

Fra sist:

Sannsynlighetsbegrepet : Abstrakt, knyttet til  $\infty$

Utvælg (det vi ser)  
 $n$

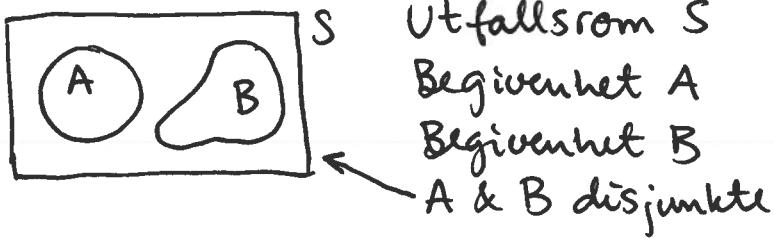
Populasjon (ukjent, det vi  
 $\infty$  ønsker å vite noe om)

$$P(\text{hendelse}) = \frac{\text{Hvor mange ganger det skjer}}{\text{Hvor mange ganger det kunne ha skjedd}}$$

← en andel.  
må være  
mellan 0 og 1.

Værligheit i statistisk forstand :  $P(\underbrace{AB}) = P(A) \cdot P(B)$   
Både A og B

### Sannsynlighetsmodell



$$P(S) = 1$$

$$0 \leq P(A) \leq 1$$

$$P(A) = 1 - P(A^c)$$

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

Hvis A & B disjunkte

Stokastisk variabel  $X$ :

har et sett med verdier ( $x$ ) den kan være,  
og en sannsynlighetsfordeling som forteller hvor sannsynlige  
de ulike verdiene er.

Variabelen skrives med stor bokstav  $X$ , verdiene med  
små bokstaver,  $x$ .

Det finnes diskrete sannsynlighets-fordelinger, og  
kontinuerlige forkorter dette til sh  
sh.-fordelinger. i notatene

Husk at dette er rammeverket for de observasjonene vi gjør,  
så når det finnes kategoriske/diskrete og kontinuerlige  
observasjoner, finnes det tilsvarende sh. fordelinger.

Eks:  $X$  er kjønn på baby, og kodes 0 eller 1. (2)

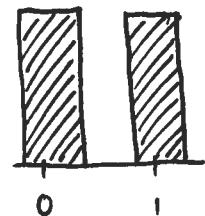
Vi lager en tabell med verdiene til  $X$ , og den tilhørende s. fordelingen:

Verdien til $X$	0	1	
$P(X=x)$			Hva skal stå her?

→ Samsynligheten for at variablen  $X$  har verdien  $x$ .

Når ganger finnes vi s. fordelingen ved å gjøre veldig mange observasjoner (evt. eksperimenter), som her. I følge ssb.no ble det født 118142 barn i 2014-2015, og 60739 var gutter. Derved er  $P(X=0) \approx \frac{60739}{118142} = 0.514$ , og  $P(X=1) \approx 1 - 0.514 = 0.486$

Verdien til $X$	0	1	sum
$P(X=x)$	0.514	0.486	1



Eks:  $X = \#$  barn en norsk kvinne føder

Verdien til $X$	0	1	2	3	4	...	sum
$P(X=x)$							1

kun hentes fra sbb eller mfr

Andre ganger kan vi bruke teori/s. regning til å finne ut hva samsynlighetsfordelingen er.

Eks La  $X = \#$  øyne på terningen i et terningkast

Verdien til $X$	1	2	3	4	5	6	sum
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

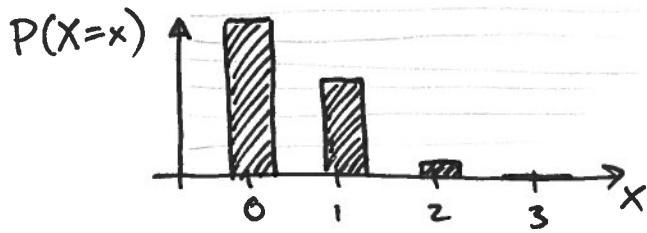
Her bruker vi gunstige/mulige-regelen for å komme fram til  $\frac{1}{6}$  se boka s. 244



Eks:  $X = \#$  kast der terningen gir en 6-er ( $\begin{smallmatrix} \text{:} & \text{:} \\ \text{:} & \text{:} \end{smallmatrix}$ ) på 3 kast

Verdiene til $X$	0	1	2	3	sum
$P(X=x)$	0.579	0.347	0.069	0.005	1
$\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$ ikke 6-er tre ganger på rad, uavhengige kast. Multiplikasjonsregelen s. 245	$\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$ $+ \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6}$ $+ \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$	$\frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$ $+ \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$ $+ \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$	$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$ $3$ 6-er på rad		
$P(6\text{-er}) = 1 - P(\text{ikke } 6)$	tre måter å få én 6-er på tre kast	tre måter å få to 6-er på tre kast			

Samsværtighetsfordelingen kan også tegnes:



Denne siste fordelingen er et eksempel på en fordeling som kan generaliseres, slik at vi får en formel for  $P(X=x)$ :

Gjør  $n$  uavhengige forsök / observasjoner, der hvert forsök har 2 mulige utfall (oftest kalt "sukcess" og "fråskø") og der  $P(\text{sukcess}) = p$  er lik i hvert forsök.  
La  $X = \#$  suksesser. Da er  $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

Vi kaller slike forsök (enhetsforsökene) Bernoulliforsök, og hele rekka for en binomial forsøksordning.

Vi sier at  $X$  har en binomial fordeling med parameter neq  $p$ ,  $X \sim \text{bin}(n, p)$

Eks:  $X = \#$  kast der vi får 6-er på 3 kast (som over).  $X \sim \text{bin}(3, \frac{1}{6})$

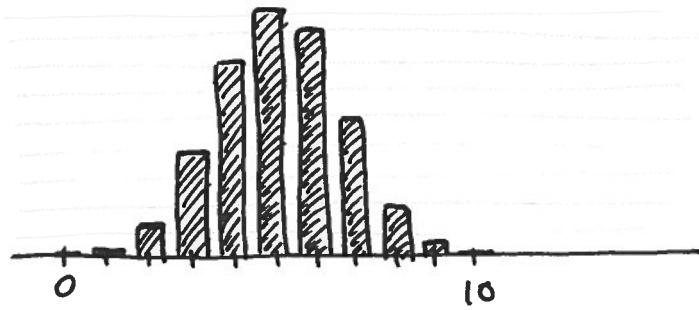
$$P(X=1) = \binom{3}{1} \cdot \left(\frac{1}{6}\right)^1 \cdot \left(1 - \frac{1}{6}\right)^{3-1} = \frac{3!}{(3-1)! 1!} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 = 0.347$$

Eks: Rik mann ønsker gutt som arving. Får 10 barn med kon  
 $X = \#\text{ gutter på 10 fødsler. } X \sim \text{bin}(10, 0.514)$

Verdiene til $X$	0	1	2	3	4	5	6	7	8	9	10
$P(X=x)$	0.000	0.008	0.027	0.104	0.193	0.245	0.216	0.131	0.052	0.012	0.001

+ Bruk R:  $\text{dbinom}(0:10, 10, 0.514)$

Tegner fordelingen:



Men heisann!

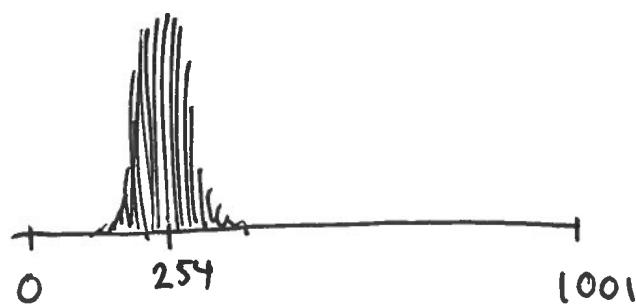
Dette ligner da på en normalfordeling?!

Eks: Meningsmåling. Spør  $n=1001$

$X = \#\text{ som stemmer Høyre. Fra fordeling: } p=0.254$

Verdiene til $X$	0, 1, 2, 3, ..., 1001
$P(X=x)$	$X \sim \text{bin}(1001, 0.254)$

dette gir ikke vi ikle å  
 regne ut eksakt. I stedet  
 bruker vi normalfordelingen  
 som tilnærming. Se ch 5. Mer senere.



Tilbake til førrige eksempel: Rik mann uten arving, og  
 $X \sim \text{bin}(10, 0.514)$

(5)

Hva er sannsynligheten for at det er like mange gutter og jenter i søskenslekten?

$$P(X=5) = (\text{se tabellen på førrige side}) \underline{0.245}$$

Sannsynligheten for at ingen er gutter?

$$P(X=0) = \underline{0.0007}$$

Sannsynligheten for at minst én er en gutt?

$$P(X \geq 1) = 0.008 + 0.037 + \dots + 0.001 =$$

eller

$$= 1 - P(X < 1) = 1 - P(X=0) = 1 - 0.0007 = \underline{0.9993}$$

---

### En digresjon (ikke pensum)

Det finnes en lignende fordeling (som den binomiske), for  $n$  uavhengige forsøk med flere mulige utfall, altså nominale data (kategoriske data med flere kategorier), der såh for de ulike utfallene er like i hvert forsøk.

Dette kalles en multinomial fordeling

Med  $n$  forsøk/observasjoner,  
 $k$  kategorier

$$X_i: \# \text{ observasjoner i kategori nr } i ; \sum_{i=1}^k X_i = n$$
$$P(X_i = x_i) = p_i ; \sum_{i=1}^k p_i = 1 ,$$

så er sannsynlighetsfordelingen for  $X_1, \dots, X_k$  gitt ved

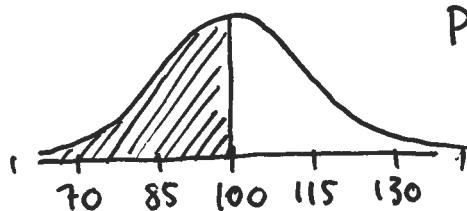
$$P(\underline{X=x}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdot \cdots \cdot p_k^{x_k}$$

Når  $X$  er kontinuerlig, kallas sh. fordelingen en sannsynlighetstetthet, og sh. finnes ved å finne arealet under kurven for det aktuelle intervallet vi er interessert i. (6)

Normalfordelingen er den mest brukte kontinuerlige sh. tetheten.

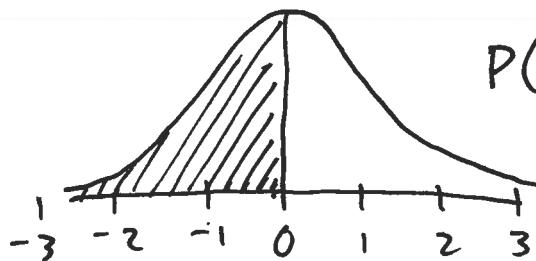
Eles:  $X = IQ$ -verdien til en student

$$X \sim N(100, 15)$$



$$P(X < 100) = 0.5$$

Husk at  $X$  er en stokastisk variabel, altså en variabel som har et sett verdier den kan vare, og en sh. fordeling som viser hvor sannsynlige de ulike verdiene er. Når  $X$  er en stokastisk variabel, er også  $Z = \frac{X-100}{15}$  en stokastisk variabel. Og  $Z \sim N(0, 1)$



$$P(Z < 0) = 0.5$$

$$P(Z < -1.645) = 0.05$$

$$P(Z < -1.96) = 0.025$$

$$\begin{aligned} P(Z > 1.96) &= 1 - P(Z < 1.96) \\ &= 1 - 0.975 = 0.025 \end{aligned}$$

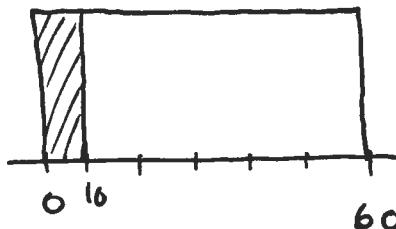
$$P(-1.96 < Z < 1.96) = 0.95$$

Altså:  $X$  er en stokastisk variabel,  $X \sim N(\mu, \sigma^2)$ , da er også  $Z = \frac{X-\mu}{\sigma}$  også en stokastisk variabel,  $Z \sim N(0, 1)$ .

Det finnes også andre sannsynlighetstettheter (jfr tettelskurver (Ch 1.4, og 4.3) enn normalfordelingen.

Eles  $X$  = det sekundet i løpet av et minutt, som en T-bane stopper på perrongen.

$X$  er et høyde som helst tall mellom 0 og 60, og selv om toget går i rute, vil alle sekunder være like sannsynlige:



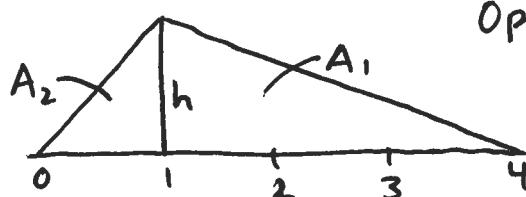
Vi sier at  $X$  er uniformt fordelt i intervallet  $(0, 60)$ :

$X \sim \text{Uniform}(0, 60)$ . Husk at arealet = 1, høyden =  $\frac{1}{60}$

Sannsynligheten for at tog stopper i løpet av de 10 første sekundene av et minutt, er

$$P(X < 10) = \int_0^{10} \frac{1}{60} dx = \left[ \frac{x}{60} \right]_0^{10} = \frac{10}{60} - \frac{0}{60} = \underline{\frac{1}{6}}$$

Eles  $X$  = tall mellom 0 og 4, en oppkonstruert situasjon fra en av ukusoppgavene, der tettelsen er gitt ved



Oppgaven var å finne høyden  $h$ :

Bruker at det totale arealet = 1

$$A_1 = \frac{h \cdot 3}{2} = \frac{3}{2}h ; A_2 = \frac{h \cdot 1}{2} = \frac{1}{2}h$$

$$\text{Totalt area} = A_1 + A_2 = \frac{3}{2}h + \frac{1}{2}h = 2h$$

$$\rightarrow 2h = 1 \rightarrow h = \underline{\frac{1}{2}}$$

$$\text{Deretter: } P(X < 1) = \text{Arealet } A_2$$

$$= \frac{1}{2} \cdot \frac{1}{2} = \underline{\frac{1}{4}}$$

Hvis vi har mange  $X$ -er, noteres det ofte  $X_1, X_2, \dots, X_n$ .<sup>(8)</sup>

Når alle  $X_i$ -ene er stokastiske variabler, vil også

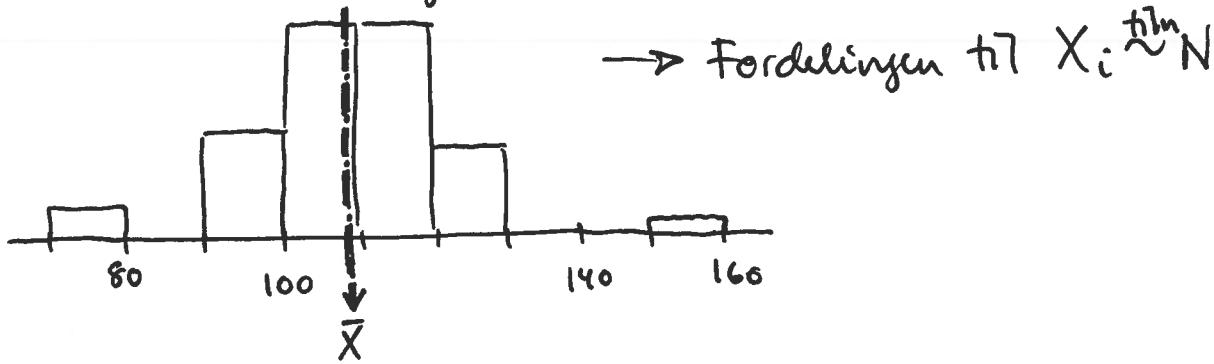
$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  være en stokastisk variabel, altså en variabel som har et sett verdier den kan anta, og en tilhørende sh.-fordeling, som viser hvor sannsynlige de ulike verdiene er.

Ch.5.1

Hvordan finnes vi sh.-fordelingen /sh.-tetheten til  $\bar{X}$ ?

Eks: En statistikkklasse får 58 svar på et spørreskjema om IQ-tester. Hvert enkelt svar er en  $X_i$ , og hun har  $X_1, X_2, \dots, X_{58}$ , altså  $n = 58$ .

Fordelingen som de 58 målingene kommer fra kan vi få et inntrykk av ved å tegne histogrammet over de 58 målingene.



Men hva med fordelingen til  $\bar{X}$ ? Vi ser jo bare én verdi av  $\bar{x}$ . Hvordan kan vi da vite hvordan fordelingen til  $\bar{X}$  ser ut?

Svar: Noen har regnet på det, og kommet fram til det som står på neste side.

Svar 2: Bruk simuleringer, for eksempel slik det er vist i R-Scriptet Sampling-distributions\_R.R

Fordelingen til  $\bar{X}$ : (fire situasjoner) (4)

$X_i \sim$  Normalfordelt  
 eller  $X_i \sim$  Hvilken-som-helst fordeling og  $\sigma$  ukjent

$\bar{X} \sim ?$

$X_1, \dots, X_n$  uavh,  
 $X_i \sim N(\mu, \sigma)$   
 $\sigma$  kjent

Da er

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

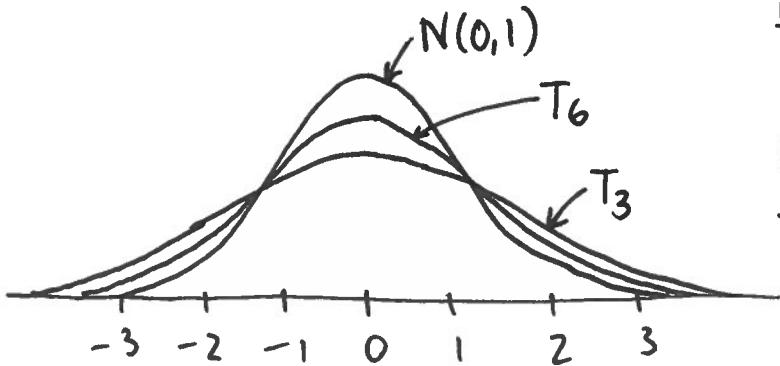
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

unsett n

$X_1, \dots, X_n$  uavh,  
 $X_i \sim N(\mu, \sigma)$   
 $\sigma$  ukjent

Da er

$$\frac{\bar{X} - \mu}{s_{\bar{X}}/\sqrt{n}} \sim T_{n-1}$$



$X_1, \dots, X_n$  uavh,  
 $X_i \sim ?(\mu, \sigma)$   
 $\sigma$  kjent

Hvis n er stor, er

$$\bar{X} \xrightarrow{t \downarrow n} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

og

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{t \downarrow n} N(0, 1)$$

n stor

$X_1, \dots, X_n$  uavh,  
 $X_i \sim ?(\mu, \sigma)$   
 $\sigma$  ukjent

Hvis n er stor, er også

$$\frac{\bar{X} - \mu}{s_{\bar{X}}/\sqrt{n}} \xrightarrow{t \downarrow n} N(0, 1)$$

Hvor stor må n være for at dette skal gjelde?  $n \rightarrow \infty$   
 Det avhenger av hvor forskjellig fordelingen "?" er fra N-fordeling

Husk (fra side ④) at når vi gjør  $n$  uavhengige forsøk/obs.  
 der hvert forsøk har to mulige utfall, suksess/firasko, og  
 $P(\text{Sukcess}) = p$  er lik i hvert forsøk, så er  $X = \#\text{ suksesser}$ ,  
 binomisk fordelt:  $X \sim \text{bin}(n, p)$

(Ch 5.2)

Eles: Undersøker 600 pensjonister for grå stær.

$X = \#\text{ med grå stær}$ . Da er  $X \sim \text{bin}(600, p)$ , og  
 når  $n$  er så stor, vil det ligne på en  $N$ -fordeling.

Noen har regnet på dette også, og vist at hvis  $n$  er stor,  
 og hvis andelen "sukcess" i utvalget er  $\frac{X}{n}$ , så er

$\hat{p} = \frac{X}{n}$  også en stokastisk variabel, altså en  
 variabel som kan ha ulike verdier, og som  
 også har en sannsynlighetsfordeling.

Før det første: når vi gjør  $n$  forsøk og teller opp  $X$ ,  
 ser vi jo bare én verdi av  $X$ . Fordelingen til  $X$  er noe vi  
 må <sup>regne/</sup>resonne oss fram til (som på side ③).

Vi må videre <sup>regne/</sup>resonne oss fram til at  
 når  $X \sim \text{bin}(n, p)$ ,

så er  $X \xrightarrow{\text{tiln}} N(np, \sqrt{np(1-p)})$ , når  $n$  er stor,

og  $\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{\text{tiln}} N(0, 1)$ ,

og  $\frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \xrightarrow{\text{tiln}} N(0, 1)$ .

Da er også  $\frac{\frac{X}{n} - \frac{np}{n}}{\sqrt{\frac{np(1-p)}{n^2}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\text{tiln}} N(0, 1)$ .

# Eksamplar (fra tidligere forelesning) på Ch 5-stoff

(11)

Forskningsoppsørsmål: Kartlegging av gangfunksjon

6 minutters gangtest: Måler hvor langt man kan gå på 6 min

$n = 25$  gir  $X_1, \dots, X_{25}$

Histogram:



Passer fint med normalfordeling.

$\Rightarrow \bar{x} = 623\text{m}$  og  $sd = 83\text{m}$  gir en god oppsummering av observasjonene.

Populasjonens gjennomsnittlige gangfunksjondistanse:  $\mu$

Hvis de  $n = 25$  er et representativt utvalg fra populasjonen, vil  $\mu$  ligne på  $\bar{x}$ , gjennomsnittet i utvalget.

Vi bruker  $\bar{x}$  som estimat for  $\mu$ .  $\bar{x} = 623$

Estimeringsusikkerheten til  $\bar{x}$ :  $S.E.(\bar{x}) = \frac{sd}{\sqrt{n}} = \frac{83}{\sqrt{25}} = 16.6$

Fordi vi nå vet (fra side 9) at når  $\sigma$  er ukjent og må estimeres med  $s_{dx}$ , så er

$$\frac{\bar{x} - \mu}{\frac{s_{dx}}{\sqrt{n}}} \sim T_{n-1} \quad \text{hvis } x_i \sim N(\mu, \sigma^2), \text{ og} \quad \frac{\bar{x} - \mu}{\frac{s_{dx}}{\sqrt{n}}} \stackrel{t \text{-distr}}{\sim} N(0, 1) \quad \text{hvis } n \text{ stor}$$

$$P\left(t_{24, 0.025} < \frac{\bar{x} - \mu}{\frac{s_{dx}}{\sqrt{n}}} < t_{24, 0.975}\right) = 0.95 \quad \text{og} \quad P(-1.96 < \frac{\bar{x} - \mu}{\frac{s_{dx}}{\sqrt{n}}} < 1.96) \approx 0.95$$

↓  
Table D i boka :  $t = 2.064$   
 $df = 24$

95% KI for  $\mu$ :

$$\bar{x} \pm 2.064 \cdot \frac{s_{dx}}{\sqrt{n}}$$

$$623 \pm 2.064 \cdot 16.6$$

$$[588.7, 657.3] \xleftarrow[\text{ganske like.}]{\text{ganske like.}} [590.5, 655.5]$$

ca 95% KI for  $\mu$ :

$$\bar{x} \pm 1.96 \cdot \frac{s_{dx}}{\sqrt{n}}$$

$$623 \pm 1.96 \cdot 16.6$$

Forskingsspørsmål : Kartlegging av oppslutningen til politiske partier, med bestemt partiet Høyre. (12)

Meningsmåling : Ringer rundt og spor ca 1000 ↗ interessert i Høyres oppslutning i pop : p

$n = 1001$  gir  $x_1, \dots, x_{1001}$  0 (ikke Høyre) eller 1 (Høyre) - svar.

$X = \#$  som stemmer Høyre. Her svarer 254 ja.

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{1001} x_i \quad \text{eller} \quad \hat{p} = \frac{X}{n} \quad \leftarrow \text{andel i utvalg}, \hat{p} = \frac{254}{1001}$$

Bruk at  $x_i \sim ?(\mu, \sigma)$   
og at  $n$  er stor, fra  
side ⑨ :

Bruk s. ⑩, som egentlig sier  
det samme som s. ⑨ :

når  $n$  er stor :

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{n \rightarrow \infty} N(0, 1)$$



$$P\left(-1.96 < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < 1.96\right) \approx 0.95$$

95% KI for p :

$$\hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.254 \pm 1.96 \cdot \sqrt{\frac{0.254 \cdot 0.746}{1001}}$$



$$\underline{[0.227, 0.281]}$$

Sannsynlighetsfordelinger og sh. tetteter vil, i likhet med observasjonene som de er modeller for, ha både en forventningsverdi (tilsvarende tyngdepunktet i tettheten/fordelingen), og et standardavvik som beskriver spredning.

De følgende sidene fortsetter etter side ④ og ⑦ i notatene fra forrige forelesning.

### Forventningsverdi

I en diskret sannsynlighetsfordeling er forventningsverdien gitt ved

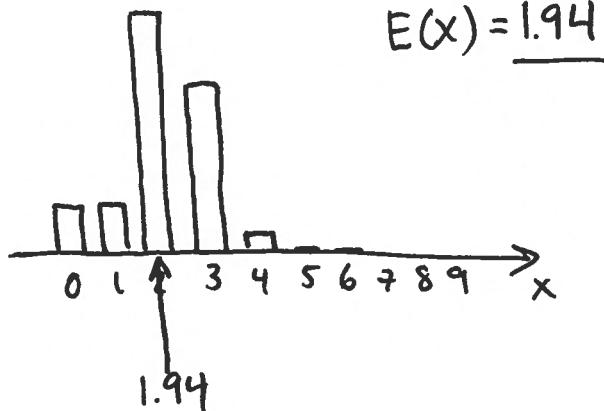
$$\mu_x = E(x) = \sum_{\text{alle } x} x \cdot P(X=x)$$

↑  
expected value

Eks:  $X = \# \text{ barn en norsk kvinne føder}$

Verdiene til $X$	0	1	2	3	4	5	6*	Sum
$P(X=x)$	0.16	0.15	0.38	0.24	0.05	0.01	0.01	1
$x \cdot P(X=x)$	0	0.16	1 \cdot 0.15	$2 \cdot 0.38$ 0.76	0.72	0.2	0.05	0.06 1.94

\* Dette betyr ikke at ingen norske kvinner føder flere enn 6 barn, men andelen er så liten at de tegnes ikke opp i tabellen.

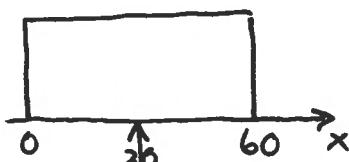


I en kontinuerlig sh. tettet må vi integrere i stedet for å summere

$$\mu_x = E(x) = \int_{\text{alle } x} x \cdot p(x) dx$$

Eks:  $X \sim \text{Uniform}(0, 60)$  (se s. ⑦ i forrige forelesning)

$$p(x) = \frac{1}{60}, x \in [0, 60]$$



$$E(x) = \int_0^{60} x \cdot \frac{1}{60} dx = \left[ \frac{1}{60} \cdot \frac{1}{2} x^2 \right]_0^{60} = \underline{30}$$

## Stortalls lov

(2)

Gjennomsnittet av et sett med observasjoner  $x_1, \dots, x_n$  blir mer og mer lik forventningsverdien, jo flere observasjoner vi har:

$$\bar{x} \xrightarrow{n \rightarrow \infty} \mu_x$$

i populasjonen som observasjonene kommer fra

Dette gjelder uansett hvilken fordeling observasjonene kommer fra, og uansett om  $\bar{x}$  ikke er et godt oppsummeringstall for det typiske for observasjonene.

Eks: Fedtslekt for norske barn	crp-verdier på et fastlegkontor
$n = 1$	$n = 1$
6	6
30	30
100	100
10000	10000

R-script: storstallslov.R

Vi sier at  $\bar{x}$  er forventningsrett for  $\mu$ , fordi  $E(\bar{x}) = \mu$

OBS: Stortalls lov & "randomness" synes først når  $n \rightarrow \infty$

## Regneregler for forventningsverdien

Hvis  $X$  er en stokastisk variabel, og  $a$  og  $b$  er faste tall, så er

$$E(a) = a$$

$$E(a+bX) = a + b \cdot E(X)$$

$$E(X) = \mu = \mu_X$$

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

$$E(bX) = b \cdot E(X) = b \cdot \mu_X$$

Eks: La  $X = \#$  gram kjøtt som stekes per person til middag.

$E(X) = 150\text{g}$ , men dette varierer med størrelsen på pakkene o.l.

Familien på 5 ønsker ofte å ha nok til restemiddag for fire (600g). Hvor mye kjøtt forventer du at de tilbereder?

$$E(600 + 5 \cdot X) = 600 + 5 \cdot E(X) = 600 + 5 \cdot 150 = \underline{\underline{1350}}$$

Hvis  $X$  og  $Y$  er to stokastiske variabler, så er (3)

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

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

Eks: Familien får 2 storspilde gjester

$Y = \#$  gram som tilberedes til hver gjest.  $E(Y) = 250$

$$E(5X + 2Y) = 5 \cdot E(X) + 2 \cdot E(Y) = 5 \cdot 150 + 2 \cdot 250 = \underline{1250}$$

### Standardavvik

Standardavviket finner vi ved å finne variansen først.

I boka: Variansen til  $X$ ,  $\text{Var}(X) = \sigma_X^2 = \sum_{\text{alle } x_i} (x_i - \mu_X)^2 \cdot p_i$

Alternativ regnemåte:

$$\begin{aligned}\text{Var}(X) &= E(X - \mu)^2 \\ &= E(X^2 - 2\mu X - \mu^2) \\ &= E(X^2) - 2\mu E(X) - \mu^2 \\ &= E(X^2) - 2\mu^2 - \mu^2 \\ &= E(X^2) - \mu^2 \\ &= \underbrace{E(X^2)}_{\sum_x x^2 \cdot P(X=x)} - \underbrace{(E(X))^2}_{\sum_x x \cdot P(X=x)}\end{aligned}$$

Eks:  $X = \#$  barn (se o. ①)

$x$	0	1	2	3	4	5	6	Sum
$P(X=x)$	0.16	0.15	0.38	0.24	0.05	0.01	0.01	1
$x \cdot P(X=x)$	0	0.15	0.76	0.72	0.2	0.05	0.06	1.94
$x^2 \cdot P(X=x)$	0	0.15	1.52	2.16	0.8	0.25	0.36	5.24

$$\sigma_X^2 = \text{Var}(X) = E(X^2) - (E(X))^2 = 5.24 - 1.94^2 = 1.48$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{1.48} =$$

## Regneregler for variansen

Vi bruker disse for å finne variansen, og tar  $\sqrt{\cdot}$  til slutt for å finne standardavviket.

Hvis  $X$  er en stokastisk variabel, og  $a$  og  $b$  er faste tall:

$$\text{Var}(a) = \text{Var}(b) = 0$$

$$\text{Var}(X) = \sigma_X^2$$

$$\text{Var}(bX) = b^2 \text{Var}(X)$$

$$\text{Var}(a+bX) = \text{Var}(a) + \text{Var}(bX) = b^2 \text{Var}(X)$$

Eks:  $X$  = vekten til en appelsin.  $\sigma_X = 38g \Rightarrow \sigma_X^2 = 1444$   
 $a$  = vekten til stor gjennbrukspose: 130g (fast)

Variasjonen, variansen til (120 appelsiner i bærapose):

$$\text{Var}(130 + 120 \cdot X) = 120^2 \text{Var}(X) = 120^2 \cdot 1444 = 20793600$$

$$\sigma_{130+120X} = \sqrt{20793600} = 4560$$

Hvis  $X$  og  $Y$  er uavhengige stokastiske var:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Eks:  $X$  = vekten til en appelsin  
 $Y$  = vekten til en annen appelsin

$$\sigma_X^2 = \sigma_Y^2 = 1444$$

$$\sigma_{X+Y}^2 = \text{Var}(X+Y)$$

$$\stackrel{\text{uavh}}{=} \text{Var}(X) + \text{Var}(Y)$$

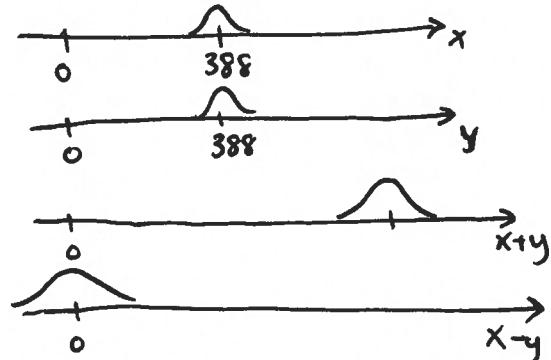
$$= 1444 + 1444 = 2888$$

$$\sigma_{X+Y} = \sqrt{2888} = 54$$

$$\sigma_{X-Y}^2 = \text{Var}(X - Y)$$

$$\stackrel{\text{uavh}}{=} \text{Var}(X) + \text{Var}(Y) = 2888$$

$$\sigma_{X-Y} = 54$$



Hvis  $X$  og  $Y$  er to avhengige stokastiske variabler med korrelasjon  $\rho$ :

$$\sigma_{x+y}^2 = \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\rho\sigma_x \cdot \sigma_y \quad (5)$$

$$\sigma_{x-y}^2 = \text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\rho\sigma_x \cdot \sigma_y$$

Eles:  $X = \text{temperaturen kl } 23 \text{ 16. mai}$        $Y = \text{---, --- kl } 17. \text{ mai}$        $\left. \begin{array}{l} \rho = 0.6 \\ \sigma_x = \sigma_y = 1.6 \end{array} \right\}$

$$\begin{aligned} \sigma_{x+y}^2 &= \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\rho\sigma_x \cdot \sigma_y \\ &= 1.6^2 + 1.6^2 + 2 \cdot 0.6 \cdot 1.6 \cdot 1.6 = 8.192 \end{aligned}$$

$$\sigma_{x+y} = \sqrt{8.192} = \underline{2.9}$$

$$\begin{aligned} \sigma_{x-y}^2 &= \text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\rho\sigma_x \cdot \sigma_y \\ &= 1.6^2 + 1.6^2 - 2 \cdot 0.6 \cdot 1.6 \cdot 1.6 = 2.048 \end{aligned}$$

$$\sigma_{x-y} = \underline{1.43}$$