Kapittel 4 Sannsynlighet: Studiet av tilfeldighet

4.1 Tilfeldighet

4.2 Sannsynlighetsmodeller

4.3 Tilfeldige variabler

4.4 Forventningsverdi og varians til tilfeldige variabler

4.5 Generelle sannsynlighetsregler

Opprinnelsen til sannsynlighetsregninga er knytta til gamblingspill på 1600-tallet.

- Enkle spill som bygger på tilfeldighet (myntkast, terningspill,..) er fortsatt gode illustrasjoner på prinsippene for sannsynlighet.

- Etter mer nøyaktige målinger innen astronomi og landskapsmåling ble teorien utvidet på 17- og 1800-tallet til å inkludere fordelinger av tilfeldige variable.

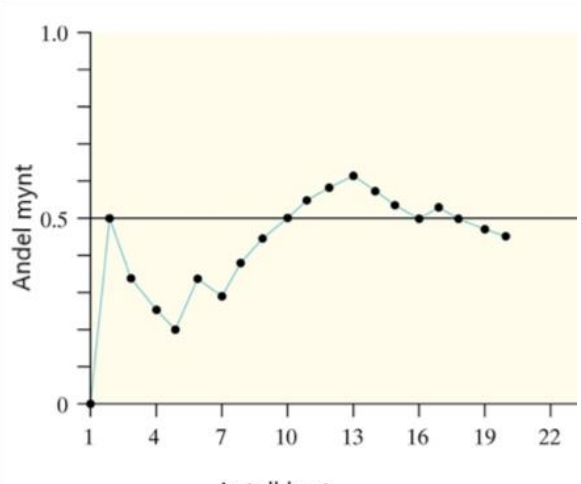
- Sannsynlighet brukes i dag i flerfoldige fagfelt blant annet
  - Trafikkflyt på veier
  - Trafikk på Internett
  - Gensammensetningen i en befolkning
  - Energitilstanden til elementærpartikler
  - Spredningen av epidemier og Twitter-meldinger
  - Avkastning på finansinvesteringer
  - *Statistiske metoder*

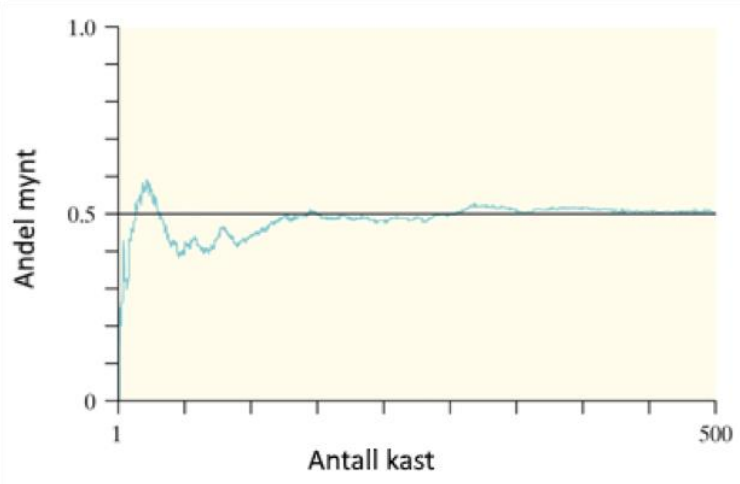
Tilfeldigheter er umulig å forutsi på kort sikt, men har et regulært og forutsigbart mønster i det lange løp.

Vi kaller et fenomen **tilfeldig** hvis de enkelte utfallene er usikre, men det er likevel en regelmessig fordeling av utfall når vi har et større antall repetisjoner.

Sannsynligheten for et utfall av et tilfeldig fenomen er andelen dette utfallet inntreffer når vi har mange repetisjoner.

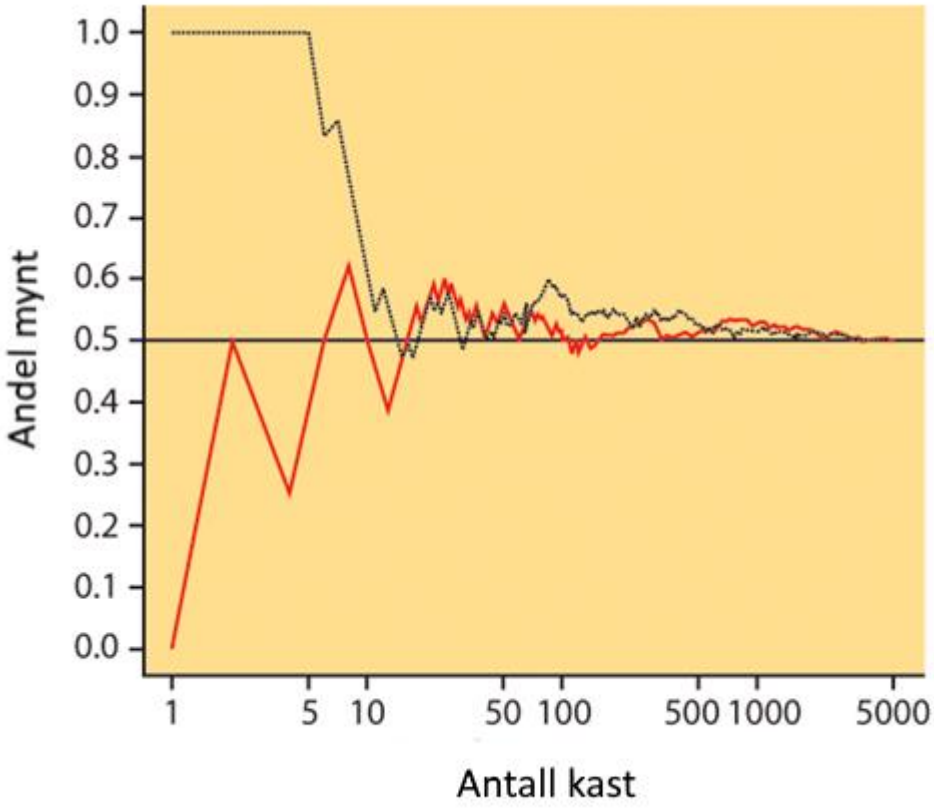
Eksempel: Myntkast- det regulære mønsteret sees etter mange kast. (Statistical Applets: Probability)





Hvordan tenke rundt tilfeldighet

Resultatet av ett enkelt myntkast er tilfeldig.
Men resultatet over flere kast er forutsigbart, så lenge kastene er **uavhengige** (utfallet til et nytt kast blir ikke påvirka av resultatet i forrige kast).



Sannsynligheten for mynt er 0.5, og dette er lik andelen mynt i en lang nok rekke av myntkast.

Første rekke av kast
Andre rekke

4.2 Sannsynlighetsmodeller

- Utfallsrom
- Sannsynlighetsregler
- Tildeling av sannsynlighet: endelige utfall
- Tildeling av sannsynlighet: like stor sannsynlighet for utfall
- Uavhengighet og produktregelen
- Bruk av reglene

Sannsynlighetsmodell: S og p(s)

Beskrivelsen av tilfeldigheter inneholder to deler: en liste med mulige utfall og en sannsynlighet for hvert utfall.

Utfallsrommet *S* av et tilfeldig fenomen er mengden av alle mulige utfall.

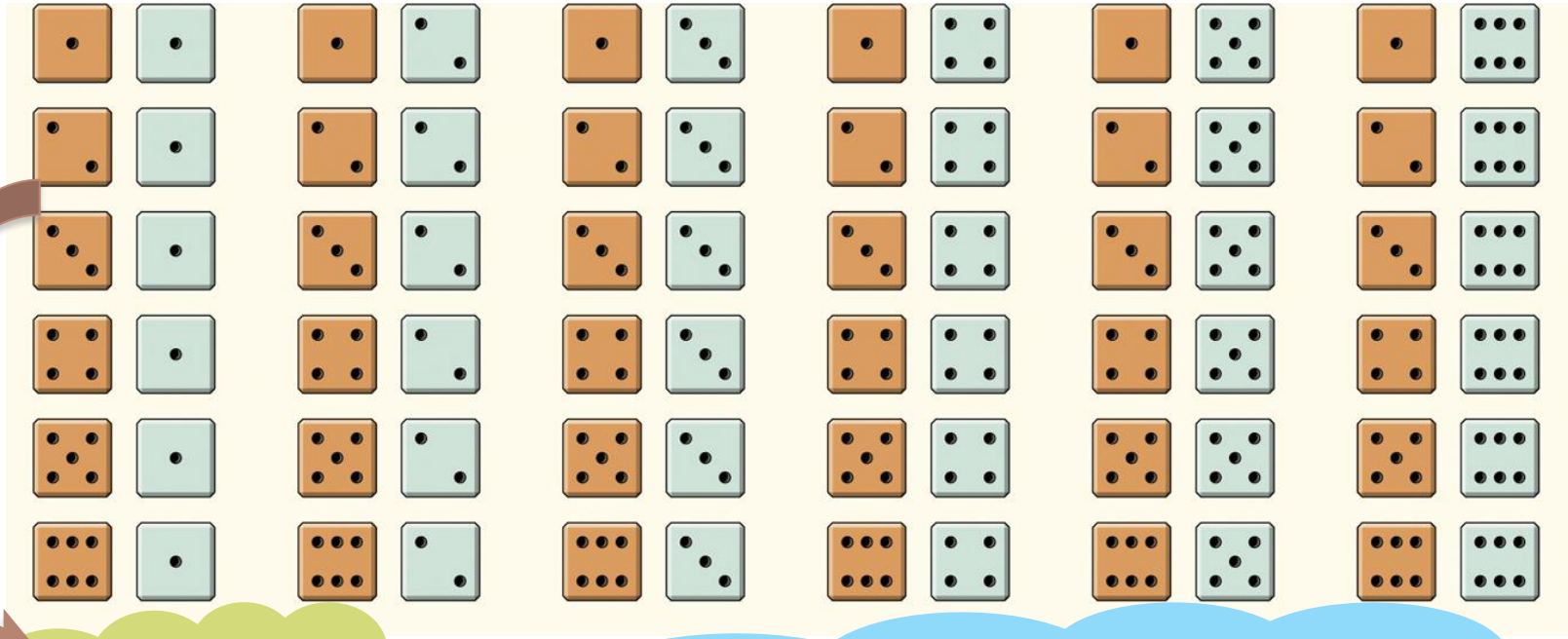
En **sannsynlighetsmodell** er en beskrivelse av et tilfeldig fenomen, og består av to deler: et utfallsrom *S* og sannsynligheten p(s) for hvert utfall s.

Eksempel: Kasting av en mynt. Kan ikke si på forhånd helt sikkert hva et spesifikt utfall av ett kast vil være. Men vi kan si noe om mulige utfall (kron (K) eller mynt (M)), og hvis vi antar en balansert mynt er det rimelig å tro at sannsynligheten for hvert utfall er 1/2. Dette er et eksempel på en sannsynlighetsmodell

$$S=\{K,M\},\ P(K) = P(M) = 1/2$$

Sannsynlighetsmodell for 2 terningkast

Eksempel: Beskriv en sannsynlighetsmodell for kast av to sekssidede terninger, en rød og en grønn. Ingen av dem jukseterninger.



Utfallsrom
36 utfall

Med at terningene ikke er jukseterninger, menes det at alle utfallene like sannsynlige (slik det skal være).
Det vil si: Hvert av utfallene har sannsynlighet 1/36.
(Dette er andelen etter mange repetisjoner)

Utfallsrommet S må inneholde alle mulige utfall av fenomenet

-Men vi har ofte valgmuligheter ved spesifisering av utfallsrommet *S*, avhengig av hvilket detaljnivå som er av interesse.
Eksempel: Kast av mynt 3 ganger, hvert kast resulterer i kron (K) eller mynt (M).

Kanskje er man interessert i hvilken side av mynten som kommer opp i hvert kast, da er utfallsrommet alle mulige utfall av dette fenomenet:

$$S = \{KKK,KKM,KMK,MKK,KMM,MKM,MMK,MMM\}$$

Antall elementer i utfallsrommet er 8.

Alternativt kan man være interessert i for eksempel antall kron i hvert kast. Da er utfallsrommet alle mulige utfall av dette fenomenet:

$$S=\{0,1,2,3\}$$

Antall elementer i utfallsrommet er 4.

En hendelse er et utfall eller en samling utfall av et tilfeldig fenomen.

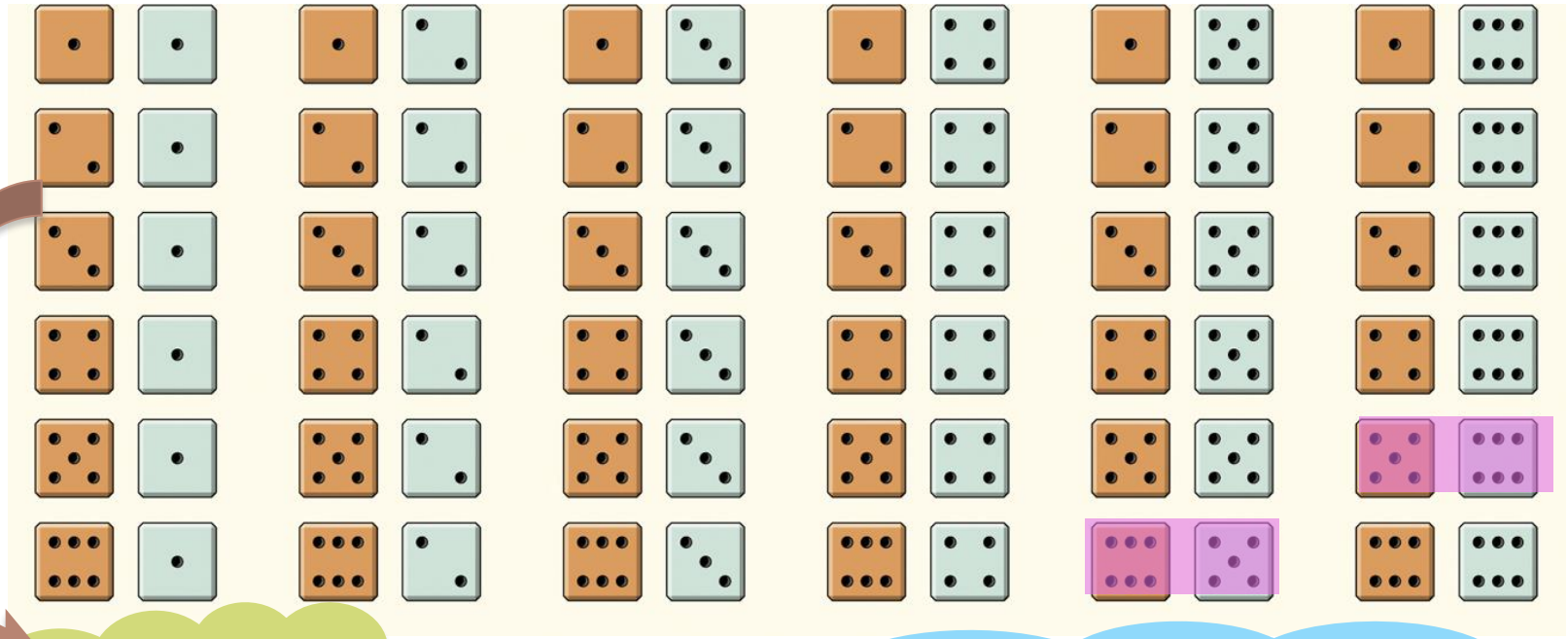
Vi har et tilfeldig fenomen med utfallsrom ***S***.

Ofte ønsker vi å se på sannsynligheten ikke bare for et enkelt utfall, men for en samling av utfall.

En hendelse er en del (undermengde) av hele utfallsrommet.

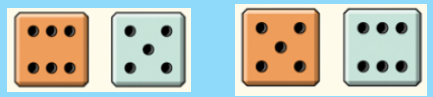
Hendelser for 2 terningkast

Eksempel: Kast av to sekssidede terninger, en rød og en grønn. Ingen av dem jukseterninger. **Vi er interesserte i følgende hendelse: En femmer og en sekser.**



Utfallsrom
36 utfall

Hendelsen vi er interesserte i er altså *A*={56,65}



Sannsynligheten for hendelsen blir 2/36

Hva kan sannsynligheten være?

1. Alle sannsynligheter er et tall mellom 0 og 1 (som en andel).
2. Alle mulige utfall må tilsammen ha sannsynlighet lik 1.
3. Hvis to hendelser ikke har noen utfall felles (disjunkte), er sannsynligheten for at den ene eller den andre skjer, lik summen av de individuelle sannsynlighetene.
4. Sannsynligheten for at en hendelse *ikke* inntreffer er 1 minus sannsynligheten for at hendelsen inntreffer.

1 flervalgsspørsmål

Et Venn-diagram viser utfallsrommet S som et rektangulært område, og hendelser i S som områder inne i rektangelet.

Et bilde som viser forholdet mellom to eller flere hendelser.

To hendelser *A* og *B* er **disjunkte** hvis de ikke har noen felles utfall. Sagt på en annen måte: De kan ikke begge inntreffe samtidig.

Disjunkte hendelser:

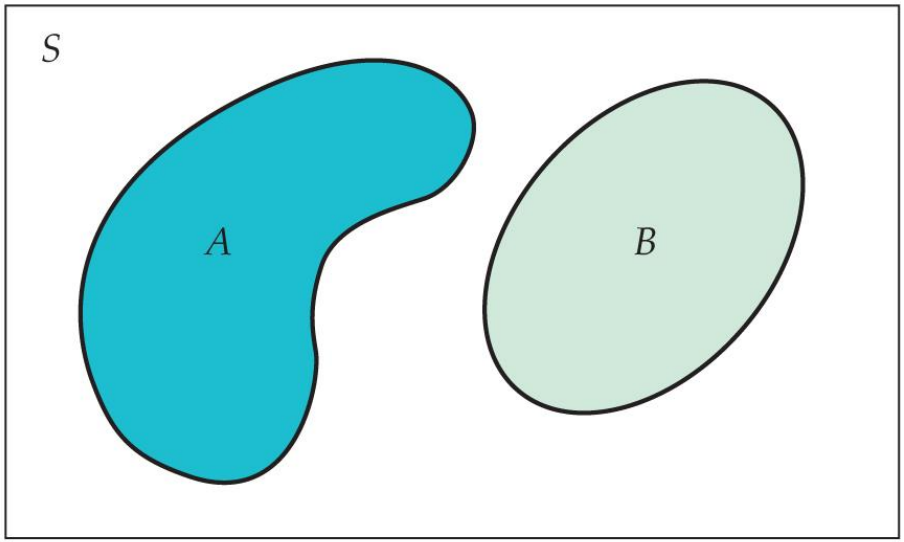


Figure 4.2
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Ikke-disjunkte hendelser:

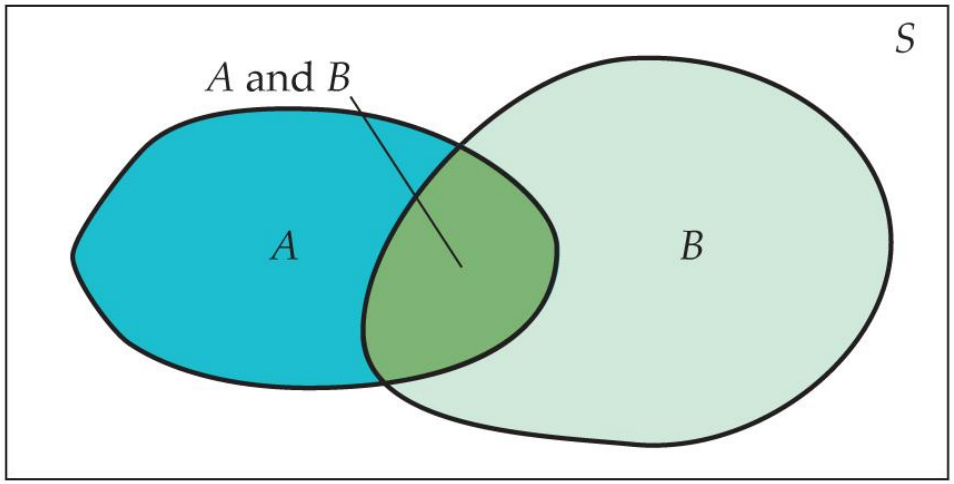


Figure 4.4
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Hvis ikke A, så komplimentet til A: A^C

La *A* være en begivenhet i utfallsrommet *S*. Komplementet til *A*, som vi kaller A^C , er den begivenhet som ikke inntreffer hvis *A* inntreffer.

Sagt på en annen måte: A^C er den mengden av elementer i *S* som ikke er med i *A.*

Enten skjer A, eller så skjer A^C

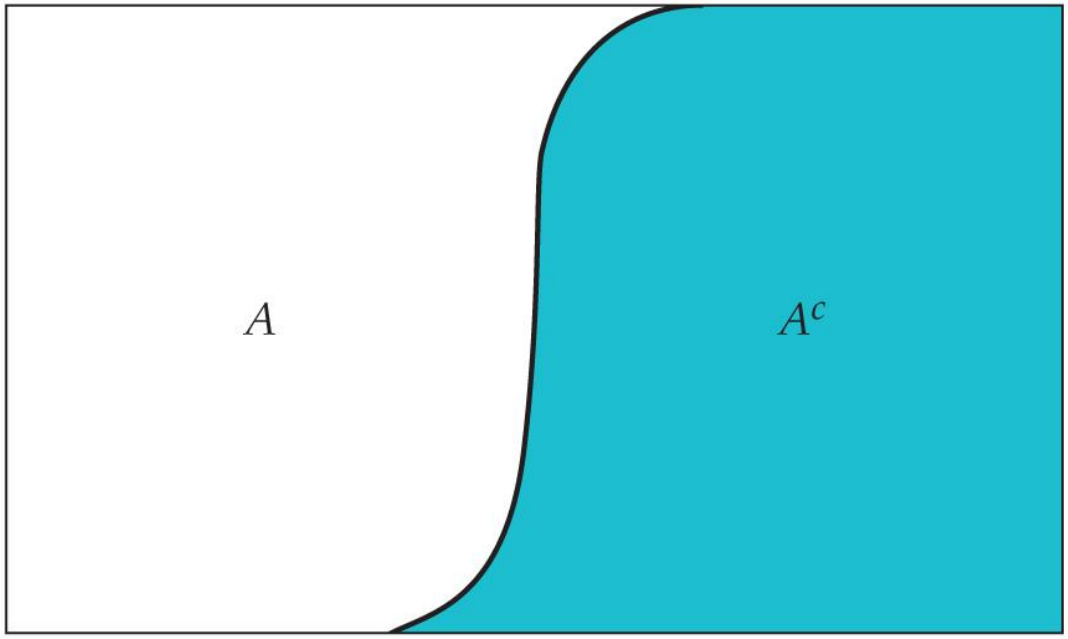


Figure 4.3
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Regler for sannsynligheter for hendelser

Regel 1. Sannsynligheten $P(A)$ for en hendelse A: $\mathbf{0 \le P(A) \le 1}$.

Regel 2. Hvis S er utfallsrommet i en sannsynlighetsmodell,

så er $\mathbf{P(S) = 1}$.

Rule 3. Hvis A og B er **disjunkte,** $\mathbf{P(A \text{ eller } B) = P(A) + P(B)}$.

Dette er **addisjonsregelen for disjunkte hendelser.**

Rule 4: Vi har at $\mathbf{P(A^C) = 1 - P(A)}$.

Regneeksempel: Sannsynlighetsregler

Eksempel: Favorittfarge på bil. Forskjellige personer har gjerne ulike preferanser, blant annet hvilken farge de vil ha på bilen sin. Her er en sannsynlighetsmodell for dette:

Farge	**Hvit**	**Svart**	**Sølvfarget**	**Grå**
Sannsynlighet	0.24	0.19	0.16	0.15
Farge	**Rød**	**Blå**	**Brun**	**Annen**
Sannsynlighet	0.10	0.07	0.05	0.04

(a) Vis at dette er en sannsynlighetsmodell.

For at det skal være en sannsynlighetsmodell, må hele utfallsrommet S være beskrevet, med tilhørende sannsynligheter. For å sjekke dette kan vi sjekke at P(S)=1: 0.24 + 0.19 + 0.16 + 0.15 + 0.10 + 0.07 + 0.05 + 0.04 = 1

(b) Hva er sannsynligheten for at en person har svart eller rød som favorittfarge?

$$P(\text{svart eller rød}) = P(\text{svart}) + P(\text{rød}) = 0.19 + 0.10 = 0.29$$

(c) Finn sannsynligheten for at favorittfargen ikke er blå.

$$P(\text{ikke blå}) = 1 - P(\text{blå}) = 1 - 0.07 = 0.93$$

Sannsynligheten til en hendelse i endelige sannsynlighetsmodeller

En sannsynlighetsmodell med et endelig utfallsrom kalles **endelig**.

For å tildele sannsynligheter til hendelser i en endelig modell, list opp sannsynlighetene til alle utfallene. Disse sannsynlighetene må være tall mellom 0 og 1, som summerer seg til nøyaktig 1.

Sannsynligheten til en hendelse er summen av sannsynlighetene til utfallene som er en del av hendelsen.

Eksempel: Benfords lov

- Første siffer i reelle tall som rapporteres inn (skattemelding, regnskap, utgiftsrefusjon etc) følger ofte en regelmessighet som kalles Benfords lov. Den sier at første siffer i slike tall har følgende sannsynlighetsmodell:

Første tall	1	2	3	4	5	6	7	8	9
Sannsyn-lighet	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

- For å oppdage juks med innrapporteringer, kan myndighetene sammenligne første siffer med disse sannsynlighetene. Dersom andelene for de første sifrene i innrapporterte tall er veldig forskjellig fra sannsynlighetene fra Benfords lov, kan det være antydning til juks, og gi grunnlag til videre undersøkelser.

Eksempel: Benfords lov

- Hendelse A={første tall er 1}
  - P(A)=P(1)=0.301, $P(A^c)$=1-0.301=0.699
- Hendelse B={første tall er ≥6}
  - P(B)=P(6)+P(7)+P(8)+P(9)=0.222
- P(A eller B)=P(A)+P(B)=0.523
- Hendelse C={første tall er odde}
  - P(C)=P(1)+P(3)+P(5)+P(7)+P(9)=0.609
- P(B eller C)
  =P(1)+P(3)+P(5)+P(6)+P(7)+P(8)+P(9)=0.727
  ≠P(B)+P(C)=0.831
  (fordi B og C er ikke disjunkte)

Dersom en kriminell lager oppkonstruerte data med tilfeldige tall ved at alle muligheter for første siffer er like sannsynlige (P(1)=P(2)=...=P(9)=1/9), blir P(B)=1/9+1/9+1/9+1/9=0.444 ≠0.222

Første tall	1	2	3	4	5	6	7	8	9
Sannsyn-lighet	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

I noen tilfeller kan vi med rimelighet anta at de ulike utfallene er like sannsynlige.

Dette er fordi det er balanse i fenomenet i seg selv. Eksempler på dette kan være myntkast eller terningkast.

Like stor sannsynlighet for utfall

Hvis et tilfeldig fenomen har *k* utfall, som alle er like sannsynlige, så har hvert enkelt utfall sannsynlighet 1/*k*. Sannsynligheten for en hendelse *A* er da

$$P(A) = \frac{antall\ utfall\ i\ A}{antall\ utfall\ i\ S}$$

$$= \frac{antall\ utfall\ i\ A}{k}$$

- Dersom alle første siffer 1-9 er like sannsynlige, kunne vi på forrige slide alternativt skrevet rett ned at P(B)= P(første tall er ≥6)=4/9 (i stedet for å regne ut P(B)=1/9+1/9+1/9+1/9)

Uavhengighet og produktregelen

Hvis to utfall *A* og *B* ikke påvirker hverandre, og sannsynligheten for det ene ikke endres av at vi vet utfallet på det andre, er utfallene **uavhengige** av hverandre.

Regel 5: Produktregelen for uavhengige utfall

To utfall *A* og *B* er **uavhengige** hvis sannsynligheten for det ene ikke påvirkes av kunnskap om utfallet til det andre. Hvis *A* og *B* er uavhengige har vi at

$$P(A \text{ og } B) = P(A) \times P(B)$$

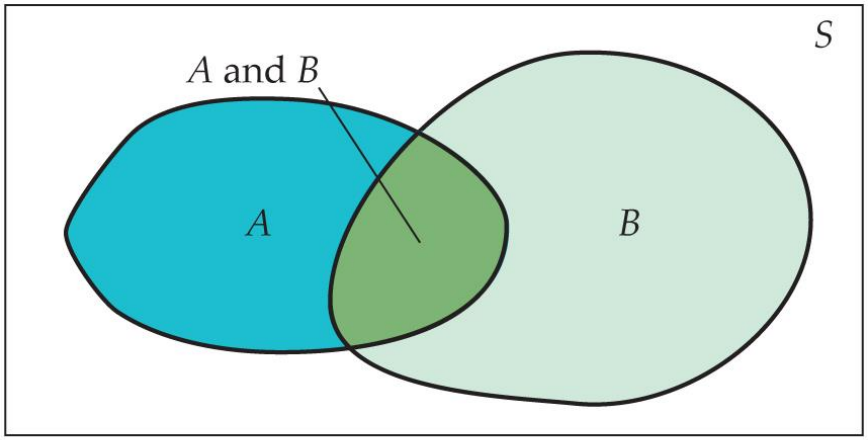


Figure 4.4
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

2 flervalgsspørsmål

Myntkast: Uavhengighet og produktregel

- Anta vi har en rettferdig mynt, dvs P(mynt)=P(kron)=1/2
- Se på hendelsene

  A=«første kast gir kron»
  B=«andre kast gir kron»

- A og B er **ikke disjunkte**- de kan begge inntreffe
- A og B er **uavhengige**: dersom vi vet hva som er utfallet av første kast (om A skjer eller ikke), endrer det ikke sannsynligheten for om B skjer
- Sannsynligheten for at både A og B inntreffer, fås fra produktregelen:

$$P(A \text{ og } B) = P(A) \times P(B) = 1/2 \times 1/2 = 1/4$$

Mendels lov: Uavhengighet og produktregel

Arv er en tilfeldig mekanisme:

- Erter kan være grønne eller gule, og hver plante bærer to gener for frøfarge
  - –Hvert gen: G (green) eller Y (yellow)
  - –Gul (Y) er dominant: GY, YG eller YY gir gule erter, mens kun GG gir grønne erter

Profimedia.CZ a.s./Alamy

- Arver ett gen fra hver foreldreplante, uavhengig av hverandre
- Anta «far» = GY og «mor» = GY, sannsynlighet 0.5 for G fra mor, sannsynlighet 0.5 for G fra far
- M={G fra «far»}, F={G fra «mor»}
- Sannsynlighet for grønt frø (uavhengighet):

$$P(M \text{ og } F)=P(M)P(F)=0.5*0.5=0.25$$

- Kan se at P(Grønn)=0.25 etter å ha observert mange generasjoner

Utdeling av kort fra kortstokk: Avhengighet

- 52 kort i en kortstokk
  - 26 røde, 26 svarte
- P(første kort rødt) = 26/52 = 0.5
- P(andre kort rødt hvis første rødt)=25/51 < 0.5
- P(andre kort rødt hvis første sort) = 26/51 > 0.5

- **Ikke uavhengighet** mellom
  A=første kort rødt og
  B=andre kort rødt hvis første rødt.
  Å vite om A skjer påvirker sannsynligheten for om B skjer

(!) Produktregelen gjelder bare ved uavhengighet

- Krybbedød: 1 av 8500 dør uforklarlig, sannsynlighet 0,000118
- Sannsynligheten for at to barn dør av krybbedød i samme familie
    - Ved å anta uavhengighet beregnes

      P(To barn dør)=0.000118*0.000118=1/72 250 000
- Men det er **urimelig å anta uavhengighet**!!! Og dermed blir tallet over for lavt

- Både genetiske og miljø-faktorer kan øke sannsynligheten i en familie, noe som gjør at *sannsynligheten for at et nytt barn dør av krybbedød er høyere dersom man vet at det allerede har vært en krybbedød i samme familie*

Addisjonsregelen og produktregelen

- **Addisjonsregelen**: Hvis *A* og *B* er **disjunkte**

$$P(A \text{ eller } B) = P(A) + P(B)$$

- **Produktregelen**: Hvis *A* og *B* er **uavhengige**

$$P(A \text{ og } B) = P(A) \times P(B)$$

- Observer: **Disjunkte hendelser *A* og *B* kan ikke være uavhengige!**

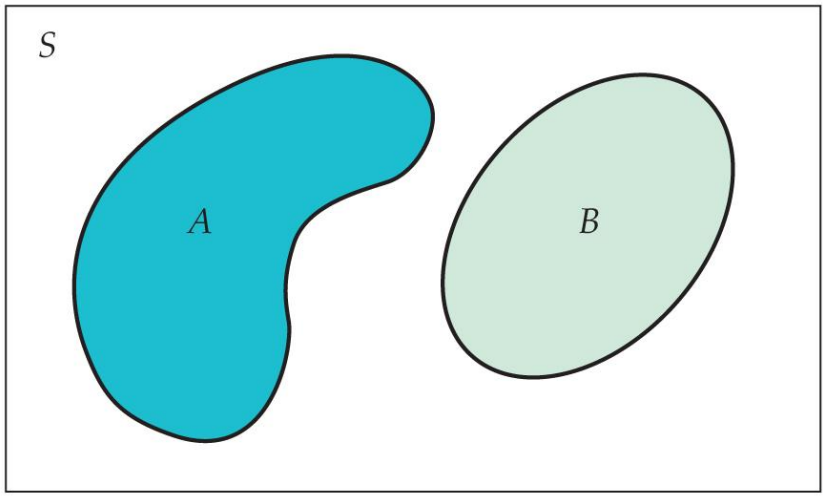


Figure 4.2
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

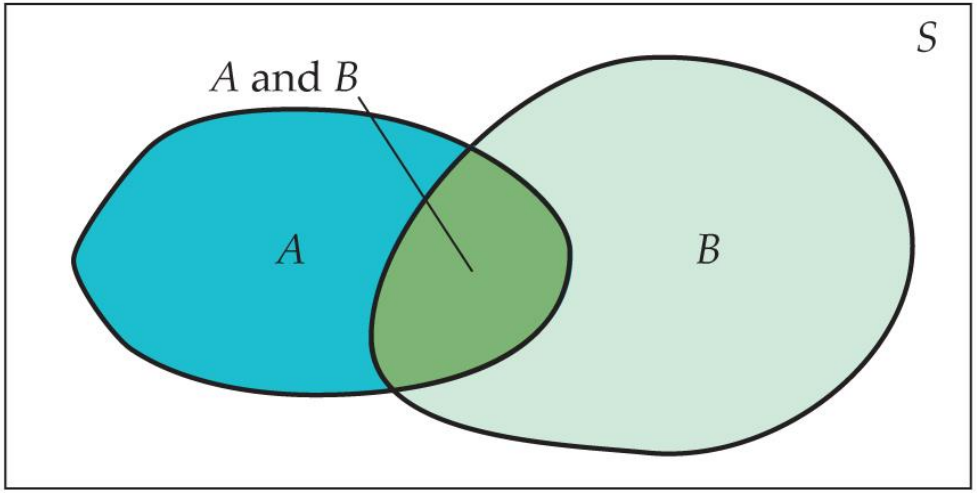


Figure 4.4
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

- Og uavhengige hendelser kan ikke være disjunkte!

- Hvis *A* og *B* er **disjunkte** kan de **ikke være uavhengige**: Hvis du vet at *A* skjer, kan ikke *B* skje. Kunnskap om *A* skjer eller ikke påvirker altså sannsynligheten for om *B* skjer.
- Vi kan se det matematisk fra produktregelen for hendelser *A* og *B* med P(*A*) > 0 og P(*B*) > 0 :
  - Hvis *A* og *B* er **disjunkte** kan de ikke begge inntreffe, og dermed P(*A* og *B*) = 0
  - Men P(*A*) x P(*B*)>0 når både P(*A*) > 0 og P(*B*) > 0, dermed kan ikke produktregelen gjelde, og *A* og *B* er **ikke uavhengige**

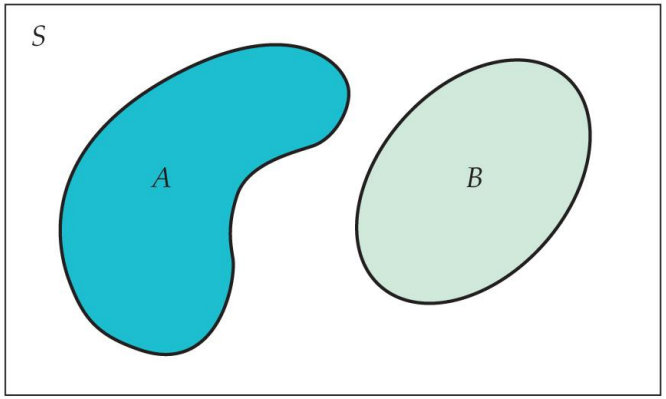


Figure 4.2
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e. © 2017 W. H. Freeman and Company

4.3 Tilfeldige variable

- Tilfeldige variable
- Diskrete tilfeldige variable
- Kontinuerlige tilfeldige variable
- Normalfordelinger som sannsynlighetsfordelinger

En numerisk variabel som beskriver utfallene til en tilfeldig prosess kalles en tilfeldig variabel

En sannsynlighetsmodell beskriver både de mulige utfallene av en tilfeldig prosess, og sannsynligheten for at hver av disse vil inntreffe

Utfallsrom består ikke nødvendigvis består av tall, for eksempel ved kast av tre mynter:

S = {KKK,KKM,KMK,MKK,KMM,MKM,MMK,MMM}

Men i statistikk er vi er oftest interesserte i numeriske størrelser, for eksempel den tilfeldige variabelen ‘antall kron i de tre myntkastene’. Utfallsrommet for den tilfeldige variabelen blir S={0,1,2,3}.

Sannsynlighetsmodellen til en tilfeldig variabel er sannsynlighetsfordelinga dens

En **sannsynlighetsmodell** beskriver de mulige utfallene av en tilfeldig prosess og sannsynligheten for at hver av disse vil inntreffe.

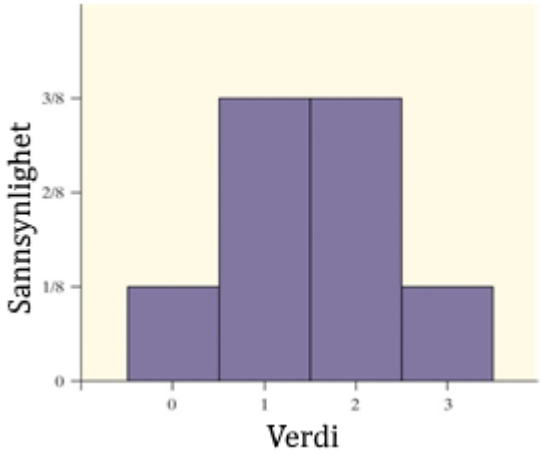
En **tilfeldig variabel** er en numeriske variabel som beskriver utfallene av en tilfeldig prosess.

Sannsynlighetsfordelinga til en tilfeldig variabel angir mulige verdier og deres sannsynligheter.

Eksempel: En mynt kastes tre ganger.
Definer X = antall mynt

X = 0: KKK
X = 1: MKK KMK KKM
X = 2: MMK MKM KMM
X = 3: MMM

Verdi	0	1	2	3
Sannsynlighet	1/8	3/8	3/8	1/8



Notasjon for tilfeldige variable

- **Tilfeldige variable** betegnes vanligvis med store bokstaver som *X* og *Y*.
- De tilsvarende små bokstavene, *x* og *y*, betegner de **faktiske (observerte) utfallene**
- Unntak:
  - $\bar{x}$ for gjennomsnitt
  - *s* for standardavvik
  - $\hat{p}$ for andel

  For disse unntakene brukes betydningene (tilfeldig variabel eller utfallet av den) om hverandre, selv om små bokstaver brukes.

En diskret tilfeldig variabel X har et antall mulige verdier.

Det finnes to hovedtyper av tilfeldige variable: *diskrete* og *kontinuerlige*. Hvis vi kan liste opp alle mulige utfall til en tilfeldig variabel og gi sannsynligheten til hver av dem, har vi en **diskret tilfeldig variabel**.

Sannsynlighetsfordelinga til en diskret tilfeldig variabel *X* gir verdiene x_i og deres sannsynligheter p_i:

Verdi:	x_1	x_2	x_3	…
Sannsynlighet:	p_1	p_2	p_3	…

Sannsynlighetene p_i må tilfredsstille to krav:

1. Hver sannsynlighet p_i er et tall mellom 0 and 1.

2. Summen av sannsynlighetene er 1.

For å finne sannsynlighetene for en hendelse, summer sannsynlighetene p_i for de verdiene x_i som er med i hendelsen.

Vi kan fremstille diskrete sannsynlighets-fordelinger grafisk med sannsynlighets-histogrammer.

Eksempler: Diskret tilfeldig variabel *X* beskriver første siffer i et tall.

(a) Sannsynlighetene er like for alle verdier

(b) Sannsynlighetene følger Benfords lov

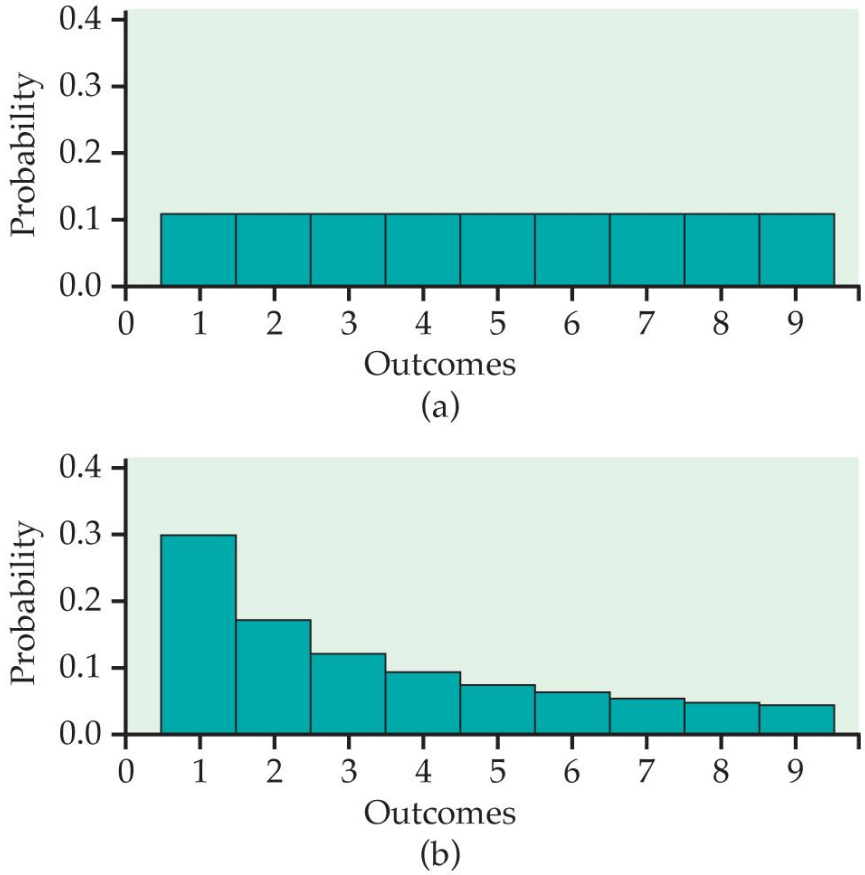


Figure 4.5
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e,

Eksempel: Antall kron i 4 myntkast

Antar: (1) kron (H) og mynt (T) er like sannsynlige utfall i hvert kast, og
(2) kastene er uavhengige.

La *X* være antall kron i fire myntkast. Utfallsrommet er *S*={0,1,2,3,4}

Resultatet etter fire kast er en sekvens av kron og mynt som for eksempel HHTH. Det er i alt 16 mulige slike sekvenser, og sannsynligheten for hver sekvens er *antall gunstige/antall mulige*=1/16

Sannsynligheten for hvert av utfallene til *X* er *ikke like sannsynlige*

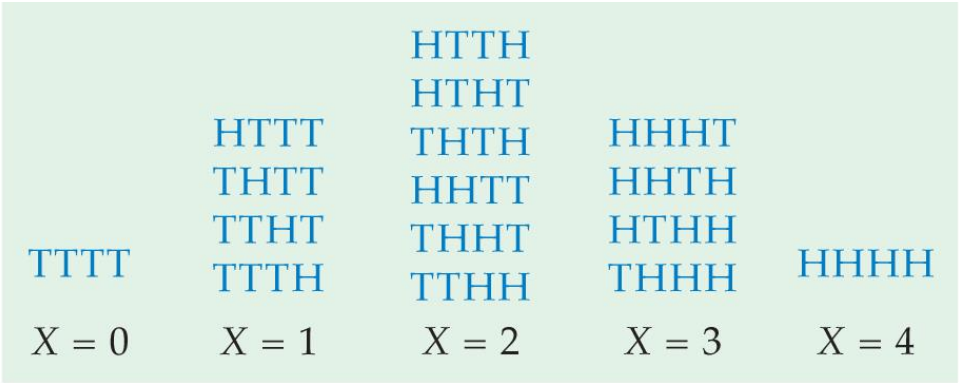


Figure 4.6
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e,

1 flervalgsspørsmål

Eksempel: Antall kron i fire myntkast.

Antar (1) kron (H) og mynt (T) er like sannsynlige utfall i hvert kast og (2) kastene er uavhengige.

La *X* være antall kron i fire myntkast. Utfallsrommet er *S*={0,1,2,3,4}

Sannsynligheten for hvert av utfallene til *X* er *ikke like sannsynlige*

- P(X=0) = P(X=4) = antall gunstige/antall mulige = 1/16 = 0.0625
- P(X=2) = antall gunstige/antall mulige = 6/16 = 0.375
- P(X=1) = P(X=3) = antall gunstige/antall mulige = 4/16 = 0.25

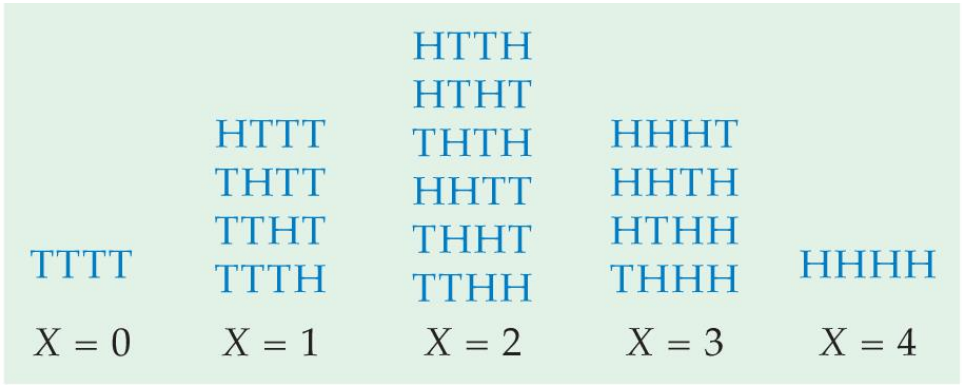


Figure 4.6
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e,
© 2017 W. H. Freeman and Company

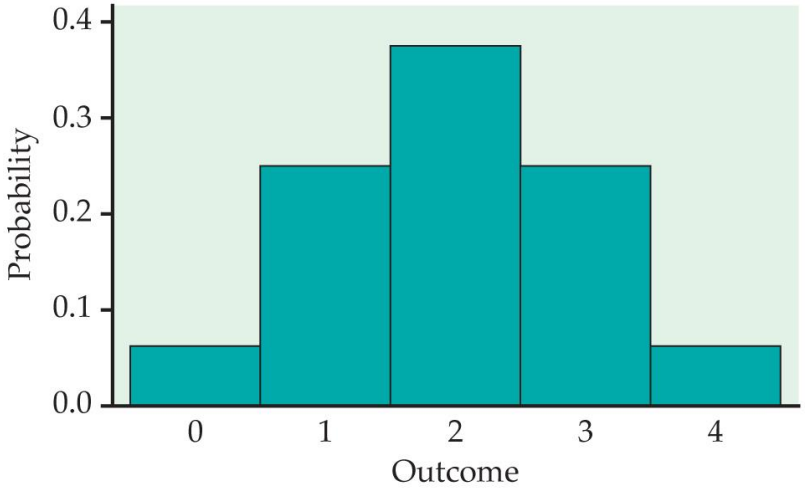


Figure 4.7
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e,

Bokstavkarakterer kan angi tilfeldig variabel

Eksempel

Anta at fordelingen til karakterene i STK1000 er:
32% får A, 42% får B, 18% får C, 3% får D, 1% får E og 4% stryker (F).
Utfallsrommet av bokstavkarakterer S={A,B,C,D,E,F} kan lett skrives om til en tilfeldig variabel X med et numerisk utfallsrom S={5,4,3,2,1,0}, hvor 5 tilsvarer A, 4 tilsvarer B

Her er sannsynlighetsfordelingen til den diskrete tilfeldige variabelen X, hvor 5 tilsvarer karakteren A, 4 tilsvarer karakteren B, osv.:

Verdi:	0	1	2	3	4	5
Sannsynlighet:	0.04	0.01	0.03	0.18	0.42	0.32

Hva er sannsynligheten for hendelsen at en tilfeldig valgt student får B eller bedre?

$$P(X \geq 4) = P(X = 4) + P(X = 5)$$
$$= 0.42 + 0.32 = 0.74.$$

En kontinuerlig tilfeldig variabel Y tar verdier i et intervall av tall.

Diskrete tilfeldige variabler kommer vanligvis fra situasjoner hvor man teller noe. Situasjoner der man måler noe resulterer ofte i en **kontinuerlig tilfeldig variabel.**

Sannsynlighetsfordelingen til en tilfeldige variabel Y blir ofte omtalt som en **tetthetsfunksjon**. Sannsynligheten for en hendelse er her arealet under tetthetsfunksjonen i det aktuelle området.

Sannsynlighetsmodellen til en diskret tilfeldig variabel *X* gir en sannsynlighet mellom 0 og 1 til hver mulig verdi *X* kan ha.

En kontinuerlig tilfeldig variabel *Y* har *uendelig mange* mulige verdier. Enkeltutfall har sannsynlighet 0 i alle kontinuerlige sannsynlighetsmodeller. Det er kun *intervaller* av verdier som har sannsynlighet større enn 0.

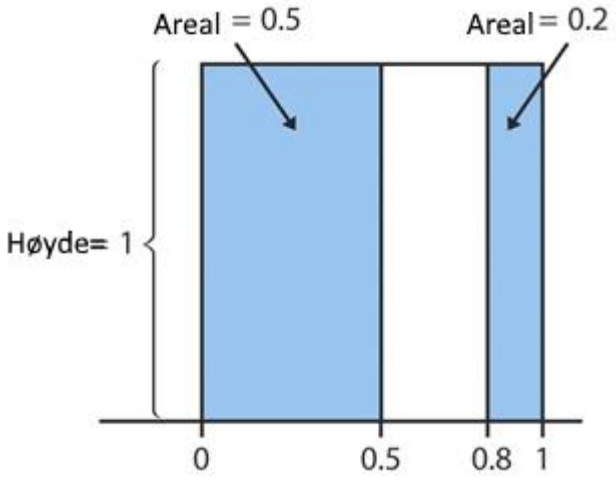
Kontinuerlige sannsynlighetsmodeller

La oss si at vi vil velge et tilfeldig tall mellom 0 og 1. Siden vi kan ta absolutt hvilket som helst reelt tall, kan vi ikke ta sannsynligheten til ett enkelt tall. Hvis hvert enkelt tall hadde en sannsynlighet, og det er uendelig mange tall, ville summen av sannsynlighetene deres bli uendelig stor og ikke 1. Derfor må vi heller regne med intervaller.

En **kontinuerlig sannsynlighetsmodell** gir sannsynligheter som arealer under en tetthetsfunksjon. Sannsynligheten for en hendelse er arealet under tetthetsfunksjonen over verdiene som utgjør hendelsen.

Eksempel: Uniform fordeling mellom 0 og 1. Hva er sannsynligheten for å få et tilfeldig tall som er mindre enn eller lik 0.5 eller større enn 0.8.

$$P(X \leq 0.5 \text{ eller } X > 0.8)$$
$$= P(X \leq 0.5) + P(X > 0.8)$$
$$= 0.5 + 0.2$$
$$= 0.7$$



Uniform fordeling

Kontinuerlige sannsynlighetsmodeller 2

En **kontinuerlig sannsynlighetsmodell** gir sannsynligheter som arealer under en tetthetsfunksjon. Sannsynligheten for en hendelse er arealet under tetthetsfunksjonen over verdiene som utgjør hendelsen.

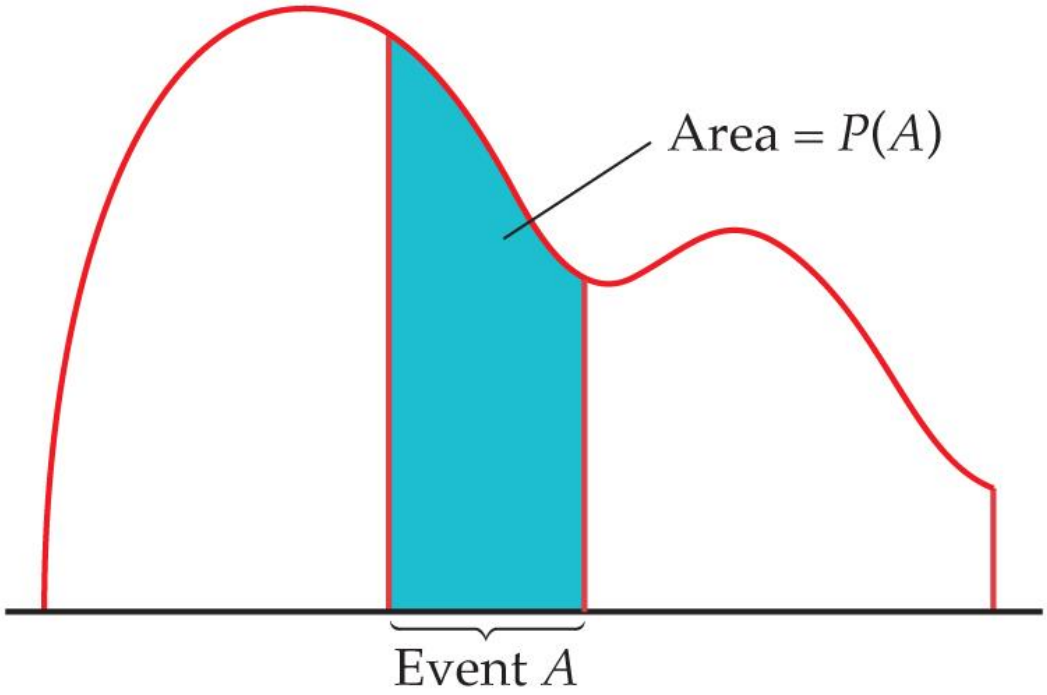


Figure 4.10
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

1 flervalgsspørsmål

Normalfordelinger er sannsynlighetsfordelinger

Vi har sett i kapittel 1 hvordan vi finner arealer under en normalkurve, og det representerte andelen observasjoner i det tilsvarende verdiområdet.

Nå vil vi snakke om disse arealene som sannsynligheter for at en normalfordelt tilfeldig variabel *X* ligger innenfor det spesifiserte verdiområdet.

Vi kan finne sannsynlighetene til intervaller av utfall ved å bruke standard normal-sannsynligheter fra Tabell A (eller ved å bruke R).

Vi **standardiserer** en $N(\mu,\sigma)$-fordelt tilfeldig variabel *X* ved å standardisere

$$Z = \frac{X - \mu}{\sigma}$$

Da er *Z* standard normalfordelt ($N(0,1)$-fordelt)

Normalfordelinga angir sannsynligheter til intervaller av verdier

Kvinners høyde er normalfordelt med gjennomsnitt på 164.0 cm og standardavvik på 6.35 cm. Hvis vi velger en kvinne tilfeldig, hva er da sannsynligheten for at høyden hennes (som vi kaller *X*) er mellom 172.9 og 178.0? Dette vil si hva er *P*(172.9 < *X* < 178.0)? Siden kvinnen blir valgt tilfeldig, er *X* en tilfeldig variabel.

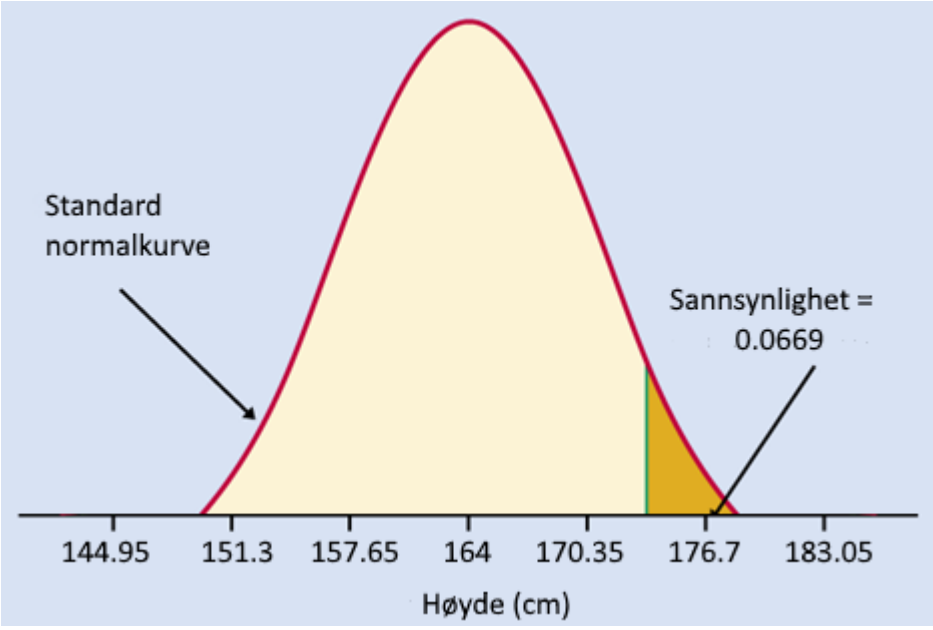
$$z = \frac{(x - \mu)}{\sigma}$$

Vi regner ut z-scoren for 172.9 og 178.0.

For *x* = 172.9 cm, $z = \frac{(172.9-164)}{6.35} = 1.4$

For *x* = 178.0 cm, $z = \frac{(178.0-164)}{6.35} = 2.2$

P(173 < *X* < 178) = P(1.4 < *Z* < 2.2) = P(*Z* < 2.2) - P(*Z* < 1.4) = (se tabell neste side)



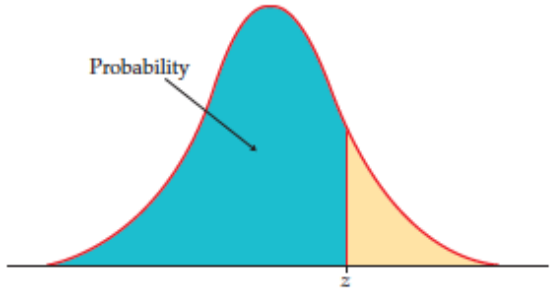


Table entry for z is the area under the standard Normal curve to the left of z.

TABLE A **Standard Normal Probabilities (continued)**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Normalfordelinga angir sannsynligheter til intervaller av verdier

Kvinners høyde er normalfordelt med gjennomsnitt på 164.0 cm og standardavvik på 6.35 cm. Hvis vi velger en kvinne tilfeldig, hva er da sannsynligheten for at høyden hennes (som vi kaller *X*) er mellom 172.9 og 178.0? Dette vil si hva er *P*(172.9 < *X* < 178.0)? Siden kvinnen blir valgt tilfeldig, er *X* en tilfeldig variabel.

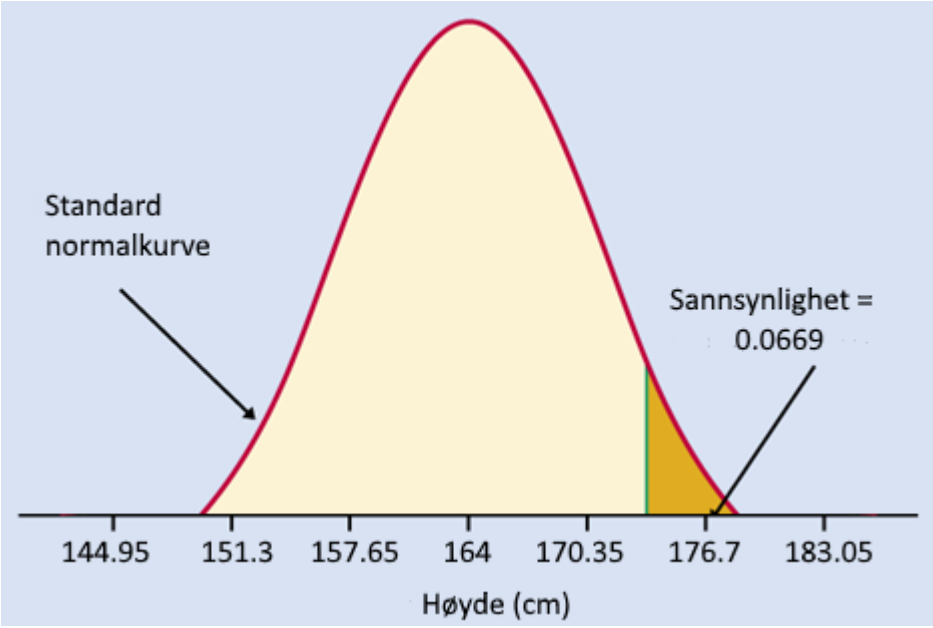
$$z = \frac{(x - \mu)}{\sigma}$$

Vi regner ut z-scoren for 172.9 og 178.0.

For *x* = 172.9 cm, $z = \frac{(172.9-164)}{6.35} = 1.4$

For *x* = 178.0 cm, $z = \frac{(178.0-164)}{6.35} = 2.2$

P(173 < *X* < 178) = P(1.4 < *Z* < 2.2)
= P(*Z* < 2.2) - P(*Z* < 1.4) = (se tabell neste side)
0.9861-0.9192 = 0.0669



4.4 Forventningsverdi og varians til tilfeldige variabler

- Forventningsverdien til en tilfeldig variabel
- Statistisk estimering og store talls lov
- Regler for forventning
- Variansen til en tilfeldig variabel
- Regler for varians og standardavvik

Form, senter og spredning for en diskret tilfeldig variabel

Når vi analyserer sannsynlighetsfordelingen til diskrete tilfeldige variable, gjør vi *det samme som med kvantitative data*:
Vi beskriver formen, og angir sentralmål og spredning.

Forventningsverdien til en diskret tilfeldig variabel er et vektet gjennomsnitt av mulige utfall; hvert utfall er vektet med sin sannsynlighet.

Forventningsverdien til en diskret tilfeldig variabel

La oss si at *X* er en diskret tilfeldig variabel med sannsynlighetsfordeling lik

Verdi:	x_1	x_2	x_3	…
Sannsynlighet:	p_1	p_2	p_3	…

For å finne **forventningsverdien** til *X*, multipliser hver mulig verdi med sin sannsynlighet og summer produktene:

$$\mu_x = E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \cdots$$
$$= \sum x_i p_i$$

Eksempel: Helsa til nyfødte

Sannsynlighetsfordelingen til X = Apgar-score for nyfødte er vist nedenfor:

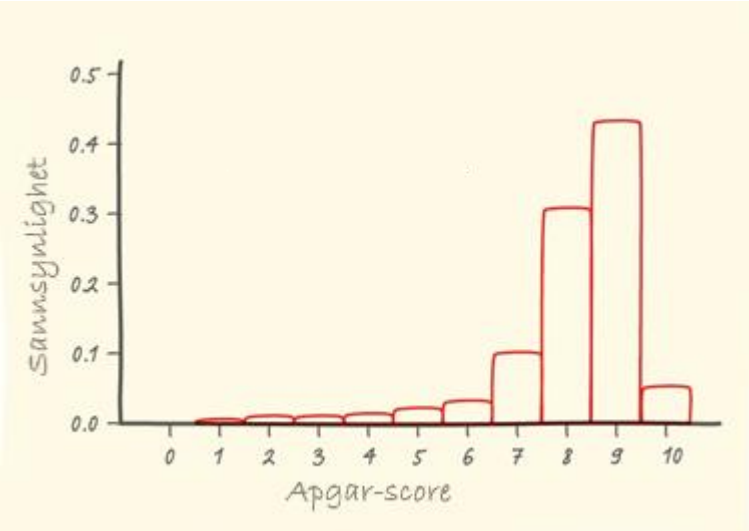
a. Vis at sannsynlighetsfordelingen til X er en legitim sannsynlighetsfordeling.

b. Lag et histogram av sannsynlighetsfordelingen. Beskriv hva vi ser.

c. Apgar-score på 7 eller høyere indikerer et friskt barn. Hva er $P(X \geq 7)$?

Value:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

(a) Alle sannsynligheter er mellom 0 og 1, samt at de summeres til 1. Dette gjør den til en legitim sannsynlighets-fordeling.



(c) $P(X \geq 7) = .908$. Det er 91% sannsynlighet for at en tilfeldig valgt baby er frisk.

(b) den venstreskjeve formen på fordelingen antyder at en tilfeldig valgt nyfødt kommer til å ha en Apgar-score i den høye enden av skalaen. Det er bare en liten sannsynlighet for å få en baby med score på 5 eller lavere.

Hva er typisk Apgar-score?

La oss se på den tilfeldige variabelen *X* = Apgar-score.

Regn ut forventninga til den tilfeldige variabelen *X* og tolk den i sin kontekst.

Verdi:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sannsynlighet:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

$$\mu_x = E(X) = \sum x_i p_i$$
$$= (0)(0.001) + (1)(0.006) + (2)(0.007) + \cdots + (10)(0.053)$$
$$= 8.128$$

Forventet Apgar-score for en tilfeldig valgt nyfødt er 8.128. Dette er det gjennomsnittet Apgar-scoren får i det lange løp når vi tar Apgar-scoren av veldig mange tilfeldig valgte babyer.

Merk: Forventningsverdien trenger ikke å være en mulig verdi av *X* eller et heltall! Det er gjennomsnittet i det lange løp etter mange repetisjoner.

Forventningsverdien til en kontinuerlig tilfeldig variabel

For **kontinuerlig tilfeldige variable** defineres forventninga ved bruk av tetthetskurven. Den kan beskrives som **balansepunktet**.

Eksakt beregning av forventningsverdien gjøres matematisk ved et integral. Hvis tettheten til den kontinuerlig tilfeldige variabelen *X* beskrives av funksjonen *f(x)*, gis forventninga til *X* av

$$\mu = \mu_x = \int_{-\infty}^{\infty} x\, f(x) dx$$

Regneregler for forventning

Regel 1: Hvis X er en tilfeldig variabel og a og b er gitte tall, så er

$$\mu_{a+bX} = a + b\mu_X.$$

Regel 2: Hvis X og Y er tilfeldige variabler, så er

$$\mu_{X+Y} = \mu_X + \mu_Y.$$

Regel 3: Hvis X og Y er tilfeldige variabler, så er

$$\mu_{X-Y} = \mu_X - \mu_Y.$$

Eksempel: Noen gresshopper som lever på en åker har forventa lengde på 1.2 tommer. Hva er forventninga i centimeter?
Det er 2.54 cm i en tomme, så forventninga i tommer må ganges med 2.54: 1.2 x 2.54 = 3.05 cm.
(Merk at vi brukte regel 1 med b = 1.2 og a = 0 her.)

Variansen til en tilfeldig variabel

Siden vi bruker forventninga som sentralmål for en diskret tilfeldig variabel, bruker vi standardavvik som mål på spredning.

Variansen til en diskret tilfeldig variabel

La oss si at *X* er en diskret tilfeldig variabel med sannsynlighetsfordeling

Verdi:	x_1	x_2	x_3	…
Sannsynlighet:	p_1	p_2	p_3	…

og μ_X er forventninga til *X*.

Variansen til *X* er da

$$\sigma_X^2 = (x_1 - \mu_X)^2\, p_1 + (x_2 - \mu_X)^2\, p_2 + (x_3 - \mu_X)^2\, p_3 + \cdots$$
$$= \sum (x_i - \mu_X)^2\, p_i$$

For å få **standardavviket til en tilfeldig variabel**, ta kvadratroten av variansen.

For en kontinuerlig tilfeldig variabel må vi igjen bruke integralregning for å finne variansen

Eksempel: Hvor variable er Apgar-scorene?

Vi har den tilfeldige variabelen X = Apgar-score

Regn ut standardavviket til den tilfeldige variabelen *X* og tolk dette inn i konteksten til variabelen.

Verdi:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sannsynlig-het:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 \, p_i$$

$$= (0 - 8.128)^2(0.001) + (1 - 8.128)^2(0.006) + \cdots + (10 - 8.128)^2(0.053)$$

$$= 2.066$$

Varians

$$\sigma_X = \sqrt{2.066} = 1.437$$

Standardavviket til *X* er 1.437. Dette vil da si noe om hvor mye vi regner med at Apgar-score for en tilfeldig nyfødt er forskjellig fra forventningsverdien.

Regneregler for varians

Regler for varianser og standardavvik

Regel 1: Hvis X er en tilfeldig variabel og a og b er gitte tall, så er

$$\sigma^2_{a+bX} = b^2\sigma^2_X.$$

Regel 2: Hvis X og Y er *uavhengige* tilfeldige variable, så er

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y,$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y.$$

Regel 3: Hvis X og Y har korrelasjon ρ, så er

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y,$$

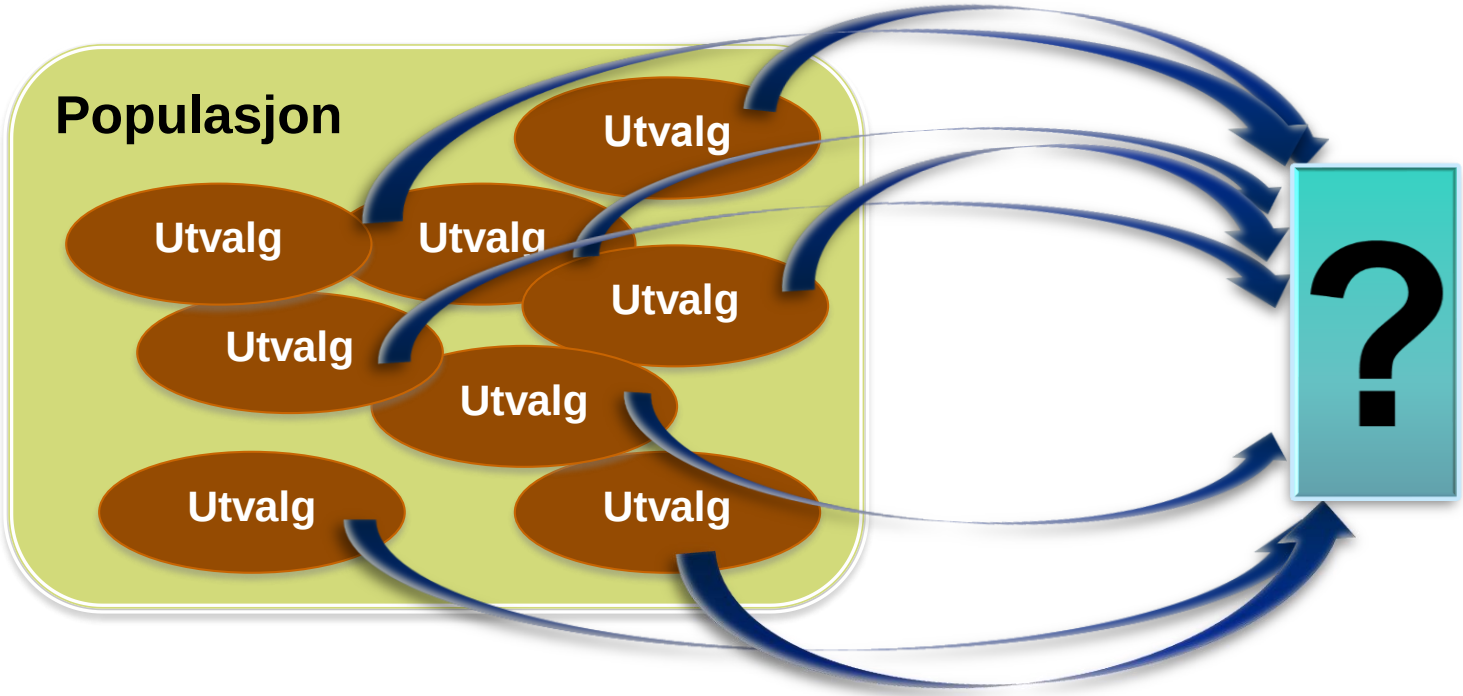
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y.$$

Verdien til observatoren vil variere når vi gjentar trekning av tilfeldige utvalg

Vi vil prøve å estimere en ukjent forventningsverdi μ. Vi kan trekke et tilfeldig utvalg og basere estimatet på utvalgsgjennomsnittet. Men da vil forskjellige utvalg antageligvis gi oss forskjellige utvalgsgjennomsnitt.

Dette kalles **utvalgsvariabilitet**: Verdien til observatoren vil variere når vi gjentar tilfeldige utvalg.

Hva ville skjedd dersom vi tok mange utvalg?



Store talls lov

Hvordan kan $\bar{x}$ være et godt estimat for μ? Man vil jo få ulike verdier på $\bar{x}$ av forskjellige tilfeldige utvalg.

Hvis vi tar større og større utvalg, vil observatoren $\bar{x}$ garantert komme nærmere og nærmere parameteren μ.

Trekk uavhengige observasjoner tilfeldig fra en populasjon med forventning μ. **Store talls lov** sier da at, når antallet observasjoner øker, vil utvalgsgjennomsnittet av observasjonene nærme seg mer og mer forventninga μ til populasjonen.

4.5 Generelle sannsynlighetsregler

- Sannsynlighetsregler
- Generelle addisjonsregler
- Betinga sannsynlighet
- Generelle produktregler
- Trediagrammer
- Bayes' regel
- Uavhengighet

Lovene som styrer sannsynlighetsregninga

Til nå har vi konsentrert oss om tilfeldige variable og deres fordelinger.
Lover og regler for sannsynlighetsregning:

Regel 1. Sannsynligheten $P(A)$ for hendelse A tilfredsstiller $\mathbf{0 \le P(A) \le 1}$.

Regel 2. Hvis S er utvalgsrommet i en sannsynlighetsmodell, så er $\mathbf{P(S) = 1}$.

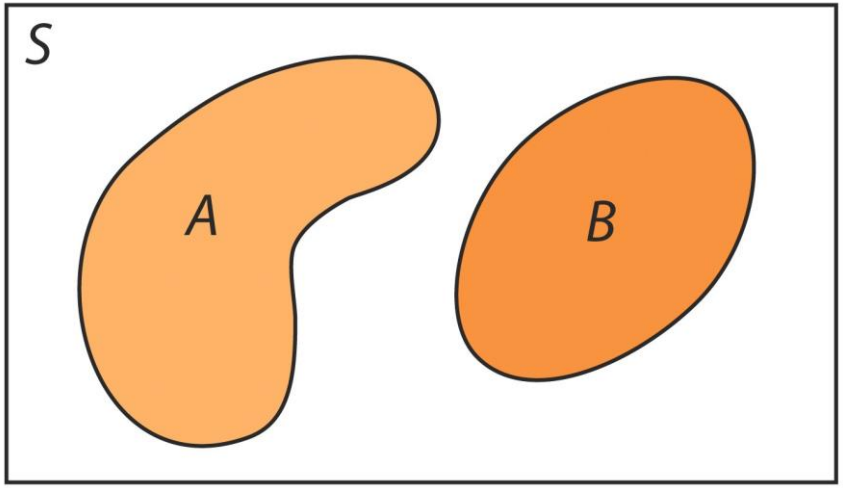
Regel 3. Hvis A og B er **disjunkte**, $\mathbf{P(A \text{ eller } B) = P(A) + P(B)}$.

Regel 4. For enhver hendelse A, $\mathbf{P(A^C) = 1 - P(A)}$.

Regel 5. Hvis A og B er **uavhengige**, $\mathbf{P(A \text{ og } B) = P(A)P(B)}$.

Et intuitivt verktøy: Venn-diagram

To disjunkte hendelser:



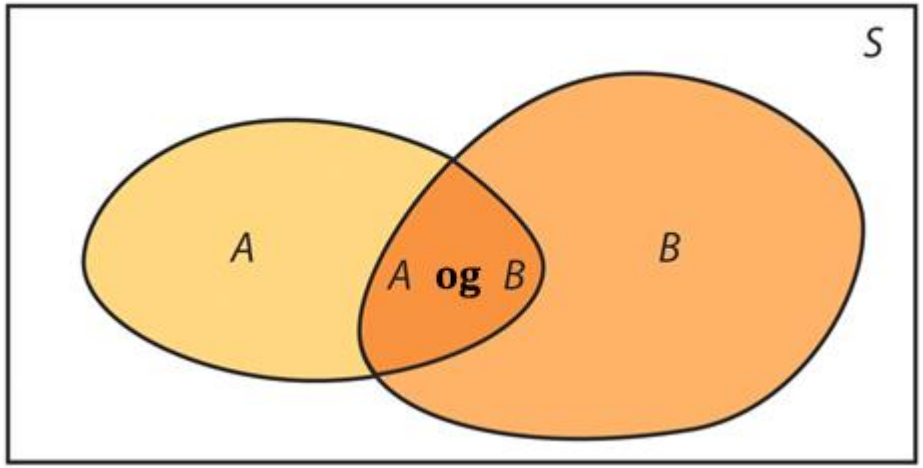
Vi kan utvide addisjonsregelen for disjunkte hendelser

Hvis *A*, *B,* og *C* er **disjunkte** slik at ingen av dem har noen felles utfall, så er

***P*(en eller flere av *A*, *B, C*) = *P*(*A*) + *P*(*B*) + *P*(*C*).**

Den generelle addisjonsregelen

To hendelser som ikke er disjunkte, og hendelsen "*A* og *B*" som inneholder utfall de har felles:



Addisjonsregelen for unioner av to hendelser

For to hendelser *A* og *B*:

$$P(A \text{ eller } B) = P(A) + P(B) - P(A \text{ og } B)$$

Merk: Dersom A og B er disjunkte er siste leddet P(A og B)=0, og dermed følger addisjonsregelen for disjunkte hendelser

Den generelle addisjonsregelen, eks:

Eksempel: Anta at 40% av voksne får nok søvn, 46% trener jevnlig, og 24% både får nok søvn og trener jevnlig. Hva er sannsynligheten for at en voksen person får nok søvn og/eller trener regelmessig?

La A=«nok søvn», B=«trener jevnlig». Vi vet P(A)=0.4, P(B)=0.46 og P(A og B)=0.24.

Da er

P(A eller B)=P(A)+P(B)-P(A og B)
= 0.40+0.46-0.24=0.62

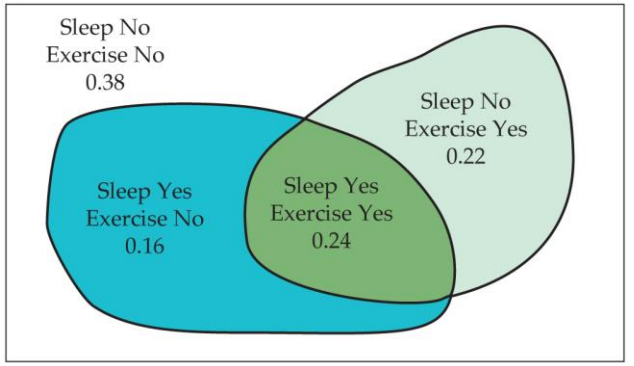


Figure 4.18
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Hvis vi tegner opp det vi vet i Venn-diagrammet,
ser vi at P(A og B^c)=0.4-0.24=0.16 og P(B og A^c)=0.46-0.24=0.22

Dessuten er sannsynligheten for (A eller B)^c (dvs verken nok søvn eller trener jevnlig) = 1-0.62=0.38

Betinga sannsynlighet

Sannsynligheten for en hendelse kan endre seg hvis vi vet at en annen hendelse har skjedd. Dette er et viktig aspekt i mange situasjoner der man bruker sannsynlighetsregning

Når vi skal prøve å finne sannsynligheten for en hendelse på bakgrunn av at en annen hendelse har skjedd, prøver vi å finne ut en **betinga sannsynlighet**.

Sannsynligheten for en hendelse skjer *gitt* at en annen hendelse har allerede skjedd kalles en **betinga sannsynlighet**.

Når $P(A) > 0$, så finner vi sannsynligheten for at hendelsen B skjer *gitt* at hendelsen A har skjedd, ved formelen

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

Alle sannsynligheter, også betinga, kan bli funnet fra tildelinga av sannsynligheter for hendelser

Definisjonen av betinga sannsynlighet kan skrives om til en regel for sannsynligheten for at begge hendelsene skjer.

Sannsynligheten for at begge hendelsene *A* og *B* skjer kan finnes ved hjelp av den **generelle produktregelen**:

$$P(A \text{ og } B) = P(A) \bullet P(B \mid A)$$

hvor $P(B \mid A)$ er den betingede sannsynligheten for at hendelsen *B* skjer gitt at hendelsen *A* allerede har skjedd.

Eksempel: Den generelle produktregelen

Eksempel:

- Kortspill med en vanlig kortstokk (52 kort, av dem 13 hjerter, 13 ruter, 13 kløver og 13 spar)
- 11 kort er allerede trukket fra kortstokken, hvorav 4 ruter
- To kort til skal trekkes. Hva er sannsynligheten for at begge disse er ruter?
  - Definerer
    - hendelse A: «første kort er ruter»
    - hendelse B: «andre kort er ruter»
  - P(A) = gunstige/mulige = (13-4)/(52-11) = 9/41
  - P(B | A) = gunstige/mulige = (13-5)/(52-12) = 8/40
  - P(A og B) = P(A) x P(B | A) = (9/41) x (8/40) = 0.044

Den utvida produktregelen

Snittet (‘intersection’ på engelsk) til en samling hendelser er hendelsen at alle hendelsene inntreffer.

For å utvide produktregelen til sannsynligheten at alle av mange hendelser inntreffer, er nøkkelen å bruke produktregelen flere ganger, ved først å betinge en hendelse på at alle de andre hendelsene inntreffer, osv.

For eksempel, for snittet av tre hendelser blir produktregelen

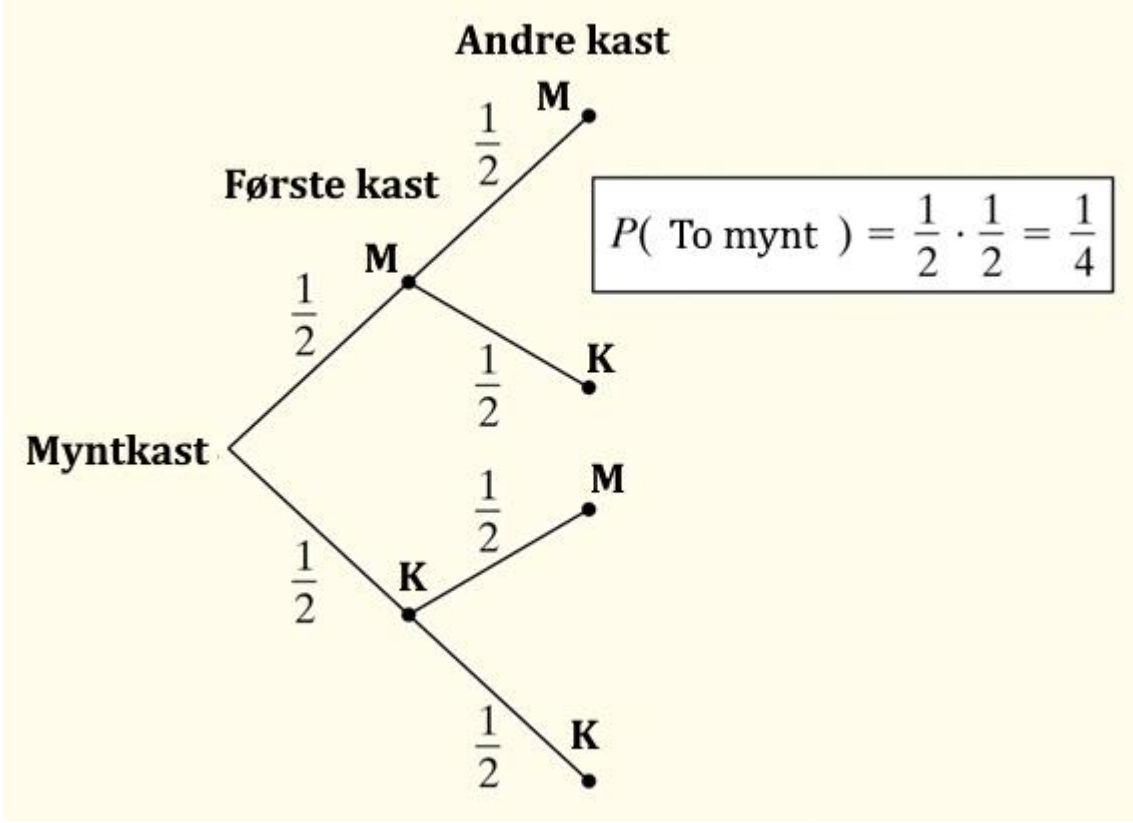
$$P(A \text{ og } B \text{ og } C) = P(A \text{ og } B) \cdot P(C \mid A \text{ og } B)$$
$$= P(A) \cdot P(B \mid A) \cdot P(C \mid A \text{ og } B)$$

Trediagram

Oppgaver i sannsynlighet krever ofte at vi må kombinere to eller flere av de grunnleggende reglene til mer innfløkte utregninger. Én måte å modellere tilfeldigheter som skjer etter hverandre, er å lage et **trediagram**.

Vi kaster en mynt to ganger

Hva er da sannsynligheten for å få to mynt?



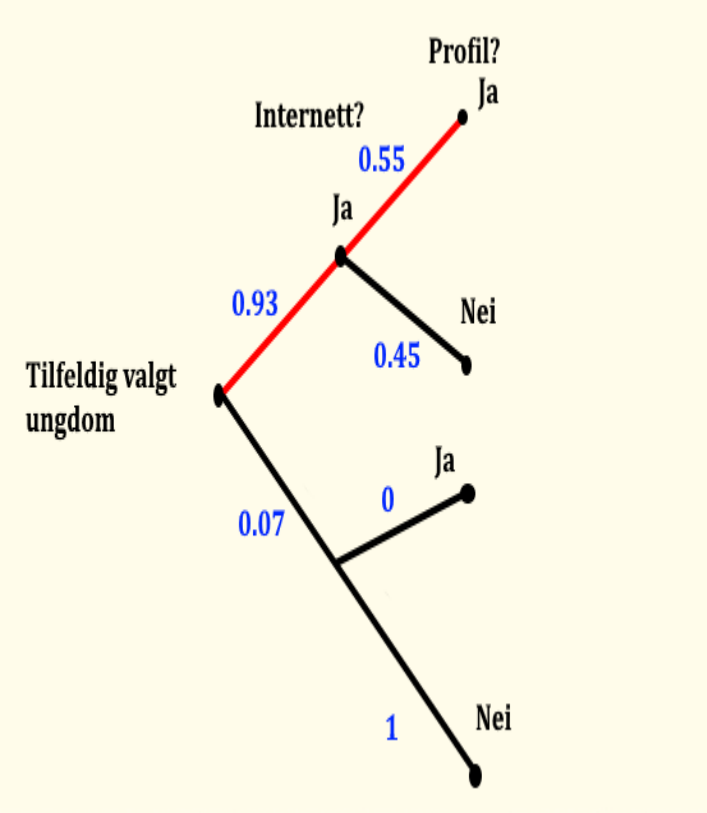
Utfallsrom
MM MK KM KK

Så, *P*(to mynt) = *P*(MM) = 1/4

Eksempel med trediagram

En amerikansk studie fra 2010 viste at 93% av alle ungdommer mellom 12 og 17 brukte internett. Av disse hadde 55% en profil på en eller annen form for sosiale medier.

Hvor stor andel brukte internett og var på sosiale medier?



P(internett) = 0.93

P(profil | internett) = 0.55

P(internett og profil) = P(internett)·P(profil | internett)

$$= (0.93)(0.55)$$
$$= 0.5115$$

51.15% av ungdommene brukte internett *og* hadde en profil.

Bayes' regel: enkleste utgave

Et viktig bruksområde for betinga sannsynlighet er Bayes' regel. Den er en viktig byggestein i mange statistiske problemstillinger, også i mer avanserte modeller som er utenfor pensum i STK1000.

- Først skal vi se på den enkleste formen for to hendelser *A* og *B* slik at $0<P(A)<1$ og $0<P(B)<1$
- Utledning av Bayes' regel for to hendelser: Fra den generelle produktregelen har vi at

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

- Vi kan skrive telleren: $P(A \text{ og } B) = P(B|A)P(A)$, og nevneren (fordi *B=((A og B) eller (A^c og B))*): $P(B) = P(A \text{ og } B) + P(A^c \text{og } B) = P(B|A)P(A) + P(B|A^c)P(A^c)$
- Da får vi Bayes' regel:

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

Bayes' regel: Generell form

Generell form

- Anta at et utfallsrom er delt inn i *k* disjunkte hendelser $A_1, A_2, \ldots, A_k$ — slik at $P(A_1) + P(A_2) + \ldots + P(A_k) = 1$, og at ingen av hendelsene har sannsynlighet lik 0

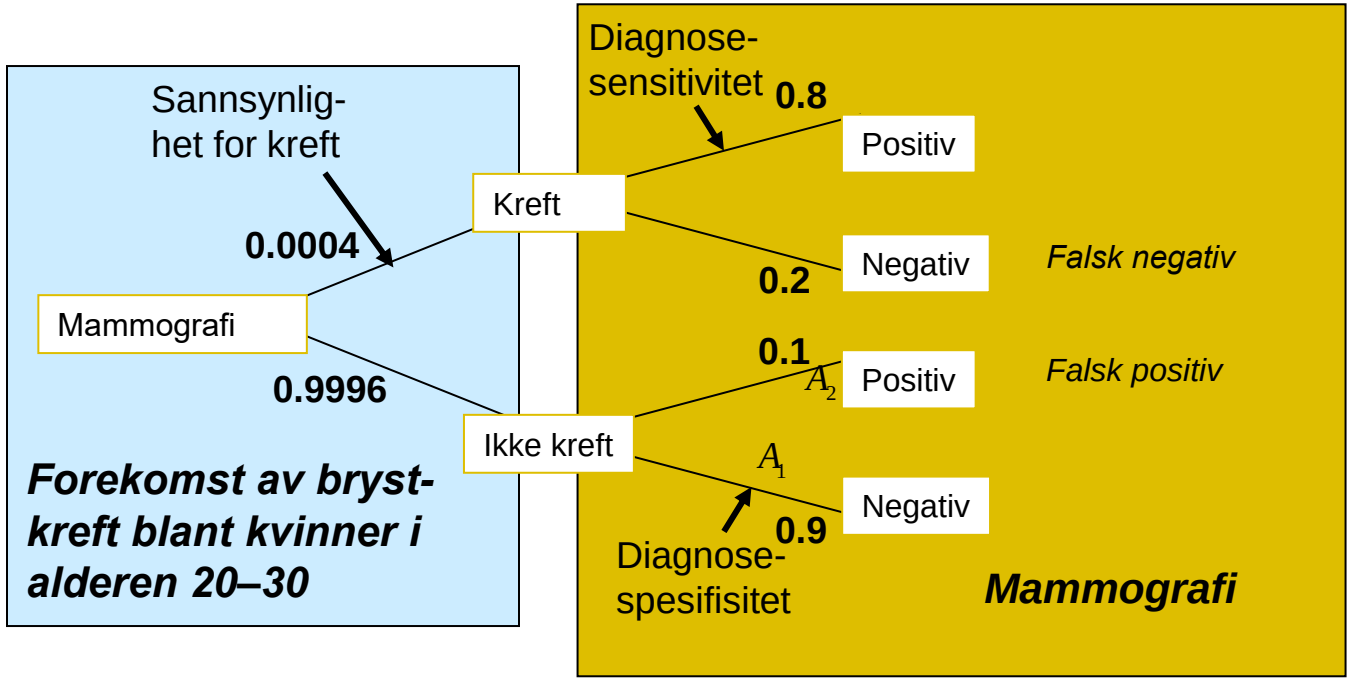
- La *C* være en annen hendelse slik at *P*(*C*) ikke er 0 eller 1. Da er

$$P(A_i|C) = \frac{P(C|A_i)P(A_i)}{P(C|A_1)P(A_1) + P(C|A_2)P(A_2) + \cdots + P(C|A_k)P(A_k)}$$

Det er ofte intuitivt enklere å finne ut svarene ved å bruke et trediagram enn å bruke denne lange formelen.

Bayes' regel eksempel

Hvis en kvinne i tjueårene blir scannet for brystkreft og testene viser positivt testresultat, hva er da sannsynligheten for at hun da faktisk har brystkreft?



A_1 er kreft, A_2 er ikke kreft, og *C* er positivt testresultat.

$$P(kreft|pos) = \frac{P(pos|kreft)P(kreft)}{P(pos|kreft)P(kreft) + P(pos|ikke\ kreft)P(ikke\ kreft)}$$

$$= \frac{0.8 * 0.0004}{0.8 * 0.0004 + 0.1 * 0.9996} \approx 0.003 = 0.3\%$$

Uavhengighet

Hvis to hendelser *A* og *B* ikke påvirker hverandre, og sannsynligheten for det ene ikke endres av at vi vet utfallet på det andre, er hendelsene **uavhengige** av hverandre.

To hendelser *A* og *B* som begge har positiv sannsynlighet er **uavhengige** hvis:

$$P(B|A) = P(B)$$

Vi ser nå at **produktregelen for uavhengige hendelser**

$$\boldsymbol{P(A \text{ og } B) = P(A) \times P(B)}$$

er et spesialtilfelle av den **generelle produktregelen**

$$\boldsymbol{P(A \text{ og } B) = P(A) \times P(B \mid A)}$$

Kapittel 4
Sannsynlighet: Studiet av tilfeldighet (tilbakeblikk)

4.1 Tilfeldighet

4.2 Sannsynlighetsmodeller

4.3 Tilfeldige variabler

4.4 Forventningsverdi og varians til tilfeldige variabler

4.5 Generelle sannsynlighetsregler