STK1100 våren 2023

Introduksjon til sannsynlighetsbegrepet

Svarer til avsnittene 2.1 og 2.2 i læreboka

Matematisk institutt
Universitetet i Oslo

Deterministiske fenomener

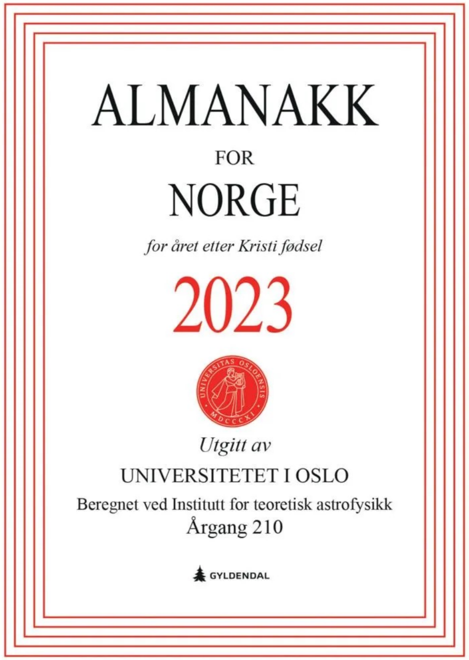
Almanakk for Norge viser:

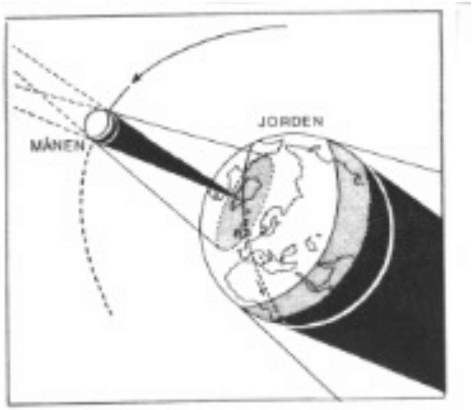
- når det er fullmåne
- når det er soloppgang og solnedgang

Det er mulig siden astronomene kan gi en matematisk beskrivelse av himmellegemenes bevegelser.

De kan også regne ut når vi vil få solformørkelse.

Fullmåne, soloppgang/nedgang og solformørkelse kan forutsies. De er determinstiske fenomener.





Fra Store Norske Leksikon

Stokastiske forsøk

Vi vet ikke hva resultatet vil bli:

- når vi kaster en terning
- i neste ukes lottotrekning

Norsk Tipping

Terningkast og lottotrekning er eksempler på stokastiske forsøk (tilfeldige forsøk).

Det er også et stokastisk forsøk å se hvilket kjønn et nyskapt barn får.

Kjennetegnet på et stokastisk forsøk er at vi *ikke* kan forutsi resultatet.

Sannsynlighetsregning er matematikken for stokastiske forsøk.

Litt historikk

Menneskene har hatt terninger i tusenvis av år:

– Astragalus: lagd av en liten knokkel fra foten til en sau eller hund

– Sekssidet terning lagd for eksempel av bein, stein eller bronse

(www.historicgames.com)

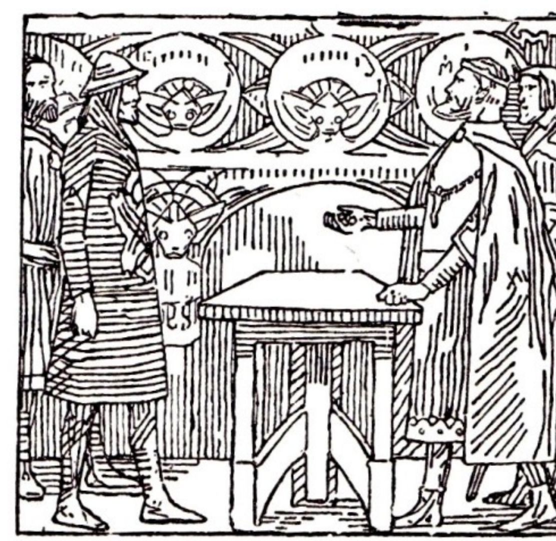
(www.gambletribune.org)

Terningkast og loddtrekning ble brukt for å komme fram til riktig avgjørelse på et vanskelig problem. Avgjørelsen ble da overlatt til høyere makter.

Apg. 1, 23-26:

To menn ble kalt fram, Josef Barsabbas med tilnavnet Justus, og Mattias. Så bad de: «Herre, du som kjenner alles hjerter, vis oss hvem av disse to du har utvalgt til å ha den tjeneste og det apostelembete som Judas forlot for å gå til sitt sted.» De kastet lodd mellom dem, og loddet falt på Mattias. Fra nå av ble han regnet som apostel sammen med de elleve.

Torstein Frode sier at det var en bygd på Hisingen som snart hadde fulgt med Norge og snart med Gøtaland. Nå avtalte kongene med hverandre at de skulle kaste lodd om hvem som skulle eie den; de skulle kaste terninger, og den som fikk størst, skulle ha den. Sveakongen kastet to seksere, og så sa han at kong Olav trengte ikke kaste. Han ristet terningene i hånden og sa: «Det er to seksere på terningen ennå, og det er ingen sak for Gud min herre å la dem komme opp.» Han kastet, og det kom opp to seksere. Så kastet Olav sveakonge, og det ble to seksere igjen. Så kastet Olav Norges konge, og da kom det opp seks på den ene, men den andre gikk i stykker, så det kom opp sju på den. Da fikk han bygda. Vi har ikke hørt noe annet å fortelle fra dette møtet. Kongene skiltes som forlikte.



Kongene kaster terninger om en bygd på Hisingen.

(Snorre: Olav den helliges saga)

Blaise Pascal (1623-1662)

Pierre de Fermat (1601-1665)

Det historiske gjennom-bruddet for sannsynlighets-regningen kom først i 1654 i en brevveksling mellom Pascal og Fermat om noen problemer knyttet til spill.

(Bilder fra www.york.ac.uk/depts/maths/histstat)

Pascal og Fermat var særlig interessert i delingsproblemet («problem of points»):

To spillere legger 50 kr hver i en pott. Spillerne kaster ett kronestykke flere ganger. Hvis det blir krone får spiller A ett poeng, hvis det blir mynt får spiller B ett poeng. Førstemann til 10 poeng vinner hele potten. Spillet må imidlertid avbrytes når spiller A har fått 8 poeng og spiller B har fått 7 poeng. Hvordan skal de dele potten?

Etter hvert rettet interessen seg også mot andre områder enn spill. Blant annet ble livsforsikring vanligere blant de velstående på 1700-tallet. Og da måtte en kunne beregne overlevelses-sannsynligheter med utgangs-punkt i empiriske data.

A generall Bill for this present year, ending the 19 of *December* 1665. according to the Report made to the KINGS most Excellent Majesty.

By the Company of Parish Clerks of *London*, &c.

The Diseases and Casualties this year.

Abortive and Stilborne	617	Executed	21	Palsie	30
Aged	1545	Flox and Small Pox	655	Plague	68596
Ague and Feaver	5257	Found dead in streets, fields, &c.	20	Plannet	6
Appoplex and Suddenly	116	French Pox	86	Plurisie	15
Bedrid	10	Frighted	23	Poysoned	1
Blasted	5	Gout and Sciatica	27	Quinsie	35
Bleeding	16	Grief	46	Rickets	557
Bloody Flux, Scowring & Flux	185	Griping in the Guts	1288	Rising of the Lights	397
Burnt and Scalded	8	Hangd & made away themselves	7	Rupture	34
Calenture	3	Headmouldshot & Mouldfallen	14	Scurvy	105
Cancer, Gangrene and Fistula	56	Jaundies	110	Shingles and Swine pox	2
Canker, and Thrush	111	Impostume	227	Sores, Ulcers, broken and bruised Limbs	82
Childbed	625	Kild by severall accidents	46	Spleen	14
Chrisomes and Infants	1258	Kings Evill	86	Spotted Feaver and Purples	1929
Cold and Cough	68	Leprosie	2	Stopping of the stomack	332
Collick and Winde	134	Lethargy	14	Stone and Strangury	98
Consumption and Tissick	4808	Livergrown	20	Surfet	1251
Convulsion and Mother	2036	Meagrom and Headach	12	Teeth and Worms	2614
Distracted	5	Measles	7	Vomiting	51
Dropsie and Timpany	1478	Murthered and Shot	5	Wenn	8
Drowned	50	Overlaid & Starved	45		

Christned: Males 5114, Females 4853, In all 9967

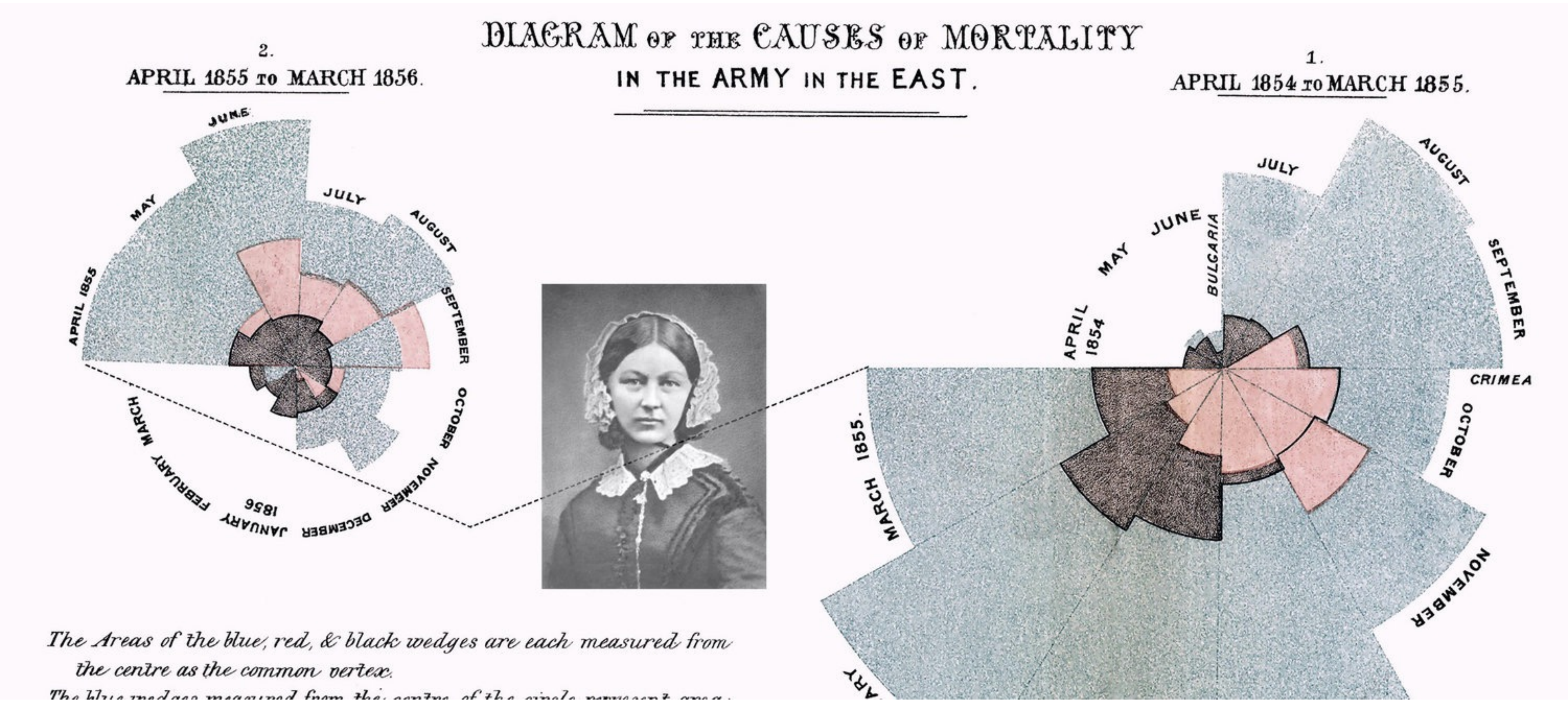
Buried: Males 48569, Females 48737, In all 97306 — Of the Plague 68596

Increased in the Burials in the 130 Parishes and at the Pest-house this year — 79009

Increased of the Plague in the 130 Parishes and at the Pest-house this year — 68550

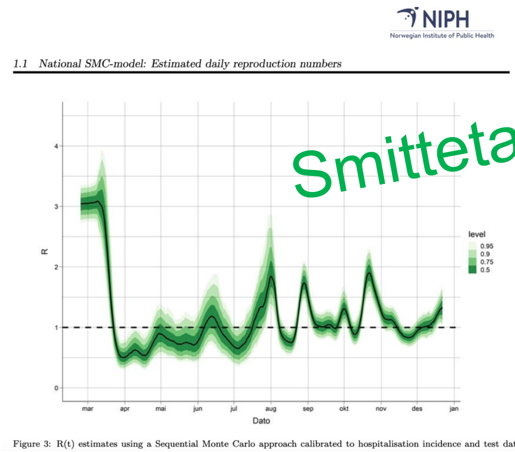
(www.york.ac.uk/depts/maths/histstat)

En av de første som systematisk samlet inn store mengder data og visualiserte dem med revolusjonerende grafikk, var Florence Nightingale (1820– 1910). Dette startet hun med under Krimkrigen midt på 1800-tallet. Hun brukte statistiske metoder for å vise at det var nødvendig å endre medisinske rutiner, og regnes som en pioner innen dataanalyse.



https://www.dataceutics.com/blog/2018/4/27/a-passionate-statistician-florence-nightingale (ill.)
https://www.sciencemuseum.org.uk/objects-and-stories/florence-nightingale-pioneer-statistician

I dag brukes sannsynlighetsregning og statistiske modeller og metoder for stokastiske forsøk på en rekke områder.



Smittetallet R

For eksempel innen:

- Forsikring og finans
- Medisin og genetikk
- Klima-studier
- Forskning, på alle fakulteter
- Bilde- og talegjenkjenning
- Analyse av kundedatabaser (og fra mobil, bank, klikkedata) og anbefalings-systemer
- Web-applikasjoner (Google, Facebook)
- AI, kunstig intelligens
- ++

Data Science, Statistikk, Maskinlæring

Utfall og utfallsrom

Generelt er resultatet av et stokastisk forsøk *ikke* gitt på forhånd. Men vi kan angi de mulige resultatene, eller *utfallene* for forsøket.

Mengden av alle utfall kaller vi *utfallsrommet* S *.*

Eksempel 1: Når vi kaster en terning vet vi *ikke* hvor mange øyne vi vil få. Men vi vet at antall øyne vil bli 1, 2, 3, 4, 5 eller 6.

Utfallsrommet er $S = \{1, 2, 3, 4, 5, 6\}$.

Eksempel 2: Kast et kronestykke til første gang du får «mynt»:

Utfallsrom: $S = \{M, KM, KKM, KKKM, \ldots\}$

Eksempel 3: Registrer vekten til en nyfødt jente.

Kan ikke her liste opp alle mulige fødselsvekter.

Utfallsrommet er er intervall på tallinja, for eksempel $S = (0, 6)$.

I eksempel 1 er utfallsrommet *endelig*, mens det er *tellbart uendelig* i eksempel 2.

En fellesbetegnelse for disse to situasjonene er at utfallsrommet er *diskret.*

I eksempel 3 er utfallsrommet *kontinuerlig.*

Begivenheter

Vi vil ofte være interessert i et resultat av et forsøk som svarer til flere utfall.

Et resultat av et forsøk som svarer til ett eller flere utfall kaller vi en *begivenhet* (hendelse).

Eksempel 1 (forts): Ved terningkast kan vi være interessert i om vi får *minst* fem øyne.

Begivenheten A = «minst fem øyne» kan vi skrive $A = \{5, 6\}$.

Eksempel 2 (forts): Kast kronestykke til første mynt.

A = «høyst tre kast» = $\{M, KM, KKM\}$

B = «minst tre kast» = $\{KKM, KKKM, KKKKM, \ldots\}$

Eksempel 3 (forts): Registrer vekten til en nyfødt jente.

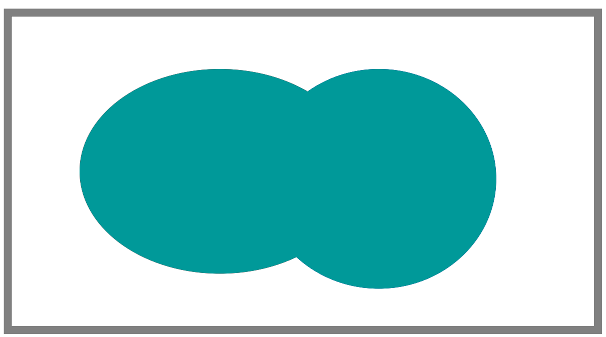
A = «veier mellom 3 kg og 4 kg» = $(3, 4)$

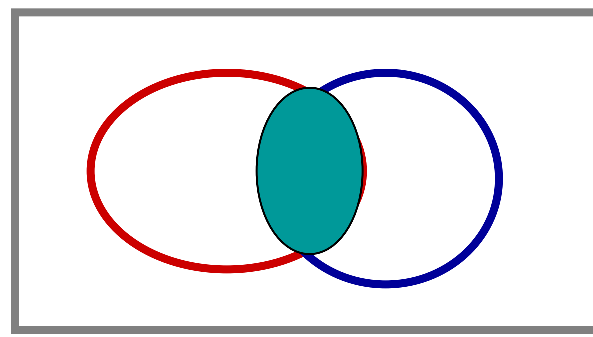
Union og snitt

La A og B være to begivenheter ved et forsøk
Ut fra disse kan vi lage to nye begivenheter:

$A \cup B$ omfatter alle utfall som er med i A eller i B eller i begge («A union B»)

$A \cap B$ omfatter alle utfall som er med i både A og B («A snitt B»)





Hvis det ikke fins noen utfall som er med i både A og B (dvs. $A \cap B = \varnothing$) er begivenhetene *disjunkte.*

Komplementære begivenheter

La A være en begivenhet ved et forsøk.

Det at A *ikke* inntreffer er en begivenhet vi kaller «ikke A» og skriver A' .

A' inneholder alle utfall som *ikke* er med i A.

A og A' er *komplementære* begivenheter.

Merk at A og A' er *disjunkte.*

Det klassiske sannsynlighetsbegrepet

De første arbeidene om sannsynlighet var om problemer i spill, og sannsynlighetsbegrepet de brukte var tilpasset dette.

Et enkelt eksempel illustrerer tankegangen.

Vi kaster en terning.

Hva er sannsynligheten for at vi får minst fem øyne?

Alle sidene på terningen har *samme sannsynlighet* for å vende opp når terningen kastes.

Sannsynligheten er lik 1 for at en eller annen side vil vende opp.

Derfor er sannsynligheten 1/6 for hver av de seks utfallene 1, 2, 3, 4, 5 og 6.

Begivenheten «minst fem øyne» består av de to utfallene 5 og 6.

Sannsynligheten for minst fem øyne er derfor 2/6 = 1/3.

Generell formulering av argumentet:

Et stokastisk forsøk har N utfall.
Det er de *mulige utfallene* for forsøket.

Vi antar at de N utfallene er *like sannsynlige.*

Da har hvert utfall sannsynlighet $1/N$.

En begivenhet A består av $N(A)$ utfall.
Det er de *gunstige utfallene* for begivenheten A.

Sannsynligheten for begivenheten A er

$$\boxed{P(A) = \frac{N(A)}{N} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}}}$$

Sannsynlighet og relativ frekvens

Hva betyr det at sannsynligheten er 1/6 for sekser i et terningkast?

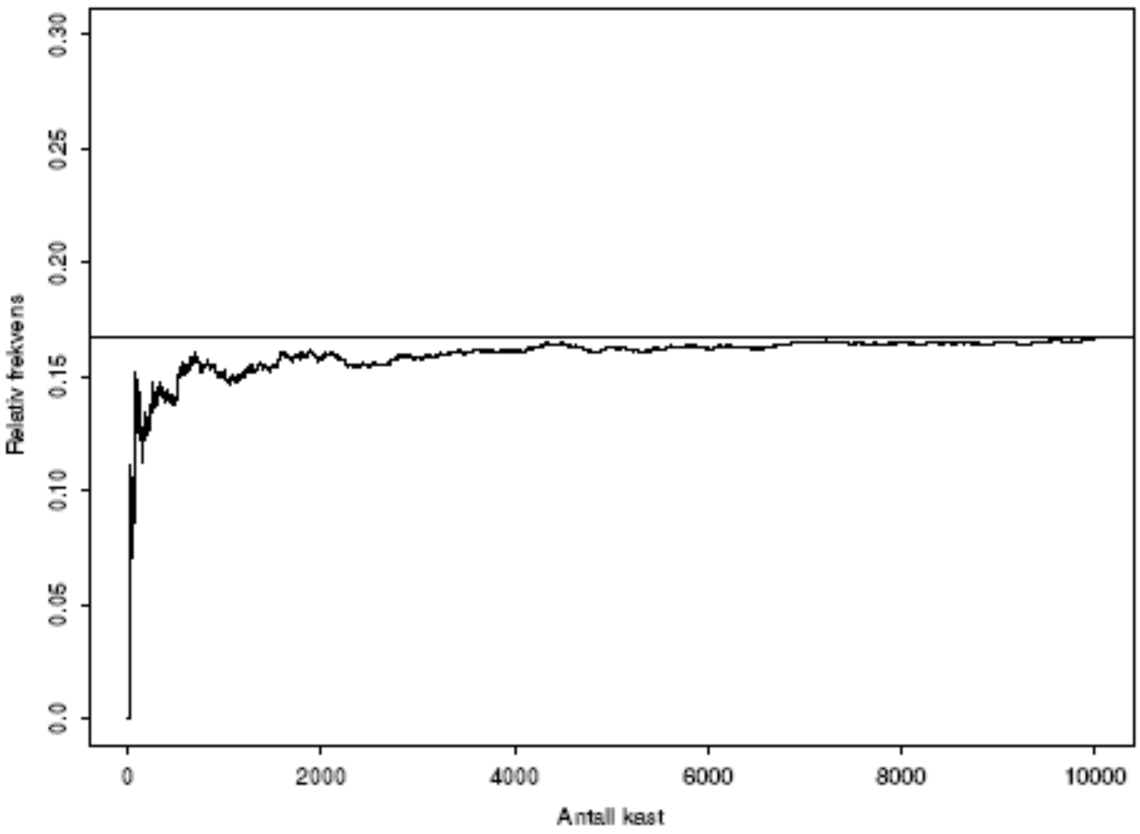
Vi kan ikke forutsi resultatet av ett kast, men vi ser et mønster når terningen kastes mange ganger.

Kaster først en terning 10 ganger:

Relativ frekvens av seksere er 1/10 = 0.10.

Kaster så en terning om og om igjen. Etter *n* kast er den relative frekvensen av seksere:

$$r_n(6) = \frac{\text{antall seksere i } n \text{ kast}}{n}$$



Merk at den relative frekvensen nærmer seg 1/6=0.167 når *n* blir stor.

Hva er sannsynligheten for at et nyfødt barn er en jente?

År	Fødte (årsgjennomsnitt)		
	I alt	Gutter	Jenter
1951-1955	62478	32182	30296
1956-1960	63021	32374	30647
1961-1965	63989	32992	30997
1966-1970	66697	34368	32329
1971-1975	61393	31487	29906
1976-1980	51744	26619	25125
1981-1985	50660	26030	24629
1986-1990	56862	29154	27708
1991-1995	60196	30993	29202
1996-2000	59522	30598	29043
2001-2005	56459	28925	27534
2006-2010	60150	30885	29265
2011-2015	59522	30563	28959

(Kilde: Statistisk sentralbyrå)

Den relative frekvensen av jentefødsler varierer lite fra femårsperiode til femårsperiode.

I perioden 1951-2015 ble det født 3864050 barn i Norge. Av disse var 1878200 jenter.

Relativ frekvens av jenter i hele perioden er 48.6%

Den relative frekvensen av seksere er omtrent 1/6 når vi kaster en terning mange ganger.

Den relative frekvensen av jenter blant alle nyfødte er omtrent 48.6% hvert år.

Det er to eksempler på et fenomen som observeres gang på gang:

> Vi er interessert i en begivenhet *A* i et stokastisk forsøk. Forsøket gjentas under like betingelser. Da vil den relative frekvensen av *A* nærme seg en grenseverdi når forsøket gjentas mange ganger. Denne grenseverdien er sannsynligheten $P(A)$.

Sannsynlighet er relativ frekvens «i det lange løp».

Merk at det er en forutsetning at forsøket gjentas under *like betingelser*.

Et eksempel hvor dette ikke er tilfelle er flerfødsler:

År	Fødte	Tvillinger (årsgj.snitt)	Trillinger	Relativ frekvens tvillinger (%)	Relativ frekvens trillinger (%)
1951-1955	62478	787	9	1.3	0.015
1956-1960	63021	731	7	1.2	0.011
1961-1965	63989	700	8	1.1	0.013
1966-1970	66697	663	7	1.0	0.011
1971-1975	61393	568	4	0.9	0.007
1976-1980	51744	494	4	1.0	0.008
1981-1985	50660	495	8	1.0	0.016
1986-1990	56862	634	19	1.1	0.034
1991-1995	60196	821	24	1.4	0.040
1996-2000	59522	957	24	1.6	0.041
2001-2005	56459	1034	17	1.9	0.031
2006-2010	60150	1021	15	1.7	0.025
2011-2015	59522	960	15	1.6	0.025

Andelen tvilling- og trillingfødsler økte fra midten av 1980 årene på grunn av assistert befruktning.

Sannsynlighet forstått som relativ frekvens i det lange løp kalles enkelte ganger *objektiv sannsynlighet*.

Merk at denne forståelsen av sannsynlighet forutsetter at forsøket kan gjentas flere ganger under like betingelser.

Noen ganger brukes sannsynlighetsbegrepet i situasjoner hvor forsøket ikke kan gjentas under like betingelser.

«Det er minst 50 % sjanse for at vi klarer å reise til Mars i løpet av dette tiåret!».

Sannsynlighet som uttrykker en personlig vurdering kalles *subjektiv sannsynlighet*.

Aksiomer for sannsynlighet

Vi kan ikke bruke fortolkningen av sannsynlighet som relativ frekvens i det lange løp til å lage en matematisk teori for sannsynlighet.

For å lage en matematisk teori for sannsynlighet, tar vi utgangspunkt i noen aksiomer som sannsynligheter må oppfylle.

Aksiomene er motivert ut fra det klassiske sannsynlighetsbegrepet og fortolkningen av sannsynlighet som relativ frekvens i det lange løp.

Aksiom 1

For enhver begivenhet A er $P(A) \geq 0$.

Aksiom 2

For utfallsrommet S er $P(S) = 1$.

Aksiom 3

Hvis $A_1, A_2, A_3, ...$ er en uendelig følge av disjunkte begivenheter, så er $P(A_1 \cup A_2 \cup A_3 \cup .....) = \sum_{i=1}^{\infty} P(A_i)$.

Aksiomene sier ikke noe om hvilke verdier sannsynlighetene skal ha i en gitt situasjon. For å angi konkrete verdier for sannsynlighetene, må vi bruke kunnskaper om den situasjonen vi studerer.

Med utgangspunkt i aksiomene, kan vi utlede regler for å regne med sannsynligheter (utledningen blir gjort på forelesningen)

1) $P(\varnothing) = 0$

2) Hvis $A_1,\ A_2\ , ...,A_k$ er en endelig følge av disjunkte begivenheter, så er $P(A_1 \cup A_2 \cup .... \cup A_k) = \sum_{i=1}^{k} P(A_i)$

3) $P(A) = 1 - P(A')$

4) $P(A) \leq 1$

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

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

Sannsynligheter for diskrete utfallsrom

Anta at utfallsrommet er diskret:

$$S = \{E_1, E_2, E_3, \ldots\ldots\}$$

Det kan vises at vi får sannsynligheter som tilfredsstiller aksiomene ved å angi en sannsynlighet til hvert utfall E_i slik at

$$P(E_i) \geq 0$$

$$\sum_{\text{alle } i} P(E_i) = 1$$

Videre finner vi sannsynligheten for en begivenhet *A* ved å legge sammen sannsynligheten for alle utfallene som er med i *A*, dvs.

$$P(A) = \sum_{\substack{\text{alle } i \text{ der } E_i \\ \text{er med i } A}} P(E_i)$$

Uniform sannsynlighet

Anta at utfallsrommet er endelig og består av N utfall som alle er like sannsynlige.

Da har vi at $P(E_i) = \dfrac{1}{N}$ for hvert utfall.

Hvis begivenheten A består av $N(A)$ utfall, er

$$P(A) = \sum_{\substack{\text{alle } i \text{ der } E_i \\ \text{er med i } A}} P(E_i) = \sum_{\substack{\text{alle } i \text{ der } E_i \\ \text{er med i } A}} \frac{1}{N} = \frac{N(A)}{N}$$

Vi får altså regelen om «gunstige delt på mulige» som en konsekvens av aksiomene.