slik disse eksemplene viser.

M – minst fire i et terningkast

B – trekke en knekt fra en kortstokk

Disse hendingene blir da

$$M = \{4, 5, 6\}$$

$$B = \{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\}$$

En hending kan omfatte ingen, ett eller flere av utfallene i utfallsrommet.

Definisjon 4 Hending

En *hending* er en delmengde av utfallsrommet

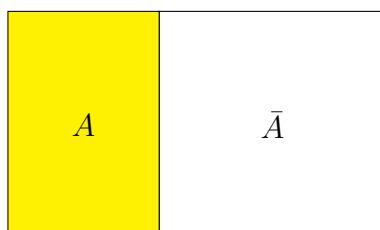
Fortsetter vi med eksemplet vårt kan vi definere denne hendingen:

A - hendingen spiss opp

Vi vil også få behov for å skrive det motsatte av at A inntreffer. Det kalles den komplementære hendinga og leses som «ikke A ». Vi skriver det slik

$\bar{A}$ - hendingen ikke spiss opp

I tilfellet med tegnestiften kan vi bruke symbolene når vi tegner opp de to hendingene.

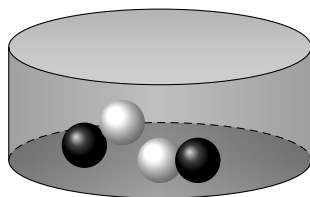


Disjunkte utfall

Disjunkte utfall er utfall som utelukker hverandre. Et eksempel er om vi kaster en mynt. De to utfallene kron og mynt kan ikke inntreffe samtidig. De utelukker hverandre og er derfor disjunkte.

Avhengige eller uavhengige hendinger

Hendinger kan enten være avhengige av hverandre eller så er de uavhengige. Kaster vi en terning, tegnestift eller en mynt er utfallene i hvert kast uavhengige av hverandre. Får vi en sekser i et terningkast er det akkurat like sannsynlig å få en sekser neste gang. Uansett hva vi innbiller oss så brukes ikke sekserne opp. Slik er det ikke alltid. Se bare på hva som skjer når vi trekker fra ei urne.



Vi trekker kuler fra urna uten å legge tilbake. Etter å ha trukket ei kule vil neste utfall være påvirket av det forrige. Trekker vi ei svart kule blant de fire som er i urna, vil sannsynligheten for å trekke svart neste gang være noe annet enn tidligere. De to hendingene er avhengige av hverandre.

1.6 Symboler

I sannsynlighetsregning får vi behov for å skrive «sannsynligheten for» gjentatte ganger. Da benytter vi symbolet P som er en forkortelse av probability. Et annet symbol en også kan støte på er Pr .

Definerer vi hendingen

G - å få en gutt ved en fødsel

kan vi nå skrive at

$$P(G) = 0.514$$

Det leser vi som «sannsynligheten for å få en gutt ved en fødsel er 0.514. Denne sannsynligheten er henta fra SSB og artikkelen: [Blir det gutt eller jente?](#)

1.7 Sannsynlighetsmodeller

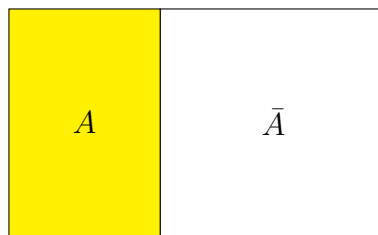
Når vi angir sannsynlighetene for alle utfallene for et stokastisk forsøk gir vi en sannsynlighetsmodell for forsøket. For en slik sannsynlighetsmodell stilles det visse krav.

Den russiske matematikeren Kolmogorov¹ (1934) har gitt betingelsene for en sannsynlighetsmodell som aksiomer.

Definisjon 5

- (i) Sannsynligheten for et utfall er fra og med 0 til og med 1
- (ii) Summen av sannsynlighetene til alle utfallene i utfallsrommet er lik 1
- (iii) Sannsynligheten til en hending er lik summen av sannsynlighetene til utfallene som utgjør begivenheten
- (iv) Sannsynligheten for en begivenhet som ikke kan skje, den umulige begivenheten $\emptyset$, skal være lik 0

At summen av sannsynlighetene skal være 1 har vi vært innom tidligere da vi tegnet figuren under.



Hele arealet er 1. Det betyr at sannsynligheten for komplementære hendinger skal være like 1.

Teorem 1

For en hendelse A og komplementærhendelsen $\bar{A}$ gjelder

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

Av teorem 1 følger det at $A \cap \bar{A} = \emptyset$. Kaller vi hele utfallsrommet for U har vi også at $A \cup \bar{A} = U$ og at $P(U) = 1$

¹Andrej Nikolajevitsj Kolmogorov (1903 – 1987) kjent russisk matematiker

En mer matematisk definisjon av sannsynlighetsmodell

Definisjon 6 Sannsynlighetsmodell

La funksjonen p ha utfallsrommet $U = \{u_1, u_2, \dots, u_n\}$ som definisjonsmengde og la verdiene $p(u_1), p(u_2), \dots, p(u_n)$ være slik at

- $0 \leq p(u) \leq 1$ for alle utfall i U
- $p(u_1) + p(u_2) + \dots + p(u_n) = 1$
- Hvis A er en begivenhet, så er: $p(A) = \sum_{u \in A} p(u)$
- $p(\emptyset) = 0$

Vi kaller funksjonen p for en sannsynlighetsfunksjon og verdien $p(u)$ for sannsynligheten for utfallet u .

Vi sier at U sammen med p utgjør en sannsynlighetsmodell for forsøket

Uniform sannsynlighet

Kaster vi en ideell terning vil sannsynligheten for at den lander på ei gitt side være lik $\frac{1}{6}$. Sannsynligheten gjelder for alle sider og er like stor: sannsynligheten er *uniform*

Definisjon 7 Uniform sannsynlighetsmodell

I en *uniform* sannsynlighetsmodell er alle enkeltutfallene like sannsynlige. Hvis utfallsrommet er $U = \{u_1, u_2, \dots, u_n\}$ betyr det at

$$P(u_1) = P(u_2) = \dots = P(u_n)$$

Egentlig er dette en diskret uniform sannsynlighetsmodell siden den bare gjelder for enkeltutfallene og ikke i et kontinuerlige område.

Teorem 2

I en uniform sannsynlighetsmodell med n mulige utfall er

$$P(u_1) = P(u_2) = \dots = P(u_n) = \frac{1}{n}$$

Definisjon 8

Hvis A er en delmengde av utfallsrommet U definerer vi

$$n(A) = \text{antall elementer i mengden } A$$

Legg merke til at n er en funksjon som gir antallet som utverdi.

Et mye brukt forhold

Teorem 3

I en uniform sannsynlighetsmodell med m mulige utfall er

$$P(A) = \frac{n(A)}{m}$$

Skriver vi om det over på en mer generell form kommer vi fram til denne brøken for uniform sannsynlighet

$$P(\text{ at noen skal skje}) = \frac{\text{antall gunstige for at det skjer}}{\text{antall mulige}}$$

Sannsynlighet som forholdet mellom antall gunstige og antall mulige er sentral i sannsynlighetsregninga.

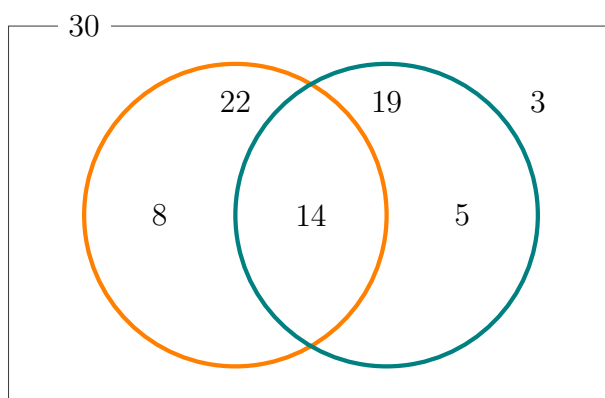
2 Illustrasjon av utfallsrom

2.1 Venndiagram

En måte å illustrere utfallsrom på er å tegne venndiagram. De er oppkalt etter en engelsk matematiker som het John Venn (1843-1923). Venndiagram forteller ikke noe om sannsynlighet, men viser sammenhengen mellom hendinger på en fin måte.

Vi ser på et eksempel: I en klasse med 30 elever ble det gjennomført en undersøkelse om hvor mange elever som liker saftis og hvor mange elever som liker fløteis. I undersøkelsen svarte 22 elever at de likte fløteis, 19 svarte at de likte pæreis, mens tre elever svarte at de ikke likte noen av istypene. Det er en enkel undersøkelse, men antall i hver kategori kan være vanskeig å illustrere.

Legger vi sammen alle antallene i undersøkelsen: $22 + 19 + 3$ får vi 44. Det er langt flere elever enn det er i klassen, så det regnestykket må være feil. Problemet er at vi regner med elevene som både liker fløte- og pæreis to ganger. Det kan vi illustrere i et venndiagram.



I venndiagrammet kan vi få fram hvor mange som liker begge de to istypene. Vi tegner sirkler for å vise mengden av alle fløteislikere. Så tegner vi en sirkel for alle pæreislikerne. De to sirklene må overlappe hverandre siden mange liker begge deler. Et venndiagram viser helheten og hvordan antall fordeler seg.

2.2 Union og snitt

For å illustrere union og snitt er det vanlig å bruke venndiagram. De er oppkalt etter en engelsk matematiker som het John Venn (1843-1923). Venndiagram forteller ikke noe om sannsynlighet, men viser sammenhengen mellom hendinger på en fin måte. La oss se på et eksempel. Vi kaster den samme terningen som så mange ganger før. Da kan vi få disse hendingene

A – antall øyne er et oddetall

B – antall øyne er mindre enn fem

Skriver vi ned elementene i utfallene i de to hendingene får vi

$$A = \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \quad \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \quad \bullet \quad \bullet \\ \hline \end{array}$$

$$B = \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \quad \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \quad \bullet \quad \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \quad \bullet \\ \hline \end{array}$$

Eller skrevet matematisk med elementer

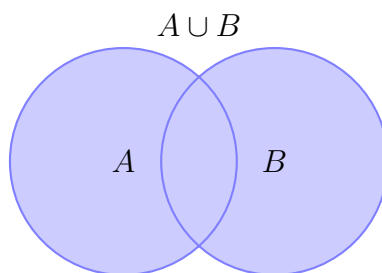
$$A = \{1, 3, 5\}$$

$$B = \{1, 2, 3, 4\}$$

Union

Tidligere var Norge i en union med Sverige. De to landene var slått sammen til ett land. Det er nettopp det en union er: En sammenslåing av to mengder.

En union kan vi vise i et venndiagram på denne måten



Venndiagrammet minner også litt om Norge og Sverige?

Hva blir unionen mellom de to hendingene A og B ?

Bruker vi de matematiske symbolene kan vi skrive:

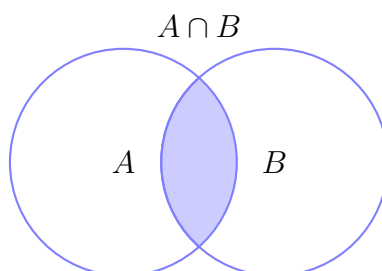
$$A \cup B = \{1, 3, 5\} \cup \{1, 2, 3, 4\} = \{1, 2, 3, 4, 5\}$$

$\cup$ er det matematiske symbolet for union

«Eller» er et stikkord når det gjelder union siden unionen vil være alle elementene som er med i mengden A eller i mengden B .

Snitt

Et snitt er det som er felles for to mengder. I et venndiagram kan vi framstille et snitt slik



Nå kan vi skrive snittet av de to mengdene A og B

$$A \cap B = \{1, 3, 5\} \cap \{1, 2, 3, 4\} = \{1, 3\}$$

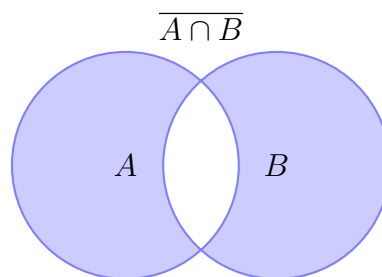
Bare tallet 1 og 3 er felles for de to.

$\cap$ er det matematiske symbolet for snitt

«og» er et stikkord siden snittet består av de elementene som fins i både den ene og den andre mengden.

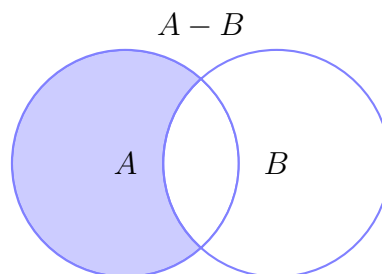
Noen andre eksempler

Alt annet enn snittet av A og B kan illustreres slik



I logikken kan $\overline{A \cap B}$ også skrives som $A \text{ xor } B$. På norsk er også *neller* brukt for *xor*

Venn diagrammet som viser mengde A, men ikke mengde B blir da slik



2.3 Krysstabell

Vi kan starte med et eksempel hvor det er gjennomført en spørreundersøkelse

Eksempel 1

Det ble gjennomført en spørreundersøkelse blant elevene som deltar i et matematikkurs. Det ble spurt om kjønn og om egevaluering alltid ble levert. Til sammen ble 120 elever spurt. 42 av dem svarte at de alltid skrev egenvurdering. 61 av elevene var gutter. 29 av jentene svarte at de alltid leverte egevaluering.

Vi ønsker å framstille dette resultatet og alle opplysningene vi har på en oversiktlig måte. Da er en krysstabell et nyttig verktøy. Opplysningene vi får i eksemplet er

120 – summen av alle elevene

61 – antall gutter

29 – antall jenter som alltid leverte egevaluering

42 – antall elever som alltid leverte egevaluering

Opplysningene gjør at vi kan sette opp en krysstabell og legge til alle verdiene som mangler.

	Gutter	Jenter	Sum
Har alltid levert	13	29	42
Har ikke levert	48	30	78
Sum	61	59	120

I tabellen er de manglende verdiene markert med **denne fargen**.

En slik framstilling gjør at vi finne betinga sannsynlighet for alle mulige hendinger.

3 Empirisk sannsynlighet

3.1 Introduksjon

For å finne en empirisk sannsynlighet må vi gjenta et stokastisk forsøk mange ganger og så se på den relative frekvensen av utfallene.

Vi kan ta et eksempel som er gjennomført med noen elever. De ble delt i åtte grupper og hver gruppe kastet en tegnestift to hundre ganger. For hvert kast noterte de om tegnestiften landet med spissen opp eller spissen ned. Her er resultatet:

	A	B	C	D	E	F	G	H	Sum
opp	123	107	120	125	105	126	120	126	952
ned	77	93	80	75	95	74	80	74	648

Nå kan vi regne ut sannsynligheten for at denne tegnestiften skal lande med spissen opp ved å se på forholdet antall ganger den lander med spissen opp og antall ganger den ble kastet:

$$\frac{952}{952 + 648} = 0.595$$

Hvor nøyaktig vi kan tallfeste denne sannsynligheten er et annet spørsmål, men vi kan runde av til 0.60.

Vi har nå funnet sannsynligheten som en relativ frekvens:

$$\text{Sannsynligheten for spiss opp} = \frac{\text{Antall ganger med spissen opp}}{\text{Totalt antall kast}}$$

Det er viktig at dette gjelder for den samme type tegnestift som ble benyttet i dette forsøket. Vi kan tenke oss at vi ville fått et annet resultat om tegnestiften var annerledes utformet.

3.2 De store talls lov

Forsøket illustrer de store talls lov – et svært viktig grunnlag for sannsynlighetsregningen.

Teorem 4 Store talls lov

Ved økende antall forsøk hvor utfallene er uavhengige vil den relative frekvensen til hendingen A gå mot sannsynligheten for A

$$\frac{f_n(A)}{n} \rightarrow P(A) \text{ når } n \rightarrow \infty$$

3.3 Simuleringer

Går vi til ordboka finner vi denne definisjonen

simulere **simulere** v2 (fra latin; beslektet med *simili-*)
1 forestille seg, late som
2 etterligne et (konstruert) hendelsesforløp (ved hjelp av datamaskin)

Her gjelder den siste definisjonen. Vi etterlikner et fenomen. Ofte kan det være strevsomt å gjennomføre forsøk som krever gjentatte kast eller annen aktivitet. De må gjentas mange ganger før vi kan være sikre på resultatet. Da kan det være greit å simulere omtrent det samme. Gjør vi det er det viktig å huske at vi ikke utfører forsøket. Det blir bare simulert ved hjelp at en algoritme genererer tilfeldige tall.

I GeoGebra kan vi simulere et terningkast ved å be om et tilfeldig heltall, fra og med en og til og med seks, ved kommandoen

```
TilfeldigMellom[1,6]
```

På den måten kan vi simulere et kast med en ideell terning.

Regneark inneholder samme muligheter for simuleringer.

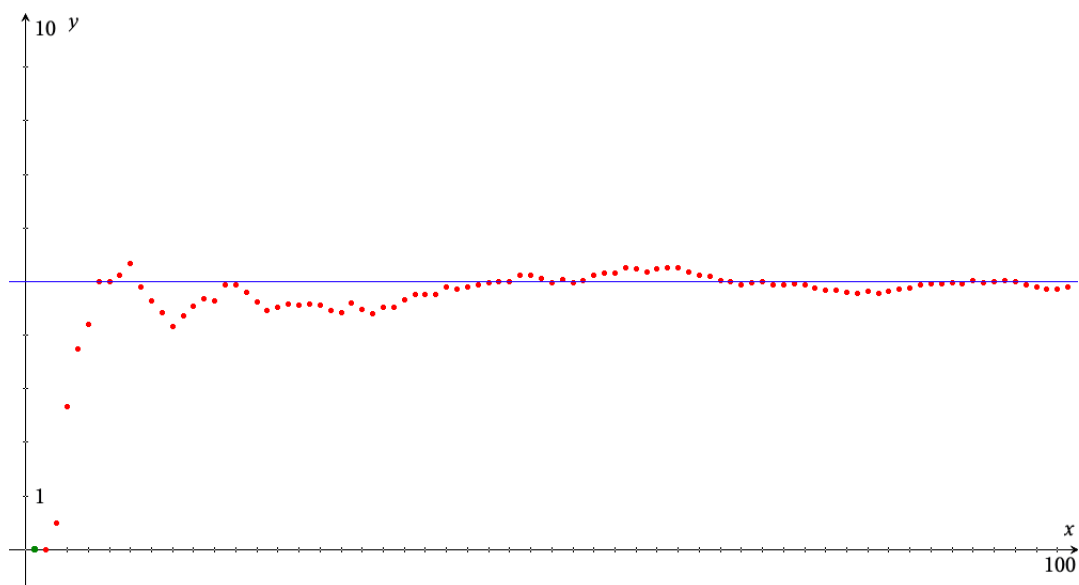
En del programvare, og noen kalkulatorer, har muligheten for å definere rekursive formler. Et eksempel er programvaren TI-Nspire hvor det er mulig å definere rekursive formler. Legger vi inn denne kommandoen:

```
u1(n)=u1(n-1)+randInt(0,10)
```

betyr det at $u1_n$ skal være lik det forrige leddet $u1_{n-1}$ addert med et tilfeldig tall mellom null og ti. Lager vi en ny rekursiv formel

```
u2(n)=((u1(n-1))/(n-1))
```

kan det gi oss den relative frekvensen.



En programvare som benyttes mye i statistikk er R. Programvaren kan være krevende å benytte, men gir gode muligheter.

Simuleringer kan vi også gjennomføre ved programmering i Python. I dette kompendiet kommer vi til å benytte Python.

3.4 Simulering med Python

Når vi skal simulere stokastiske forsøk med Python vil vi få bruk for tilfeldig genererte tall. Da må vi importere en modul som inneholder algoritmer som gjør det mulig. Modulen `random` har to funksjoner som genererer «tilfeldige tall» eller pseudo-tilfeldige tall. Når vi må være litt forsiktig med å kalle dem helt tilfeldige tall skyldes det at de er generert ved en algoritme. Det fins flere måter å gjøre det på og i noen tilfeller, særlig på enkle kalkulatorer, kan en oppleve at det de samme tallene som gjentar seg. I Python er det nok mer avansert, men samtidig viktig å være klar over at vi nettopp *simulerer*. Vi kaster ikke en terning eller en mynt, men vi får ei datamaskin til å generere pseudo-tilfeldige tall.

La oss se på en programkode som kaster en terning 100 ganger.

```
1 from random import randint #Importerer modulen random
2
3 for i in range(100): #gjentar det som er rykket inn under 100 ganger
4     kast = randint(1,6) #genererer et tall fra og med 1 til og med 6
5     print(kast)
```

Programkode 3.1: Terningkast

Programkode 3.1 består av ei `for`-sløyfe hvor det som er rykket inn under gjentas flere ganger. Variabelen `i` holder styr på antall ganger. Det står at den skal være i området (`range`) fra og med null og opp til 100. For hver gjentakelse, eller iterasjon, i sløyfa økes `i` med én. Kommandoen `randint(1,6)` genererer et tilfeldig heltall fra og med 1 til og med 6. Denne verdien tilordnes til variabelen `kast`. Til slutt skrives verdien til variabelen ut til skjermen.

Tilfeldige tall?

Datamaskina kan verken kaste terninger eller mynter. For å generere et tilfeldig tall utføres en kode med en algoritme. Algoritmen som benyttes i dette tilfellet kalles *The Mersenne Twister*. Du kan lese mer her: [Mersenne Twister](#). Navnet kommer fra Mersenne-primtall som er $2^{19937}-1$. Denne verdien er lengden på den lengste perioden, antall tall før det hele begynner å gjenta seg. Tilfeldige tall fra en slik algoritme kalles derfor *pseudotilfeldige tall* siden de ikke egentlig er helt tilfeldige. Det er teorien – for våre simuleringer gjør disse verdiene jobben. Algoritmen tar utgangspunkt i en startverdi. Den kan vi spesifisere, men om vi ikke gjør det hentes den fra operativsystemet. Da er det som regel klokka som benyttes. Hvis vi vet denne startverdien kan vi også finne ut hvilke tall som genereres. Les mer om modulen random i dokumentasjonen <https://docs.python.org/2/library/random.html>

Vi kan utvide programmet til å finne sannsynligheten for å få en sekser når vi kaster en terning flere ganger.

```
1 from random import randint
2 s = 0 # antall seksere
3
4 for i in range(1000):
5     terning = randint(1,6)
6     if terning == 6:
7         s = s + 1
8     r = s / (i+1)
9
10 print(r)
```

Programkode 3.2: Simulering av terningkast

I programkode 3.2 teller vi opp antall seksere. En hvis-setning gjør det: `if terning == 6:`. Hvis det tilfeldige tallet er lik 6, så skal vi øke telleren med 1. For hver gjentakelse regnes den relative frekvensen ut. Etter at `for`-sløyfa er gjennomført, skrives den relative frekvensen ut.

I mitt tilfelle ble resultatet

0.172

Vi observerer at det ikke er langt unna den antatte sannsynligheten for en ideell terning $\frac{1}{6} \approx 0.166666\dots$

Det hadde vært fint å fått illustrert dette grafisk også. Vi kan utvide programmet vårt for å få til det.

```
1 # Store tall's lov: Terningkast
2 # Relativ frekvens for å få en sekser
3 from random import randint
4 import matplotlib.pyplot as plt
5 s = 0 # antall seksere
6 xverdier = []
7 rverdier = []
8 for i in range(1000):
```

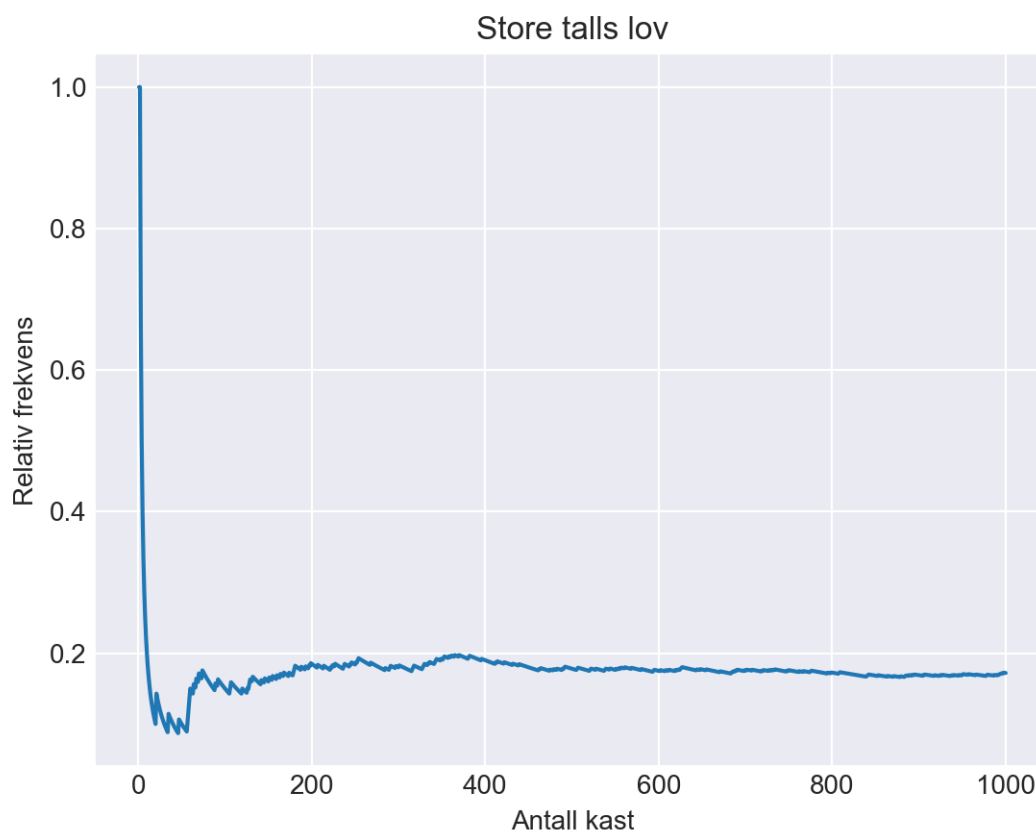
```

9     terning = randint(1,6)
10    if terning == 6:
11        s = s + 1
12    r = s / (i+1)
13    xverdier.append(i+1)
14    rverdier.append(r)
15
16    plt.style.use("seaborn-darkgrid")
17    plt.xlabel("Antall kast")
18    plt.ylabel("Relativ frekvens")
19    plt.title("Store talls lov")
20    plt.plot(xverdier,rverdier)
21    plt.show()
22    print(r)

```

Programkode 3.3: Grafisk framstilling

Kort forklart må vi importere en modul som kan vise grafikk og vi må benytte lister. Figur 3.1 viser resultatet når jeg kjørte programmet. Hver oppmerksom på at slike simuleringer gir forskjellig resultat for hver gang.



Figur 3.1: Relativ frekvens seksere

4 Teoretisk sannsynlighet

4.1 Introduksjon

Vi skal stort sett se på teoretisk sannsynlighet. En teoretisk sannsynlighet kan vi finne ved teoretiske betraktninger: Vi prøver å gjøre noen antakelser for så å finne en sannsynlighet for at noe skal inntreffe. Dette er en annen innfallsvinkel enn den empiriske, hvor vi utfører stokastiske forsøk for å finne ut en sannsynlighet.

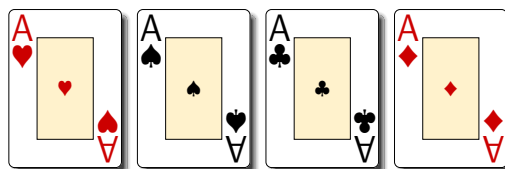
Et mye brukt forhold

Skriver vi om det over på en mer generell form kommer vi fram til denne brøken for sannsynlighet

$$P(\text{at noen skal skje}) = \frac{\text{antall gunstige for at det skjer}}{\text{antall mulige}}$$

Sannsynlighet som forholdet mellom antall gunstige og antall mulige er sentral i sannsynlighetsregninga.

Et eksempel på hvordan vi kan beregne en sannsynlighet er å trekke et kort fra en godt stokka kortstokk. La oss si at vi ønsker å finne sannsynligheten for å trekke et ess fra kortstokken. Totalt er det 52 kort i kortstokken. Kortene som vil være gunstige for oss er



Det er fire kort. Vi definerer hendingen

A - trekke et ess

Sannsynligheten blir nå

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

4.2 Enten den ene eller den andre

I sannsynlighetsregninga er det noen situasjoner som går igjen. Noen situasjoner kan stille flere krav. Slike situasjoner kan vi stort sett dele i to: «enten eller» eller «både og». Vi starter med å se på den første.

Vi kaster en terning

Et enkelt eksempel et terningkast. Vi kaster en ideell terning og skal finne sannsynligheten for at vi enten får fire eller fem.

Utfallrommet for forsøket er da:



Vi definerer hendingene

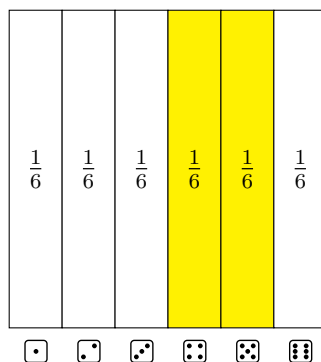
A – terningen lander med fire øyne opp

B – terningen lander med fem øyne opp

Da kan vi benytte symbolene til å skrive sannsynligheten for at vi enten får fire eller fem.

$$P(A \cup B)$$

Tenker vi oss sannsynlighetene som arealer kommer vi fram til denne figuren.



Dette kunne vi også tegnet som et sektordiagram

Hvor stor areal utgjør det som er gunstig for hendingen $A \cup B$? Jo, det må bli det gule arealet:

$$\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

En annen måte å finne det på er å se på utfallrommet og hva som er gunstig for utfallet. Teller vi opp finner vi at to av totalt seks hendinger er gunstige: $\frac{2}{6} = \frac{1}{3}$

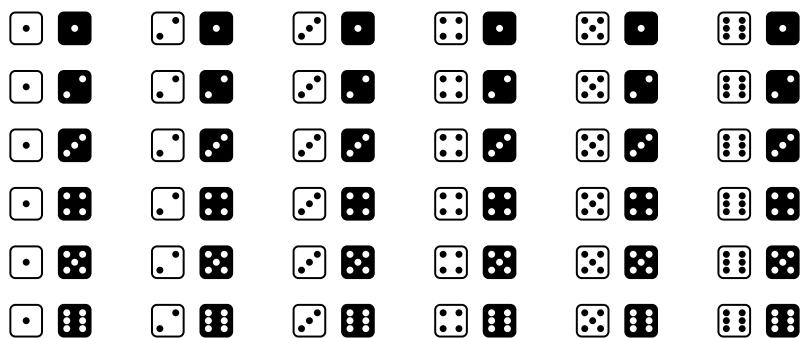
Da har vi at

$$P(A \cup B) = \frac{1}{3}$$

Vi ser at resultatet blir akkurat det samme!

Kast med to terninger

Vi vil benytte kast med to terninger som et eksempel framover. Tenker vi oss at vi har to terninger, en hvit og en svart, og kaster disse vil utfallsrommet være:



La oss se på et eksempel: Hva er sannsynligheten for at terningene enten har summen tre eller 6?

Vi definerer hendingene

A – summen av øynene er tre

B – summen av øynene er seks

Vi skal finne $P(A \cup B)$

La oss først se på utfallene og telle antall mulige og antall gunstige. Her er de som gir summen tre:



Her er de gunstige for å få seks:



Teller vi opp ser vi at vi har sju gunstige utfall. Totalt er det 36 mulige utfall. Sannsynligheten blir da:

$$P(A \cup B) = \frac{7}{36}$$

La oss finne sannsynligheten ved å tenke se på sannsynlighetene $P(A)$ og $P(B)$.

$$P(A) = \frac{2}{36}$$

$$P(B) = \frac{5}{36}$$

Hva blir summen av de to? Jo, $\frac{2}{36} + \frac{5}{36} = \frac{7}{36}$ Legg merke til at de to hendingene ikke hadde noe felles utfall. Utfallene er disjunkte.

Teorem 5

For to disjunkte hendelser A og B gjelder

$$P(A \cup B) = P(A) + P(B)$$

Hva om de to hendingene har felles utfall?

Vi kaster to terninger igjen og definerer disse to hendingene:

C – den hvite lander med 1 opp

D – den svarte lander med 1 opp

Her er utfallene som er gunstige for C



Her er utfallene som er gunstige for D



Et utfall er felles for C og D



Hvis vi skriver ned sannsynlighetene får vi

$$P(C) = \frac{6}{36}$$

$$P(D) = \frac{6}{36}$$

$$P(C \cap D) = \frac{1}{36}$$

Når vi nå skal finne sannsynligheten for $P(C \cup D)$ må vi passe på at vi ikke tar med et utfall flere ganger. Teller vi opp de gunstige utfallene får vi 11. Vi har at $P(C \cup D) = \frac{11}{36}$

Det samme finner vi slik

$$P(C \cup D) = \frac{6}{36} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}$$

Vi må trekke fra sannsynligheten for det felles utfallet $P(C \cup D) = \frac{11}{36}$.

Det gir oss addisjonsregelen for to utfall, A og B , som ikke er disjunkte

Teorem 6

For to ikke disjunkte utfall A og B har vi at

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

Python

Simulering med Python

Python kan benyttes for å simulere terningkast. Med kommandoen `randint(1,6)` får vi generert et tilfeldig tall i mengden $\{1, 2, 3, 4, 5, 6\}$. Kommandoen er ikke innebygd i Python, så vi må hente den inn fra et bibliotek som heter `random`. Det gjør vi med linja `from random import randint`. Det gir oss muligheten til å generere pseudo-tilfeldige tall. Ved å lage ei `for`-løkke kan vi gjenta «kastene» og telle opp resultatet for hver gang.

Ved simuleringer er det igjen viktig å være klar over at vi ikke kaster en terning, men lar teknologien prøve å etterlikne et terningkast. Her er hvordan vi kan gjøre det med Python-kode:

```
1 from random import randint
2
3 ant_kast = 1000
4 teller_1 = 0
5 teller_2 = 0
6 teller_3 = 0
7
8 for i in range (ant_kast):
9     terning_1 = randint(1,6)
10    terning_2 = randint(1,6)
11    if terning_1 == 1:
12        teller_1 = teller_1 + 1
13    if terning_2 == 1:
14        teller_2 = teller_2 + 1
15    if terning_1 == 1 or terning_2 == 1:
16        teller_3 = teller_3 + 1
17
18 print("Sannsynlighet for at terning 1 blir 1: ", teller_1/ant_kast)
19 print("Sannsynlighet for at terning 2 blir 1: ", teller_2/ant_kast)
20 print("Sannsynlighet for at en av terningene blir 1: ", teller_3/
    ant_kast)
```

Programkode 4.1: Kast med to terninger

I programkode 4.1 telles det opp hvor mange ganger vi får 1 når vi «kaster» 1000 ganger. Resultatet skrives ut. I ett tilfelle ble dette resultatet:

```
Sannsynlighet for at terning 1 blir 1: 0.169
Sannsynlighet for at terning 2 blir 1: 0.153
Sannsynlighet for at en av terningene blir 1: 0.298
```

Her $\frac{11}{36} \approx 0.305556$. Det er ikke langt unna resultatet av simuleringen.

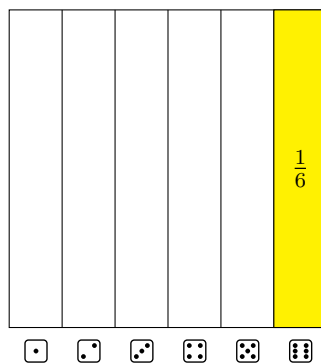
4.3 Både og

I de tilfellene vi har sett på tidligere fant vi sannsynligheten når enten det ene, eller det andre, kriteriet måtte oppfylles. Vi kunne addere sannsynlighetene. Nå skal vi se på situasjoner hvor det stilles to krav og begge må oppfylles. La oss finne sannsynligheten for å få to seksere på rad når vi kaster en ideell terning. Her er det to krav: vi må få sekser på første kast og på andre kast. Desto fler krav som stilles, desto mindre blir sannsynligheten. Vi definerer hendingene

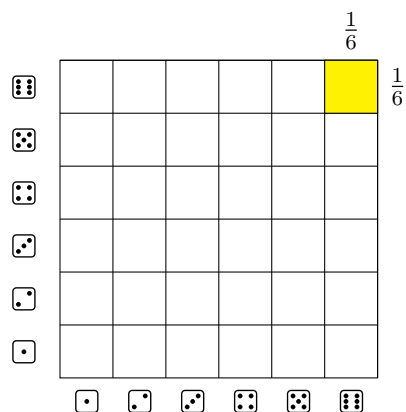
S – terningen lander med sekser opp

$\bar{S}$ – terningen lander ikke med sekser opp

Ut fra geometrien til terningen kan vi anta at $P(S) = \frac{1}{6}$. Kaster vi bare en gang har vi tidligere framstilte det som $\frac{1}{6}$ av et kvadrat. Sannsynligheten representeres som en del av et areal.



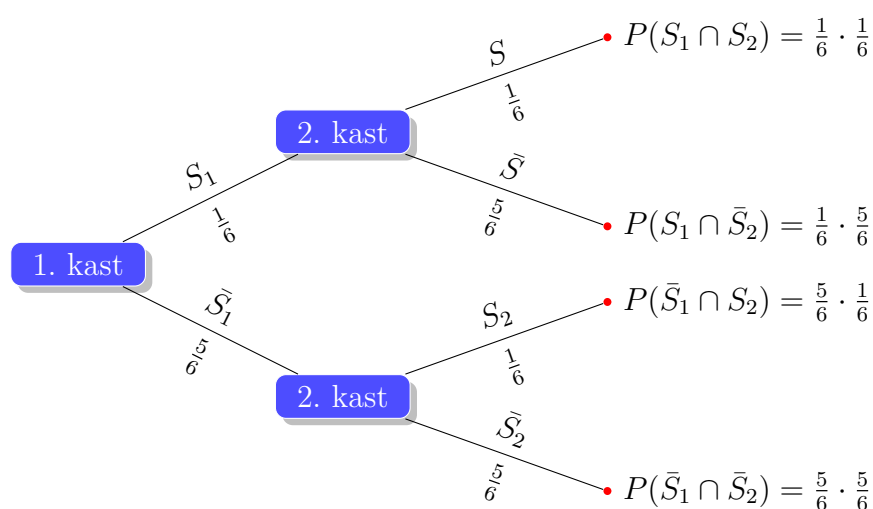
Med et tilleggskrav vil denne sannsynligheten reduseres. Vi skal kaste terningen en gang til og må få det med i arealmodellen vår. Sannsynligheten for at vi skal få en sekser til vil være $\frac{1}{6}$ av den forrige. Det kan vi framstille i denne figuren.



En slik arealmodell har sine begrensninger. Hva om vi kaster en gang til og skal finne sannsynligheten for tre sekser på rad? Vi kan kanskje representere det med en kube, men det viser at mer komplekse stokastiske forsøk ikke kan representeres med areal. Sånn er det ofte med arealmodeller, men for enkle situasjoner kan de bidra til å forklare hvorfor produktet av delsannsynlighetene blir framgangsmåten for å finne sannsynligheten.

Valgtrær kan være lure

I et valgtre, eller et sannsynlighetstre, tegner vi opp hva vi skal gjøre og alle mulige utfall. I tillegg skriver vi på sannsynlighetene. Her er et eksempel for de to terningkastene hvor det er benyttet en indeks for å markere rekkefølgen på sekserene



For å finne sannsynligheten for å få to sekser på rad må vi først få en sekser på første kast, så må vi få en sekser på andre kast. På veien gjennom valgtreet minker sannsynligheten for hvert steg. Valgtreet kan også bidra til å forklare at vi må multiplisere de to sannsynlighetene for å finne svaret. Vi kan skrive det som

$$P(S_1 \cap S_2) = \frac{1}{6} \cdot \frac{1}{6}$$

Vi kan også simulere det samme med regneark eller programmering og vise at de teoretiske beregningene stemmer.

Python

```
1 from random import randint
2 teller = 0
3 antall = 10000
4 for i in range(antall):
5     kast = randint(1,6) #første kast
6     if kast == 6:
7         kast = randint(1,6) #andre kast
8         if kast == 6:
9             teller = teller + 1
10
11 print("Reativ frekvens: ", teller/antall)
```

Programkode 4.2: Simulering av to kast med terning

Etter å ha kjørt programmet fikk jeg

```
Reativ frekvens: 0.0268
```

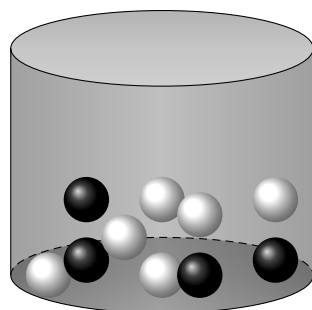
Et resultat som er svært nært $\frac{1}{36} \approx 0.0277$ og som viser at den teoretiske sannsynligheten stemmer meget godt overens med den simulerte modellen.

Et tilsvarende forsøk vil være å kaste to terninger en gang. Tenk litt på det...det blir det samme!

Betinge sannsynlighet

I de tidligere eksemplene har hvert forsøk vært uavhengig av det andre: Å få en sekser på ett kast er ikke avhengig av hva vi fikk på forrige kast. Det er ikke alltid tilfelle. La oss se på et nytt eksempel.

Vi trekker kuler fra ei urne hvor det er 6 hvite og 4 svarte. Vi trekker uten å legge tilbake kulene. Hva er sannsynligheten for å trekke først ei hvit kule og så ei svart?



Vi definerer hendingene

H_1 – trekker ei hvit kule i første trekkingen

S_2 – trekker ei svart kule i andre trekkingen

Nå kan vi skrive sannsynligheten vi skal finne som:

$$P(H_1 \cap S_2)$$

Siden vi ikke legger tilbake kulene vi trekker vil utfallet av det første trekket påvirke sannsynligheten for det andre.

Sannsynligheten for å trekke ei hvit kule ved første trekk finner vi ved antall gunstige delt på antall mulige. Det gir:

$$P(H_1) = \frac{6}{10}$$

Sannsynligheten for å trekke ei svart kule i andre trekk når vi vet at det er trukket ei hvit kule først er:

$$P(S_2 | H_1) = \frac{4}{9}$$

Legg merke til skrivemåten. Tegnet $|$ leser vi som «gitt at». Denne sannsynligheten er betinget av hva som skjedde i forrige trekk

Begge kravene må tilfredstilles og vi må finne produktet av de to sannsynlighetene

$$P(H_1 \cap S_2) = P(H_1) \cdot P(S_2 | H_1) = \frac{6}{10} \cdot \frac{4}{9}$$

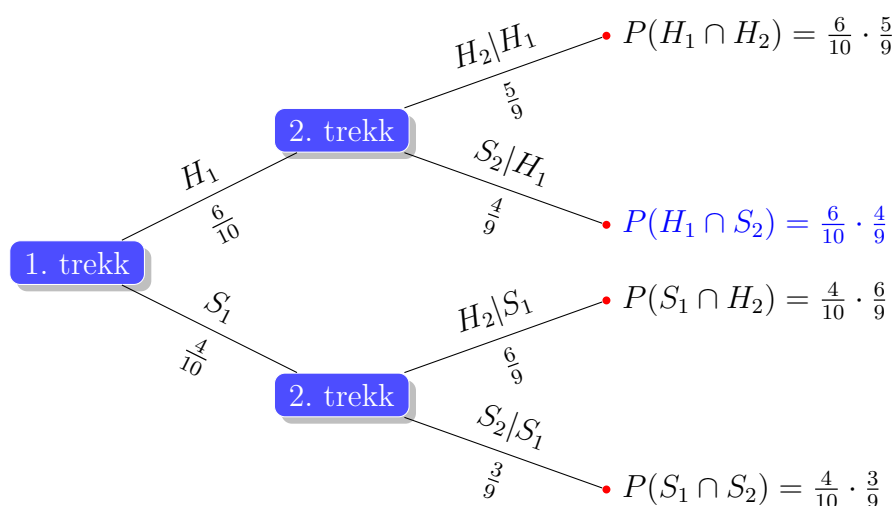
Da har vi funnet svaret $P(H_1 \cap S_2) = \frac{6}{10} \cdot \frac{4}{9} = \frac{4}{15}$

Vi kan tegne et valgtre. Treet gir svar på denne oppgaven og alle andre varianter av de to trekkene. Ut fra de hendingene vi allerede har definert får vi:

H_2 – trekker ei hvit kule i andre trekkingen

S_1 – trekker ei svart kule i første trekkingen

Valgtreet



Kan du lese hva sannsynligheten for å trekke ei svart kule i første trekk og så ei svart i andre? Svaret må bli $P(S_1 \cap S_2) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}$.

Produktsetningen

Vi lar A og B igjen være to hendinger. Da kan vi skrive produktsetningen som

Teorem 7 Den generelle produktsetningen

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

Hvis hendingerne er uavhengige vil $P(B) = P(B \mid A)$ og vi har:

Teorem 8

For to uavhengige hendinger, A og B , gjelder

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

En del eksempler på sannsyn Ikke binomisk og hypergeoemtrisk

4.4 Bayes teorem

Thomas Bayes

Mannen som har gitt navnet til den setningen vi skal se på er engelskmannen Thomas Bayes (1702 - 1761). Han var både matematiker, statistiker og prest. Nå diskuteres det i hvor stor grad han fortjener å ha navnet sitt forevigt i denne setningen. Han skrev den aldri slik vi nå kjenner den i alle fall. Det han gjorde var en del eksperimenter og tanker knytta til det den moderne varianten av setningen beskriver.



Figur 4.1: Thomas Bayes

La oss se nærmere på det. Ut fra produktsetningen (teorem 7) for avhengige hendinger har vi at

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B \mid A) \\ P(B \cap A) &= P(B) \cdot P(A \mid B) \end{aligned}$$

Nå har vi også at $P(A \cap B) = P(B \cap A)$ og det betyr at

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

Det gir oss en sammenheng mellom de to betingte sannsynlighetene.

Litt omformulert får vi setningen som er oppkalt etter Thomas Bayes.

Teorem 9 Bayes' teorem

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

Dette teoremet er nyttig i mange sammenhenger hvor det kan være enkelt å bestemme $P(B | A)$ og vi ønsker å finne $P(A | B)$.

Vi kan se på det med et eksempel.

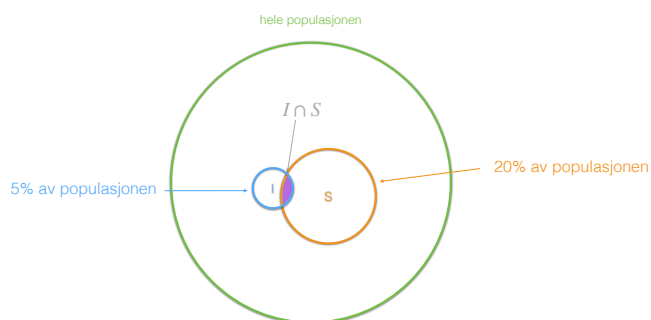
Et eksempel

En dag kjenner du deg syk. Det er ondt i hodet, halsen kjennes sår og nesen er tett. Fra statistikken får vi opplyst at 90 prosent av alle som har influensa har akkurat de samme symptomene som du nå har.

Videre kan vi lese at fem prosent av hele befolkningen vil få influensa dette året. Vi får også opplyst at 20 prosent vil ha disse symptomene i løpet av året.

Spørsmålet er nå om hvor sannsynlig det er at du har fått influensa. I utgangspunktet virker det ganske så sannsynlig - 90 prosent har jo disse symptomene! Det er her Bayes kommer inn og forandrer litt på hva vi intuitivt opplever som sannsynlig.

Vi starter med å tegne et Venn-diagram hvor vi starter med hele populasjonen. Så kan vi tegne en sirkel som markerer de fem prosentene som vil få influensa. Vi markerer også alle som vil ha symptomene. Det gir en oversikt over opplysningene.



Før vi går videre definerer vi disse hendingene

I – har influensa

S – har symptomene

Ut fra det vi allerede vet har vi disse sannsynlighetene

$$\begin{aligned}P(I) &= 0.05 \\P(S) &= 0.2 \\P(S \mid I) &= 0.9\end{aligned}$$

Vi ønsker nå å finne sannsynligheten for at du har influensa når du har disse symptomene: $P(I \mid S)$

Her er det at teoremet til Bayes kan være til hjelp. Litt omskrevet med hendingene vi definerte har vi at

$$P(I \mid S) = \frac{P(I) \cdot P(S \mid I)}{P(S)}$$

Sannsynlighetene kjenner vi og kan sette inn

$$\begin{aligned}P(I \mid S) &= \frac{P(I) \cdot P(S \mid I)}{P(S)} \\&= \frac{0.05 \cdot 0.9}{0.2} \\&= 0.225\end{aligned}$$

Ut fra opplysningene i eksemplet vårt har vi nå funnet ut at sannsynligheten bare er 0.225 for at du skal ha influensa. Det er en sannsynlighet det kan være vanskelig å tenke seg til eller se ut fra opplysningene.

Paradokset med falske positive

«The False Positive Paradox» oppstår ved all form for testing hvor testen ikke er helt sikker. Er det en minste mulighet for feiltesting vil det være noen som tester positivt uten å være det. Testes mange vil denne andelen bli større. Et argument mot massetesting av hele populasjonen er derfor å unngå for mange falske positive.

Vi kan ta for oss en tenkt test som skal avgjøre om en person har en sykdom. Eksempler på slike tester kan være både mammografi, prostatakreft, koronavirus eller andre sykdommer.

Vi starter med å definerer disse hendingene

S – har sykdommen
 T – tester positivt

La oss også gå ut fra at én prosent av alle har sykdommen. På fagspråket heter dette *prevalens*. Det betyr at 99 prosent ikke har den. I hele populasjonen har vi da disse sannsynlighetene

$$\begin{aligned}P(S) &= 0.01 \\P(\bar{S}) &= 0.99\end{aligned}$$

I medisinske tester forteller *sensitiviteten* hvor god testen er til å oppdage sykdommen og er definert som sannsynligheten for at prøven er positiv gitt at den som testes har sykdommen. For en hurtigtest for koronavirus er det typisk 80 prosent, så vi kan gå ut fra den verdien i eksemplet vårt.

Testens *spesifisitet* er sannsynligheten for at testen er negativ gitt av den som testes ikke har sykdommen. La oss gå ut fra at spesifisiteten er 90 prosent. Det betyr at $P(\bar{T} | \bar{S}) = 0.9$ og at $P(T | \bar{S}) = 0.1$.

Da har vi

$$\begin{aligned}P(T | S) &= 0.8 \\P(T | \bar{S}) &= 0.1\end{aligned}$$

Hva blir sannsynligheten for å ha sykdommen ved positiv test? Her er det Bayes kommer til hjelp igjen.

$$P(S | T) = \frac{P(S) \cdot P(T | S)}{P(T)}$$

Nå må vi finne ut sannsynligheten for å teste positivt. Det kan enten skje for de som har sykdommen eller for de som ikke har den. Sannsynligheten for de som ikke har den er

$$P(\bar{S}) \cdot P(T | \bar{S}) = 0.99 \cdot 0.1 = 0.099$$

Sannsynligheten for å teste positivt blant de som har sykdommen er

$$P(S) \cdot P(T | S) = 0.01 \cdot 0.8 = 0.008$$

Det gir

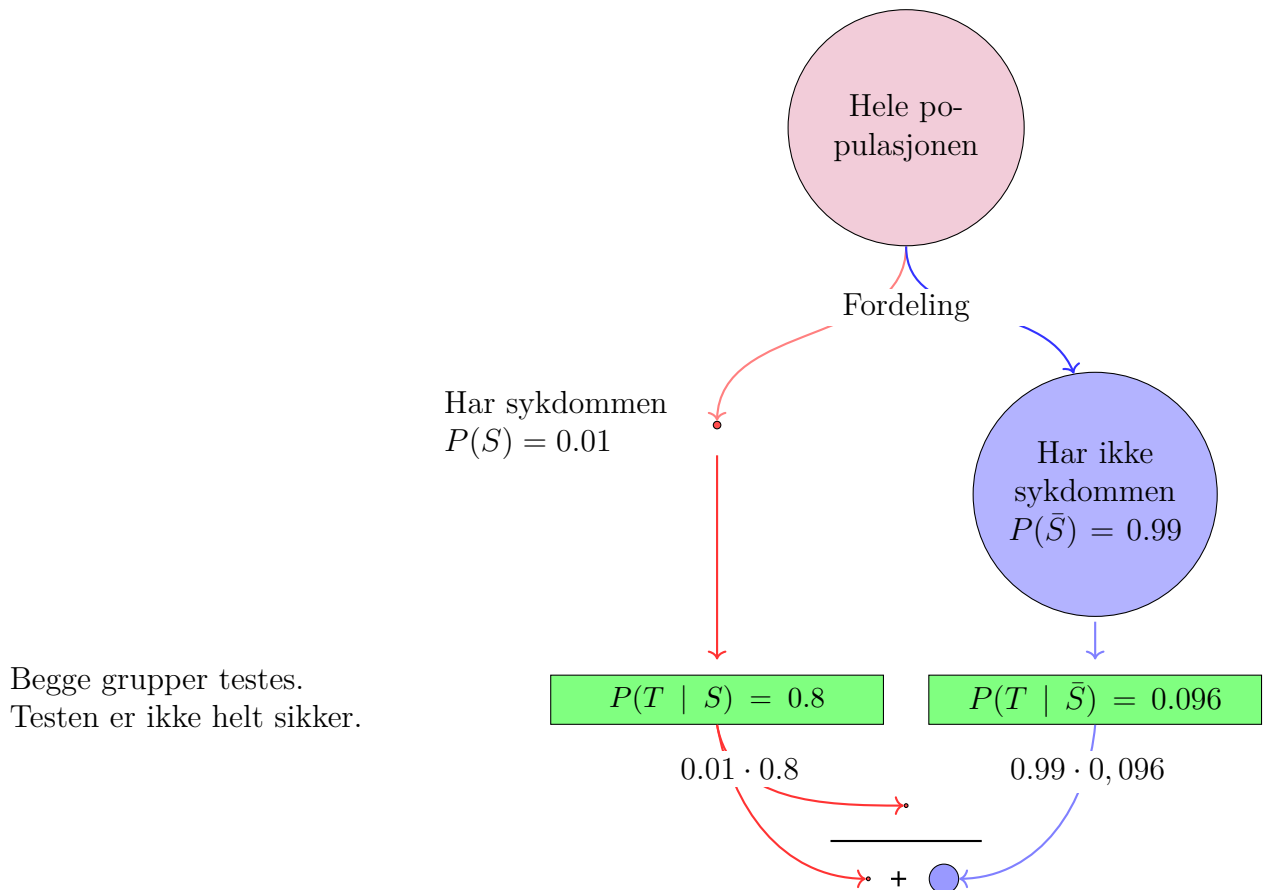
$$P(T) = P(\bar{S}) \cdot P(T | \bar{S}) + P(S) \cdot P(T | S) = 0.099 + 0.008 = 0.107$$

Nå kan sette inn i formelen og regne ut

$$P(S | T) = \frac{P(S) \cdot P(T | S)}{P(T)} = \frac{0.01 \cdot 0.8}{0.107} = 0.07477$$

Sannsynligheten for å ha sykdommen gitt en positiv test er altså bare 0.07477 eller 7.477 prosent. Med de antatte verdiene ser vi at en overvekt med falske positive gjør sannsynligheten liten.

Vi kan illustrere det samme med denne figuren hvor både teller og nevner kommer fram.



Ved å benytte algebra er det mulig å eksperimentere. Vi kaller

p – prevalens
 k – klinisk sensitivitet
 s – spesifisitet

Da får vi

$$P(S | T) = \frac{p \cdot k}{p \cdot k + (1 - p) \cdot (1 - s)}$$

Det gir muligheten for eksperimentering ved å legge inn i et regneark

Verdier	
prevalens i befolkningen	0,01
klinisk sensitivitet	0,8
spesifitivitet	0,9

$$P(S | T) = \frac{p \cdot k}{p \cdot k + (1 - p) \cdot (1 - s)}$$

	Prosent	
Sannsynlighet for at et positivt svar er sant	0,0748	7,48 %

Vi kan undersøke verdiene. Benytter vi det Folkehelseinstituttet oppgir for hurtigtester:

Gitt en god hurtigtest med klinisk sensitivitet på 80 % og spesifisitet på 99,96 %.

kan vi komme fram til samme tabell som de oppgir på sidene sine:

Prevalens i befolkningen som testes	30 %	5 %	1 %	0.1 %	0.01 %
Sannsynlighet for at et positivt svar er sant (positiv prediktiv verdi)	99,9 %	99,1 %	95,3 %	66,7 %	16,7 %

<https://www.fhi.no/nettpub/coronavirus/testing/antigen-hurtigtester/>

Legg merke til hvordan sannsynligheten endres med prevalens.

5 Kombinatorikk

5.1 Kombinatorikk

Kombinatorikk er en egen del av matematikken som tar for seg kombinasjoner og opptellinger.

Oppdagelsen av DNA-molekylet¹ er en av de store vitenskaplige gjennombruddene i vår tid. DNA (deoksiribonukleinsyre) inneholder arvematerialet vårt og har fått stor betydning for identifikasjon av individer. Molekylet består blant annet av fire baser: Adenin (A), guanin (G), thymin (T) og cytosin (C). Menneskets genetiske kode er bygd opp av omtrent seks milliarder av disse basene som en lineær sekvens. Tenker vi oss en sekvens med to baser fins det disse variantene

AA	GG	TT	CC
AG	GA	TA	CA
AT	GT	TC	CG
AC	GC	TG	CT

For hver plassering er det fire baser det kan velges blant. Det gir $4^2 = 16$ mulige kombinasjoner. Med en sekvens på tre baser vil det være $4^3 = 64$ muligheter. Når vi vet at hele sekvensen er på seks milliarder baser vil det kunne gi $4^{6\,000\,000\,000}$ mulige kombinasjoner. Nå vil ikke alle disse kombinasjonene kan forekomme i et menneske. Mindre enn én prosent av den genetiske koden er individuell. Alle mennesker er nesten genetisk like.

Eksemplet viser hvordan kombinatorikk kan benyttes for å finne mulige kombinasjoner. I sannsynlighetsregning må vi ofte finne antall gunstige, og antall mulige, kombinasjoner. Da får vi bruk for kombinatorikken.

5.2 Multiplikasjonsprinsippet

La oss se på denne oppgaven:

Oppgave 1



Karoline skal på syforening og har plukket ut tre overdelar og to skjørt og ei bukse. Hvor mange antrekk kan hun velge å ha på seg?

For å finne svaret har vi flere representasjoner som kan benyttes. Vi kan tegne en multiplikasjonstabell slik figur 5.1 viser. Da setter vi opp alle de ulike alternativene i rader og kolonner.

¹1953 av James Watson og Francis Crick. De to fikk Nobelprisen i fysiologi og medisin i 1962

I hver rute finner vi kombinasjonene. En slik tabell er en god representasjon når vi har to mengder med elementer som skal kombineres.



Figur 5.1: Multiplikasjonstabell

Alle de kombinasjonene som oppstår kan også kalles *det kartesiske produkt*. Den matematiske definisjonen er slik:

Definisjon 9 Kartesiske produkt

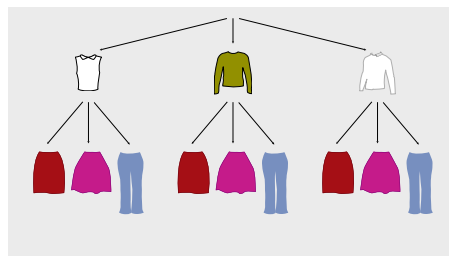
La A og B være to mengder. Da vil $A \times B$ være mengden av alle kombinasjoner av elementer i de to mengdene. $A \times B$ kalles *det kartesiske produktet* av de to mengdene.

Det dukker opp i forskjellige sammenhenger i både matematikk og informatikk, så det kan være greit å vite hva det heter. Vi kan se det i multiplikasjonstabellen vi satte opp hvor hver rute viser kombinasjonene.

Tar vi utgangspunkt i to mengder A og B hvor $A = \{a, b, c\}$ og $B = \{d, e, f\}$, vil altså det kartesiske produktet være alle kombinasjoner av elementene. Vi kan skrive det slik:

	a	b	c
d	(a,d)	(b,d)	(c,d)
e	(a,e)	(b,e)	(c,e)
f	(a,f)	(b,f)	(c,f)

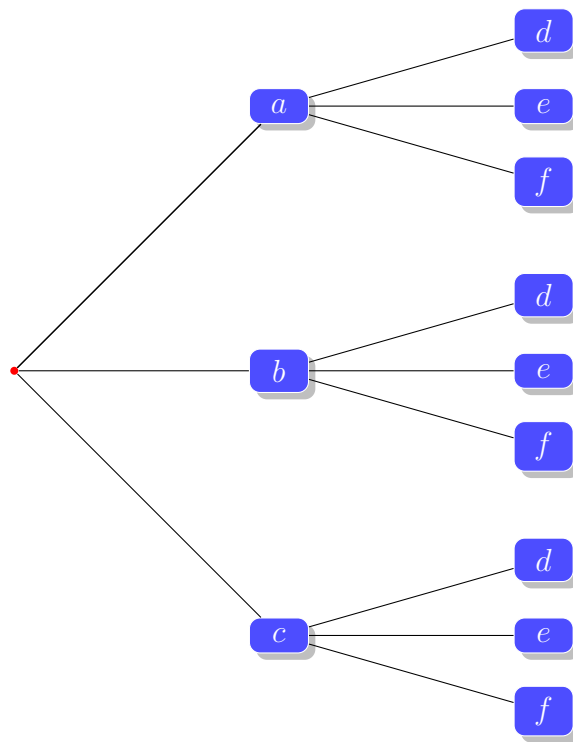
En annen representasjon er å tegne et tredigram. I figur 5.2 ser vi først et valg for alle toppene, og så kan vi se valgene for hver topp.



Figur 5.2: Tredigram

For de to mengdene $A = \{a, b, c\}$ og $B = \{d, e, f\}$ vil vi kunne tegne et tredigram ved å starte med alle elementene i den ene mengden og så tegne alle elementene i den andre ut fra de første.

Disse representasjonene viser at det totale antall antrekk Karoline kan ha på seg er ni. Vi kan multiplisere antall elementer i hver av de to mengdene som skal kombineres med hverandre.



Figur 5.3: Tredigram

Det kalles *multiplikasjonsprinsippet*

Teorem 10 Multiplikasjonsprinsippet

Ved et sammensatt utvalg blir det totale antall kombinasjoner lik produktet av antallet i hvert delvalg.

Python

Vi kan se på hvordan vi kan skrive ut et kartesisk produkt av to mengder med Python. To `for`-sløyfer hjelper oss med det. Elementene i mengdene er lagt inn i lister og alle kombinasjoner blir skrevet ut i programkode 5.1.

```
1 A = ["a", "b", "c"]
2 B = ["d", "e", "f"]
3 # Kartesisk produkt
4 for i in A:
5     for j in B:
6         print(i + ", " + j)
```

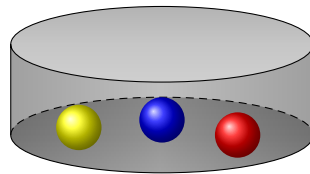
Programkode 5.1: Kartesisk produkt

Resultatet av dette programmet blir

a, d
a, e
a, f
b, d
b, e
b, f
c, d
c, e
c, f

5.3 Vi trekker kuler fra ei urne

Vi skal se på forskjellige utvalg framover og vi skal prøve å knytte alle situasjonene til at vi trekker kuler fra ei urne. I praksis vil vi finne kombinasjoner av personer, koder, lottotrekninger, tipping og mye annet rart. Uansett vil det være mulig å trekke en analogi til det å trekke lapper eller kuler fra ei urne. Vårt utgangspunkt er derfor at vi trekker kuler fra ei urne. I urna er det ei gul, ei blå og ei rød kule. Figuren under viser urne med kulene.



Når vi trekker fra ei urne kan vi gjøre det på to forskjellige måter. Vi kan enten legge tilbake kula vi trekker eller ikke. Det kaller vi med og uten tilbakelegging. En annen ting er at rekkefølgen for hva vi trekker kan spille en rolle eller ikke. Det kaller vi ordna eller uordna utvalg. La oss se litt mer på det. Vi har altså tre kuler vi kan trekke.



Vi skal trekke ut to kuler. Det gjør vi ved å trekke ei kule og så trekke ei ny kule.

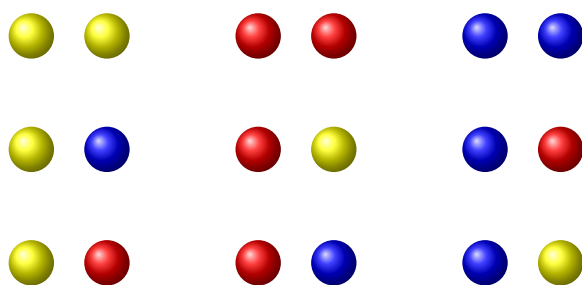
5.4 Ordna utvalg med tilbakelegging

Vi starter med å tilfellet hvor vi trekker ei kule, observerer resultatet, legger kula tilbake og trekker ei ny kule. Nå er vi også opptatt av rekkefølgen kulene trekkes i. Det betyr at trekker vi ei rød kule først og så ei blå, så er det et annet resultat enn om de ble trukket i omvendt rekkefølge.

Før du gå videre kan du tenke på hvilke resultat dette forsøket kan få.

Et spesielt tilfelle

Rekkefølgen spiller nå en rolle og vi legger tilbake kula vi trekker. Da kan vi ende opp med disse kombinasjonene



Hver gang vi trekker kan vi velge mellom tre kuler. Trekker vi to kuler vil antall kombinasjoner bli $3 \cdot 3 = 9$.

Hvis det var 10 kuler med nummer fra 1 til 10 og vi skulle trekke tre ganger ville vi kunne få

$$10 \cdot 10 \cdot 10 = 1000 \text{ kombinasjoner.}$$

Et generelt tilfelle

I det generelle tilfellet vil vi ha n kuler i urna og så trekke et utvalg på r . For hver kule kan vi velge mellom n . Da vil antall kombinasjoner være

$$\underbrace{n \cdot n \cdot n \cdots n}_{r \text{ ganger}} = n^r$$

Et ordna utvalg med tilbakelegging

Trekker vi et utvalg på r elementer fra n elementer er antall kombinasjon for ordna utvalg med tilbakelegging

$$n^r$$

La oss se på et eksempel.

Eksempel 2

Et eksempel på et ordna utvalg med tilbakelegging er når vi tenker oss at vi ønsker å sette opp ei tipperekke ved å trekke en lapp fra en boks. I den har vi tre lapper, en med H, en med B og en med U på. Vi trekker, noterer på kupongen og legger tilbake lappen. Et slikt ordna utvalg med tilbakelegging gir ei tilfeldig tipperekke.

Her ser vi at rekkefølgen spiller rolle, fordi hver lapp er resultatet på en spesiell kamp. Vi må legge tilbake lappen vi trekker hver gang siden flere kamper kan få samme resultat.

Vi kan dele dette opp i flere trinn. For hvert trinn har vi tre muligheter. Trekker vi 12 kamper har vi da

$$3^{12}$$

mulige utvalg.

Python

Kombinasjonene kan programmeres som to nøsta **for**-sløyfer. En annen måte er å importere modulen **itertools** og benytte kommandoen **product**. Navnet kjenner vi igjen fra *kryssprodukt*. Vi kan si fra om hvor mange vi velger med **repeat**. Programkode 5.2 viser

hvordan.

```
1 import itertools as it
2 kuler = [1,2,3]
3
4 om = it.product(kuler,repeat=2)
5
6 for i in list(om):
7     print (i)
```

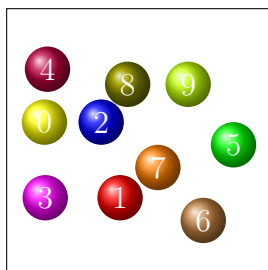
Programkode 5.2: Ordna utvalg med tilbakelegging

Vi får da denne utskrifta

```
(1, 1)
(1, 2)
(1, 3)
(2, 1)
(2, 2)
(2, 3)
(3, 1)
(3, 2)
(3, 3)
```

5.5 Ordna utvalg uten tilbakelegging

Vi forandrer litt på situasjonen og trekker kulene uten å legge tilbake. Den første kula vi trakk beholder vi og så trekker vi ei ny kule. Et praktisk eksempel kan være at vi skal trekke kuler hvor sifrene skal danne et tresifra tall hvor ingen av sifrene skal gjentas. Da kan vi tenke oss at vi trekker ei kule og noterer oss nummeret, så trekker vi neste kule uten å legge tilbake kula. Da får vi et ordna utvalg uten tilbakelegging.

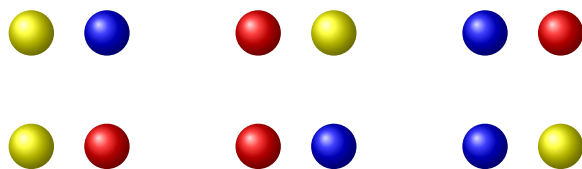


Her har vi ti kuler og trekker tre kuler. Hvor mange kombinasjoner kan vi få? Først kan vi velge blant ti kuler. Neste gang er det bare ni igjen. Antall kombinasjoner blir da

$$10 \cdot 9 \cdot 8 = 720$$

Et spesielt tilfelle

Går vi tilbake til å trekke fra urna med ei blå, ei rød og ei gul kule kan vi trea du trakk. Hvor mange kombinasjoner kan vi da trekke disse kombinasjonene av et ordna utvalg uten tilbakelegging



I figuren er to og to kuler plassert sammen og alle kombinasjonene er vist. Siden vi ikke har lagt tilbake etter hver trekning er dette et utvalg *uten tilbakelegging*. Siden rekkefølgen på kulene avgjørende kaller vi dette *et ordna utvalg*.

Vi kan finne antall kombinasjoner på samme måte: $3 \cdot 2 = 6$. Ved første trekk kan vi trekke en av tre forskjellige kuler. I andre trekk kan vi velge mellom to.

Et generelt tilfelle

I det generelle tilfellet er det n kuler i urna og vi trekker ut r kuler. Hvor mange kombinasjoner blir det da? Tankegangen blir den samme som i det spesielle tilfellet og vi kan skrive uttrykket slik

$$\underbrace{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}_{r \text{ faktorer}}$$

På akkurat samme måte som i det spesielle tilfellet skal r faktorer multipliseres sammen. Vi følger multiplikasjonsprinsippet. Når første kule skal plukkes er det n mulige som kan velges. Den legges ikke tilbake. Da er det $n-1$ mulige kuler. Slik fortsetter vi helt til vi kommer til faktor nummer r , som vil være $n-r+1$.

Nå er det vanlig å skrive om det på denne måten:

$$\frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) \cdot (n-r) \cdots 3 \cdot 2 \cdot 1}{(n-r) \cdots 3 \cdot 2 \cdot 1}$$

Legg merke til at brøken kan forkortes slik at vi får uttrykket vi startet med. Grunnen til det er at vi kan skrive uttrykket kortere hvis vi benytter fakultet. I matematikken har vi innført en skrivemåte for produktet av alle hele tall fra 1 til n og det er $n!$. Det leser vi som n fakultet. Med det mener vi:

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$$

når $n \in \mathbb{N}$

Definisjon 10 Fakultet

La $n \in \mathbb{N}$. Da vil n *fakultet* skrives som

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$$

Det gjør at vi kan skrive antall kombinasjoner for ordna utvalg uten tilbakelegging som

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

Andre skrivemåter for det uttrykket er $n^{(r)}$ ($= n$ i r -faktoriell) eller ${}_nP_r$

Et ordna utvalg uten tilbakelegging

Trekker vi et utvalg på r elementer fra n elementer er antall kombinasjon for ordna utvalg uten tilbakelegging

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

Benyttes digitale verktøy er kommandoen `nPr(n,r)` (kan variere litt, men et eller annet med `nPr`). Bokstaven P kommer fra ordet *permutasjon*

Definisjon 11 Permutasjon

En permutasjon er ei ordna rekkefølge av elementer

Da kan vi også kalle det ordna utvalget for antall k -permutasjoner, eller ordninger, av en r -mengde.

Her er noen eksempler på ordna utvalg med tilbakelegging

Eksempel 3

The Beatles besto av fire musikere: George, Ringo, Paul og John. To skal skrive en sang sammen. Den første som trekkes skriver melodien og den neste skriver teksten. Hvor mange slike par kan bandet danne?

I dette tilfellet spiller rekkefølgen en rolle og vi trekker uten tilbakelegging. Antall par kan vi finne ved

$$\begin{aligned}\frac{n!}{(n-r)!} &= \frac{4!}{(4-2)!} \\ &= \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} \\ &= 12\end{aligned}$$

Python

Vi kan finne permutasjoner med Python, men det krever at vi importerer en modul som heter `itertools`. Programkode 5.3 viser hvordan vi kan få skrevet ut permutasjonene.

```
1 import itertools as it
2 beatles = ["Ringo", "George", "Paul", "John"]
3 par = it.permutations(beatles,2)
4 for i in list(par):
5     print(i)
```

Programkode 5.3: Permutasjoner av The Beatles

Her er det kommandoen `par = it.permutations(beatles,2)` som trekker ut to elementer fra lista `beatles` og legger de i variabelen `par`. Resultatet blir da

```

('Ringo', 'George')
('Ringo', 'Paul')
('Ringo', 'John')
('George', 'Ringo')
('George', 'Paul')
('George', 'John')
('Paul', 'Ringo')
('Paul', 'George')
('Paul', 'John')
('John', 'Ringo')
('John', 'George')
('John', 'Paul')

```

Hvis det er ønskelig med en litt penere utskrift kan utskrifta endres til `print (i[0] + " og " + i[1])`.

Eksempel 4

På langrennslaget er det seks langrennsløpere. Nå skal det tas ut et stafettlag. Det skal være fire på stafettlaget. På hvor mange måter kan det gjøres hvis det trekkes vilkårlig?

Vi tenker oss at vi skriver navnene på $n = 6$ lapper. Så trekker vi ut $r = 4$ lapper. Rekkefølgen spiller en stor rolle siden det er forskjell på å bli trukket ut til første eller siste etappe. Hver løper kan trekkes ut bare en gang.

$$\begin{aligned}
 \frac{n!}{(n-r)!} &= \frac{6!}{(6-4)!} \\
 &= \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} \\
 &= 360
 \end{aligned}$$

Benytter vi den andre skrivemåten kan vi skrive

$$n^{(r)} = 6^{(4)} = 6 \cdot 5 \cdot 4 \cdot 3 = 360$$

Python

For å finne antall r - permutasjoner av en r -mengde må vi definere en egen funksjon som gjør det. For å kunne regne ut fakultet uten for mye egen programmering importerer vi modulen `math`.

```

1 import math as m
2 n = 6
3 r = 4
4
5 def nPr(n,r):
6     perm = int(m.factorial(n)/m.factorial(n - r))

```

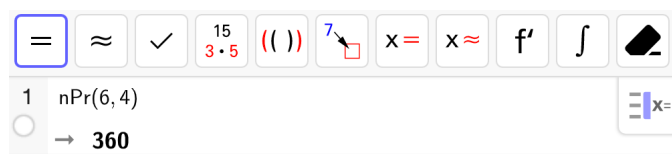
```

7   return perm
8
9   print("Antall ordna utvalg uten tilbakelegging er ", nPr(n,r))

```

Programkode 5.4: Egendefinert nPr

GeoGebra kan også benyttes. Da må vi velge CAS-modulen og benytte kommandoen nPr



Figur 5.4: GeoGebra

5.6 Uordna utvalg uten tilbakelegging

En annen variant er å trekke uten å legge tilbake hver kule. Samtidig er vi bare opptatt av resultatet: rekkefølgen kulene trekkes i spiller ingen rolle. Vi sier da at utvalget er *uordna*. Om vi rekker først ei rød og så ei blå kuler, eller omvendt – resultatet er at vi sitter igjen med ei rød og ei blå.

Et spesielt tilfelle

Da er situasjonen at vi trekker ei og ei kule uten å legge tilbake. Hvor mange kombinasjoner av to kuler vil vi kunne få hvis rekkefølgen på kulene ikke spiller noen rolle?

Se på kulene i forrige figur. Hvis rekkefølgen ikke spiller noen rolle er det ikke forskjell på disse to kombinasjonene:



Begge inneholder ei gul og ei blå kule. Går vi gjennom alle kombinasjonene i den forrige figuren ender vi opp med disse:



Antallet kan vi finne ved å se på et ordna utvalg og se på hvor mange som inneholder de samme to kulene. To kuler kan vi arrangere på to forskjellige måter. Antallet i et uordna utvalg blir derfor halvdelen av det antallet vi hadde i forrige eksempel.

Hva om det var 10 kuler med nummer fra 1 til 10 og vi skulle trekke tre ganger? Da ville vi kunne få $10 \cdot 9 \cdot 8 = 720$ ordna kombinasjoner.

I et uordna utvalg spiller ikke rekkefølgen noen rolle. De tre kulene vi trekker kan vi arrangere på $3! = 6$ forskjellige måter. Vi har da:

$$6 \cdot \text{antall uordna utvalg} = \text{antall ordna utvalg}.$$

Noe som gir oss svaret $\frac{720}{6} = 120$ kombinasjoner.

Et generelt tilfelle

Antall kombinasjoner når det er n kuler og vi trekker ut r . Da følger vi samme resonnement som over:

$$r! \cdot \text{antall uordna utvalg} = \text{antall ordna utvalg}$$

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

Denne brøken har vi en egen skrivemåte for. Vi skriver den slik

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

Den kalles binomialkoeffisienten.

En annen måte å skrive det samme er $nCr\{n\}\{r\}$.

Et uordna utvalg uten tilbakelegging

Trekker vi et utvalg på r elementer fra n elementer er antall kombinasjon for uordna utvalg uten tilbakelegging

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

På kalkiser og datamaskiner finner vi binomialkoeffisienten ved å bruke $nCr(n,k)$

Eksempel 5

Et eksempel på et uordna uten tilbakelegging er om vi trekker 13 kort fra en kortstokk. Hvor mange hender (kombinasjoner av de fem kortene) kan vi da få?

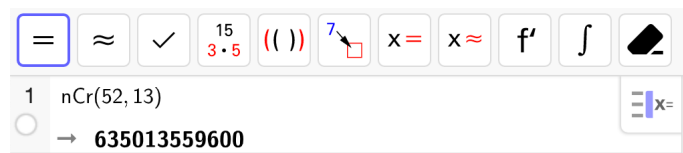
Kortstokken er allerede $n = 52$ lapper som allerede er ferdig påskrevet. Rekkefølgen spiller ingen rolle siden det er den samme handa uansett hvordan du stiller den opp. Hvert kort kan bare trekkes en gang.

Antall kombinasjoner finner vi slik

$$\begin{aligned} \binom{n}{r} &= \binom{52}{13} \\ &= \frac{52 \cdot 51 \cdot 50 \cdot \dots \cdot 40}{13 \cdot 12 \cdot 11 \cdot \dots \cdot 1} \\ &= 635\,013\,559\,600 \end{aligned}$$

Med slik store tall er det enkleste å regne ut med kommandoen $nCr(n,k)$.

GeoGebra kan hjelpe oss. Igjen er det CAS-modulen vi må bruke



Figur 5.5: GeoGebra

Eksempel 6

To av medlemmene i The Beatles må dele rom. Hvilke par kan vi danne?

Python

Python har heller ikke en innebygd kommando for å finne kombinasjonene ved et slikt utvalg, men vi kan importere modulen `itertools`, som gir denne muligheten. Da får vi tilgang til kommandoen `combinations`. I programkode 5.6 skrives også navnene ut hver for seg.

```
1 import itertools as it
2 beatles = ["Ringo", "George", "Paul", "John"]
3 par = it.combinations(beatles, 2)
4 #print(par)
5 for i in list(par):
6     print(i[0] + " og " + i[1])
```

Programkode 5.5: Uordna utvalg uten tilbakelegging

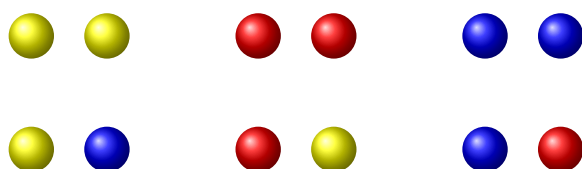
```
Ringo og George
Ringo og Paul
Ringo og John
George og Paul
George og John
Paul og John
```

5.7 Uordna utvalg med tilbakelegging

Det siste utvalget vi kan gjøre er et uordna utvalg med tilbakelegging. Ofte utelates dette, f.eks. i kombinatorikken i den videregående skolen.

Et spesielt tilfelle

Igjen trekker vi fra den samme urna som tidligere. Vi legger tilbake og rekkefølgen spiller ingen rolle. Da kan vi ta utgangspunkt i det ordna utvalget med tilbakelegging. Vi så at antall kombinasjoner, når rekkefølgen hadde betydning, var ni. Når vi ser bort fra rekkefølgen minker antallet med tre.



Et generelt tilfelle

Generelt kan antall kombinasjoner for et uordna utvalg med tilbakelegging formuleres slik

Et uordna utvalg med tilbakelegging

Trekker vi et utvalg på r elementer fra n elementer er antall kombinasjon for uordna utvalg med tilbakelegging

$$\binom{n+r-1}{r}$$

Noen eksempler

Her er noen eksempler som viser situasjoner med uordna utvalg med tilbakelegging og hvordan vi kan finne antall kombinasjoner.

Eksempel 7

Hvor mange utfall kan vi få ved å kaste to terninger når vi ikke tar hensyn til rekkefølgen? For oss som har spilt yatzy er vi vant med å ikke tenke på rekkefølgen.

Vi har med andre ord tilbakelegging siden vi kan få samme resultat neste gang vi kaster en terning.

$$\begin{aligned} \binom{n+r-1}{r} &= \binom{6+2-1}{2} = \binom{7}{2} \\ &= \frac{7 \cdot 6}{2 \cdot 1} = 21 \end{aligned}$$

Eksempel 8

En kiosk selger bananer, epler og appelsiner. Du skal kjøpe fire frukter. Hvor mange kombinasjoner kan du kjøpe?

$$\binom{n+r-1}{r} = \binom{3+4-1}{4} = \binom{6}{4} = 15$$

Eksempel 9

Det skal trekkes ut to fra The Beatles som skal intervjues i to aviser. Hvem kan få de to jobbene?

Python

Python kan også finne disse kombinasjonene med kommandoen: `combinations_with_replacement`. Programkode 5.6 viser bruken av kommandoen.

```
1 import itertools as it
2 beatles = ["Ringo", "George", "Paul", "John"]
3
4 um = it.combinations_with_replacement(beatles,2)
5
6 for i in list(um):
7     print (i[0] + " og " + i[1])
```

Programkode 5.6: Uordna utvalg med tilbakelegging

Resultatet når vi kjører dette programmet blir

```
Ringo og Ringo
Ringo og George
Ringo og Paul
Ringo og John
George og George
George og Paul
George og John
Paul og Paul
Paul og John
John og John
```

5.8 En oppsummering

Et utvalg er noe vi velger ut. I disse eksemplene har vi gjort det ved å trekke fra ei urne. Vi har nå trukket både med og uten tilbakelegging og vi har sett på ordna og uordna utvalg. Hvis vi trekker ut r elementer fra en mengde med n elementer, har vi fire hovedtilfeller. De kan vi summere opp i denne tabellen.

Antall kombinasjoner

Gjør vi et utvalg av r enheter fra en mengde n vil antall kombinasjoner være

	Med tilbakelegging	Uten tilbakelegging
Ordna	n^r	$\frac{n!}{(n-r)!}$
Uordna	$\binom{n+r-1}{r}$	$\binom{n}{r}$

5.9 Oppgaver

I oppgavene er det gitt noen situasjoner som kan omskrives til urnemodellen. Analyser situasjonen ved å stille spørsmålene

- Er det med eller uten tilbakelegging?
- Spiller rekkefølgen noen rolle?

Prøv å løse oppgavene uten å se på løsningen.

Oppgave 2

Melodikonkurranse

Til en melodikonkurranse er det plukket ut 20 melodier. Hvor mange resultatlistes med 10 melodier er det mulig å lage?

I slike oppgaver er det være lurt å finne ut om det er med tilbakelegging eller ikke og om rekkefølgen spiller noen rolle.

	Med tilbakelegging	Uten tilbakelegging
Ordna	☐	☒
Uordna	☐	☐

Her finner vi at det er et ordna utvalg uten tilbakelegging (OUT). Da kan vi benytte

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

Her er $n = 20$ og $r = 10$

$$20^{(10)} = \frac{20!}{10!} = 670442572800$$

Bruker vi et digitalt verktøy er kommandoene ofte `Pnr()` eller `nPr()`. I GeoGebra vil vi finne svaret ved `nPr(20,10)`

Oppgave 3

Middag

Du skal ut å spise middag sammen med vennene dine. Til sammen blir dere 12 personer. Dit dere kommer kan dere velge mellom hamburgere, kylling og pizza. Hver bestiller en rett. Hvor mange ulike bestillinger kan dere gjøre?

Hver av middagsgjestene kan velge mellom tre retter. For at hver person skal få det som blir bestilt spiller rekkefølgen en rolle. Det er en forutsetning vi kan ta som kanskje ikke gjelder i virkeligheten.

	Med tilbakelegging	Uten tilbakelegging
Ordna	☒	☐
Uordna	☐	☐

Dette er et ordna utvalg med tilbakelegging (OMT) og vi finner svaret med

$$n^r$$

Her er $n = 3$ og $r = 12$

$$3^{12} = 531441$$

Oppgave 4

Kodelås

En kodelås har seks taster som hver kan plasseres i fire forskjellige stillinger. Vi kaller disse stillingene 1, 2, 3 og 4. Hvor mange tallkoder kan en lage til denne kodelåsen?

Her spiller rekkefølgen av sifrene en avgjørende rolle og vi bruker ikke opp tallene.

	Med tilbakelegging	Uten tilbakelegging
Ordna	☒	☐
Uordna	☐	☐

Dette er et ordna utvalg med tilbakelegging (OMT) og vi finner svaret med

$$n^r$$

Her er $n = 4$ og $r = 6$

$$4^6 = 4096$$

Oppgave 5

Elevråd

I en klasse er det 12 jenter og 10 gutter. Det skal velges tre representanter til elevrådet. Representantene skal bestå av 2 jenter og 1 gutt. Hvor mange ulike utvalg kan dannes?

En representant kan ikke velges to ganger, så her er det ikke tilbakelegging. Så lenge det er snakk om et utvalg spiller heller ikke rekkefølgen noen rolle.

	Med tilbakelegging	Uten tilbakelegging
Ordna	☐	☐
Uordna	☐	☒

Da er det et uordna utvalg uten tilbakelegging (UUT) hvor vi kan finne antallet ved

$$\binom{n}{r}$$

Antall kombinasjoner for jentevalg er: $\binom{12}{2} = 66$

Antall kombinasjoner for guttevalg er: $\binom{10}{1} = 10$

Disse kan igjen kombineres og vi finner totalt antall kombinasjoner ved produktet av de to.

Antall mulige utvalg er 660

Oppgave 6

Vaktmesteren

En vaktmester har to kartonger med 12 lyspærer i hver kartong. I kartong A er tre av pærene defekte, mens to av pærene i kartong B ikke fungerer. Vaktmesteren henter to pærer.

Hva er sannsynligheten for at begge pærene er defekte når:

- a) han trekker en lyspære fra hver kartong
- b) han trekker begge lyspærene fra kartong A
- c) han tilfeldig velger en av kartongene og tar to pærer fra den kartongen

Ved å trekke fra en kartong for så å finne ut om han sitter igjen med defekte pærer spiller ikke rekkefølgen noen rolle. Vi har heller ikke tilbakelegging.

	Med tilbakelegging	Uten tilbakelegging
Ordna	☐	☐
Uordna	☐	☒

Dette er et uordna utvalg uten tilbakelegging hvor vi finner kombinasjonene med

$$\binom{n}{r}$$

Her definerer vi to hendinger

A – trekker ei defekt pære fra kartong A

B – trekker ei defekt pære fra kartong B

Hva er sannsynligheten for at begge pærene er defekte når han trekker en lyspære fra hver kartong?

$$\begin{aligned}P(A) &= \frac{3}{12} = \frac{1}{4} \\P(B) &= \frac{2}{12} = \frac{1}{6} \\P(A \cap B) &= P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24} \approx 0.041667\end{aligned}$$

Hva er sannsynligheten for at begge pærene er defekte når han trekker begge lyspærene fra kartong A?

$$P(\text{trekke to defekte pærer fra A}) = \frac{\text{gunstige}}{\text{mulige}} = \frac{\binom{3}{2}}{\binom{12}{2}} = \frac{1}{22} = 0.045455$$

Svaret kunne vi også funnet på en annen måte. Vi skal trekke to ganger og vi kan finne sannsynligheten for hvert trekk

$$P(\text{trekke to defekte pærer fra A}) = \frac{3}{12} \cdot \frac{2}{11} = \frac{1}{22} = 0.045455$$

Hva er sannsynligheten for at begge pærene er defekte når han tilfeldig velger en av kartongene og tar to pærer fra den kartongen?

$$P(\text{trekke to defekte pærer fra B}) = \frac{\text{gunstige}}{\text{mulige}} = \frac{1}{\binom{12}{2}} = \frac{1}{66}$$

Vi regner med at det er lik sannsynlighet mellom valgene av kartong A eller B. Da kan vi finne svaret som

$$\frac{1}{2} \cdot \frac{1}{22} + \frac{1}{2} \cdot \frac{1}{66} = \frac{1}{33} = 0.030303$$

Egentlig kunne vi brukt Bayes teorem

$$\begin{aligned} P(\text{trekke to defekte pærer}) &= P(2 \text{ pærer} \mid \text{kasse A}) \cdot P(\text{kasse A}) + P(2 \text{ pærer} \mid \text{kasse B}) \cdot P(\text{kasse B}) \\ &= \frac{\binom{3}{2} \binom{9}{0}}{\binom{12}{2}} \cdot \frac{1}{2} + \frac{\binom{2}{2} \binom{10}{0}}{\binom{12}{2}} \cdot \frac{1}{2} = \frac{1}{33} \end{aligned}$$

6 Binomisk forsøk

6.1 Hva er et binomisk forsøk?

Ordet binomisk kommer fra latin og betyr opprinnelig «å ha to navn». Seinere, på 1600-tallet, endra betydningen seg til å bety «å bestå av to deler». Det er i den betydningen ordet brukes her. Et binomisk forsøk er et forsøk hvor vi kan ha to utfall. Ofte kaller vi det en suksess eller en fiasko. Videre er det et forsøk hvor sannsynligheten for å få disse utfallene er det samme hver gang. Alle delforsøkene skal også være uavhengige av hverandre.

La oss se på noen eksempler.

6.2 Et ektepar bestemmer seg for å få 6 barn

Dette er kanskje ikke et eksempel henta fra virkeligheten. Her er det vel snakk om en pseudo-virkelighet? Uansett vil eksemplet illustrere et binomisk forsøk eller en binomisk forsøksserie. Situasjonen er at en mann og ei dame bestemmer seg for å starte en familie. De gir seg ikke før de har fått seks barn.

Hva er sannsynligheten for at de får to gutter og fire jenter?

Dette tilfredsstiller kravene til å kalles et binomisk forsøk. De gjentar forsøket seks ganger. Det er to utfall: gutt eller jente. Det er samme sannsynlighet hver gang. Delforsøkene er ikke avhengige siden kjønn ved en fødsel ikke påvirker neste ¹.

Sannsynligheten for å få en gutt er statistisk 0.513. Vi gir den sannsynligheten navnet p og skriver: $p = 0.513$. Sannsynligheten for å få ei jente blir da: $(1 - p) = 1 - 0.513 = 0.487$

Vi definerer disse hendingene

G – de får en gutt

J – de får ei jente

Hvis oppgaven hadde vært å finne sannsynligheten for at de fikk en gutt, en gutt og så fire jenter kunne vi regne ut det som

$$P(GGJJJJ) = 0.513 \cdot 0.513 \cdot 0.487 \cdot 0.487 \cdot 0.487 \cdot 0.487 = 0.014803$$

Her viser rekkefølgen på bokstavene til rekkefølgen på hendingene. To gutter og fire jenter kan de få ved flere kombinasjoner. Her er en tabell som viser alle kombinasjonene.

GGJJJJ	JGGJJJ	JJGGJJ	JJJGGJ	JJJJGG
GJGJJJ	JGJGJJ	JJGJGJ	JJJGJG	
GJJGJJ	JGJJGJ	JJGJJG		
GJJJGJ	JGJJJG			
GJJJJG				

¹Dette kan diskuteres og at det er langt fra sikkert at det stemmer, men la ikke det forstyrre eksemplet. Her er vi ute etter å forstå binomiske forsøk.

Til sammen er det 15 kombinasjoner. Sannsynligheten for hver av de 15 er 0.014803 Sannsynligheten for å få to gutter og fire jenter blir da

$$15 \cdot 0.014803 = 0.22205$$

Kanskje vi kan komme fram til en mer generell løsning? Det som ga litt problemer var å komme fram til at det var 15 kombinasjoner?. En annen måte å spørre på er: Hvor mange måter kan vi velge ut to gutter blant seks barn hvor resten er jenter. Vi kan spørre oss om hvilken type utvalg er dette? Er det tilbakelegging? Nei, trekker vi en gutt er det bare en igjen. Spiller rekkefølgen noen rolle? Nei, det er likegyldig bare vi får trukket to gutter. Da har vi et uordna utvalg uten tilbakelegging. Antall kombinasjoner blir da

$$\binom{6}{2} = 15$$

La oss gå over til å betrakte dette som en binomisk forøksserie. Med det mener vi at vi gjentar et forøk med to utfall flere ganger. Binomisk betyr altså at vi har to utfall: Suksess eller fiasko. Her er det gutt eller jente, så det passer ikke så godt med suksess og fiasko, men ofte benyttes de to begrepene.

Bruker vi symboler kan vi skrive utregningen vi gjorde slik:

$$P(G) \cdot P(G) \cdot P(J) \cdot P(J) \cdot P(J) \cdot P(J) = P(G)^2 \cdot P(J)^4 = p^2 \cdot (1 - p)^4$$

Den samme sannsynlighet gjelder for kombinasjonen GJGJJJ

$$P(G) \cdot P(J) \cdot P(G) \cdot P(J) \cdot P(J) \cdot P(J) = P(G)^2 \cdot P(J)^4 = p^2 \cdot (1 - p)^4$$

Slik kan vi fortsette, men vi vet nå at vi må gjøre det for alle 15 kombinasjoner. Det kan vi skrive som

$$\binom{6}{2} \cdot p^2 \cdot (1 - p)^4$$

6.3 Tegnestifter

La oss se på et annet eksempel. Vi kaster en tegnestift fire ganger, sannsynligheten for at den lander med stift opp kaller vi igjen p .

Vi har nå to hendinger:

S – tegnestiften lander med stift opp
 $\bar{S}$ – tegnestiften lander med stift ned

En oppgave kan være å finne ut sannsynligheten for at ingen, 1, 2, 3 eller 4 kast hvor stift lander med spissen opp. I tabellen er de forskjellige kombinasjonene satt opp og sannsynligheten for hver av dem er regnet ut.

hendinger	antall suksess	sannsynlighet
$\bar{S}\bar{S}\bar{S}\bar{S}$	0	$(1-p)^4$
$\bar{S}\bar{S}\bar{S}S$	1	$p(1-p)^3$
$\bar{S}\bar{S}S\bar{S}$	1	$p(1-p)^3$
$\bar{S}S\bar{S}\bar{S}$	1	$p(1-p)^3$
$S\bar{S}\bar{S}\bar{S}$	1	$p(1-p)^3$
$\bar{S}\bar{S}SS$	2	$p^2(1-p)^2$
$\bar{S}S\bar{S}S$	2	$p^2(1-p)^2$
$S\bar{S}\bar{S}S$	2	$p^2(1-p)^2$
$\bar{S}SS\bar{S}$	2	$p^2(1-p)^2$
$S\bar{S}S\bar{S}$	2	$p^2(1-p)^2$
$SS\bar{S}\bar{S}$	2	$p^2(1-p)^2$
$\bar{S}SSS$	3	$p^3(1-p)$
$S\bar{S}SS$	3	$p^3(1-p)$
$SS\bar{S}S$	3	$p^3(1-p)$
$SSS\bar{S}$	3	$p^3(1-p)$
$SSSS$	4	p^4

Hvis vi skriver sannsynlighetene som $P(S = \text{antall})$, kan vi sette opp disse utregningene

$$P(S = 0) = (1-p)^4$$

$$P(S = 1) = 4p(1-p)^3$$

$$P(S = 2) = 6p^2(1-p)^2$$

$$P(S = 3) = 4p^3(1-p)$$

$$P(S = 4) = p^4$$

Dette kan vi sette sammen til

$$P(S = k) = \binom{4}{k} \cdot p^k \cdot (1-p)^{4-k}, \text{ der } k = 0, 1, 2, 3, 4$$

6.4 Et generelt binomisk forsøk

Da kan vi trekke en generell konklusjon for alle binomiske forsøk.

Binomisk forsøk

Et binomisk forsøk gjenkjenne ved at det

- har to utfall
- identiske forsøk med samme sannsynlighet hver gang
- delforsøkene er uavhengige

Et binomisk forsøk gjentas n ganger. Alle disse forsøkene har et binomisk utfallsrom $U = \{S, F\}$. Utfallet er enten en suksess eller en fiasko. Sannsynlighetene er $P(S) = p$ og $P(F) = 1 - p$.

Da kan vi finne sannsynligheten for at S inntreffer k ganger ved

$$P(S = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

Lar vi X være antall suksesser i en binomisk forsøksserie sier vi at X er binomisk fordelt. Det kan vi skrive som at X er $\text{bin}(n; p)$ -fordelt.

6.5 Andre eksempler på binomiske forsøksserier

Noen eksempler på binomiske forsøk er disse

- Vi kaster en mynt n ganger og registrerer hvor mange «kron» vi får
- En sår n frø og registrerer hvor mange som spirer etter ei viss tid.
- En kvalitetskontroll av n artikler hvor en registrerer hvor mange som er defekte.

6.6 Eksempel: Straffespark

Vi kan se på et eksempel litt grundigere. Først kan du prøve det som en oppgave

Oppgave 7

Anta at ved en straffesparkkonkurranse i fotball er sannsynligheten $p = 0.7$ for at en spiller scorer mål. Anta også at sannsynligheten er den samme for hvert eneste spark. Spilleren tar fem straffespark. Hva er sannsynligheten for at hun

- a. scorer alle fem gangene
- b. bommer alle fem gangene
- c. scorer minst tre mål

Vi tar for oss denne oppgaven som et eksempel og finner løsningene. Med de forutsetningene som er gitt i oppgaven er dette et binomisk forsøk. To utfall, samme sannsynlighet hver gang og forsøkene er uavhengige. Forsøket gjentas fem ganger og da har vi at

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} = \binom{5}{x} \cdot 0.7^x \cdot (1 - 0.7)^{5-x}$$

a) Spilleren tar fem straffespark. Hva er sannsynligheten for at hun scorer alle fem gangene

Denne oppgaven kunne vi løst ganske enkelt. Her er det bare en kombinasjon og sannsynligheten vil bli $0.7^5 = 0.16807$. Benytter vi hele formelen får vi

$$P(X = 5) = \binom{5}{5} \cdot 0.7^5 \cdot (1 - 0.7)^{5-5} = 0.16807$$

Svar: Sannsynligheten for at hun scorer alle fem gangene er 0.1681

b) Hun bommer alle fem gangene

Her er det også bare en kombinasjon og svaret blir $0.3^5 = 0.00243$. Det samme kommer vi fram til om vi benytter formelen

$$P(X = 0) = \binom{5}{0} \cdot 0.7^0 \cdot (1 - 0.7)^{5-0} = 0.3^5 = 0.00243$$

Sannsynligheten for at hun bommer alle fem gangene er 0.0024

c) Hun scorer minst tre mål

For å score minst tre mål er det tre muligheter. Enten blir det tre, fire eller fem mål. Vi har allerede funnet sannsynligheten for det siste alternativet. Da må vi bare finne de vi mangler

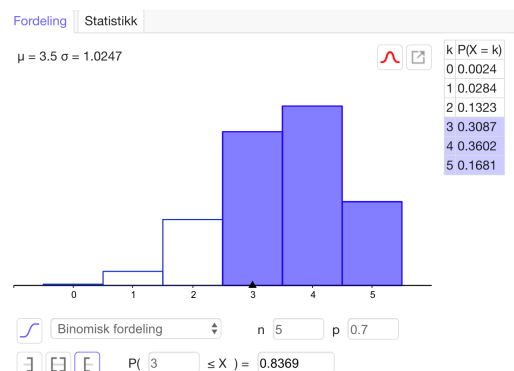
$$P(X = 3) = \binom{5}{3} \cdot 0.7^3 \cdot (1 - 0.7)^{5-3} = 0.3087$$

$$P(X = 4) = \binom{5}{4} \cdot 0.7^4 \cdot (1 - 0.7)^{5-4} = 0.36015$$

Da kan vi legge sammen sannsynlighetene og får

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.3087 + 0.36015 + 0.16807 \\ &= 0.83692 \end{aligned}$$

Her går det greit å regne ut, men ofte kan slike oppgaver gi lange utregninger. Da er det greit å ha et verktøy. I GeoGebra kunne vi løst hele denne oppgaven slik



Svar: Sannsynligheten for at hun scorer minst tre mål er 0.8369

Python

Vi kan få gjort det samme med Python også. For å gjøre det bør vi lage to funksjoner. Den ene finner binomialkoeffisienten og den andre regner ut den binomiske sannsynligheten. Ved å lage ei **for**-sløyfe kan vi summere opp sannsynlighetene. Programkode 6.1 viser en måte å gjøre det på.

```
1 import math as m
2
3 def binomial(a,b):
4     bin = m.factorial(a)/(m.factorial(b)*m.factorial(a-b))
5     return bin
6
7 def binomisk(n, p, x):
8     bin = binomial(n, x)*p**x*(1 - p)**(n-x)
9     return bin
10
11 p = 0.7
12 n = 5
13 sum = 0
14 for i in range(3,6):
15     sum = sum + binomisk(n,p,i)
16
17 print( sum)
```

Programkode 6.1: Straffespark

For slike utregninger er det også mulig å bruke ferdige moduler for å slippe å lage funksjoner på egen hånd. I programkode 6.2 er det benyttet en modul som heter **scipy.stats** hvor vi finner to nyttige funksjoner. **binom.cdf** finner den kumulative binomiske sannsynligheten. For å løse oppgaven vår må vi da finne den komplementære sannsynligheten. Det kan vi også gjøre med funksjonen **binom.sf**. De to svarene som skrives ut er helt like.

```
1 import scipy.stats as st
2
3 p = 0.7
4 n = 5
5 sum_1 = 1 - st.binom.cdf(2,n,p)
6 sum_2 = st.binom.sf(2,n,p)
7
8 print(sum_1, sum_2)
```

Programkode 6.2: Straffespark

I noen tilfeller kan vi også simulere stokastiske forsøk for å finne en sannsynlighet. Det kan vi gjøre med forskjellige verktøy som regneark. Programmering egner seg også.

Python

Vi kan prøve å simulere det samme med Python.

```
1 import random as rd
2 ant_maal = [0,0,0,0,0,0]
3 ant_forsok = 100000
4 for i in range(0,ant_forsok):
5     maal = 0
6     for i in range(0,5):
7         t = rd.random()
8         if t < 0.7:
9             maal = maal + 1
10    ant_maal[maal] = ant_maal[maal] + 1
11
12 print(ant_maal)
```

Programkode 6.3: Straffespark

I mitt tilfelle ble dette skrevet ut

```
[239, 2855, 13266, 30834, 35989, 16817]
```

Setter vi dette opp i en tabell og regner ut sannsynlighetene får vi

Mål	Antall	$P(X)$
0	239	0.00239
1	2855	0.02855
2	13266	0.13266
3	30834	0.30834
4	35989	0.35989
5	16817	0.16817

Det manuelle arbeidet kunne vi naturligvis unngått ved å la Python regne ut det samme for oss. Hekter vi på ei **for**-sløyfe kan vi få skrevet ut det samme.

```
1 for i in range(5):
2     print("Sannsynligheten for ", i , "mål er: ",ant_maal[i]/
          ant_forsok)
```

Programkode 6.4: **for**-sløyfe

6.7 Newton-Pepys problem

[Wikipedia](#)

I 1693 skrev Samuel Pepys til Isaac Newton og lurte på hva som var mest sannsynlig av

- A: Seks ideelle terninger kastes og en får minst en sekser
- B: 12 ideelle terninger kastes og en får minst to seksere

- C: 18 idelle terninger kastes og en får minst tre seksere

Pepys mente sjøl at C måtte ha høyest sannsynlighet, men Newton ga som svar at A var mest sannsynlig.

Nå vet vi at det kan regnes ut på denne måten

$$P(A) = 1 - \left(\frac{5}{6}\right)^6 = \frac{31031}{46656} \approx 0.6651$$

$$P(B) = 1 - \sum_{x=0}^1 \binom{12}{x} \cdot \left(\frac{1}{6}\right)^x \cdot \left(\frac{5}{6}\right)^{12-x} = \frac{1346704211}{2176782336} \approx 0.6187$$

$$P(C) = 1 - \sum_{x=0}^2 \binom{18}{x} \cdot \left(\frac{1}{6}\right)^x \cdot \left(\frac{5}{6}\right)^{18-x} = \frac{15166600495229}{25389989167104} \approx 0.5973$$

Python

Skal vi gjøre dette med Python må vi importere modulen `scipy.stats`. Da har vi en kommando tilgjengelig for å finne den kumulative binomiske sannsynligheten: `binom.cdf`. Programkode 6.5 gjør akkurat det samme som i utregningene over.

```
1 import scipy.stats as st
2
3 for i in range(1,4):
4     n = 6*i
5     p = 1- st.binom.cdf(i-1,n,1/6)
6     print("Sannsynligheten for å få minst ",i, " seksere med ", n, "
        ideelle terninger er ", p)
```

Programkode 6.5: Newton-Pepys med Python

Svaret blir det samme

```
Sannsynligheten for å få minst 1 seksere med 6 ideelle terninger er
0.6651020233196159
Sannsynligheten for å få minst 2 seksere med 12 ideelle terninger er
0.6186673737323085
Sannsynligheten for å få minst 3 seksere med 18 ideelle terninger er
0.5973456859477229
```

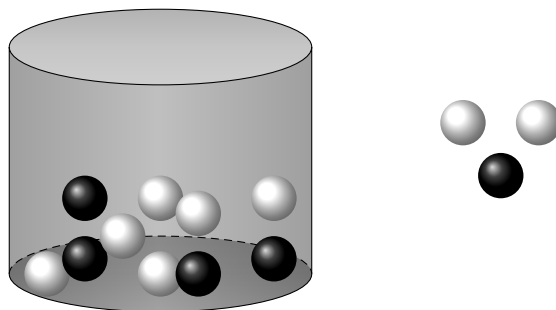
Newton hadde ikke hjelp av teknologien, men konklusjonen hans kan vi se holder mål.

7 Hypergeometrisk forsøk

7.1 Hypergeometrisk forsøk

Mens det i et binomisk forsøk er samme sannsynlighet for alle delforsøkene, skal vi nå se på en annen måte å gjøre et utvalg på. Fortsatt kan vi holde oss til urnemodellen. Der tenker vi oss at det er n kuler og at noen av disse kulene er spesielle. Trekker vi ei spesiell kule kan vi kalle det en suksess og trekker vi ikke ei slik kule så er det en fiasko. Også her er det bare to utfall $U = \{S, F\}$. Til forskjell fra det binomiske legger vi ikke tilbake kula mellom hvert trekk i et hypergeometrisk forsøk.

7.2 Et klassisk eksempel



Vi kan tenke oss at vi har ei urne med ti kuler hvor fire er svarte og seks er hvite. La oss si at vi skal trekke ut tre kuler. Hva er da sannsynligheten for å få ei svart kule og to hvite?

Sannsynligheten for å trekke ei svart kule først og så to hvite er

$$\frac{4}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

Sannsynligheten for å trekke ei hvit kule først, så ei svart og så ei hvite er

$$\frac{6}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

Den siste muligheten er å trekke to hvite kuler først og så ei svart til slutt

$$\frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} = \frac{1}{6}$$

Da kan vi finne sannsynligheten for å trekke ei svart og to hvite kuler

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Vi kan se at hvis oppgaven blir litt mer komplisert, vil det bli ganske lange utregninger og i tillegg mange kombinasjoner. Da kan det være enklere å finne gunstige og mulige kombinasjoner. I vårt tilfelle kan vi starte med å se på hvor mange gunstige kombinasjoner det blir. Hvor mange

måter kan vi trekke den ene svarte kula? I urna er det fire svarte og da kan vi trekke den på fire måter. Her er det enkelt å se, men skulle vi regne ut det kunne vi skrive

$$\binom{4}{1} = 4$$

Hvor mange måter kan vi trekke ut to hvite fra urna. Vi følger samme resonnement og finner

$$\binom{6}{2} = 15$$

Så var det antall muligheter for å trekke tre kuler fra urna med ti kuler

$$\binom{10}{3} = 120$$

Nå kan vi finne sannsynligheten for å trekke ei svart og to hvite kuler

$$\frac{4 \cdot 15}{120} = \frac{1}{2}$$

Hele utregninga blir da

$$\frac{\binom{4}{1} \cdot \binom{6}{2}}{\binom{10}{3}} = \frac{1}{2}$$

Svaret blir det samme, men nå har vi heller sett på antall kombinasjoner. Eksemplet viser et hypergeometrisk utvalg. I ei urne er det et visst antall kuler. Vi sier at noen er spesielle. I vårt eksempel ser vi på at det de svarte kulene. Vi skal trekke et visst antall kuler og finne sannsynligheten for at noen av de er spesielle. Kaller vi antall spesielle kuler i urna for s og antallet i utvalget vi gjør for r , ønsket vi å finne sannsynligheten for at antall spesielle kuler i utvalget var $x = 1$. I vårt tilfelle er da

$$\begin{array}{ll} n = 10 & r = 3 \\ s = 4 & x = 1 \end{array}$$

Vi skulle finne sannsynligheten for dette resultatet

	antall trukket	antall igjen	sum
S	1	3	4
F	2	4	6
sum	3	7	10

7.3 Et generelt utvalg

Et hypergeometrisk utvalg kjennetegnes ved at vi velger ut r elementer fra n elementer. Mengden n inneholder s spesielle elementer. Vi ønsker å finne sannsynligheten for at det er x spesielle i utvalget vi gjør. Da kan vi skrive det som

	antall trukket	antall igjen	sum
S	x	$s - x$	s
F	$r - x$	$n + x - r - s$	$n - s$
sum	r	$n - r$	n

Det generelle uttrykket for hvor mange utvalg kan vi gjøre for å oppnå x spesielle blir da

$$\binom{s}{x} \cdot \binom{n-s}{r-x}$$

$\binom{s}{x}$ er antall måter vi kan velge x spesielle av s

$\binom{n-s}{r-x}$ er antall måter vi kan velge $r-x$ vanlige fra $n-s$ vanlige

Sannsynligheten for å velge x av de spesielle:

$$P(X = x) = \frac{\binom{s}{x} \cdot \binom{n-s}{r-x}}{\binom{n}{r}}$$

Hypergeometrisk sannsynlighetsfordeling

Et hypergeometrisk utvalg kjennetegnes ved at vi velger ut r elementer fra n elementer. Mengden n inneholder s spesielle elementer. Sannsynligheten for at et utvalg inneholder x spesielle er da

$$P(X = x) = \frac{\binom{s}{x} \cdot \binom{n-s}{r-x}}{\binom{n}{r}}$$

En hypergeometrisk sannsynlighetsmodell kjennetegnes ved at det er avhengighet mellom trekningene siden vi trekker uten tilbakelegging

Legg merke til at hvis n er veldig stor så vil sannsynligheten holder seg omtrent konstant uansett. Da vil et hypergeometrisk forsøk være tilnærma lik et binomisk.

7.4 Et eksempel

Her er et eksempel på en typisk oppgave hvor vi har et hypergeometrisk utvalg. I eksemplet er oppgaven generalisert. Variabelen x vil kunne variere mellom 0 og 5.

Eksempel 10

En klasse har $n = 28$ elever hvorav $s = 10$ er jenter. Vi trekker ut $r = 5$ elever som skal være med på en tur. Hva er sannsynligheten for at 3 av disse er jenter?

Legg merke til at dette er en typisk hypergeometrisk sannsynlighetsmodell: Avhengighet mellom trekninger – vi trekker uten tilbakelegging!

Det er ikke en binomisk modell hvor vi har uavhengighet mellom trekninger – vi trekker da med tilbakelegging. For å finne svaret kan vi benytte formelen

$$P(X = x) = \frac{\binom{s}{x} \binom{n-s}{r-x}}{\binom{n}{r}}$$

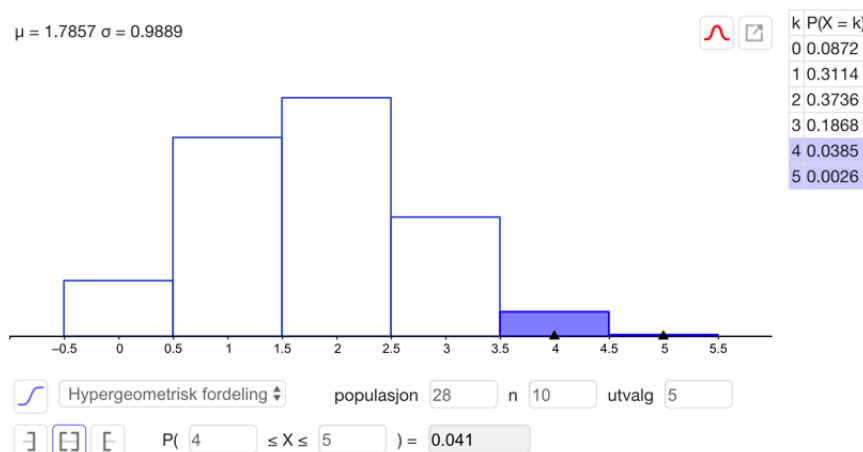
Da får vi at sannsynligheten for å trekke tre jenter blir

$$P(X = 3) = \frac{\binom{10}{3} \binom{28-10}{5-3}}{\binom{28}{5}} = \frac{17}{91} = 0.1886813$$

Ved hjelp av en kalkulator, eller programvare, kan vi benytte kommandoen $\text{nCr}(n, r)$, eller en liknende syntaks. Her er et eksempel gjort med TI-Nspire.

$\frac{\text{nCr}(10,3) \cdot \text{nCr}(18,2)}{\text{nCr}(28,5)}$	$\frac{17}{91}$
$\frac{\text{nCr}(10,3) \cdot \text{nCr}(18,2)}{\text{nCr}(28,5)}$	0.186813

GeoGebra har en sannsynlighetskalkulator vi kan benytte ved å hypergeometrisk fordeling og lese av i tabellen. Figur 7.1 viser tabellen og en grafisk framstilling. Leser vi av i tabellen for $k = 3$ får vi samme svar.



Figur 7.1: Sannsynlighetskalkulatoren i GeoGebra

Legg merke til at tabellen inneholder alle sannsynlighetene for

$$P(X = x) = \frac{\binom{s}{x} \binom{n-s}{r-x}}{\binom{n}{r}} = \frac{\binom{10}{x} \binom{28-10}{5-x}}{\binom{28}{5}}$$

Legg også merke til at vi kan også få regnet ut summen av sannsynligheter. Figur 7.1 viser også at sannsynligheten for at fire eller fem jenter velges ut er 0.041.

Python

For å få beregnet dette i Python må vi skrive en del kode eller importerer moduler. Det gjør at andre verktøy egner seg bedre for slike utregninger. Vi kan ta med noen eksempler for å vise det. Vi kan lage en funksjon som beregner sannsynligheten etter formelen som ble presentert tidligere. Programkode 7.1 viser hvordan.

```
1 import math as m
2
3 n = 28
4 s = 10
5 r = 5
6
7 def binomial(a,b):
8     bin = m.factorial(a)/(m.factorial(b)*m.factorial(a-b))
9     return bin
10
11 def hypergeometrisk(n, s, r, x):
12     hg = binomial(s,x) * binomial(n-s,r-x)/binomial(n,r)
13     return hg
14
15 print(hypergeometrisk(n,s,r,3))
```

Programkode 7.1: Egendefinert funksjon

Dette er en løsning som gjør at enkelt kan gjenbruke koden til andre oppgaver. Modulen `math` er importert for å slippe å lage en egen funksjon for fakultet. Den samme modulen har også en funksjon for å finne binomialkoeffisienten som heter `comb`. Den kunne erstattet funksjonen `binomial`.

Nå fins det flere muligheter. Modulen `scipy` kan gjøre det samme for oss, men det blir ikke så klart hva som skjer. Programkode 7.2 viser hvordan kommandoen `hypergeom` gir oss en hypergeometrisk fordeling hvor vi kan finne bruke funksjonen `pmf`, *probability mass function* for å finne det vi er ute etter.

```
1 import scipy.stats as st
2 n = 28
3 s = 10
4 r = 5
5
6 hg = st.hypergeom(n,s,r)
7 p = hg.pmf(3)
8 print(p)
```

Programkode 7.2: Scipy

Kjører vi programmene får vi samme svar som ved utregning

0.18681318681318695

7.4.1 Simulering med Python

Dette eksemplet kan vi også simulere med Python. Programkode 7.3 viser et eksempel.

```
1 import random as rd
2 ant_forsok = 10000
3 ant_gutter = 18
4 ant_jenter = 10
5 teller = 0
6
7 klassen = ["G"]*ant_gutter + ["J"]*ant_jenter
8
9 for i in range(ant_forsok):
10     utvalg = rd.sample(klassen, 5)
11     if utvalg.count("J") == 3:
12         teller = teller + 1
13
14 print ("Sannsynligheten for å trekke tre jenter: ", teller/ant_forsok)
```

Programkode 7.3: Hypergeometrisk utvalg

Det krever en del kjennskap til lister og hvordan vi kan få laget dem. Koden `klassen = ["G"]*ant_gutter + ["J"]*ant_jenter` lager ei liste med bokstaver. For å slippe å skrive inn alle bokstavene er det gitt beskjed om hvor mange av hver som skal legges inn. Lista `klassen` blir da sånn:

```
['G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'G', 'J', 'J', 'J', 'J', 'J', 'J', 'J', 'J', 'J', 'J']
```

Fra denne lista plukkes det ut fem elementer uten tilbakelegging, som legges i lista `utvalg`, med kommandoen `utvalg = rd.sample(klassen, 5)`. Vi kan telle antall "J" i utvalget med kommandoen `utvalg.count("J")`. En hvis-setning teller opp antall utvalg med tre "J". I mitt tilfelle fikk jeg dette resultatet:

Sannsynligheten for å trekke tre jenter: 0.1894

7.5 Lotto

Vinnertall							Tilleggstall
12	16	20	22	30	33	34	6
Premieoversikt							
Antall	Premie	Vinnere					
7	Gulltrekning!	0					
6+1	174 725 kr	12					
6	4 290 kr	512					
5	120 kr	19 837					
4	50 kr	284 193					

Oppgave 8

Lotto er et populært spill hvor det trekkes ut sju tall først. Så trekkes det ut tre tilleggstall. Totalt er det 34 tall det velges blant. Spillet kan gi gevinst hvis følgende oppnås

- 7 rette
- 6 rette + ett tilleggstall
- 6 rette
- 5 rette
- 4 rette + ett tilleggstall

Finn sannsynlighetene for å oppnå premie

Sannsynligheten for å få sju rette

Først finner vi antall mulige måter sju tall kan trekkes av 34 tall. Det trekkes uten tilbakelegging og rekkefølgen spiller ingen rolle: den er uordna.

Antall mulige finner vi da ved

$${}_{34}C_7 = \binom{34}{7} = 5379616$$

Av alle disse muligheten er det bare en som er gunstig for å vinne. Da finner vi at sannsynligheten blir

$$P(7 \text{ rette}) = \frac{1}{\binom{34}{7}} = \frac{1}{5379616} = 1.8589 \cdot 10^{-7}$$

Sannsynligheten for å få seks rette og ett tilleggstall

Antall gunstige kombinasjoner vil være at det seks av de sju er riktige. I tillegg kan vi ha ett av de tre tilleggstallene riktige. Antall gunstige valg blir da

$$\binom{7}{6} \cdot \binom{3}{1} = 21$$

Antall mulige kombinasjoner har vi fra før. Sannsynligheten blir

$$P(6 \text{ rette} + \text{ett tilleggstall}) = \frac{\binom{7}{6} \cdot \binom{3}{1}}{\binom{34}{7}} = \frac{21}{5379616} = 3.9036 \cdot 10^{-6}$$

Sannsynligheten for å få seks rette

I dette tilfellet er det ingen riktige tilleggstall. Når alle tallene er trukket vil det være igjen $34 - 10 = 24$. Det siste tallet må være blant de som er igjen. Da kan vi finne sannsynligheten

$$P(6 \text{ rette}) = \frac{\binom{7}{6} \cdot \binom{24}{1}}{\binom{34}{7}} = \frac{21}{672452} = 3.1229 \cdot 10^{-5}$$

Sannsynligheten for å få fem rette

Nå gjelder det å ha fem av sju riktige tall. Det spiller ikke noen rolle om vi har tilleggstall. De to vi har feil kan da velges blant de 27 øvrige

$$P(5 \text{ rette}) = \frac{\binom{7}{5} \cdot \binom{27}{2}}{\binom{34}{7}} = \frac{7371}{5379616} = 1.3702 \cdot 10^{-3}$$

7.6 Poker

I et pokerspill trekker vi også uten tilbakelegging og vi kan benytte en hypergeometrisk sannsynlighetsmodell for å finne sannsynlighetene for noen hendinger. Nå blir det noe mer komplisert enn det forrige eksemplet, men vi kan ta det med for å vise et eksempel tankegangen

Sannsynligheten for å få ett par Her er det 13 forskjellige valører og man skal ha par i ett av dem. Det er fire farger å velge mellom og vi skal ha to av dem. De andre kortene kan velges fra resten

$$P(\text{ett par}) = \frac{\binom{13}{1} \binom{4}{2} \binom{12}{3} \cdot 4^3}{\binom{52}{5}} = \frac{13 \cdot 6 \cdot 220 \cdot 64}{2598960} = 0.4226$$

Sannsynligheten for å få to par Det er 13 forskjellige verdier å velge blant for å få to kort med samme verdi $\binom{13}{2}$. Vi skal ha to par og får at det vil være $\binom{4}{2} \cdot \binom{4}{2} = 36$ mulige kombinasjoner. Da er det 11 verdier igjen å velge det siste kortet fra $\binom{11}{1} = 11$. Hvert av disse kan velges fra hver av de fire $\binom{4}{1} = 4$

$$P(\text{to par}) = \frac{\binom{13}{1} \binom{4}{2} \binom{4}{2} \binom{11}{1} \binom{4}{1}}{\binom{52}{5}} = \frac{78 \cdot 6 \cdot 6 \cdot 11 \cdot 4}{2598960} = 0.04754$$

3.1	Terningkast	18
3.2	Simulering av terningkast	19
3.3	Grafisk framstilling	19
4.1	Kast med to terninger	25
4.2	Simulering av to kast med terning	28
5.1	Kartesisk produkt	38
5.2	Ordna utvalg med tilbakelegging	41
5.3	Permutasjoner av The Beatles	43
5.4	Egendefinert <code>nPr</code>	44
5.5	Uordna utvalg uten tilbakelegging	47
5.6	Uordna utvalg med tilbkelegging	49
6.1	Straffespark	59
6.2	Straffespark	59
6.3	Straffespark	60
6.4	<code>for</code> -sløyfe	60
6.5	Newton-Pepys med Python	61
7.1	Egendefinert funksjon	66
7.2	Scipy	66
7.3	Hypergeometrisk utvalg	67