Övning 1: Normal-fördelning och linjära kombinationer

Begrepp

- En stokastisk variabel har alltid ett tal som utfall och betecknas med en stor bokstav, tex X.
- Ett utfall betecknas med en liten bokstav, tex x

- täthetsfunktionen: $f_X(x)$

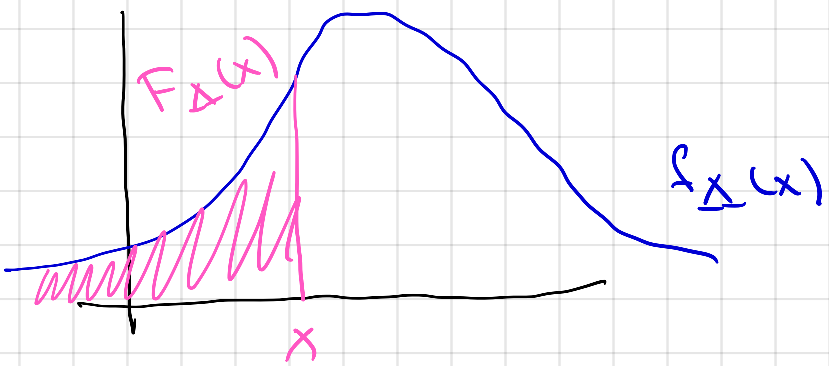
$$P(a < X \leq b) = \int_a^b f_X(t) \cdot dt = F_X(b) - F_X(a)$$

$<$, $\leq$ spelar ingen roll, för att vi har kontinuerlig fördelning.

$a < b$

- fördelningsfunktionen: $F_X(x) = P(X \leq x) = P(-\infty < X \leq x) = \int_{-\infty}^{x} f_X(t) \cdot dt$

entydigt bestämmer sannolikhetsfördelningen



Normalfördelning

$X \in N(\mu, \sigma)$ → täthetsf. $\varphi(x)$
fördelningsf. $\Phi(x)$

→ $E(X) = \mu$
→ $D(X) = \sigma \Rightarrow V(X) = \sigma^2$

- bruker använda standardiserad N-fördelning
  - tabell 1 och 2

$X \in N(0,1)$ $\quad E(X) = 0$
$\quad D(X) = 1$

- om vi har en s.v. $X \in N(\mu, \sigma)$ så kan vi standardisera den mha: $\boxed{Y = \frac{X-\mu}{\sigma} \in N(0,1)}$

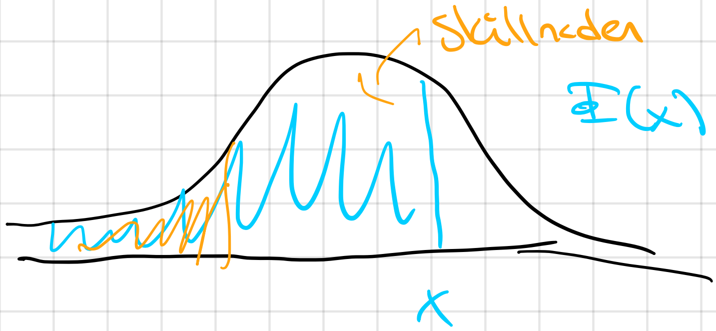
- täthetsfunktionen för N-fördelning, den är jämn

$$\varphi(-x) = \varphi(x)$$

- fördelningsfunktionen: $\boxed{\Phi(-x) = 1 - \Phi(x)}$

$$\boxed{P(a < X \leq b) = \Phi(b) - \Phi(a)}$$

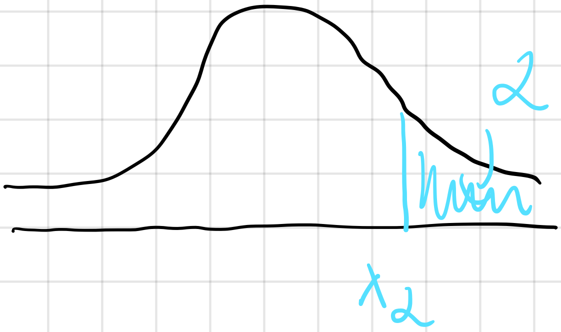
• Tabell 1: används när gränsen är känd och sannolikheten söks (läser av $\Phi(x)$)



$\Phi(x) = P(X \leq x)$

• tabell 2: standardvärden: hur mycket man har råd att förlora (statistik)

- används när sannolikheten är känd och gränsen söks



$P(X > \lambda_\alpha) = \alpha$

$\lambda_{1-\alpha} = -\lambda_\alpha$

Uppgift 6.1

$X \in N(0,1)$. Bestäm:
a) $P(X \leq 1.82)$
b) $P(X \leq -0.35)$
c) $P(-1.2 \leq X < 0.5)$
d) a så att $P(X > a) = 5\%$
e) a så att $P(|X| < a) = 95\%$

a) $P(X \leq 1.82) = \{\text{tabell } 1\} = \Phi(1.82) = 0.9656$

b) $P(X \leq -0.35) = \{\text{tabell } 1\} = \Phi(-0.35) = 1 - \Phi(0.35) =$

$= 1 - 0.6368 = 0.3632$

c) $P(-1.2 \leq X < 0.5) = \{P(a < X < b) = \Phi(b) - \Phi(a)\} =$

$= \Phi(0.5) - \Phi(-1.2) = \Phi(0.5) - [1 - \Phi(1.2)] =$

$= \Phi(0.5) - 1 + \Phi(1.2) = \{\text{tabell } 1\} = 0.5764$

d) $P(\underline{X} > a) = 5\% = 0.05$

- använder tabell 2:
- ensidig gräns: $P(\underline{X} > \lambda_{0.05}) = 0.05$

$$\lambda_{0.05} = a = 1.6449$$

e) $P(|\underline{X}| < a) = 95\% = 0.95$

tar bort $|\underline{X}| = \begin{cases} x, & x > 0 \\ -x, & x < 0 \end{cases}$

$$P(-a < \underline{X} < a) = 0.95$$

$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

- har tvåsidig gräns, tar $\lambda_{\frac{\alpha}{2}} = \lambda_{0.025} = a = 1.96$

Uppgift 6.4

$\underline{X} \in N(5, 2)$. Bestäm a) $P(\underline{X} \leq 6)$

$\mu = 5$ b) $P(1.3 < \underline{X} < 7.2)$

$\sigma = 2$ c) a så att $P(\underline{X} \leq a) = 5\%$.

a) $P(\underline{X} \leq 6) = \{\text{gör om till } N(0,1)\} =$

$$= P\left[\frac{\underline{X} - \mu}{\sigma} \leq \frac{6 - \mu}{\sigma}\right] = P\left[\underbrace{\frac{\underline{X} - 5}{2}}_{Y \in N(0,1)} \leq \frac{6-5}{2}\right] =$$

$$= P(Y \leq 0.5) = \{\text{tabell } 1\} = \Phi(0.5) = 0.6915$$

b) $P(1.8 < X < 7.2) =$

$$= P\left[\frac{1.8-5}{2} < \underbrace{\frac{X-5}{2}}_{Y \in N(0,1)} < \frac{7.2-5}{2}\right] =$$

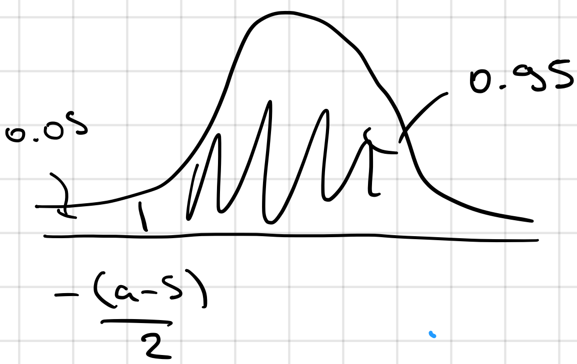
$$= P(-1.6 < Y < 1.1) = \Phi(1.1) - \Phi(-1.6) =$$

$$= \Phi(1.1) - [1 - \Phi(1.6)] = \Phi(1.1) + \Phi(1.6) - 1 =$$

$$= 0.8095$$

c) $P(X \leq a) = 5\% = 0.05$

$$P\left(\frac{X-5}{2} \leq \frac{a-5}{2}\right) = P\left(Y \leq \frac{a-5}{2}\right) = 0.05$$



$$P\left(Y \geq \frac{-a+5}{2}\right) = 0.05$$

$$\frac{-a+5}{2} = \lambda_{0.05} = 1.6449$$

$$a = 1.7102$$

Linjära kombinationer (om vi har oberoende fördelningar)

$$E(X+Y) = E(X) + E(Y)$$

$$V(X+Y) = V(X) + V(Y)$$

$$D(X+Y) = \sqrt{D^2(X) + D^2(Y)}$$

- Sats 1: om $X \in N(\mu_X, \sigma_X)$ och $Y \in N(\mu_Y, \sigma_Y)$ och X och Y är oberoende:

$$\boxed{X+Y \in N(\underbrace{\mu_X+\mu_Y}_{E(X+Y)}, \underbrace{\sqrt{\sigma_X^2 + \sigma_Y^2}}_{D(X+Y)})}$$

$$X-Y \in N(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$$

- Sats 2: om $X \in N(\mu, \sigma)$

$$Y = aX + b \in N(a\mu + b, |a|\sigma)$$

$$(Y = 2X \Rightarrow Y \in N(2\mu_X, 2\sigma_X))$$

- Centrala gränsvärdessatsen (CGS):

n är stort, X_i:na har samma fördelning och är oberoende s.v. med $E(X_i) = \mu$ och $D(X_i) = \sigma > 0$

$$\sum_{i=1}^{n} X_i \sim N(n\mu, \sigma\sqrt{n})$$

$\sqrt{n}$ pga $V(X) = \sigma^2$

$nV(X) = n\sigma^2$

$D(X) = \sqrt{n}\,\sigma^2$

Uppgift 6.15

Har 2 reläer som är inställda för utlösning enligt $X \in N(1, 0.1)$ och $Y \in N(1.5, 0.2)$. Bestäm sannolikheten att Y utlöses före X:

Lösning: $X \in N(1, 0{,}1)$
$Y \in N(1.5, 0{,}2)$ } oberoende

- Vi vill ha att Y utlöses före X:

$$Y < X \Rightarrow Y - X < 0$$

Söker: $P(Y - X < 0)$

$$Z = Y - X \in N\left(\mu_Y - \mu_X, \sqrt{\sigma_Y^2 + \sigma_X^2}\right) =$$
$$= N\left(1.5 - 1, \sqrt{0.2^2 + 0.1^2}\right) =$$
$$= N\left(0.5, \sqrt{0.05}\right)$$

$Z \in N(0.5, \sqrt{0.05})$, söker $P(Z < 0)$

$$P(Z < 0) = P\left[\frac{Z - 0.5}{\sqrt{0.05}} < \frac{0 - 0.5}{\sqrt{0.05}}\right) =$$

$$= 1 - \Phi(2.24) = 0.01287$$

Linjära kombinationer av Po, Bin

① Om vi har 2 oberoende s.v. X och Y som är respektive:

$$X \in Po(\underbrace{\mu_1}_{E(X)}) \quad \text{och} \quad Y \in Po(\underbrace{\mu_2}_{E(Y)})$$

- kan säga att $\boxed{Z = X + Y \in Po(\mu_1 + \mu_2)}$

② Låt X och Y vara 2 oberoende s.v. med samma p, sådana att

$X \in Bin(n_x, p)$ och $Y \in Bin(n_y, p)$, då gäller

att:

$$\boxed{Z = X + Y \in Bin(n_x + n_y, p)}$$

- OBS: gäller endast om X och Y har samma sannolikhet p.

Tenta 2019-01-03 #4

Låt X och Y vara 2 oberoende s.v. sådana att $X \in Po(3)$ och $Y \in Po(4)$. Beräkna $P(X+Y=2)$

$$Z = X + Y \in Po(3+4) = Po(7)$$

$Z \in Po(7)$ och vi söker $P(Z=2)$

- använder miniräknare (TI-84): (2nd) → distr →

→ poissonpdf $\begin{Bmatrix} \lambda = 7 \\ x = 2 \end{Bmatrix}$ $P(Z=2) = 0{,}0223$

tenta 2019-01-03 #5

Låt X och Y vara två oberoende s.v. där

$X \in N(2,3)$ och $Y \in N(4,2)$. Låt $Z = 2Y - X$.
Beräkna $P(Z > 4)$

Lösning: $Z = 2Y - X$

$$2Y \in N(2 \times 4,\ 2 \times 2) = N(8,4)$$

$$2Y - X \in N\left(8-2,\ \sqrt{4^2+2^2}\right) = N(6,5)$$

$$Z \in N(6,5)$$

$$P(Z>4) = 1 - P(Z \leq 4) = 1 - P\left(\frac{Z-6}{5} \leq \frac{4-6}{5}\right) =$$

$$= 1 - P(W \leq -0.4) = \Phi(0.4) = \{\text{tabell } 1\} = 0,6554$$

tenta 2019-04-17 #4

Låt X och Y vara två oberoende s.v. sådana att $X \in Bin(10, 1/5)$ och $Y \in Bin(5, 1/5)$

Beräkna $P(X+Y=3)$

Lösning: ser att sannolikheten är samma för X och Y,

$p = \frac{1}{5}$ $\Rightarrow$ kan använda satsen.

$$Z = X+Y \in Bin(n_x + n_y,\ p) = Bin\left(15, \frac{1}{5}\right)$$

$$P(Z=3) = \left\{ \begin{array}{l} \text{miniräknare: distr} \rightarrow \text{binompdf} \\ n=15 \\ p=\frac{1}{5} \end{array} \right\} x=3 = 0.250$$

Testa tenta 2019 # 4

Betrakta en kundmottagning där tiden X_i som det tar att betjäna en kund är $Exp(\lambda)$-fördelad med $\lambda = 10\ h^{-1}$. Antag att de s.v. X_i är oberoende och att kund $i+1$ börjar betjänas omedelbart efter att kund i har betjänats klart. Bestäm den approximativa fördelningen för den totala tiden som det tar att betjäna de första 100 kunderna.

$$X_i \in Exp(10)$$

$$E(X_i) = \frac{1}{\lambda} = \frac{1}{10}$$

$$V(X_i) = \frac{1}{\lambda^2} = \frac{1}{100} \Rightarrow D(X_i) = \sqrt{V(X_i)} = \frac{1}{10}$$

$n = 100 \Rightarrow$ många!

X_i oberoende och likafördelade

$\Rightarrow$ kan använda CGS

$$CGS: \quad Y = \sum_{i=1}^{100} X_i \in N(\mu n, \sigma\sqrt{n})$$

$$\mu = \frac{1}{10}$$

$$\sigma = \frac{1}{10}$$

$$n = 100$$

$$Y \sim N(10, 1)$$

Tenta 2020-05-26 #12 (Del B) (10p)

En tunnel av längd 170 m skall borras från två håll. Av erfarenhet tror man sig veta att vad som hinns med olika dagar från ett håll kan uppfattas som oberoende observationer av en stokastisk variabel med väntevärdet 5.0 m och standard avvikelsen 1.2 m

Beräkna med lämplig approximation sannolikheten att det tar längre tid än 18 dagar att borra tunnels

$X \in N(5, 1.2)$ en borr, givet i uppgiften.

- Vi borrar från 2 håll: har $2X$

$$X + X \in N\left(2 \times 5, \sqrt{1.2^2 + 1.2^2}\right) = N(10, 1.7)$$

= normalfördelning för de två borrarna tsm.

- vi vill veta sannolikheten att vi kommer att borra i mer än 18 dagar.

⇒ vi vill veta sannolikheten att på 18 dagar hinner vi inte att borra hela sträckan på 170 m

$$18 \text{ dagar} \Rightarrow \begin{cases} n = 18 \\ \mu = 10 \\ \sigma = 1.7 \end{cases} \Rightarrow \text{CGS.}$$

n är stort och vi har samma fördelning för varje dag

CGS: $Y \sim N(10 \times 18, 1.7\sqrt{18}) = N(180, 7.2)$

på 18 dagar, förväntas att borra 180 m (mer än det vi behöver)

avviker mindre än 10 m,

$\Rightarrow$ liten sannolikhet att inte hinna (kontrollera svaret)

$Y \in N(180, 7.2)$

$$P(Y < 170) = P\left(\frac{Y-180}{7.2}\right) = P(Z < -1.39) =$$

$$= \{\text{tabell } 1\} = \Phi(-1.39) = 1 - \Phi(1.39) =$$

$$= 1 - 0.9177 = 0{,}0823$$

- Svar: sannolikheten att de inte hinner borra 170 cm på 18 dagar är 8.23%.

tenta 2018-10-24 #3

Ange fördelningen för $Z = 4X - 5Y$ då $X \in N(1,2)$ och $Y \in N(1,3)$ och X och Y är oberoende:

- linjär kombinationer:

$$4X \in N(4 \times 1, 2 \times 4) = N(4, 8)$$

$$5Y \in N(5 \times 1, 5 \times 3) = N(5, 15)$$

$$Z = 4X - 5Y \in N(4-5, \sqrt{8^2 + 15^2}) = N(-1, 17)$$

$$Z \in (-1, 17)$$

tenta 2013-12-20 #7

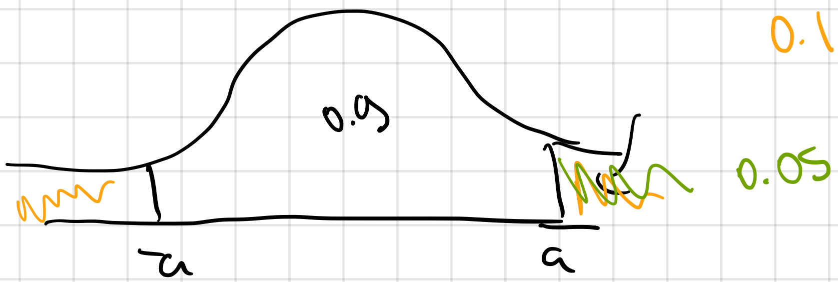
Anta att $X \in N(0,1)$. Bestäm konstanten a så att $P(|X| > a) = 0.1$

$$P(|X|>a) = 1 - P(|X| \leq a) = 0.1$$

$$P(|X| \leq a) = 1 - 0.1 = 0.9$$

$$P(-a \leq X \leq a) = 0.9$$

-tvåsidigt intervall:



$$P(X>a) = 0.05$$

tabell 2: $a = 1.6449$

Sammanfattning:

$\boxed{N: X+Y \in N(\mu_x+\mu_y, \sqrt{\mu_x^2+\mu_y^2}}$ om X, Y är oberoende och normalfördelade

Bin: $\boxed{X+Y \in Bin(n_x+n_y, p)}$ om p är samma för X och Y och båda är Bin-fördelade

Po: $\boxed{X+Y \in Po(\mu_1+\mu_2)}$ X och Y är Po-fördelade. alla är oberoende

Övning 2: Approxmationer, börja på statistik

tenta 2019-03-11 #5

Antalet spikar i en kartong kan antas vara normalfördelad med $\mu = 563.3$ och $\sigma = 33.2$. Leverantören garanterar att antalet spikar överstiger ett visst antal n med sannolikheten 99%. Hitta n.

- tabell 2 uppgift

$X \in N(563.3, 33.2)$

$P(X > n) = 0{,}99$

$$\boxed{P(Y > a) = 0.99 \iff P(Y > -a) = 0.01}$$

$$P\left(\frac{X - 563.3}{33.2} > \frac{-n + 563.3}{33.2}\right) = 0.01$$

$$P\left(Y > \frac{-n + 563.3}{33.2}\right) = 0.01$$

tabell 2: $\dfrac{-n + 563.3}{33.2} = 2.3263 \Rightarrow n = 486$

Approximationer

- **Hypergeometriska fördelningen**: Inträffar vid dragning utan återläggning

$\Rightarrow$ har N enheter där andelen enheter med egenskapen A är p och vi drar n enheter utan återläggning.

X - antalet enheter med egenskapen A som vi får

$X \in Hyp(N, n, p)$

- **Binomialfördelning** - ex gör n försök och sannolikheten att lyckas är p. X - antal lyckade försök:

$$X \in Bin(n, p)$$

om $\frac{n}{N} \leq 0.1$ kan vi approximera $Hyp(N, n, p) \sim Bin(n, p)$

$\Rightarrow$ Bin är mycket enklare att beräkna än Hyp.

- $Bin(n,p) \sim N(\underbrace{np}_{E(X)}, \underbrace{\sqrt{np(1-p)}}_{D(X)})$ om $np(1-p) \geq 10$

  - följer från CGS (gäller om n är stort)

  $p = 0.5 = 1-p$ räcker det med $n = 40$ för att kunna approximera $Bin \sim N$

  om $p = \frac{1}{1000}$ måste n vara stort.

- **Poisson-fördelning**: inträffar om vi har ett intervall och sannolikheten för en viss händelse är alltid samma. X - antal händelser som inträffar.

$Po(\underbrace{\mu}_{E(X)}) \sim N(\mu, \sqrt{\mu})$ om $\mu \geq 15$

$Bin(n,p) \sim Po(np)$ om $p \leq 0.1$

Halvkorrektion - används när vi går från en diskret fördelning till en kontinuerlig fördelning.

diskreta: Binomial, Poisson, hypergeometrisk (Σ)

kontinuerliga: Normal, exponential, likformig ($\int$)

- i det diskreta fallet: $P(a < X < b) =$

$$= P(a+1 \leq X \leq b-1)$$

- i kontinuerliga fallet: $<$ eller $\leq$ spelar ingen roll, ger samma gränser vid integration.

Ex $P(19 < X < 31) = P(20 \leq X \leq 30)$ (d)

(k) $\Rightarrow P(19.5 < X < 30.5) = \int_{19.5}^{30.5} f_X(x) \cdot dx$

$\Rightarrow$ väljer mitten halvkorrektion

tenta 2017-03-14 #2 (10p)

En butik vill sälja 2000 badbyxor, de har rea. Av erfarenhet vet man att antal byxor en kund köper är oberoende av hur många byxor andra köper och betraktas som en s.v. X med sannolikhetsfunktionen:

$$p_X(k) = \begin{cases} 0.2 & \text{om } k=0 \\ 0.5 & \text{om } k=1 \\ 0.2 & \text{om } k=2 \\ 0.1 & \text{om } k=3 \\ 0 & \text{om } k>3 \end{cases}$$

Beräkna med en lämplig och välmotiverad approx. det minsta antalet kunder som måste komma till butiken om sannolikheten ska vara 90% att butiken får alla byxor sålda.

Lösning:

X - antal badbyxor en kund köper, ej känd fördelning måste beräkna $E(X)$ och $D(X)$

n - antal kunder (söker)

Y ~ totala antalet byxor sålda. $Y = X_1 + X_2 + X_3 + \dots + X_n$

$Y \sim N(n\mu, \sigma\sqrt{n})$ (CGS)

har oberoende X_i med samma fördelning krävs många kunder (n stort)

$\mu = E(X_i)$

$\sigma = D(X_i)$

$$\mu = E(X_i) = \sum_{k=0}^{3} k \cdot p_X(k) = 0.2 \times 0 + 0.5 \times 1 + 0.2 \times 2 + 0.1 \times 3 = 1.2$$

$$\sigma^2 = V(X) = E(X_i^2) - [E(X_i)]^2$$

$$E(X_i^2) = \sum_{k=0}^{3} k^2 p_X(k) = 0^2 \times 0.2 + 1^2 \times 0.5 + 2^2 \times 0.2 + 3^2 \times 0.1 = 2.2$$

$$V(X_i) = 2.2 - 1.2^2 = 0.76 \Rightarrow D(X_i) = \sqrt{0.76}$$

$$\Rightarrow Y \sim N(1.2n, \sqrt{0.76n})$$

- behöver ta fram n som uppfyller $P(Y \geq 2000) = 0.9$

- vi hade diskreta s.v X som vi approximerade till en kontinuerlig s.v Y

- från början kunde vi räkna ut

$$P(Y > 1999) = 0.9$$

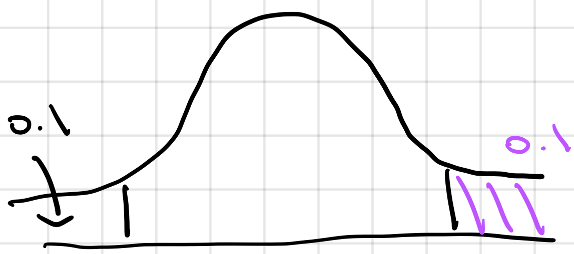
$\Rightarrow$ halvkorrektion $\Rightarrow$ $P(Y > 1999.5) = 0.90$

gör om till $Z(0,1)$ för att kunna använda tabell 2.

$$P\left(\frac{Y - 1.2n}{\sqrt{0.76n}} > \frac{1999.5 - 1.2n}{\sqrt{0.76n}}\right) = 0.9$$

$$P\left(Z > \frac{1999.5 - 1.2n}{\sqrt{0.76n}}\right) = 0.9$$

Söker gränsen $\Rightarrow$ tabell 2.



$\frac{1999.5 - 1.2n}{\sqrt{0.76n}}$ $\frac{1.2n - 1999.5}{\sqrt{0.76n}}$ pga symmetri

$$\frac{1.2n - 1999.5}{\sqrt{0.76n}} = \lambda_{0.10} = 1.2816$$

- löser ut n: $\sqrt{n} = x \Rightarrow n = x^2$

enklare att lösa än en ekv med $n, \sqrt{n}$

$$\frac{1.2x^2 - 1999.5}{\sqrt{0.76}\,x} = 1.2816$$

$$1.2x^2 - 1999.5 = 1.2816\sqrt{0.76}\,x$$

$$1.2x^2 - 1.2816\sqrt{0.76}\,x - 1999.5 = 0$$

kvadratisk ekv.

$\Rightarrow \quad x = 0.465 \pm 40.8$, tar den positiva roten

$n = x^2 = 1704.69 \Rightarrow$ behövs 1705 kunder

även om vi får 1704.1 vilket brukar avrundas ner så ska man avrunda upp för att vi söker min antal kunder som behövs.

tenta 2019-05-28 # 4 (1p)

Sannolikheten att en slumpmässigt vald hasselnöt är möglig är 0.1. Man knäcker 500 nötter. Beräkna approximativt sannolikheten att minst 60 av dem är mögliga.

notera approximativt! $\Rightarrow$ ska approximera även om det går att räkna exakt.

- bestämma fördelning: $\left.\begin{matrix} n = 500 \\ p = 0.1 \end{matrix}\right\} \Rightarrow$ Binomial-fördelning. då lika sannolikhet

$X \in Bin(500, 0.1)$, Söker $P(X \geq 60) = 1 - P(X \leq 59)$

$np(1-p) = 500 \times 0.1 \times (1-0.1) = 45 \geq 10$, ok att approximera Bin till N (viktigt att skriva på B-delen)

$$X \sim N(np, \sqrt{np(1-p)}) = N(50, \sqrt{45})$$

$$1 - P(X \leq 59) = 1 - P\left[\frac{X-np}{\sqrt{np(1-p)}} \leq \frac{59-np}{\sqrt{np(1-p)}}\right] =$$

$$= 1 - P(Y \leq 1.34) = \{\text{tabell } 1\} = 1 - \Phi(1.34) =$$

$$= 0.090$$

tenta 2020-03-10 #13 (10p)

Felkvoten vid en tillverkningsprocess är 0.03. Ur ett parti om 5000 tillverkade enheter väljer man på måfå 100 enheter och justerar processen om fler än k av dessa är defekta. Låt Y beteckna antalet defekta enheter bland de 100 utvalda. Bestäm det minsta heltal k för vilket det gäller att $P(\text{processen justeras}) = P(Y > k) < 0.01$. Välmotiverade approximationer är tillåtna.

-hypergeometrisk fördelning: $\begin{cases} N = 5000 \\ n = 100 \\ p = 0.03 \end{cases}$

$\Rightarrow Y \in Hyp(5000, 100, 0.03)$ söker $P(Y > k) < 0.01$

$\frac{n}{N} = \frac{100}{5000} = 0.02 \leq 0.1 \rightarrow$ kan approximera Y till

$\sim Bin(100, 0.03)$

$Bin(\underbrace{100}_{n}, 0.03)$ finns ej med i tabellen

$p = 0.03 < 0.1 \Rightarrow$ kan approximera Bin till Po

$np = 100 \times 0.03 = 3$ finns i tabelle

$\Rightarrow Y \sim Po(3)$

$\Rightarrow$ tabell 5: $P(Y > k) < 0.01$

$$\Updownarrow$$

$$P(Y \leq k) \geq 0.99$$

Ser i tabellen att $k \geq 8$ dvs $k = 8$ minsta värdet som funkar.

- Ser i tabellen att sannolikheten växer för större k.

Sannolikhet vs statistik

- Sannolikhet är teorin (teoretiska) och statistik är verkliga livet (praktiska)

här konf-int.

Sannolikhet

- Väntevärde μ:

  - det man får i snitt om man gör oändligt många försök

Statistik

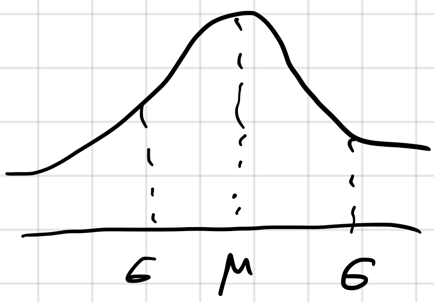
- Medelvärde, $\bar{x}$

  - det man får i genomsnitt om man gör ändligt antal försök

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

antal försök

• Varians σ^2



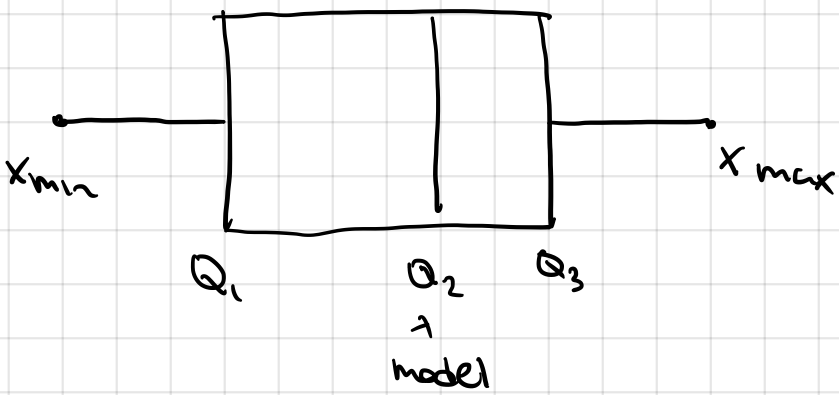
• Stickprovsvarians : s^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$n-1$ istället för n för att vi vill ha en väntevärdesriktighet ska bli σ^2 i genomsnitt.

- väljer ut några

Boxplott



• Hur tar vi fram Q1? - Q_1 är första kvantilen, 25% av data

- har n st mät data totalt: söker x_k där $\frac{k}{n} = 0.25$

Väljer k till det heltal som uppfyller olikheten

$$0.25n \leq k \leq 0.25n + 1$$

Ex: $n = 11$

$$0.25n = 0.25 \times 11 = 2.75$$
$$0.25n + 1 = 0.25 \times 11 + 1 = 3.75$$

$2.75 \leq k \leq 3.75$

k är heltal

$x_3 = Q_1$

- om $n=12$, söker Q_1:

$0.25 \times 12 = 3$

$0.25 \times 12 + 1 = 4$

$3 \leq k \leq 4$, finns 2 heltal som uppfyller olikheten, tar medelvärde

$$\frac{x_3 + x_4}{2} = Q_1$$

- fungerar metoden för alla Q_n

- kvartilavstånd: $Q_3 - Q_1$

- Variationsbredd: $x_{max} - x_{min}$

tenta 2020-05-26 # 6

Ange kvartilavstånd för dessa mätdata

7 10 10 11 12 12 12 12 13 15 19 20 21 21
21 22 23 23 26 26 26 27 30 36

x_6 x_7

← olika x

$x_1 \cdots x_{24}$

Lösning: kvartilavstånd $= Q_3 - Q_1$

$n = 24$

Q_1: $0.25n \leq k \leq 0.25n + 1$

$0.25 \times 24 \leq k \leq 0.25 \times 24 + 1$

$6 \leq k \leq 7 \Rightarrow Q_1: \frac{x_6 + x_7}{2} = \frac{12 + 12}{2} = 12$

n som funkar.

data: i växande ordning dvs minsta till största

Q_3: $0.75n \leq k \leq 0.75n+1$

$0.75 \times 24 \leq k \leq 0.75 \times 24 + 1$

$18 \leq k \leq 19$

Q_3: $\frac{X_{18} + X_{19}}{2} = \frac{23+26}{2} = 24.5$

k.a. $Q_3 - Q_1 = 24.5 - 12 = 12.5$

Punktskattning

- En skattning av θ kallas θ^*_{obs} och är ett utfall av den s.v. θ^*

observerade skattningen av θ, dvs data vi får

tex $E(X_i) = \mu$

$\mu^*_{obs} = \bar{x}$ det vi får i praktiken

och $D(X_i) = 6 \Rightarrow 6^*_{obs} = s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

skattar väntevärde med medelvärde
skattar standardavvikelse med stickprovsvariansen.

Skattningar i vanliga fördelningar

- Binomial-fördelning: $X \in Bin(n,p)$
  - n är känd (antal försök)
  - X – tex antal defekta varor
  - skattar p

  summa för Hyp

$$p^*_{obs} = \frac{x}{n}$$

- Poisson-fördelning: $X \in Po(\mu)$, $\mu^*_{obs} = \bar{x}$

- Exponentialfördelning: $X \in exp(\lambda)$, $E(X_i) = \frac{1}{\lambda}$ dvs $\lambda = \frac{1}{E(X_i)}$

$$\lambda^*_{obs} = \frac{1}{\bar{x}}$$

- Ffg-fördelning: $X \in ffg(p)$. $E(X) = \frac{1}{p} \Rightarrow p^*_{obs} = \frac{1}{\bar{x}}$

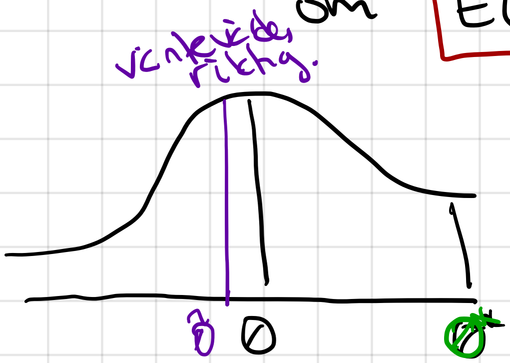
- Normal fördelning: $X \in N(\mu, \sigma)$ $\Rightarrow$ $\mu^*_{obs} = \bar{x}$

$$\sigma^*_{obs} = s$$

Väntevärdesriktighet (VVR)

- Definition: en skattning θ^*_{obs} av θ är väntevärdesriktig om $\boxed{E(\theta^*) = \theta}$

dvs om vi gör skattningen många gånger så ska vi i genomsnitt få rätt värde (väntevärde)



ej väntevärdesriktig, långt bort

↳ex. $\mu^*_{obs} = \frac{x_1 + x_2 + x_3}{3}$ dvs tre försök, tre värde.

μ^*_{obs}: utfall

x_1, x_2, x_3: lilla x (utfall)

Om $E(X_i) = \mu$, är skattningen väntevärdesriktig

$E(\mu^*) \overset{?}{=} \mu$

$$E(\mu^*) = E\left[\frac{X_1 + X_2 + X_3}{3}\right] = \frac{1}{3}\left(E(X_1) + E(X_2) + E(X_3)\right) =$$

μ^*: s.v.

X: s.v

$$= \frac{1}{3} \cdot 3\mu = \mu \Rightarrow \text{skattningen är vvr}$$

⇒ fås från satsen $E(aX + bY + c) = aE(X) + bE(Y) + c$

↳ex. $\mu^*_{obs} = \frac{5x_3 + 3x_1 + 2x_2}{10}$. Är den vvr?

$$E(\mu^*) = E\left[\frac{5X_3 + 3X_1 + 2X_2}{10}\right] = \frac{1}{10}\left(5\mu + 3\mu + 2\mu\right) =$$

$$= \frac{10}{10}\mu = \mu \Rightarrow \text{ok! vvr.}$$

Övning 3: effektivitet, ML och MK-skattningar

Repitition

• Väntevärdesriktighet: $\boxed{E(\theta^*)=\theta}$

ex $\mu^*_{obs} = \frac{x_1+x_2+x_3}{3}$ och $\hat{\mu}_{obs} = \frac{5x_3+3x_1+2x_2}{10}$

Om $E(X_i)=\mu$, är skattningar är väntevärdesriktiga?

$$E(\mu^*) = E\left(\frac{X_1+X_2+X_3}{3}\right) = \frac{1}{3}\left[E(X_1)+E(X_2)+E(X_3)\right] =$$

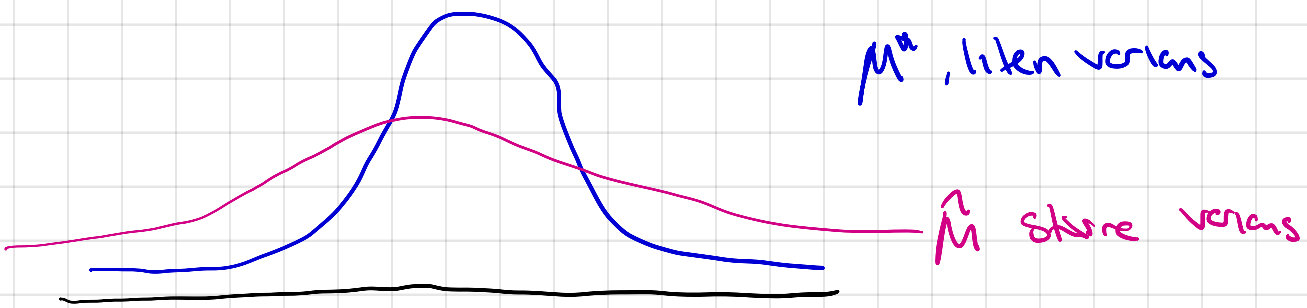
$$= \frac{1}{3}3\mu = \mu \Rightarrow \text{skattningen är vvr}$$

$$E(\hat{\mu}) = E\left[\frac{5X_3+3X_1+2X_2}{10}\right] = \frac{1}{10}(5\mu+3\mu+2\mu) = \mu$$

båda skattningar är vvr.

Effektivitet

• Om vi har flera väntevärdesriktiga skattningar av θ så är skattningen med lägst varians effektivast



μ^* är en mer effektiv skattning än $\hat{\mu}$

-Effektivitet är en jämförelse mellan olika skattningar, kan inte säga att en skattning är effektiv, utan vi kan bara säga att den är mer effektiv än en annan skattning

effektivitet = jämförelse !!

↳ tidigare exempel: $\mu^a_{obs} = \frac{X_1 + X_2 + X_3}{3}$ och

$\hat{\mu}_{obs} = \frac{5X_3 + 3X_1 + 2X_2}{10}$.

-Anta att $V(X_i) = \sigma^2$, vilken skattning är mer effektiv?

1 Måste först kolla om båda skattningar är väntevärdesriktiga (gjordes i förra exempel)

$$V(\mu^a) = V\left[\frac{X_1 + X_2 + X_3}{3}\right] = \{\text{använder } V(aX+b) = a^2V(X)\}$$

$$= \frac{1}{9}V(X_1 + X_2 + X_3) = \frac{1}{9}3V(X_i) = \frac{1}{3}\sigma^2 \approx 0.33\sigma^2$$

$$V(\hat{\mu}) = V\left[\frac{5X_3 + 3X_1 + 2X_2}{10}\right] = \frac{1}{100}\left[V(5X_3 + 3X_1 + 2X_2)\right] =$$

$$= \frac{1}{100}\left[5^2V(X_3) + 3^2V(X_1) + 2^2V(X_2)\right] =$$

$$= \frac{1}{100}\left[(25 + 9 + 4)V(X_i)\right] = \frac{38}{100}\sigma^2 \approx 0.38\sigma^2$$

$V(\mu^*) < V(\hat{\mu}) \Rightarrow \mu^a_{obs}$ är mer effektiv skattning än $\hat{\mu}_{obs}$.

tenta 2019-12-19 #7

Antag att X och Y är oberoende s.v. så att $E(X) = E(Y) = \theta$ och $V(X) = 3\sigma^2$, $V(Y) = 2\sigma^2$. Två skattningar har föreslagits

$$\theta^*_{obs} = \frac{1}{2}(x+y) \quad \text{och} \quad \hat{\theta}_{obs} = \frac{1}{5}(2x+3y)$$

effektivitet, väntevärdesriktighet (ite olika svarsalternativ)

Lösning:

1 Väntevärdesriktiga?

$$E(\theta^*) = \{=\theta?\} = E\left(\frac{1}{2}X + \frac{1}{2}Y\right) = \frac{1}{2}E(X) + \frac{1}{2}E(Y) =$$

$$= \frac{1}{2}\theta + \frac{1}{2}\theta = \theta \quad \text{vvr!}$$

$$E(\hat{\theta}) = E\left(\frac{2}{5}X + \frac{3}{5}Y\right) = \frac{2}{5}E(X) + \frac{3}{5}E(Y) = \frac{2}{5}\theta + \frac{3}{5}\theta = \theta$$

vvr!

- båda skattningar är väntevärdesriktiga, kollar varians för att bedömma effektivitet

$$V(\theta^*) = V\left[\frac{1}{2}X + \frac{1}{2}Y\right] = \frac{1}{4}V(X) + \frac{1}{4}V(Y) = \frac{1}{4}3\sigma^2 + \frac{1}{4}2\sigma^2 =$$

$$= \frac{5}{4}\sigma^2 = 1.25\sigma^2$$

$$V(\hat{\theta}) = V\left[\frac{2}{5}X + \frac{3}{5}Y\right] = \frac{4}{25}V(X) + \frac{9}{25}V(Y) =$$

$$= \frac{12}{25}\sigma^2 + \frac{18}{25}\sigma^2 = \frac{30}{25}\sigma^2 = 1.2\sigma^2$$

Slutsats: $V(\hat{\theta}) < V(\theta^*)$ $\Rightarrow$ $\hat{\theta}_{obs}$ är mer effektiv skattning än θ^*_{obs}

tenta 2019-04-17 #7

Antag att $X_1, X_2 \ldots X_5$ är oberoende s.v. sådana att $X_i \in N(\mu, \sigma)$. Två skattningar har föreslagits (av μ)

$\theta^*_{obs} = X_i$ och $\hat{\theta}_{obs} = \frac{1}{5}\sum_{i=1}^{5} X_i$. Vilken skattning är effektivast?

- kollar först väntevärdesriktighet:

$E(\theta^*) = E(X_i) = \mu$ ok!

$E(\hat{\theta}) = E\left(\frac{1}{5}\sum X_i\right) = \frac{1}{5} \cdot 5\, E(X_i) = \mu$ ok!

båda skattningar är vvr, kollar variansen

mer effektiv $\Leftrightarrow$ mindre varians.

$V(\theta^*) = V(X_i) = \sigma^2$ (pga N-fördelning)

$V(\hat{\theta}) = V\left(\frac{1}{5}\sum X_i\right) = \frac{1}{25} V\left(\sum X_i\right) = \frac{1}{25} \cdot 5\sigma^2 = \frac{\sigma^2}{5}$

$V(\hat{\theta}) < V(\theta^*)$ $\Rightarrow$ $\hat{\theta}_{obs}$ är mer effektiv skattning av μ.

- ganska logiskt att göra 5 mätningar och ta medelvärdet istället för att göra 1 mätning.

flera mätningar $\Rightarrow$ mer effektiv skattning.

Maximum-likehood-metoden (ML)

- bygger på att man antar att man kan sannolikhetsfunktionen eller täthetsfunktionen

- idén: eftersom man har fått det mätdata man fick så borde sannolikheten vara stor att få just dessa mätdata.

  - maximera denna sannolikhet map den parameter som vi vill skatta

• Det värde θ^*_{obs} för vilket $L(\theta)$ antar sitt största värde inom Ω_θ kallas för ML-skattningen av θ.

• Vid diskret fördelning:

$$L(\theta) = p_{X_1, X_2 \ldots X_n}(x_1, x_2 \ldots x_n; \theta) = \{\text{om oberoende}\} =$$

$$= p_{X_1}(x_1; \theta) \cdot p_{X_2}(x_2; \theta) \cdots p_{X_n}(x_n; \theta)$$

beror på θ

• Vid kontinuerlig fördelning:

$$L(\theta) = f_{X_1, X_2 \ldots X_n}(x_1, x_2 \ldots x_n; \theta) - \{\text{om oberoende}\} =$$

$$= f_{X_1}(x_1; \theta) \cdots f_{X_n}(x_n; \theta)$$

"min/max problem" från envariabelanalyskursen. Hitta maximum till en funktion.

Uppgift 11.10 (boken)

Den diskreta s.v. X har sannolikhetsfunktionen $p_X(k) = \theta(1-\theta)^{k-1}$ för $k = 1, 2, 3, \ldots$, $0 < \theta < 1$. Har ett slumpmässigt stickprov från denna fördelning: 4, 5, 4, 6, 4, 1

a) Skriv upp likelihood funktionen $L(\theta)$

b) Ange för vilket θ $L(\theta)$ är störst dvs vad är ML-skattningen av θ?

a) 4, 5, 4, 6, 4, 1 är värde på k, antar att vi har oberoende observationer (stickprov)

$$L(\theta) = p_X(x_1) \times p_{X_2}(x_2) \cdots p_{X_6}(x_6)$$

$$L(\theta) = \theta(1-\theta)^{4-1}\ \theta(1-\theta)^{5-1}\ \theta(1-\theta)^{4-1}\ \theta(1-\theta)^{6-1} \times$$
$$\times\ \theta(1-\theta)^{4-1}\ \theta(1-\theta)^{1-1} =$$
$$= \theta(1-\theta)^3\ \theta(1-\theta)^4\ \theta(1-\theta)^3\ \theta(1-\theta)^5\ \theta(1-\theta)^3\ \theta \cdot 1 =$$
$$= \theta^6(1-\theta)^{18} = L(\theta)$$

b) Söker maximum av funktionen, dvs vill hitta en kritisk punkt till $L(\theta)$:

$$\frac{dL(\theta)}{d\theta} = 0$$

$$L'(\theta) = \{\text{produktregeln + kedjeregeln}\} = 6\theta^5(1-\theta)^{18} +$$
$$+\ \theta^6\, 18(1-\theta)^{17}(-1) =$$
$$= 6\theta^5(1-\theta)^{18} - 18\theta^6(1-\theta)^{17} = 0$$

både $\theta = 0$ och $\theta = 1$ är lösningar till ekvationen men ingår inte i vårän definitionsmängd, vi kan dividera bort θ och $(1-\theta)$

- dividerar vänna ekvation $6\theta^5(1-\theta)^{18} - 18\theta^6(1-\theta)^{17} = 0$

med gemensamma faktor: $6\theta^5(1-\theta)^{17}$:

$$(1-\theta) - 3\theta = 0$$

$$1 - \theta = 3\theta$$

$$1 = 4\theta \Rightarrow \theta = \frac{1}{4} \Rightarrow \hat{\theta}_{obs\ ML} = \frac{1}{4}$$

Uppgift 11.11

Den s.v. X har täthetsfunktionen $f_X(x) = \theta(1+x)^{-\theta-1}$ för $x \geq 0$. θ är antingen 2, 3, eller 4. Slumpmässigt prov ger utfall 0.2 och 0.8.

a) Ange L-funktionens värde för alla θ-värdena

b) Ange ML-skattningen av θ.

a) generella funktionen: $L(\theta) = \{\text{oberoende}\} =$

$$= \theta(1+1.8)^{-\theta-1}\ \theta(1+0.2)^{-\theta-1} = \theta^2(1.8)^{-\theta-1}(1.2)^{-\theta-1}$$

$$L(2) = 2^2(1.8)^{-2-1}(1.2)^{-2-1} = 4(1.8)^{-3}(1.2)^{-3} = 0.397$$

$$L(3) = 3^2(1.8)^{-3-1}(1.2)^{-3-1} = 9(1.8)^{-4}(1.2)^{-4} = 0.40$$

$$L(4) = 4^2(1.8)^{-4-1}(1.2)^{-4-1} = 16(1.8)^{-5}(1.2)^{-5} = 0.34$$

b) $L(3)$ ger störst värde $\Rightarrow \hat{\theta}_{obs,ML} = 3$ är ML-skattningen.

exempel

Anta att vi vet att X_i:na är oberoende och Poisson-fördelade dvs $X_i \in Po(\mu)$. Vi får 5 mätdata: 10, 12, 7, 10, 4. Vad är ML-skattningen av μ?

$$L(\mu) = P(X_1 = 10 \cap X_2 = 12 \cap X_3 = 7 \cap X_4 = 10 \cap X_5 = 4) =$$

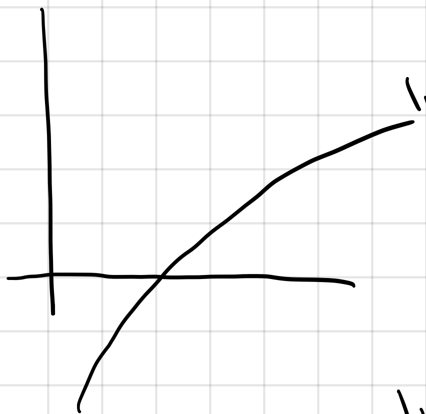
$$= \{\text{oberoende}\} = p_{X_1}(10)\, p_{X_2}(12)\, p_{X_3}(7)\, p_{X_4}(10)\, p_{X_5}(4) =$$

$$= \frac{\mu^{10}}{10!} e^{-\mu} \cdot \frac{\mu^{12}}{12!} e^{-\mu} \cdot \frac{\mu^{7}}{7!} e^{-\mu} \cdot \frac{\mu^{10}}{10!} e^{-\mu} \cdot \frac{\mu^{4}}{4!} e^{-\mu} =$$

$$= \frac{\mu^{43} e^{-\mu}}{10!\,12!\,7!\,10!\,4!}$$

- har väldigt jobbig derivata + stora tal.

- enklaste metoden att lösa är mha logaritmering:



ln(x), strängt växande funktion:

$$x_1 > x_2 \iff \ln(x_1) > \ln(x_2)$$

$$f(x_1) > f(x_2) \iff \ln[f(x_1)] > \ln[f(x_2)]$$

- ln förändrar inte

- Där $\ln[f(x)]$ har sitt max har även $f(x)$ sitt max.

då funktionen är strängt växande och kontinuerlig.

$$\ln[L(\mu)] = \ln\left[\frac{\mu^{43} e^{-5\mu}}{10!\,12!\,7!\,10!\,4!}\right]$$

logaritmlagar ger oss:

$$\ln[L(\mu)] = \ln(\mu^{43}) + \ln(e^{-5\mu}) - \ln(10!\,12!\,7!\,10!\,4!) =$$
$$= 43\ln(\mu) - 5\mu - \ln(10!\,12!\,7!\,10!\,4!)$$

$$\frac{d}{d\mu}\ln[L(\mu)] = \frac{43}{\mu} - 5 = 0$$

$$\frac{43}{\mu} - 5 = 0 \Rightarrow \mu = \frac{43}{5} = 8.6$$

$$\mu^{*}_{obs,ML} = \frac{43}{5}$$

- Vi har produkt av flera termer som beror på μ och en stor konstant. kan logaritmera och använda logaritmlagar för att förenkla derivatan.

  ⇒ konstanten försvinner, e försvinner.

  - bra att göra om vi har stora tal och e-funktionen.

Minsta-kvadrat metoden (MK)

- Går ut på att minimera skillnaden mellan praktiska och teoretiska värdena

$$Q = \sum_{i=1}^{n}(x_i - \mu_i(\theta_1, \theta_2 \ldots \theta_k))^2$$

$X_1, X_2 \ldots X_k$ är oberoende och $\theta_1, \theta_2 \ldots \theta_k$ är okända parametrar

$E(x_i) = \mu_i(\theta_1, \theta_2 \ldots \theta_k)$

x_i - praktiskt värde

μ_i - teoretiskt värde

- minde Q = mindre skillnad mellan praktiska och teoretiska värdena

- minsta Q vi kan få = minsta avvikelsen = bästa värdet eftersom den avviker minst.

- de $\theta_1 \dots \theta_k$ som minimerar Q väljs som MK-skattningar av $\theta_1 \dots \theta_k$

- precis som innan ett max/min problem och vi ska hitta minimum.

tenta 2019-12-19 # 8

Antag att $X \in Exp(\lambda)$. Vi vill skatta λ mha minsta-kvadrat metoden. Vi gör 3 oberoende försök och får stickprovet $x_1 = 18$, $x_2 = 34$, $x_3 = 12$. Bestäm MK-skattningen av λ.

$$Q = \sum_{i=1}^{n} (x_i - \mu_i(\theta))^2$$

?

$$X_i \in Exp(\lambda) \Rightarrow E(X_i) = \frac{1}{\lambda}$$

$$Q = \sum_{i=1}^{3} \left(x_i - \frac{1}{\lambda}\right)^2 = \left(18 - \frac{1}{\lambda}\right)^2 + \left(34 - \frac{1}{\lambda}\right)^2 + \left(12 - \frac{1}{\lambda}\right)^2$$

- hitta minsta värde: $\frac{dQ}{d\lambda} = 0$

$$Q' = \{\text{kedjeregeln}\} = 2\left(18 - \frac{1}{\lambda}\right)\left(\frac{1}{\lambda^2}\right) + 2\left(34 - \frac{1}{\lambda}\right)\left(\frac{1}{\lambda^2}\right) +$$

$$+ 2\left(12 - \frac{1}{\lambda}\right)\left(\frac{1}{\lambda^2}\right) = 0$$

$$\frac{2}{\lambda^2}\left[13-\frac{1}{\lambda}+34-\frac{1}{\lambda}+12-\frac{1}{\lambda}\right]=0$$

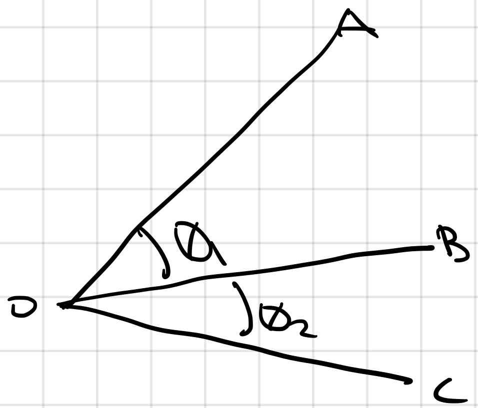
$$\frac{2}{\lambda^2}\left[64-\frac{3}{\lambda}\right]=0$$

$\lambda^2 \neq 0 \Rightarrow 64-\frac{3}{\lambda}=0$, $64=\frac{3}{\lambda} \Rightarrow \lambda=\frac{3}{64}=0.0469$

$\lambda^*_{obs,mk}=0.0469$

Uppgift 11.18

Man har gjort 3 oberoende mätningar av vinkeln AOC och erhållit mätvärdena x_1, x_2 och x_3. Vi har också gjort två oberoende mätningar av vinkeln AOB och erhållit x_4, x_5. Vid varje vinkelmätning fås ett slumpmässigt fel som har väntevärdet 0 och standardavvikelse σ. Anta att vinklarna är $\theta_1+\theta_2$ resp. θ_1



a) Bestäm MK skattningar av θ_1 och θ_2

b) Är skattningarna väntevärdesriktiga?

c) Beräkna variansen för skattningar.

a) $Q=\sum_{i=1}^{n}(x_i-\mu(\theta_1,\theta_2))^2$

AOC: x_1, x_2 och x_3 anta att vinkeln är $\theta_1+\theta_2$

AOB: x_4, x_5 anta att vinkeln är θ_1

X_1, X_2, X_3, X_4, X_5 är uppmätta vinklar och θ_1 och θ_2 är μ.

$$\Rightarrow Q = (x_1 - (\theta_1 + \theta_2))^2 + (x_2 - (\theta_1 + \theta_2))^2 + (x_3 - (\theta_1 + \theta_2))^2 +$$
$$+ (x_4 - \theta_1)^2 + (x_5 - \theta_1)^2$$

- har Q som beror på 2 okända parametrar θ_1 och θ_2. Minimum av Q fås då $\operatorname{grad} Q = \left(\frac{\partial Q}{\partial \theta_1}, \frac{\partial Q}{\partial \theta_2}\right) = \bar{0} = (0, 0)$

- flervariabel analys problem.

$$\left.\begin{array}{l} \frac{\partial Q}{\partial \theta_1} = 0 \\ \\ \frac{\partial Q}{\partial \theta_2} = 0 \end{array}\right\}$$

MK skattningar av θ_1 och θ_2, vi tar partiella derivator.

$$\frac{\partial Q}{\partial \theta_1} = 2(x_1 - \theta_1 - \theta_2)(-1) + 2(x_2 - \theta_1 - \theta_2)(-1) +$$
$$+ 2(x_3 - \theta_1 - \theta_2)(-1) + 2(x_4 - \theta_1)(-1) + 2(x_5 - \theta_1)(-1) =$$
$$= -2\sum_{i=1}^{5} x_i + 10\theta_1 + 6\theta_2 = 0 \qquad (e1)$$

$$\frac{\partial Q}{\partial \theta_2} = 2(x_1 - \theta_1 - \theta_2)(-1) + 2(x_2 - \theta_1 - \theta_2)(-1) +$$
$$+ 2(x_3 - \theta_1 - \theta_2)(-1) + 0 + 0 =$$
$$= -2\sum_{i=1}^{3} x_i + 6\theta_1 + 6\theta_2 = 0 \qquad (e2)$$

$$\Rightarrow \begin{cases} -2(x_1+x_2+x_3+x_4+x_5)+10\theta_1+6\theta_2=0 & (e1) \\ -2(x_1+x_2+x_3)+6\theta_1+6\theta_2=0 & (e2) \end{cases}$$

θ_1: substraherar (e2) från (e1): $-2x_4-2x_5+4\theta_1=0$

$$\Rightarrow \theta_1=\frac{1}{2}(x_4+x_5)=\theta^*_{1\,\text{ons mk}}$$

θ_2: substituerar θ_1 i (e1):

$$-2x_1-2x_2-2x_3-2x_4-2x_5+5(x_4+x_5)+6\theta_2=0$$

$$3x_4+3x_5-2x_1-2x_2-2x_3+6\theta_2=0$$

$$6\theta_2=2x_1+2x_2+2x_3-3x_4-3x_5$$

$$\theta_2=\frac{1}{3}(x_1+x_2+x_3)-\frac{1}{2}(x_4+x_5)=\theta^*_{2\,\text{ons mk}}$$

b) $E(\theta_1^*)\overset{?}{=}\theta_1$

$$E(\theta_1^*)=E\left[\frac{1}{2}x_4+\frac{1}{2}x_5\right]=\frac{1}{2}E(x_4)+\frac{1}{2}E(x_5)=\frac{1}{2}\theta_1+\frac{1}{2}\theta_1=\theta_1$$

$\theta^*_{1\,\text{ons mk}}$ är väntevärdesriktig

$$E(\theta_2^*)=E\left[\frac{1}{3}x_1+\frac{1}{3}x_2+\frac{1}{3}x_3-\frac{1}{2}x_4-\frac{1}{2}x_5\right]\overset{?}{=}$$

$$=\frac{1}{3}(\theta_1+\theta_2)+\frac{1}{3}(\theta_1+\theta_2)+\frac{1}{3}(\theta_1+\theta_2)-\frac{1}{2}\theta_1-\frac{1}{2}\theta_1=\theta_2$$

$\theta^*_{2\,\text{ons, mk}}$ är VVR!

c) $V(\theta_1^*) = V\left[\frac{1}{2}X_4 + \frac{1}{2}X_5\right] = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2 = 0.5\sigma^2$

$V(\theta_2^*) = V\left[\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 - \frac{1}{2}X_4 - \frac{1}{2}X_5\right] =$

$= \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{5}{6}\sigma^2$

TJABBA TJENA HALLÅ!

SANN-STAT

CRASH-COURSE

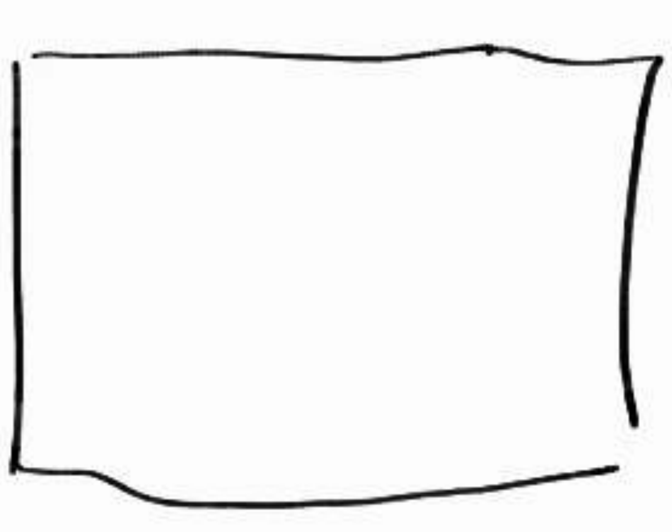
MED RICKE

$p(A) = P(B)$

$$\int_b^a x \cdot f_x(x)\,dx =$$

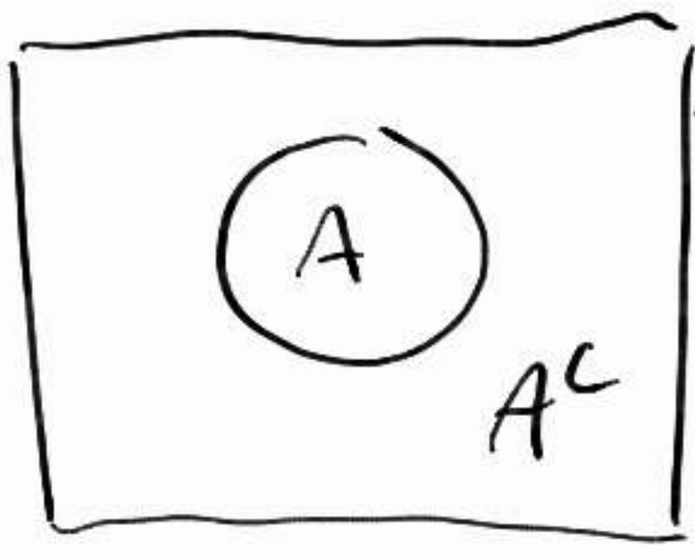
	D	I
+	x	f_x
−	1	F_x
+	0	$\int F_x$
−		

Ω = utfallsrum $\ni \omega_1, \omega_2$ = utfall



Händelser: A, B ...

= en samling av utfall



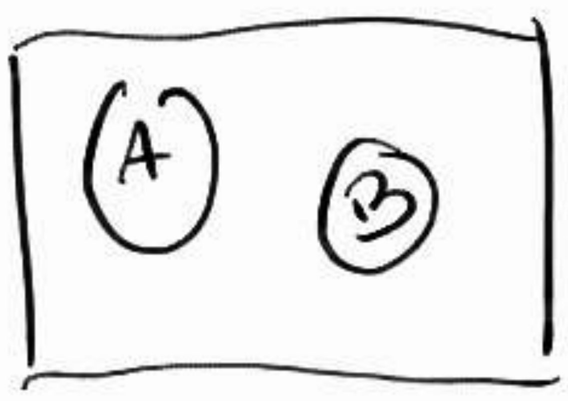
Ω UTFALLSRUMMET

A

A^C, A^* = KOMPLEMENT.

$P(A)$

$P(A^C)$



A & B ÄR DISJUNKTA.

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

LÅT A, B VA HÄNDELSER

SÅ ATT $P(A) = .6$

$P(B) = .7$

$P(A \cup B) = .8$

$P(A \cap B) = ?$

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

$.8 = .6 + .7 - x \Rightarrow x = .7 + .6 - .8 = .05$

$A \cap B$

SNITT

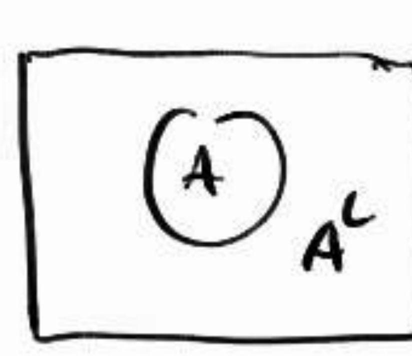
$A \cup B$

UNION A & B

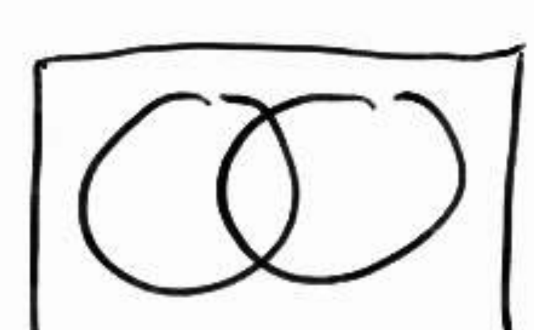
TVÅ HÄNDELSER $P(A \cup B) = ?$

ADDITIONSSATSEN:

A & B



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



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

$P(A \cap B) = .4$

$P(A) = .5$

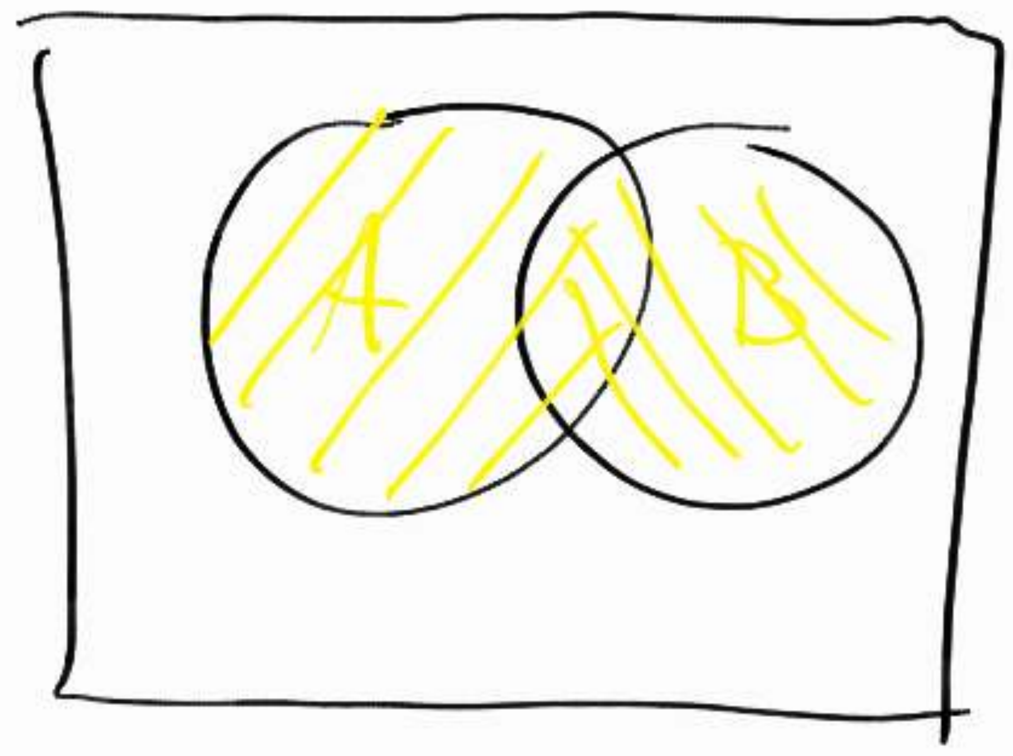
$P(A^C \cap B) = .1$ DISJUNKTA

$P(A \cup B) = P(A) + P(A^C \cap B)$

$.6 \quad .5 + .1$

7.32

A o B OBEROENDE



$P(A) = .1$

$P(B) = .05$

$P(A^c \cap B^c) = ?$

$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)]$

$= 1 - [P(A) + P(B) - P(A) \times P(B)] = (1 - P(A))(1 - P(B))$

$= (1 - .1)(1 - .05) = .855$:)

OBEROENDE:

OM A o B OBEROENDE:

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

REGN = ~30%

TORSDAG = ~14%

$\sim 0.3 \cdot .14 =$

$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2$

$3! = 3 \cdot 2 \cdot 1$

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

$\binom{10}{3} = \frac{10!}{3! \cdot 7!} = \frac{\overbrace{10 \cdot 9 \cdot 8}^{\text{3 FÖRSTA}}}{3!}$

3 FÖRRÄTTER
×
4 HUVUDRÄTT
×
2 EFTERRÄTTER
=
24 KOMBINATIONER

DRAGNING UTAN ÅTERLÄGGNING

KORTLEK (52)

$52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = \binom{52}{5}$

$$\boxed{\frac{\text{GYNNSAMMA}}{\text{MÖJLIGA}} = \frac{13}{52}}$$

2.18 KORTLEK

a) ESS, KUNG DAM, KNEKT, 10:a i SAMMA FÄRG

b) FEM KORT I FÖLJD, SAMMA FÄRG

c) FEM KORT I SAMMA FÄRG

MÖJLIGA UTFALL = ? $\boxed{\binom{52}{5}}$

a) $4 \cdot 1$ = gynnsamma utfall ⇒ SANN = $\frac{4}{\binom{52}{5}}$

b) $(2,3,4,5,6) - \underbrace{(3,4,5,6,7), (4,5,6\ldots)\ldots}_{10}$

gynnsamma $= 4 \cdot 10 = 40$ ⇒ SANNOLIKHETEN =

c) $4 \cdot \binom{13}{5}$ = gynnsamma utfall.

$$c = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}}$$

2.33

FAMILJ A, B, C BJUDNA, OBEROENDE

SANNOLIKHET ATT DE KOMMER = .8, .6, .9

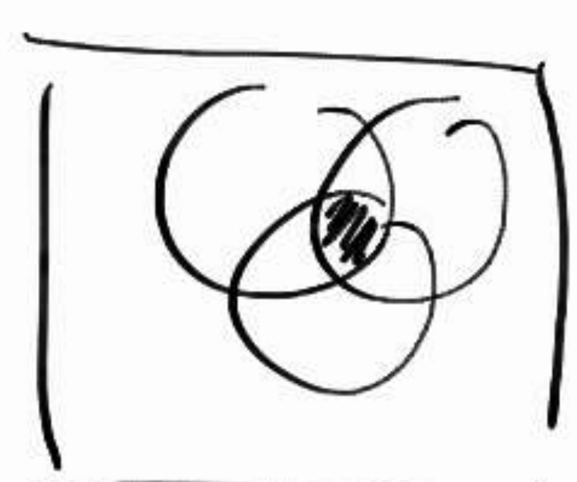
a) ALLA KOMMER

b) INGEN KOMMER

c) MINST EN KOMMER

OM A, B OBEROENDE:

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



a) $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

$\boxed{0.432} = .8 \times .6 \times .9$

b) $P(A^c \cap B^c \cap C^c) = P(A^c) \cdot P(B^c) \cdot P(C^c)$

$0.008 = .2 \cdot .4 \cdot .1$

c) $1 - 0.008 = 0.992$:)

$P(B|A)$

	RÖKER	EJ RÖK	
MAN	20	80	100
KVINNA	50	100	150
			250
			Σ

B = RÖKARE

A = MAN

$$P(B|A) = \frac{80}{100} = \frac{80/250}{100/250} = \frac{P(A \cap B)}{P(A)}$$

RAST TILL 9:15 :)

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

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

2.22 DRA SLUMPMÄSSIGT UTAN ÅTERLÄGGNING 4 KORT UR KORTLEK

a) OM 3 FÖRSTA HJÄRTER, VAD ÄR SANN ATT DEN 4:E EJ ÄR HJÄRTER?

52 - 3 = 49 KVAR

13 - 3 = 10 KVAR

$P(4{:}e \text{ HJÄRTER} \mid 3 \text{ FÖRSTA HJÄRTER}) = \frac{10}{49}$

b) SANN 3 FÖRSTA ♡ 4:E ♠

$\overset{\shortparallel}{A}$ $\overset{\shortparallel}{B}$

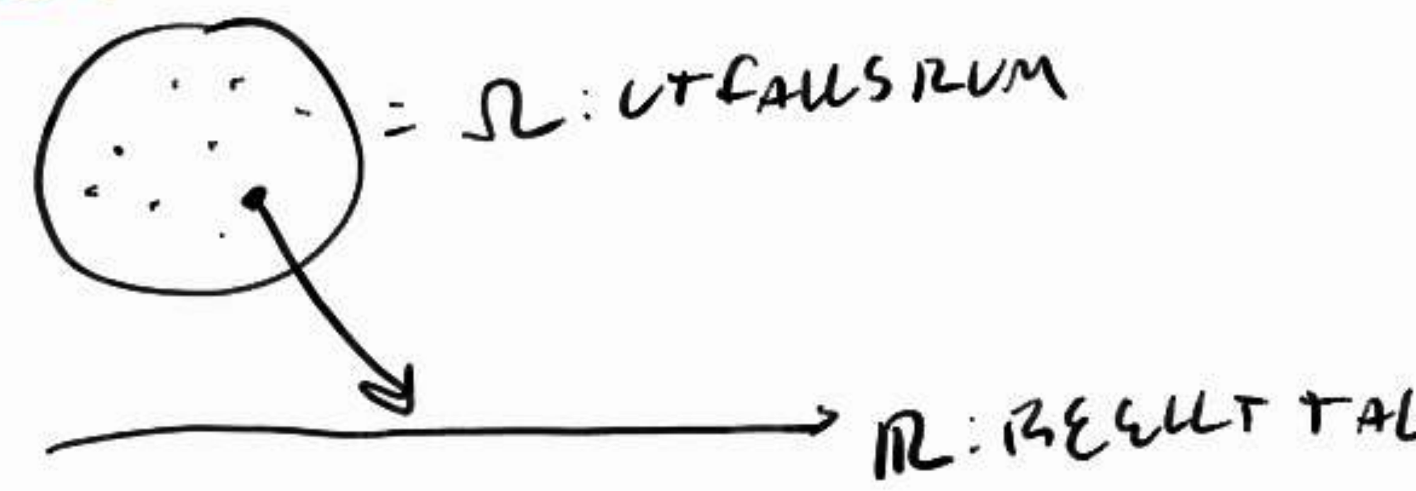
$P(A \cap B)$

$B = \frac{13}{49}$

$P(A) = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} = \frac{11}{850}$

$P(B|A)\,P(B) = P(A \cap B) = \frac{13}{49} \cdot \frac{11}{850}$

STOKASTISKA VARIABLER:
S.V.



Ω: ALLA MÄN I SVERIGE

ω_1: EN MAN

$X(\omega_1) \to \mathbb{R}$

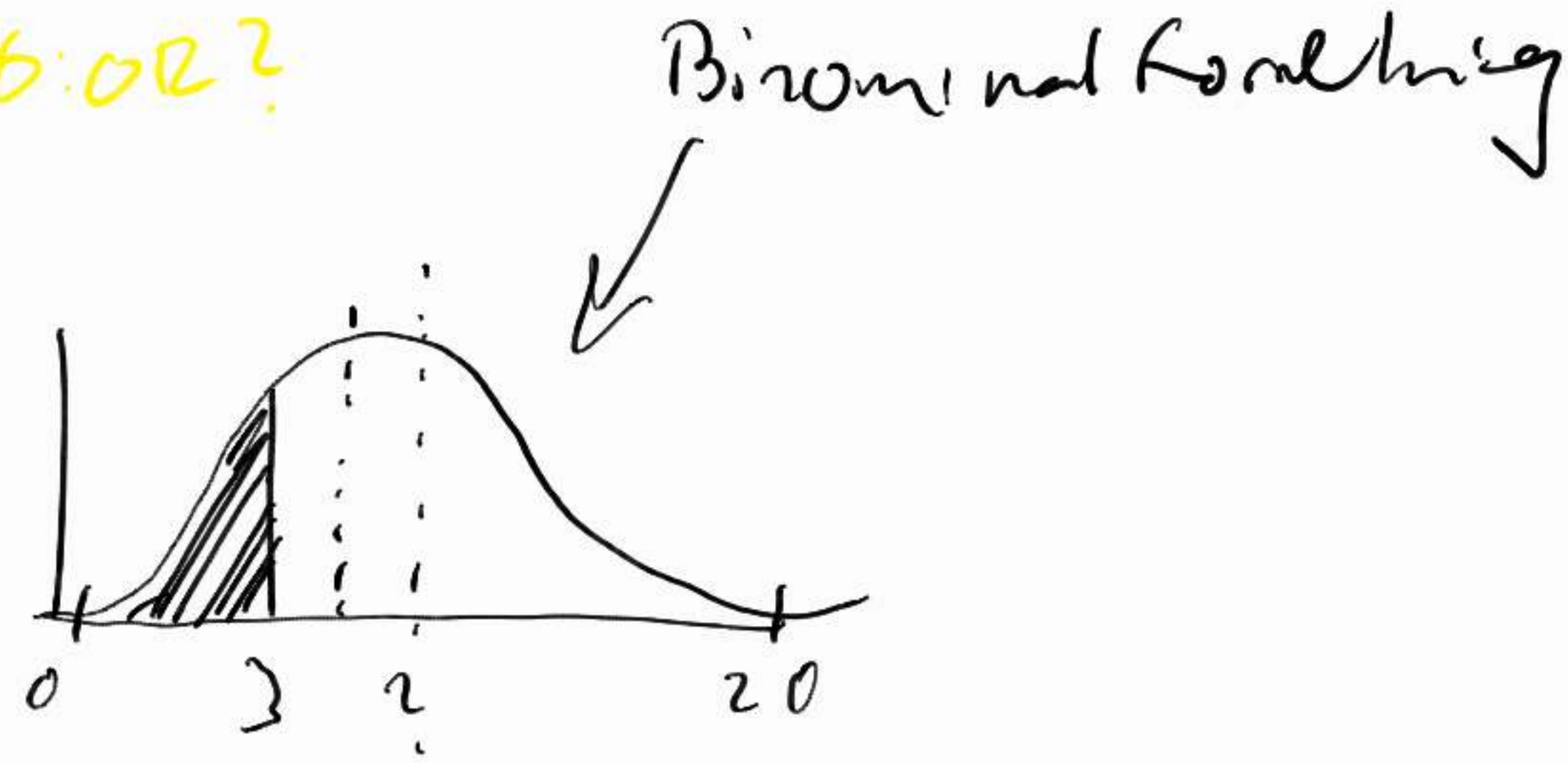
TAR EN PERSON → GENERERAR EN VIKT (KG)

$P(X > 70)$: SANNOLIKHET ATT EN SLUMPAD MAN I SVERIGE VÄGER ÖVER 70 KG

Ω: ALLA MÖJLIGA RESULTAT TÄRNINGSKAST (20 ST)

X: HUR MÅNGA 6:OR?

6^{20} i Ω



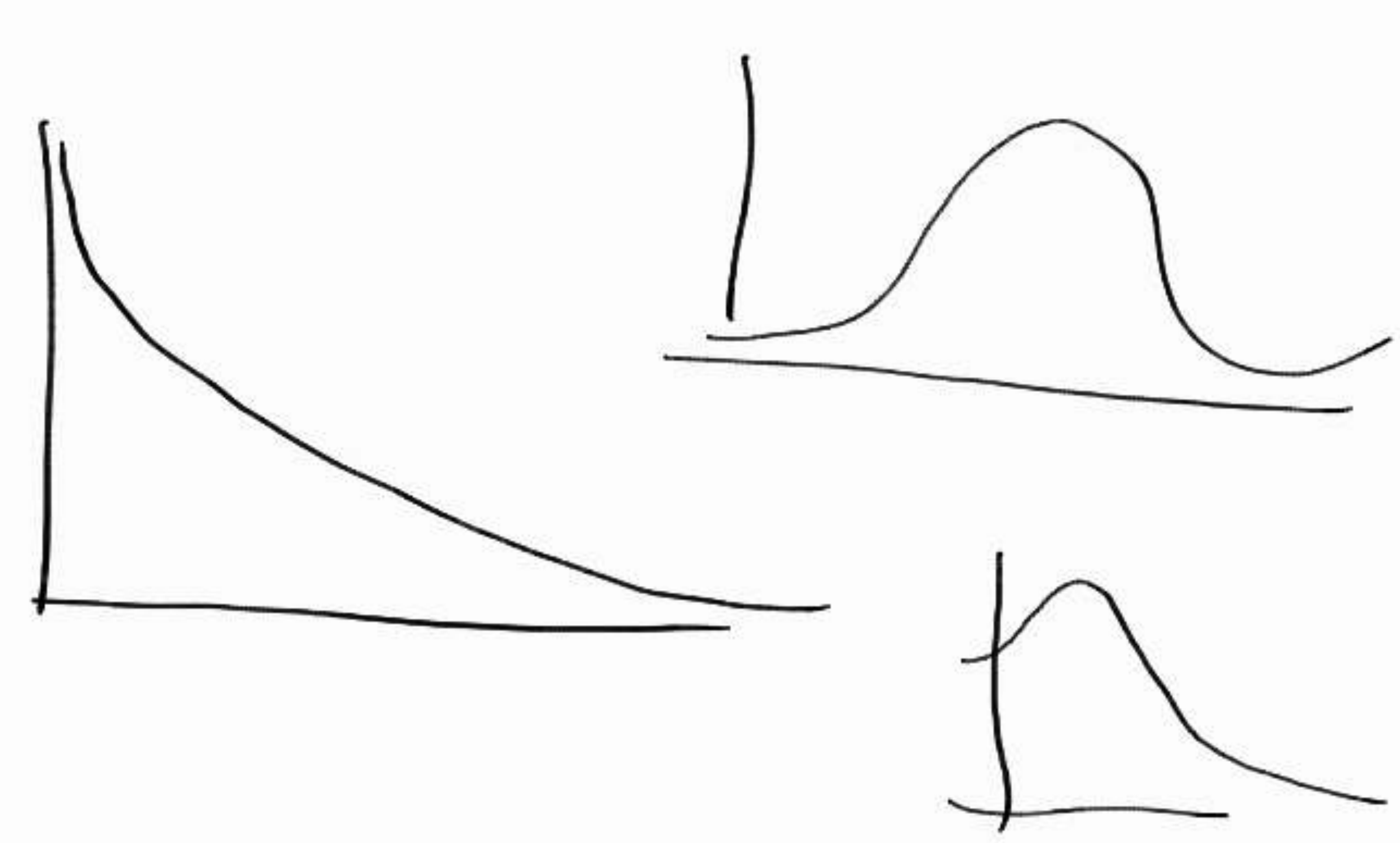
$P(X=0) =$

SANNOLIKHET ATT få 0 6:OR

$X = 0, 1, 2 \ldots 19, 20$

DISKRETA S.V.

KAN ENDAST ANTA ÄNDLIGT (ELLER UPPRÄKNELIGT OÄNDLIGT) ANTAL VÄRDEN



$P(X=0) = p_X(0)$

$P(X=k) = p_X(k)$

FÖR FÖRSTA GÅNGEN-FÖRDELNING:

$$p_X(k) = (1-p)^{k-1} \cdot p \qquad k = 1, 2, 3 \ldots \qquad 0 < p < 1 \qquad \Rightarrow X \in ffg(p)$$

UTFALLSRUM(Ω) SLÅ TÄRNINGAR

X: ANTAL SLAG TILL O MED FÖRSTA 6:an

$$P(X=1) \overset{?}{=} \frac{1}{6}$$

$$P(X=2) = \frac{5}{6} \cdot \frac{1}{6}$$

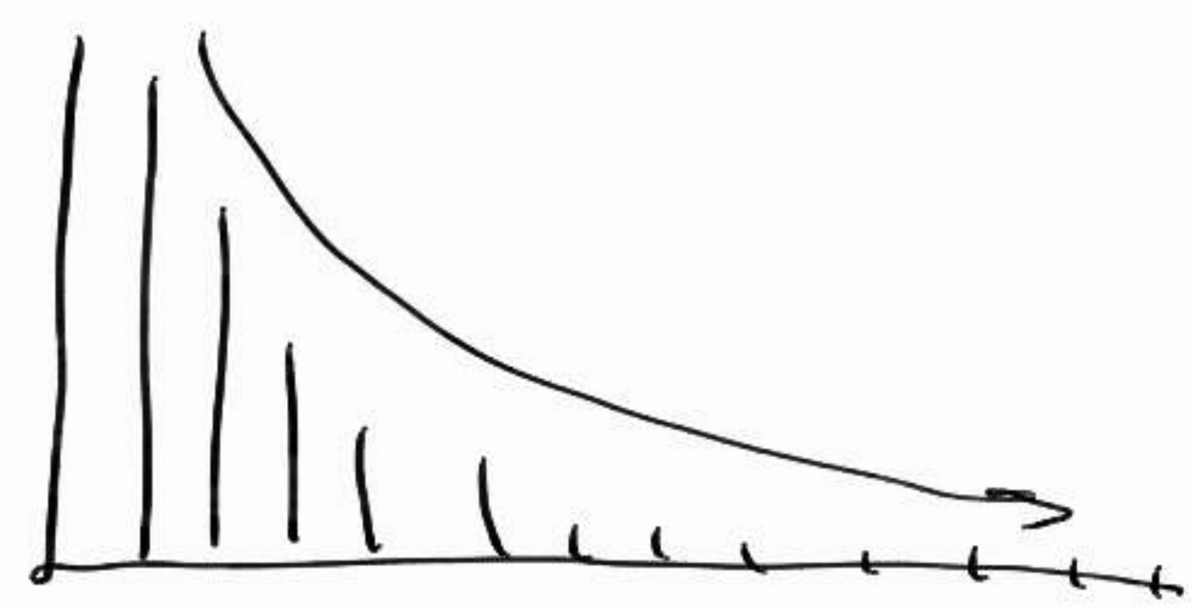
$$P(X=3) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$$

$$P(X=k) = \left(\frac{5}{6}\right)^{k} \cdot \frac{1}{6} \Rightarrow p_X(k) = \left(1-\frac{1}{6}\right)^{k} \cdot \frac{1}{6}$$

$p = \frac{1}{6}$

SINGLA SLANT

$\frac{1}{2}$



POISSON-FÖRDELNING:

OM X där $p_X(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ för $k = 0, 1, 2 \ldots$

och $\lambda > 0$

DÅ ÄR $X \in Po(\lambda)$

λ SÖNDERFALL UNDER 1 TIDSENHET (SEKUNDER)

$t \cdot \lambda$ SÖNDERFALL UNDER t SEKUNDER

OM DET I GENOMSNITT SKER

λ SÖNDERFALL UNDER EN TIDSENHET, OBEROENDE

SÅ ÄR SANNOLIKHETEN FÖR k SÖNDERFALL $= p_X(k)$

TENTA 2017 - JAN

UPPG: 1:

ANTAL TÅG SOM ÄR FÖRSENADE ANTAS VA Po(2)-förd

λ

- BESTÄM SANNOLIKHETEN ATT MINST 3 TÅG ÄR FÖRSENADE

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0{,}67 \simeq 0{,}32$$

$$\parallel$$

$$1 - (P(X=0) + P(X=1) + P(X=2)) \qquad P_X(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$$\parallel$$

$$=$$

$$P_X(0) = \frac{2^0}{0!} \cdot e^{-2} = \frac{1}{1} \cdot e^{-2} = \frac{1}{e^2}$$

$$P_X(1) = \frac{2^1}{1!} \cdot e^{-2} = \frac{2}{1} \cdot e^{-2} = \frac{2}{e^2} \quad = \sim 0{,}67$$

$$P_X(2) = \frac{2^2}{2!} \cdot e^{-2} = \frac{4}{2} \cdot e^{-2} = \frac{2}{e^2}$$

BINOMINALFÖRDELNINGEN:

Om $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, 2 \ldots n$

$0 < p < 1$

DÅ ÄR X Binominal fördelad. $= X \in Bin(n, p)$

TEX ~~X~~ KRONA NÄR MAN SINGLAR SLANT :)

UPPGIFT 7

TÄRNING KASTAS 4 GÅNGER

SANN ATT EXAKT 3 ÄR EN 6:a

$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6}$:) $\quad \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \quad \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}, \quad \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$

$4 \cdot \left(\frac{1}{6}^3 \cdot \frac{5}{6}^1\right) \quad p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$p_X(3) = \underset{\substack{\shortparallel \\ 4}}{\binom{4}{3}} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^1$$

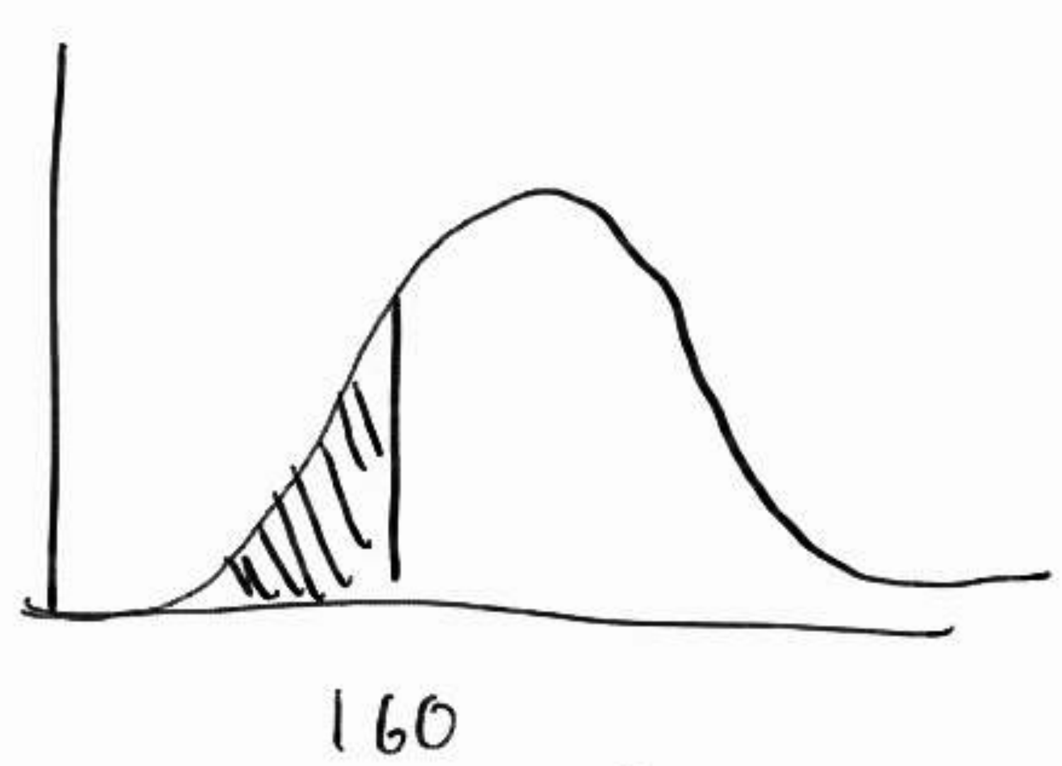
FELDFRÅGA

10 MYNT 3 KLAVE

$$p_X(3) = \underset{\substack{\shortparallel \\ 120}}{\binom{10}{3}} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^7 = \frac{120}{2^{10}} \approx 0{,}12\%$$

PAUS 10:20 :-)

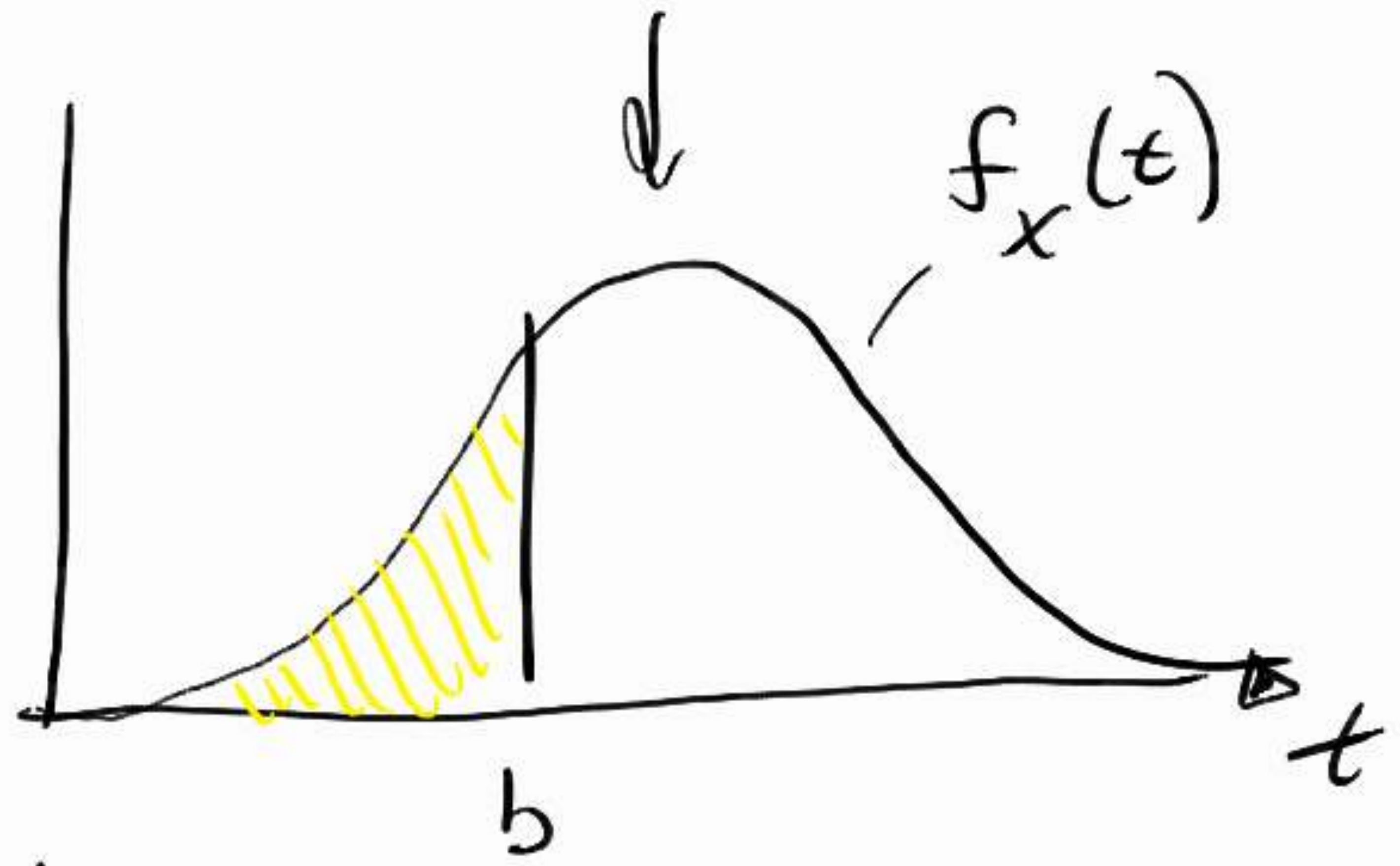
KONTINUERLIGA STOKASTISKA VARIABLER!



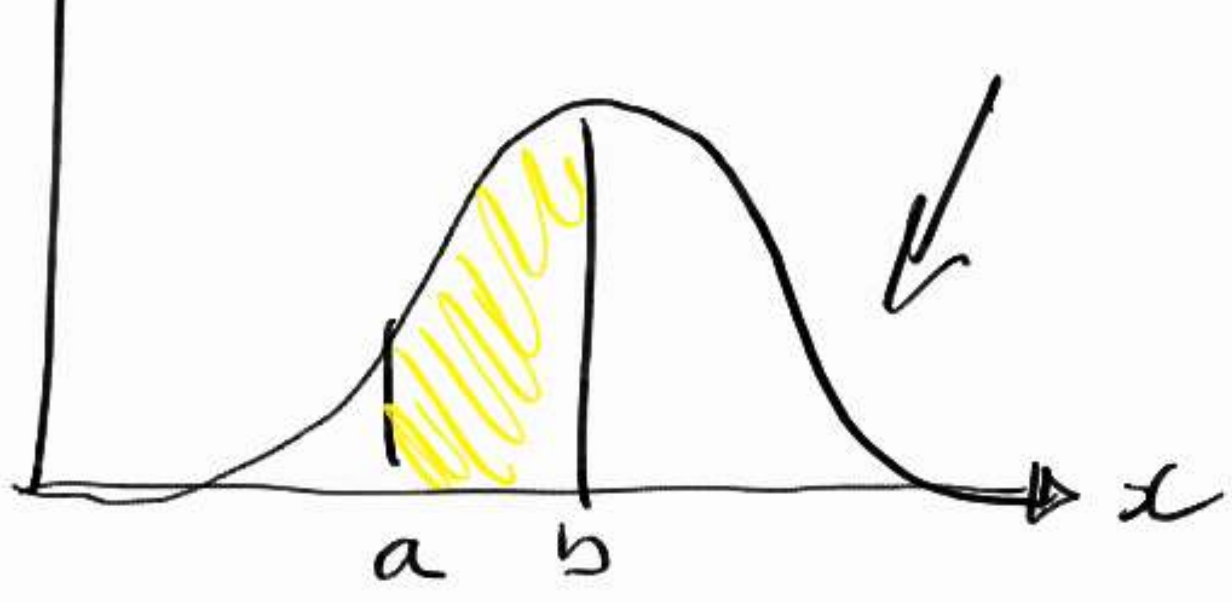
$P(X < 160) =$

$P(X = 160) + P(X = 159) + \dots$

$P(159{,}5) = ?$



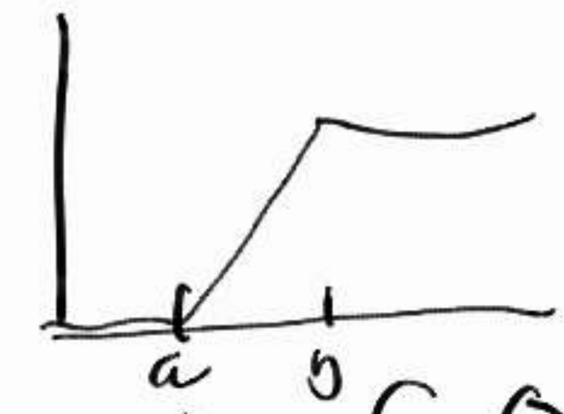
$$= \int_{-\infty}^{b} f_X(t)\,dt$$



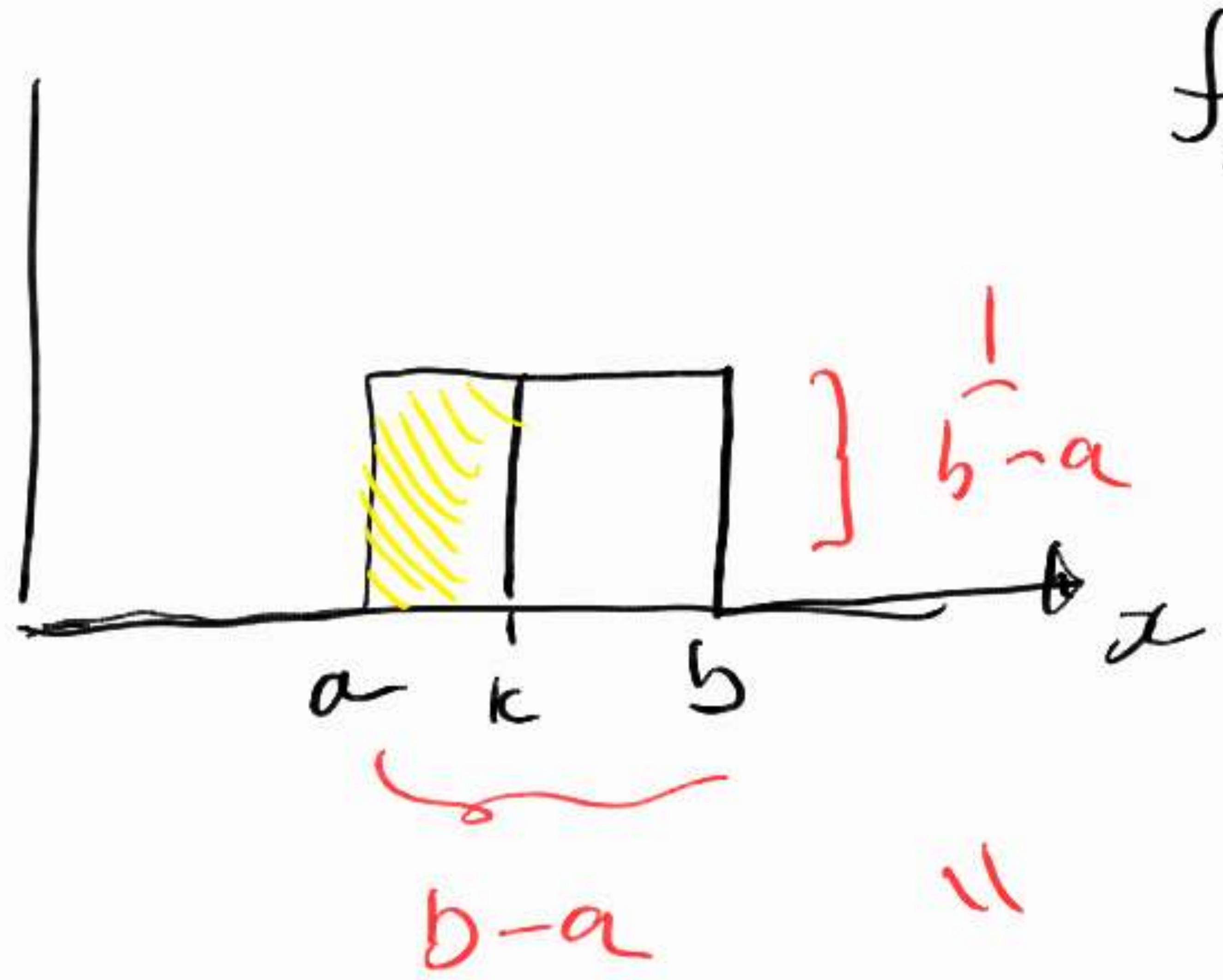
$$= \int_{a}^{b} f_X(x)\,dx$$

FÖRDELNINGAR:

LIKFORMIG FÖRDELNING:

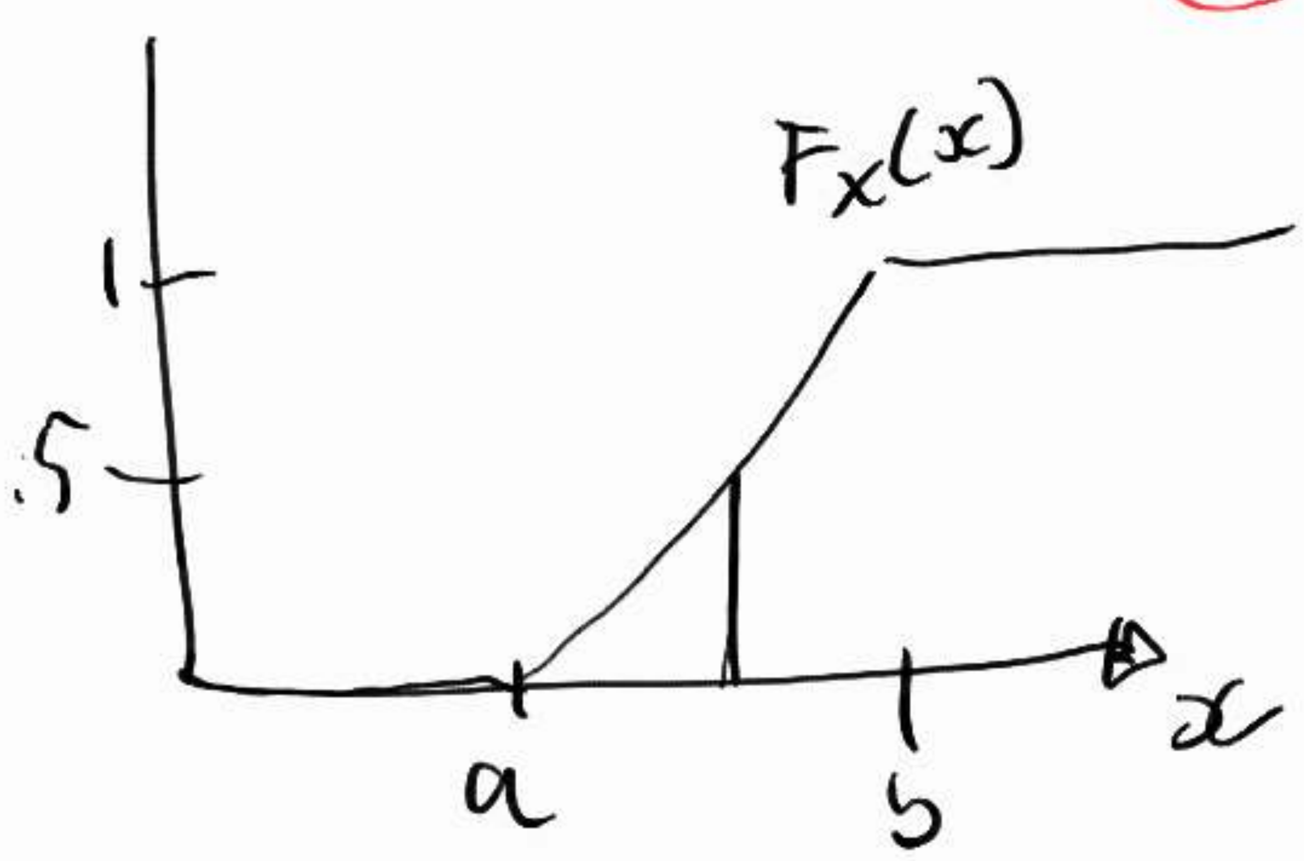


$$f_X(x) \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & x > b \end{cases} \qquad F_X \begin{cases} 0 & \text{om } x < a \\ \frac{x-a}{b-a} & \text{om } a \le x \le b \\ 1 & \text{om } x < b \end{cases}$$



$\boxed{} = \int_a^k f_X(x)\,dx$ => SANNOLIKHETEN ATT VA I INTERVAL $\boxed{}$

$$(b-a)\frac{1}{(b-a)} = 1$$

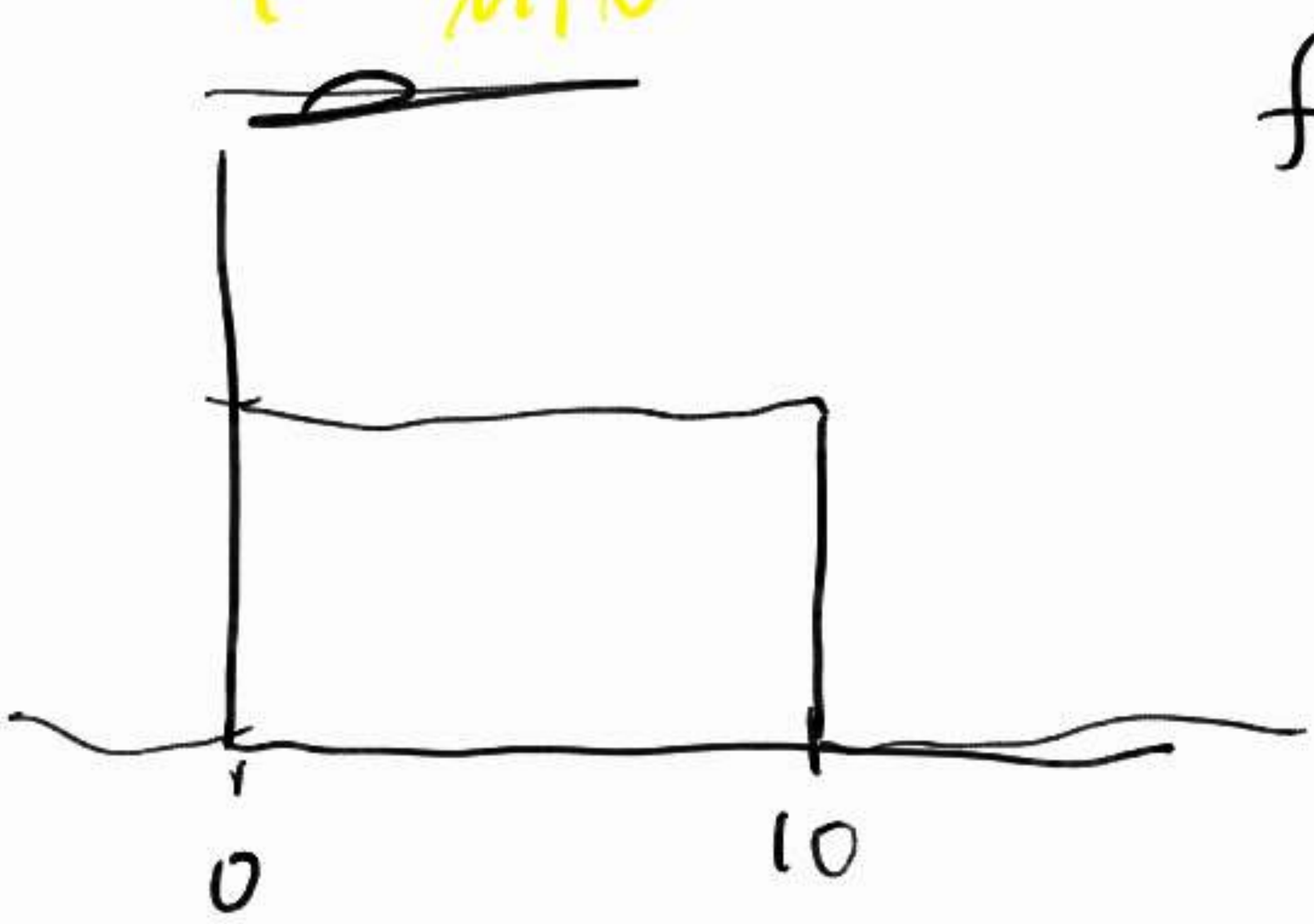


$$f_X(x) \rightarrow F_X(x)$$

täthets funktion — fördelnings funktion

BUSSAR KOMMER VAR 10:E MIN

VAD ÄR SANNOLIKHETEN ATT VÄNTA MINDRE ÄN 4 MIN



$$f_X(x) = \begin{cases} 0 & \text{om } x < 0 \\ \frac{1}{10-0} & \text{om } 0 \le x \le 10 \\ 0 & \text{om } x > 10 \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0 \\ \frac{x-0}{10-0} & \text{om } 0 < x < 10 \\ 1 & \text{om } x > 10 \end{cases}$$

$$F_X(4) = \frac{4-0}{10-0} = \frac{4}{10} = 40\%$$

EXPONENTIALFÖRDELNING:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{om } x > 0 \\ 0 & \text{annars} \end{cases}$$

täthets-
funktion

DÅ ÄR X Exp-fördelad
$X \in Exp(\lambda)$

$f_X(x)$ — $F_X(x)$

täthet funk — Förd funk

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \leq x \leq 1 \\ 1 & \text{om } x \geq 1 \end{cases}$$

$\boxed{f_X(x) = ?}$

$$f_X(x) = \frac{\partial}{\partial x} F_X(x) \Leftrightarrow F_X(x) = \int f_X(x) \partial x$$

$$f_X(x) \begin{cases} 0 & x < 0 \\ 3x^2 & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

$$\frac{\partial}{\partial x} x^3 = 3x^2$$

UPPGIFTER:

3.3 (bok) $p_X(3) = \frac{1}{3}, p_X(4) = \frac{1}{4}, p_X(7) = \frac{1}{6}, p_X(8) = \frac{1}{6}$

X kan anta $\underline{3, 4, 7, 8, 9}$

a) $p_X(9) = ?$

$$\sum_{-\infty}^{\infty} p(x) = 1$$

$$P_x(3)+P_x(4)+P_x(7)+P_x(8)+P_x(9)=1$$
$$P_x(3)=\frac{1}{3},\quad P_x(4)=\frac{1}{4},\quad P_x(7)=\frac{1}{6},\quad P_x(8)=\frac{1}{6}$$

$$1-\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{6}+\frac{1}{6}\right)=\frac{1}{12}$$

b) $F_x(5)$

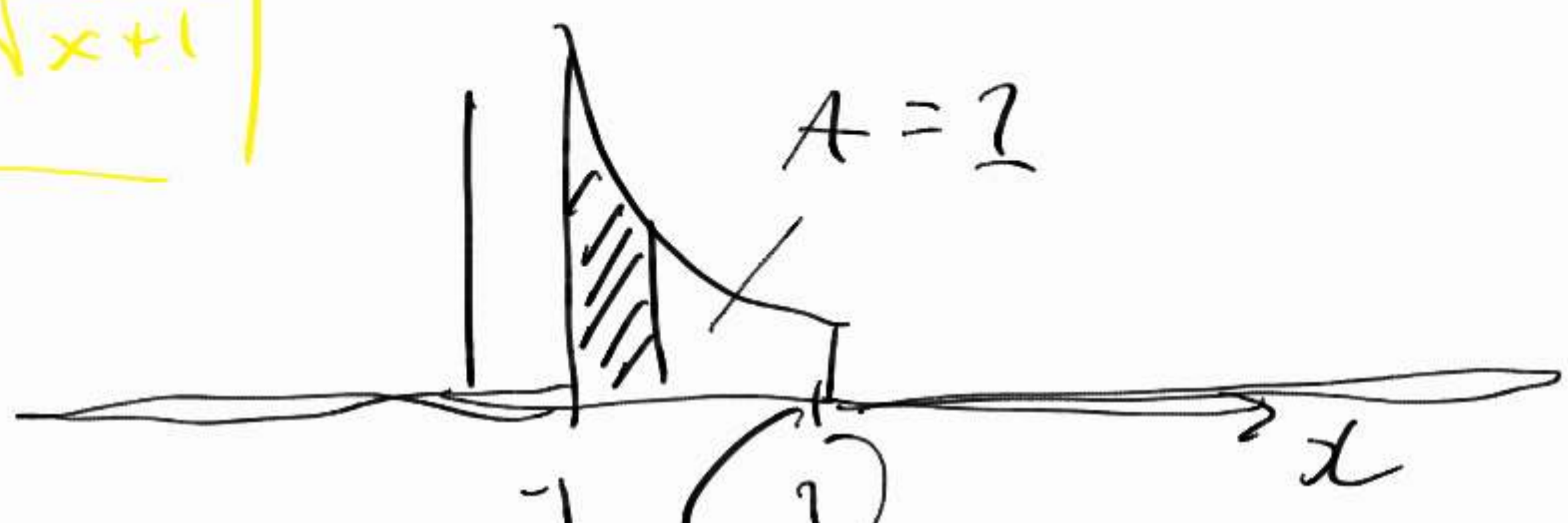
$$F_x(5)=P_x(3)+P_x(4)=\frac{1}{3}+\frac{1}{4}=\frac{7}{12}$$

c) $P(4\le X\le 8)=P_x(4)+P_x(7)+P_x(8)=\frac{1}{3}+\frac{1}{4}+\frac{1}{6}=\frac{7}{12}$

d) $P(X\ge 8)=P_x(8)+P_x(9)=\frac{1}{6}+\frac{1}{12}=\frac{3}{12}=\boxed{\frac{1}{4}}$

3,14 BESTÄM C

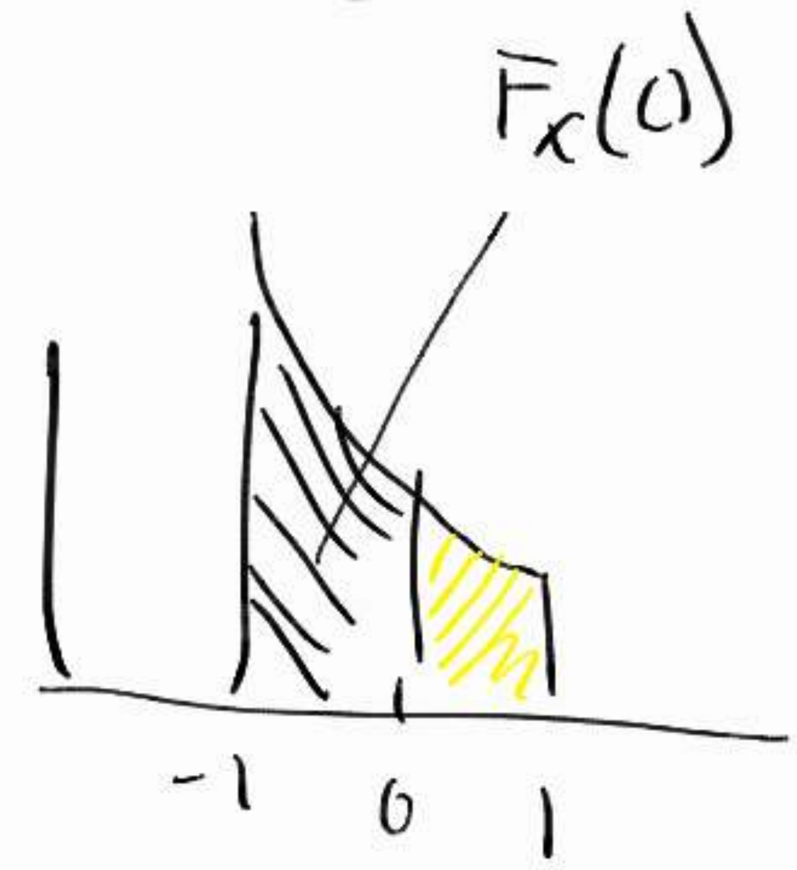
$\boxed{f(x)=\frac{c}{\sqrt{x+1}}}$ om $-1<x<1$, $f(x)=0$



$$\underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{=1} = \int_{-1}^{1} f(x)\,dx = c\int_{-1}^{1}\underbrace{\frac{1}{\sqrt{x+1}}}_{=(x+1)^{-\frac{1}{2}}}dx = 2c\left[\sqrt{x+1}\right]_{-1}^{1}$$

$$=2c\left[\sqrt{x+1}\right]_{-1}^{1}=2c\left[\sqrt{2}-\sqrt{0}\right]=2c\cdot\sqrt{2}=1$$

$$\Rightarrow 2c\cdot\sqrt{2}=1 \Rightarrow \boxed{\frac{1}{2\sqrt{2}}=c}$$



$$P(X>0)=1-F_x(0)$$

$$P(X>0) = \int_0^\infty f_X(x)\,dx = \int_0^1 \frac{c}{\sqrt{x+1}} = 2c\left[\sqrt{x+1}\right]_0^1 =$$

$$= 2c(\sqrt{2}-\sqrt{1}) = 2c(\sqrt{2}-1)$$

$$\approx \underline{0.3} \quad \blacksquare$$

RAST TILL 11.15 :)

FLERDIMENSIONELLA S.V.

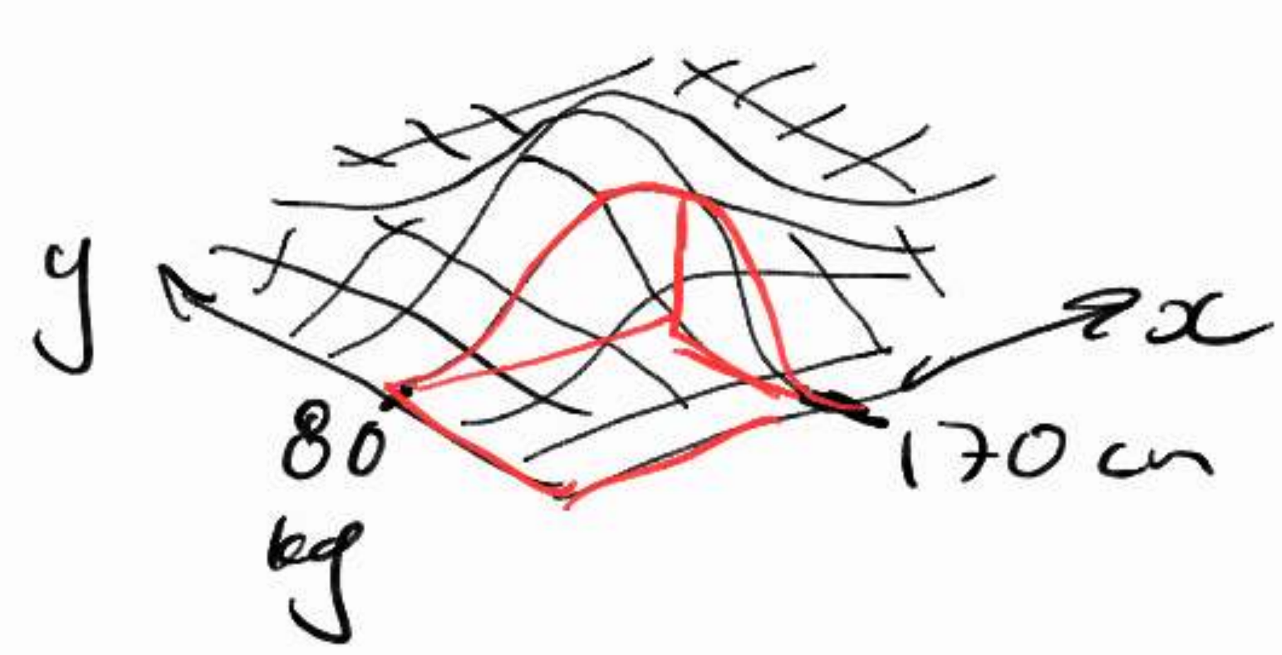
$F_X(x) \to \mathbb{R}$ då $0 \leq \mathbb{R} \leq 1$

BÅDE LÄNGD & VIKT FÖR MAN I SVERIGE

X = LÄNGD

Y = VIKT

$P(X<170, Y<80) =$



$=$

$F_{XY}(170, 80)$

$P(X<a) = F_X(a)$

$=$

FÖRDELNINGSFUNKTIONEN FÖR DEN TVÅDIMENSIONELLA S.V (X, Y)

BOSTADSUNDERSÖKNING: (BOK; 4.1)

S.V. X = # BARN I EN LÄGENHET

S.V. Y = # RUM I ETT LÄGENHET

BARN \ RUM	1	2	3	4
0	.11	.09	.07	.01
1	.07	.12	.12	.02
2	.02	.05	.17	.05
3	0	.02	.04	.02
4	0	0	.01	.01

a) SANNOLIKHET BARN ≤ 1, RUM ≤ 3

$= .11 + .09 + .07 + .07 + .12 + .12$

$= .58 = 58\%$

b) VARJE LÄGENHET HAR 2 VUXNA + BARN

TRÅNGBODD = PERSONER/RUM > 2

VAD ÄR SANNOLIKHETEN ATT EN FAMILJ ÄR TRÅNGBODD?

$\frac{(X+2)}{Y} > 2$

$= .07 + .02 + 0 + 0 + .02 \ldots = .11 = 11\%$

UPPGIFT 4.7 BOK:

$$f_X(x) \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & x > b \end{cases} \quad F_X \begin{cases} 0 & \text{om } x < a \\ \frac{x-a}{b-a} & \text{om } a \leq x \leq b \\ 1 & \text{om } x < b \end{cases}$$

TVÅ PERSONER SKA TRÄFFAS:

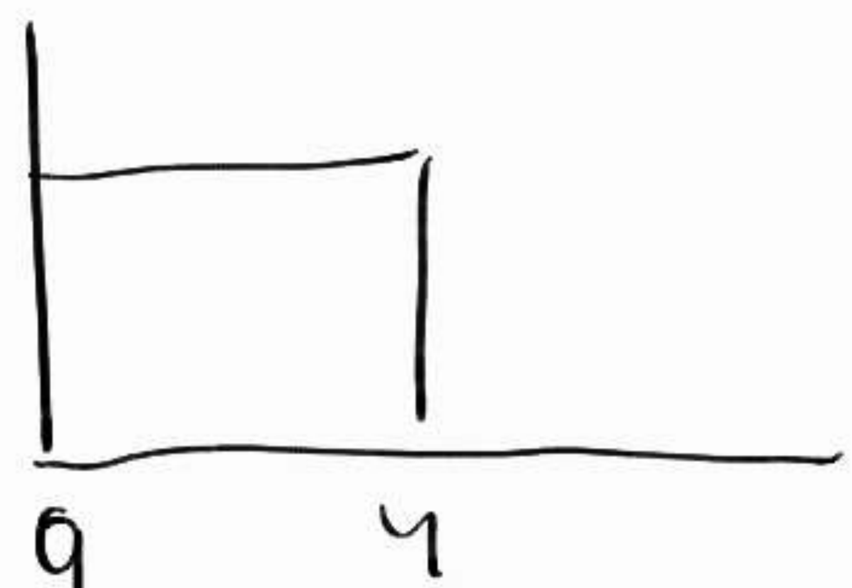
P_1 X min sent, DÄR X ÄR LIKFORMIGT FÖRDELAT MELLAN (0;4)

P_2 Y min sent — " — " —
MELLAN (0;6)

VAD ÄR $P(X<Y)$ // SANNOLIKHET ATT X BEHÖVER VÄNTA PÅ Y

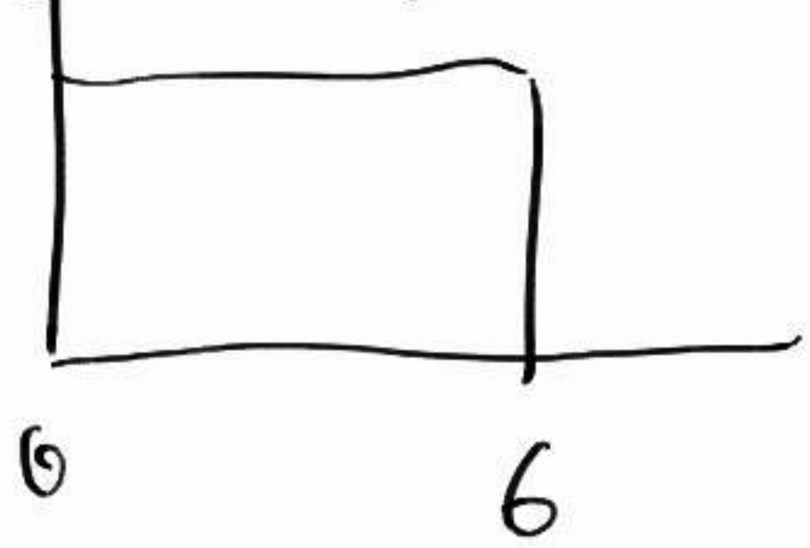
$$f_Y(y) = \frac{1}{6}$$

$X =$



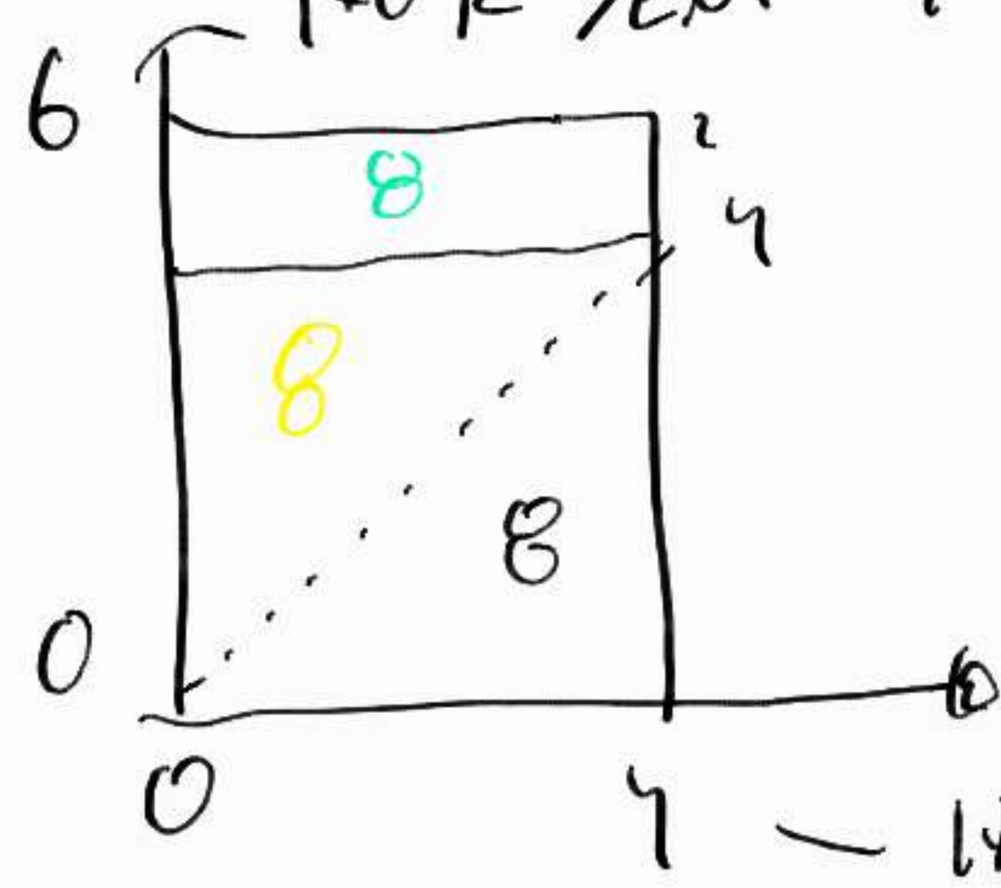
$$f_X(x) \begin{cases} 0 & \text{om } x<0 \\ \boxed{\frac{1}{4-0}} & \text{om } 0 \le x \le 4 \\ 0 & \text{om } x>4 \end{cases}$$

$Y =$



$$F_X \begin{cases} 0 & \text{om } x<0 \\ \frac{x-0}{4-0} & \text{om } 0 \le x \le 4 \\ 1 & \text{om } x>4 \end{cases}$$

HUR SENT Y KOMMER



HUR SENT X KOMMER

$$P(Y>X) = \frac{8+8}{24} = \frac{2}{3}$$

$$f_{XY}(x;y) \uparrow\downarrow F_{XY}(x;y)$$

$$\underline{f_{XY}(x;y) = \frac{\partial}{x}\left(\frac{\partial}{y} F_{XY}(x;y)\right)}$$

$$F_{XY}(x;y) = \boxed{\iint f_{XY}(x;y)\,dx\,dy}$$

$$P(X<Y) = \iint_{x<y} f_{XY}(x;y)\,\partial y\,dx = \underline{\int_0^4 \int_x^6 f_{XY}(x;y)\,dy\,\partial x}$$

$$\downarrow$$

$$f_{XY}(x;y) = f_X(x) f_Y(y) = \frac{1}{4} \cdot \frac{1}{6}$$

$$\int_0^4\left(\int_x^6 \frac{1}{24}\,dy\right)dx = \ldots = \int_0^4 \frac{6-x}{24}\,dx = \frac{1}{24}\int_0^4 6-x$$

$$= \frac{1}{24}\left[6x-\frac{x^2}{2}\right]_0^4 = \frac{1}{24}\left[24-\underbrace{\frac{16}{2}}_{8}\right] = \boxed{\frac{2}{3}}$$

4.3 BOK

DEN S.V (X,Y) HAR TÄTHETSFUNKTION

$$f_{XY}(x,y) = \left(\frac{c}{x}\right)\cdot e^{-x^2y} \quad \text{om } x>1\ y\geq 0$$

Vad är c?

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx = \int_1^{\infty}\int_0^{\infty}\left(\frac{c}{x}\right)\cdot e^{-x^2y}\,dy\,dx$$

$$= c\int_1^{\infty}\left(\int_0^{\infty}\left(\frac{1}{x}\right)\cdot e^{-x^2y}\,dy\right)dx \to c\int_1^{\infty}\frac{1}{x}\left[\cancel{e^{-x^2y}}\cdot\frac{1}{x^2}\right]_0^{\infty}dx$$

$$\Rightarrow \int_1^{\infty}\frac{1}{x}\cdot\frac{-1}{x^2}\,dx \to \int_1^{\infty}\frac{-1}{x^4}\,dx \to \ldots c\left[\frac{-1}{2x^2}\right]_1^{\infty} = \frac{c}{2}$$

$$\Rightarrow \boxed{c=2}$$

(LUNCH 13:15)

RAST TILL 13:20

TJIHO!

VÄNTEVÄRDE, VARIANS, STANDARDAVVIKELSE

1, 2, 3, 4, 5, 6

1kr 2kr 3kr

$$\frac{1}{6}\cdot 1 + \frac{2}{6} 2 + \frac{3}{6}\cdot 3 = \frac{1+4+9}{6} = \frac{14}{6} = \frac{7}{3}$$

$$\sum_{k=1}^{3} k\cdot P_X(k) = \frac{7}{3}$$

DEFINITION: VÄNTEVÄRDE

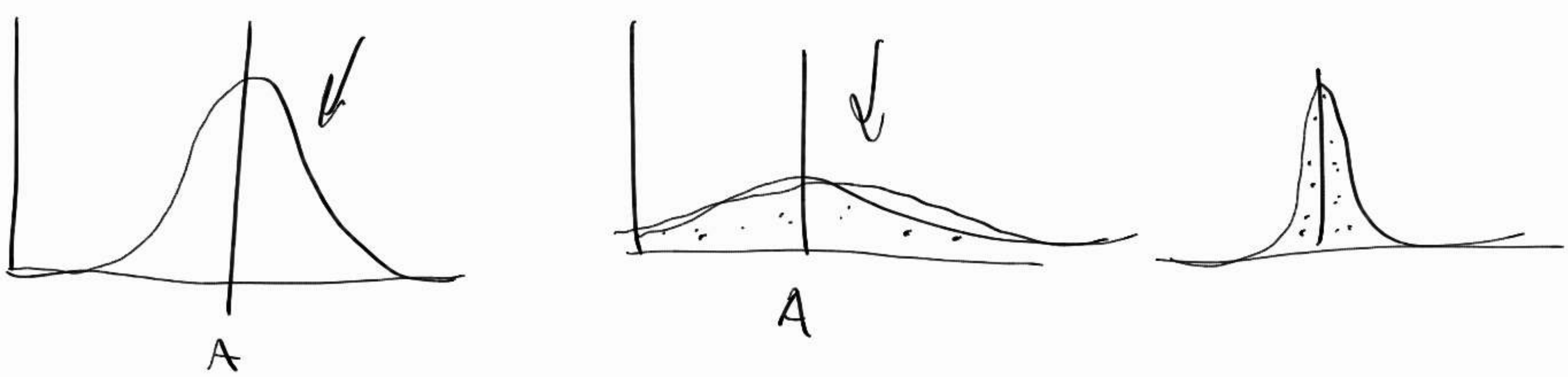
$$E(X) = \begin{cases} \sum_k k\cdot P_X(k) & : \text{(DISKRET)} \\ \int_{-\infty}^{\infty} x\cdot f_X(x)\,dx & : \text{(KONTINUERLIG)} \end{cases}$$

1, 2, 3 ... 6 = UTFALL (Ω)

X : S.V.

$X \to$ 1, 2 eller 3 kr

$$E(X) = \frac{7}{3}$$



$$E(X) = \mu_X$$

$$V(X) \sum_k (k-\mu)^2 \cdot P_X(k) \quad : \text{DISKRETA}$$

$V(x) \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \qquad V(x) = E(x^2) - (E(x))^2$

UPG 5.2 BOK

SLÅ TÄRNING. FLYTTA ANTAL STEG SOM TÄRNING VISAR

VISAR DEN 1 TAR MAN 6 STEG

a) VÄNTEVÄRDE

b) VARIANS

$\sum_k k \cdot P_X(k)$: (DISKRET)

$\sum_{k=2}^{6} k P_X(k)$

1	2	3	4	5	6
6	2	3	4	5	6

$2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{2}{6} = \frac{\overbrace{2+3+4}^{6}+5+6+6}{6} = \frac{26}{6} = \frac{13}{3}$

$E(x) = \frac{13}{3} \Rightarrow (E(x))^2 = \frac{13^2}{3^2} = \frac{169}{9}$

$V(x) = \boxed{E(x^2)} - (E(x))^2$

1	2	3	4	5	6	Σ
6	2	3	4	5	6	26
36	4	9	16	25	36	126

$\boxed{\sum_k k^2 \cdot P_X(k)}$

$4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{2}{6}$

$\shortparallel$

$21 = E(x^2)$

$V(x) = \underset{21}{\underset{\shortparallel}{E(x^2)}} - \underset{\frac{169}{9}}{\underset{\shortparallel}{(E(x))^2}} \approx \boxed{2,22\ldots}$

$\underline{\underline{D(X) = \sqrt{V(X)}}}$

$V(X) = \sum_k (k-\mu)^2 \cdot p_X(k)$

$= E(X^2) - E(X)^2$

$D(X) = \sqrt{2.22} \simeq 1,5$

$E(X) = \frac{13}{3}$

$E(Y) = \frac{26}{3} = E(2X)$

$E(Z) = \frac{26}{3} + 3$

	1	2	3	4	5	6	Σ
X =	6	2	3	4	5	6	26
Y =	12	4	6	8	10	12	
Z =	15	7	9	11	13	17	

$Z = 2X + 3$

$E(aX+b) = aE(X) + b$

$V(aX+b) = a^2 V(X)$

$D(aX+b) = |a| D(X)$

5.22 BOK:

3 S.V. X_1, X_2 & X_3. OBEROENDE $E(X_i) = 2$, $\boxed{V(X_i) = 3}$

$Y = 3X_1 - 2X_2 + X_3 - 6$

$E(Y)$, $D(Y)$

$E(X_1 + X_2)$, X_1 o X_2 OBEROENDE

$= E(X_1) + E(X_2)$

$V(X_1 + X_2) = V(X_1) + V(X_2)$

$E(Y) = E(3X_1 - 2X_2 + X_3 - 6)$

$= E(3X_1) + E(-2X_2) + E(X_3 - 6)$

$= 3 \cdot E(X_1) + (-2)E(X_2) + E(X_3) - 6$

$3 \cdot 2 + (-2)(2) + 2 - 6 = \cancel{6} - 4 + 2 - \cancel{6} = \boxed{-2}$

$V(Y) = V(3X_1 - 2X_2 + X_3 - 6) = V(3X_1) + V(-2X_2) + V(X_3 - 6)$

$$V(3X_1) + V(-2X_2) + V(X_3 - 6) = 3^2 V(X_1) + (-2)^2 V(X_2) + V(X_3) = 3^2 \cdot 3 + (-2)^2 \cdot 3 + 3 = 9 \cdot 3 + 4 \cdot 3 + 3 = \cancel{126}\ 42$$

$V(\cancel{X}),\ D(Y)$

$$D(Y) = \sqrt{V(Y)} = \boxed{\sqrt{42}} \approx 6{,}5$$

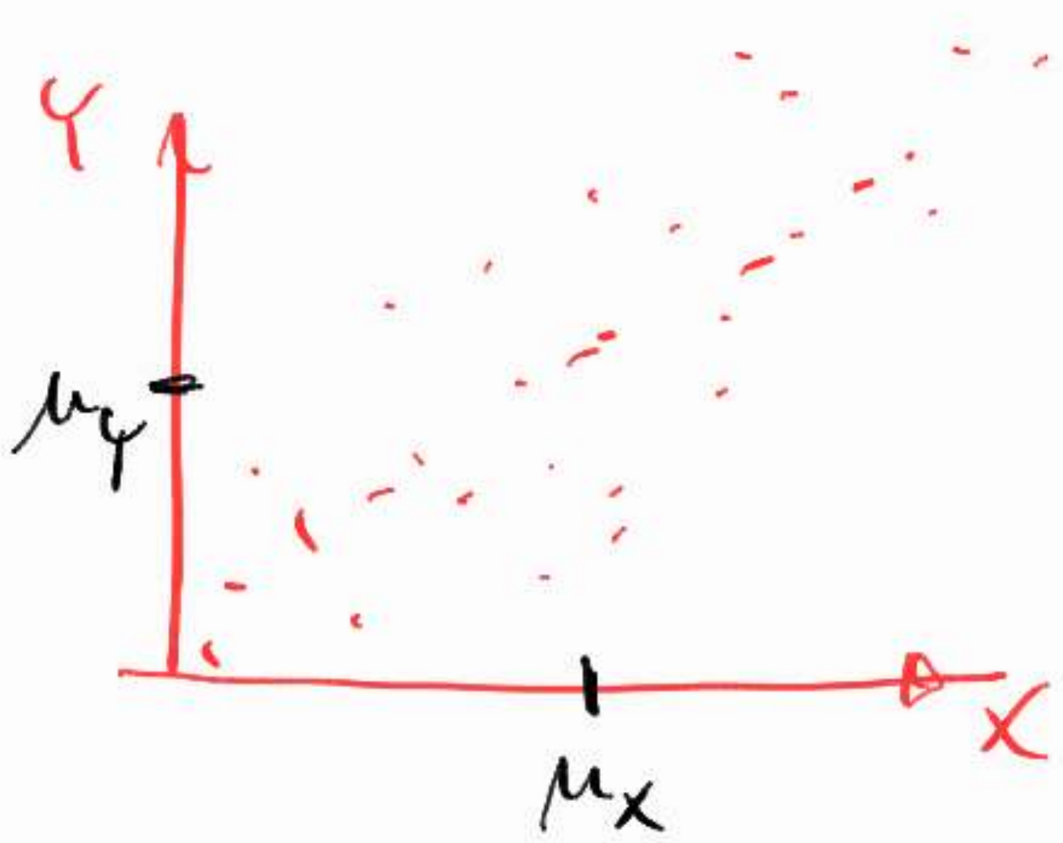
RAST 14,15 :)

X: LÄNGD

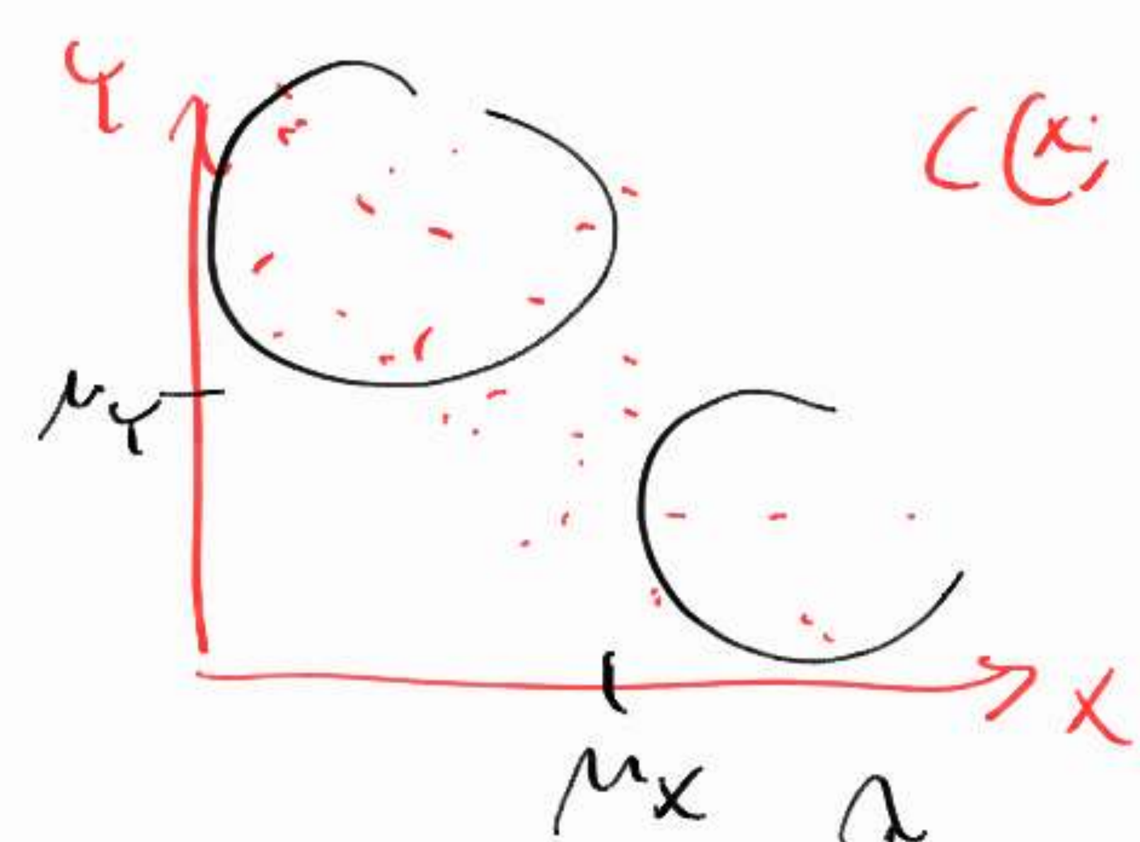
Y: VIKT

KOVARIANS

$$C(X;Y) = E((X-\mu_X)(Y-\mu_Y))$$



$C(X;Y) > 0$



$C(X;Y) < 0$

$$C(X;Y) = \begin{cases} \sum_j \sum_k (j-\mu_X)(k-\mu_Y)\, p_{XY}(j;k) & : \text{(DISKRET)} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) \cdot f_{XY}(x;y)\, dx\, dy \end{cases}$$

$$C(X;Y) = \underbrace{E(X \cdot Y)}_{kr^2} - \underbrace{E(X)}_{kr^2}\underbrace{E(Y)}_{kr^2}$$

$E(X)$: m

$V(X)$: m^2 $\quad D(X) = \sqrt{m^2} = m$

$C(X;Y) = m^2$

KORRELATIONSKOEFFICIENT :)

$$\rho = \frac{m^2}{m \cdot m} = 1$$

$$\rho(X;Y) = \frac{C(X;Y)}{D(X)D(Y)}$$

DIMENSIONSLÖS :)

5.17

2 1-KRONA, 1 5-KRONA

MAN DRAR SLUMPMÄSSIGT UTAN ÅTERLÄGG 2 MYNT

X = FÖRSTA MYNTET, Y = ANDRA MYNTET

$E(X), E(Y), V(X), V(Y); E(X \cdot Y); C(X;Y); \rho(X;Y)$

		2	
		1	5
1	1	1/3	1/3
	5	1/3	0

$E(X) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 5 = \boxed{\frac{7}{3}}$

$\frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$

$\frac{1}{3}\cdot\frac{2}{2} = \frac{1}{3}$

$\frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$

$E(Y) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 5 = \boxed{\frac{7}{3}}$

$(E(X))^2 = \left(\frac{7}{3}\right)^2 = \frac{49}{9}$

$E(X^2) = \frac{1}{3}\cdot 1^2 + \frac{1}{3}\cdot 1^2 + \frac{1}{3}\cdot 5^2 = \frac{27}{3} = \boxed{9}$

$V(X) = E(X^2) - E(X)^2$

$V(X) = 9 - \frac{49}{9} = \boxed{\frac{32}{9}}$, $9 = 81/9$

$V(Y) = \boxed{\frac{32}{9}}$

$E(X \cdot Y) = \sum_{j\,k} j \cdot k \cdot p_{XY}(j;k) = (1\cdot 1)\cdot\frac{1}{3} + (1\cdot 5)\cdot\frac{1}{3} + (5\cdot 1)\cdot\frac{1}{3} + (5\cdot 5)\cdot\frac{0}{3} = \boxed{\frac{11}{3}}$

$C(X;Y) = E(XY) - E(X)E(Y)$

$= \frac{11}{3} - \frac{7}{3}\cdot\frac{7}{3} = \frac{11}{3} - \frac{49}{9} = \frac{33}{9} - \frac{49}{9} = \boxed{-\frac{16}{9}}$

$\rho(X;Y) = \frac{C(X;Y)}{D(X)D(Y)} = \frac{-16/9}{\frac{4\sqrt{2}}{3}\cdot\frac{4\sqrt{2}}{3}} = \frac{\frac{-16}{\cancel{9}}}{\frac{32}{\cancel{9}}} = \boxed{-\frac{1}{2}}$

$D(X) = \sqrt{V(X)} = \sqrt{\frac{32}{9}} = \frac{4\sqrt{2}}{3}$

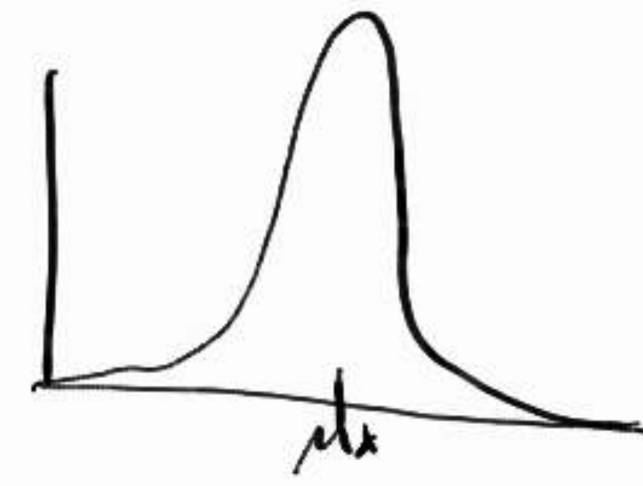
$D(Y) = $ — " — $=\frac{4\sqrt{2}}{3}$

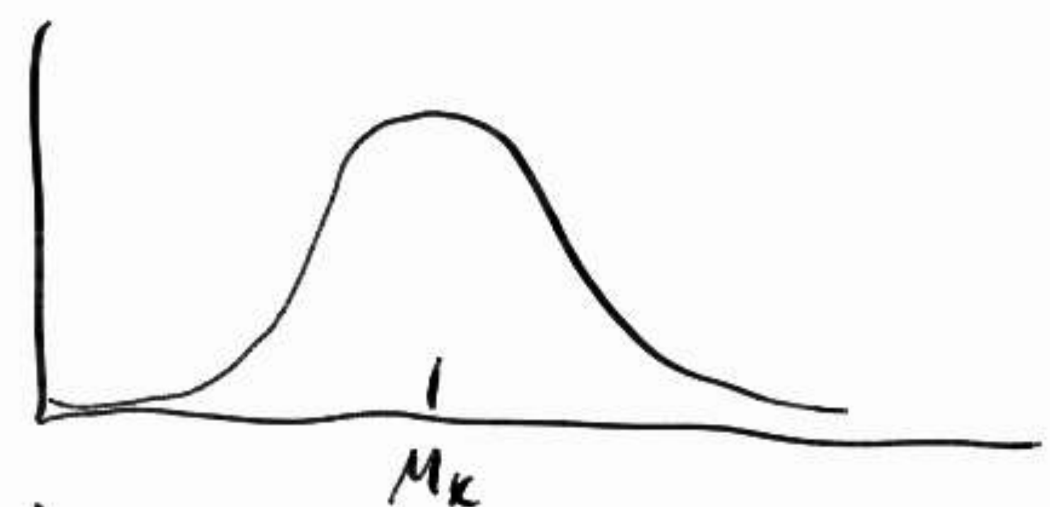
5.28 BOK

VISA ATT OM $V(X)=1$ o $V(Y)=16$

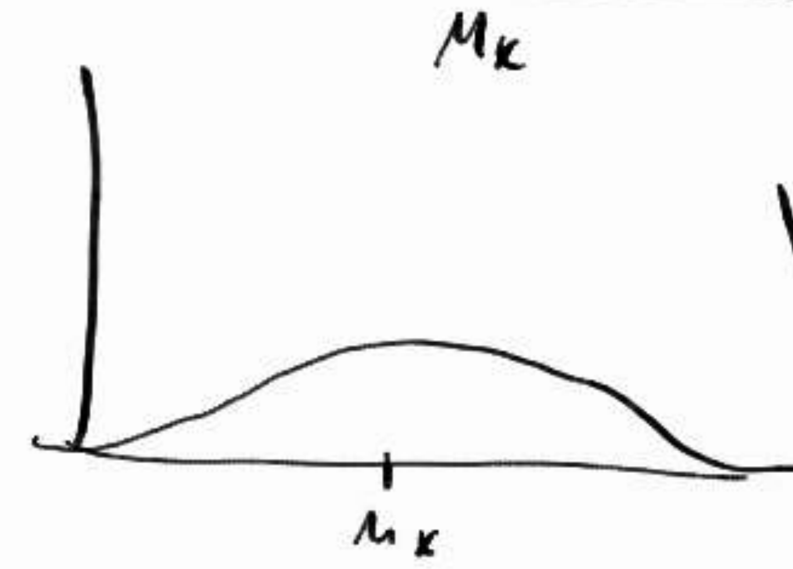
SÅ ÄR $9 \le V(X+Y) \le 25$

$V(X)$: liten





$V(X)$: stor



$$-1 < \rho(X;Y) < 1$$

$$\rho(X;Y) = \frac{C(X;Y)}{D(X)D(Y)=4}$$

$D(X)=\sqrt{1}$

$D(Y)=\sqrt{16}=4$

$D(X)D(Y)=4$

$$-1 < \frac{C(X;Y)}{4} < 1$$

$$\boxed{-4 < C(X;Y) < 4}$$

$V(X_1+X_2) = V(X_1)+V(X_2)$ om X_1, X_2 OBEROENDE

$V(X_1+X_2) = V(X_1)+V(X_2)+2\cdot C(X;Y)$

$C(X;Y) = E(X\cdot Y) - E(X)\cdot E(Y) = 0$ om $X;Y$ OBEROENDE

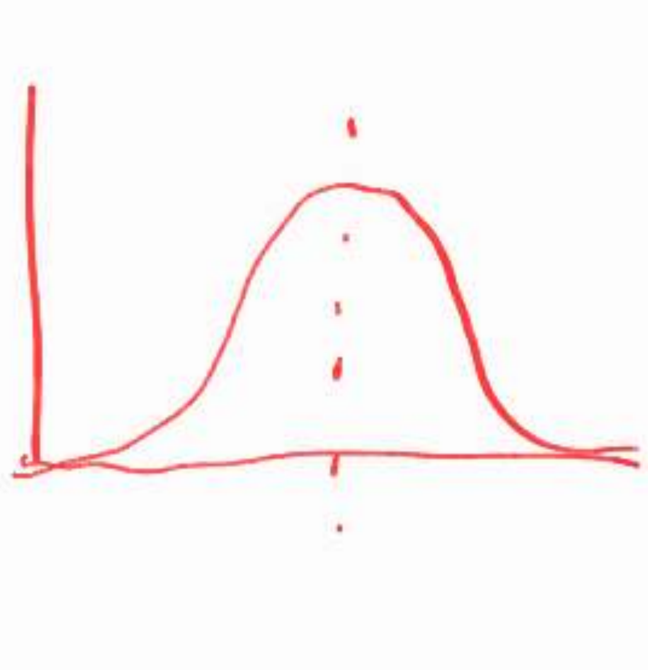
$$V(X+Y) = \underbrace{V(X)}_{1} + \underbrace{V(Y)}_{16} + \underbrace{2C(X;Y)}_{\text{mellan } -8 \text{ o } 8}$$

$1 + 16 = 17$

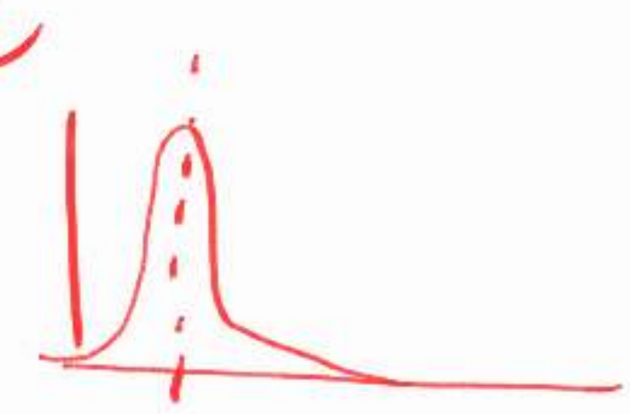
$V(X+Y) =$ mellan $(17-8, 17+8) = \underline{(9;25)}$

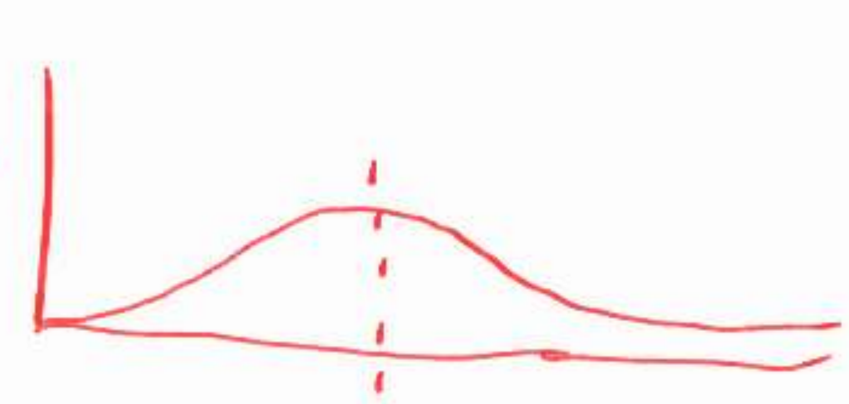
RAST TILL 15:15

NORMALFÖRDELNING ☺



← NORMAL FÖRDELNING





BOK: SANNOLIKHET & STATISTIKTEORI MED TILLÄMPNINGAR

CENTRALA GRÄNSVÄRDESSATSEN:

SATS: SUMMAN AV TILLRÄCKLIGT MÅNGA LIKAFÖRDELADE OBEROENDE FÖRDELNINGAR → NORMALFÖRDELNING

1000 TÄRNINGAR,

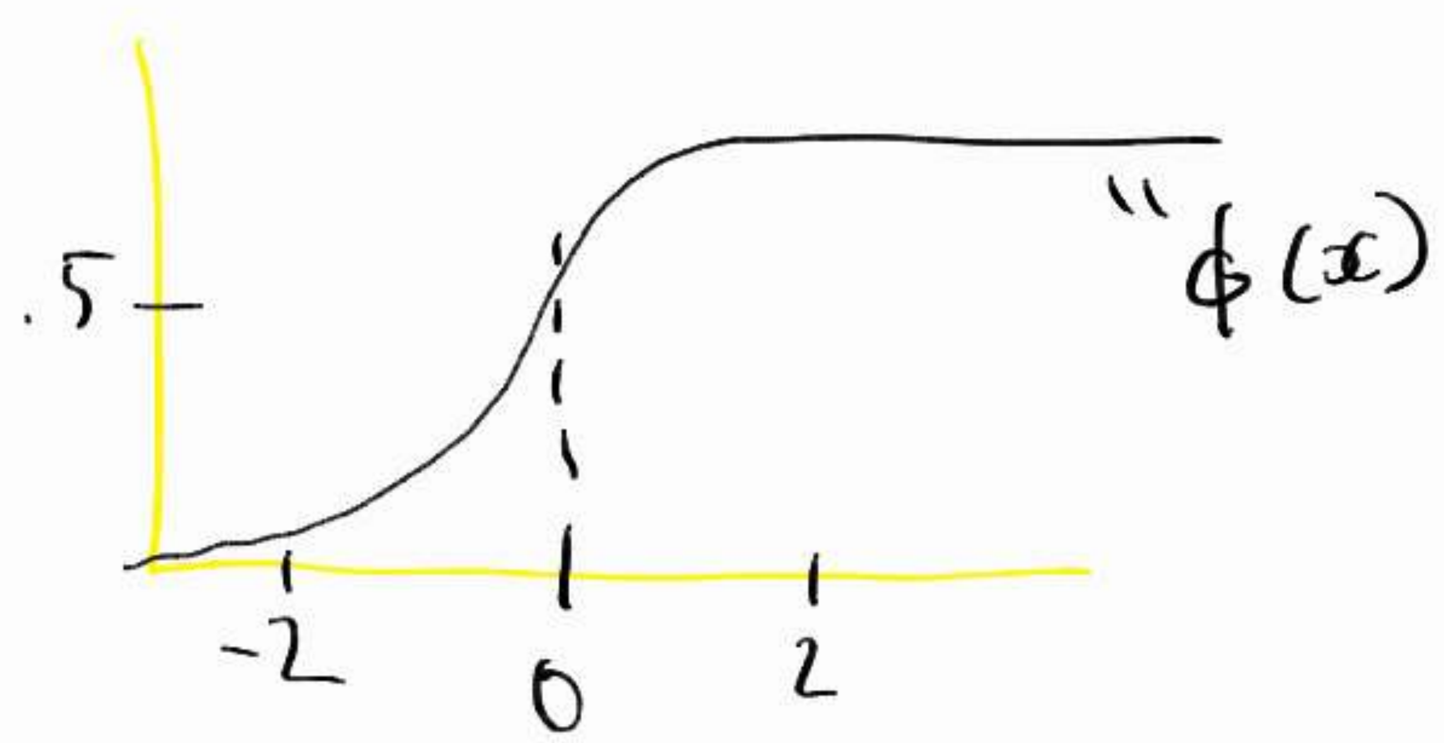
SANNOLIKHET ATT 200 - 250 ST 6:OR

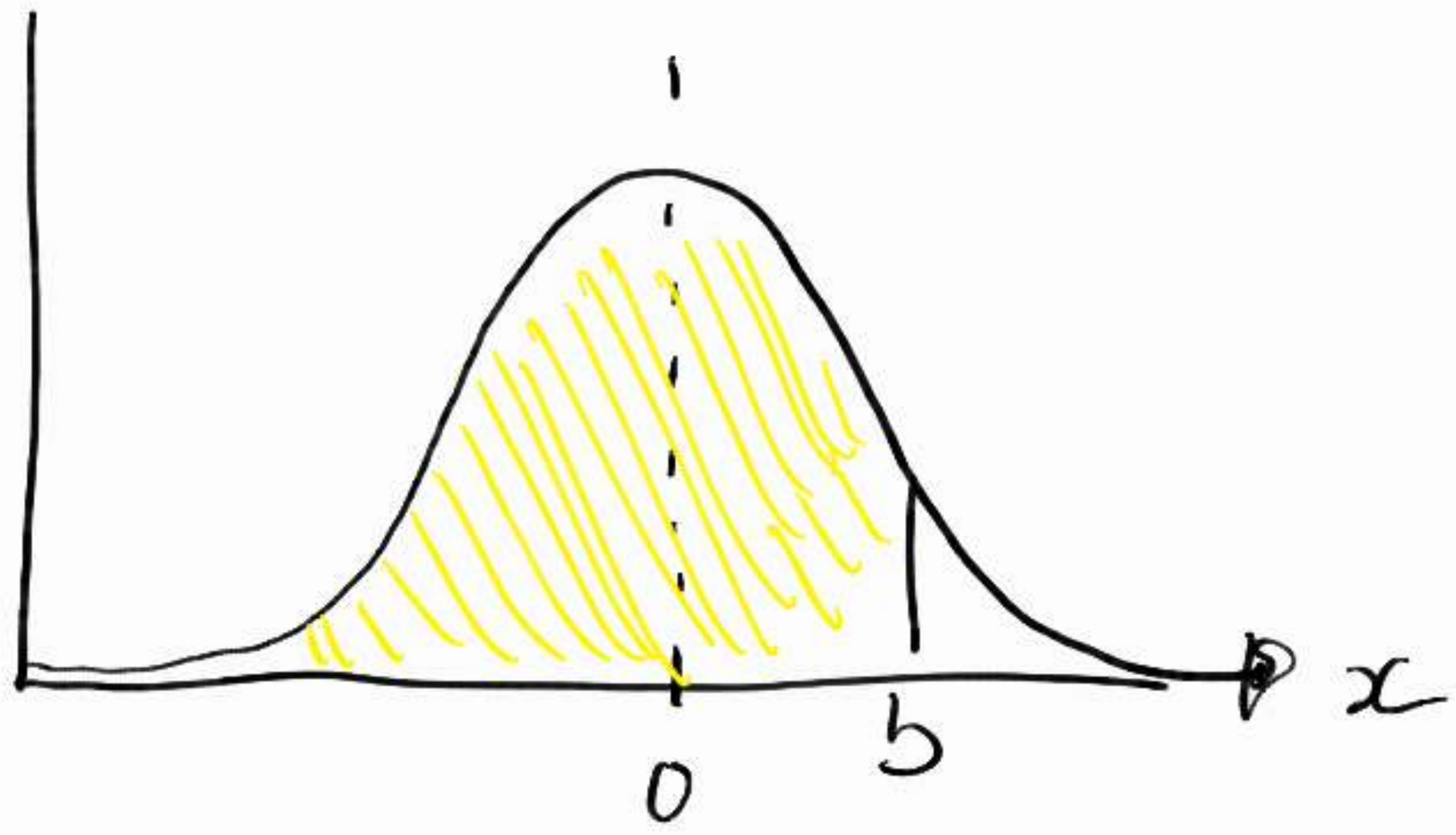
NORMALFÖRDELNING:

$$F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$\downarrow$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{t^2}{2}} dt$$

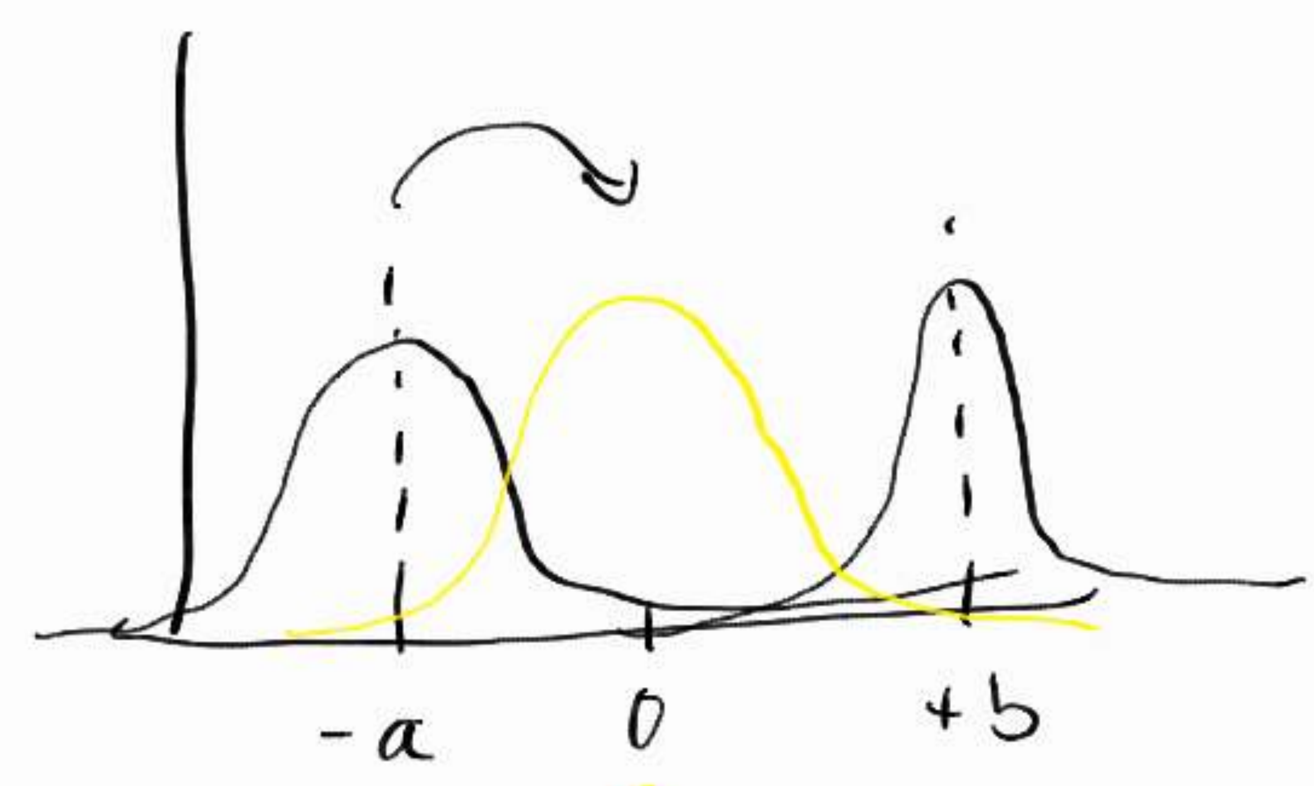




$$P(X \le b) = \phi(b)$$

$$\phi(0,5) \simeq .6915$$

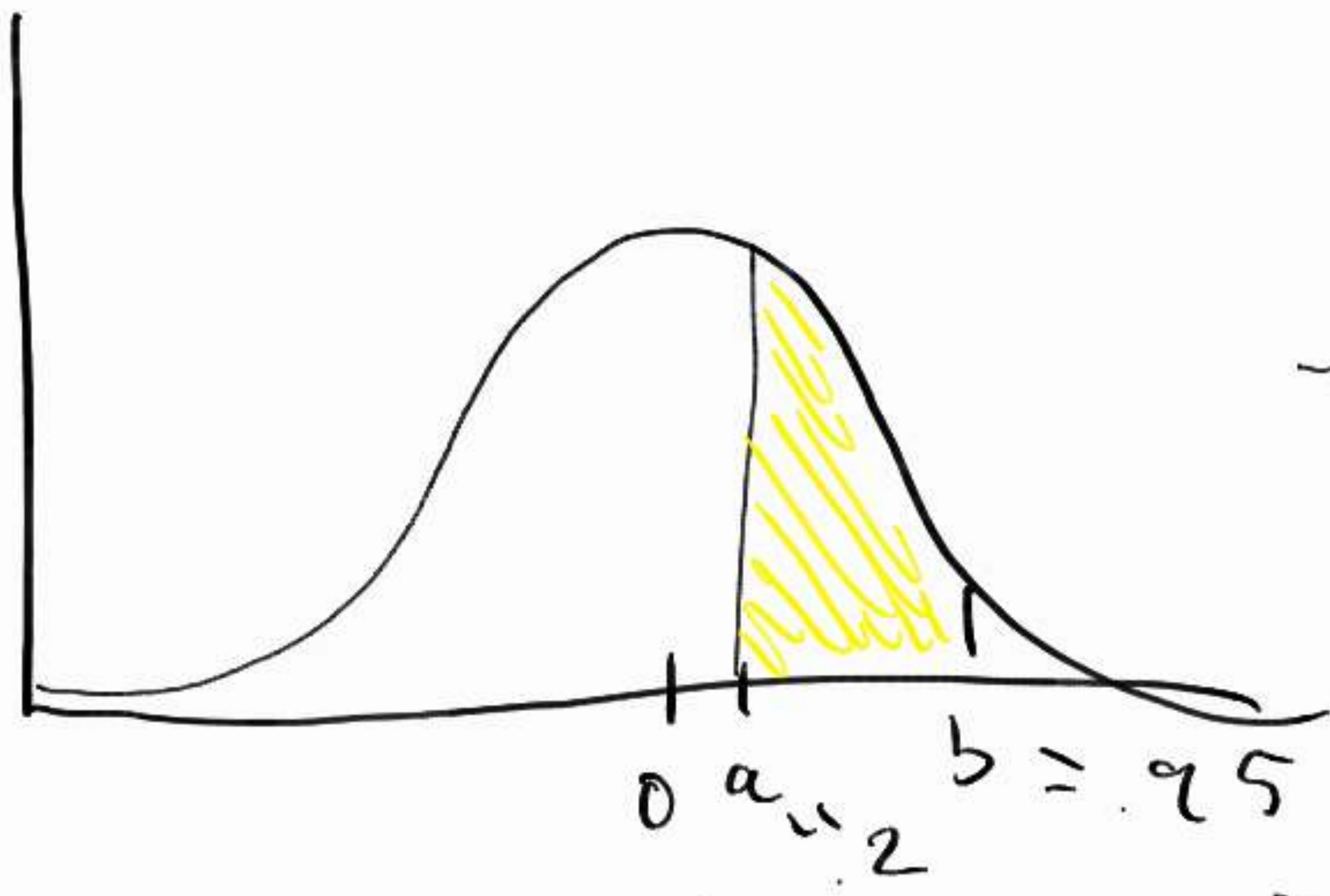
NORMALFÖRDELNING

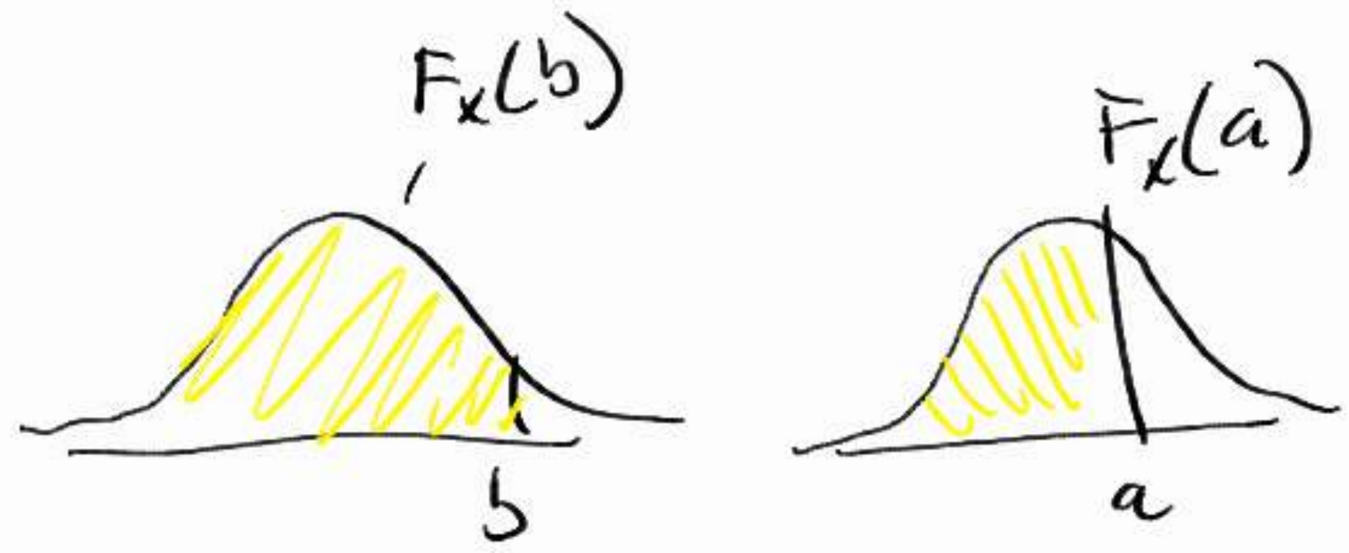


$b = .5$

$b = \underline{.53}$

$\phi(.53) = .7019$





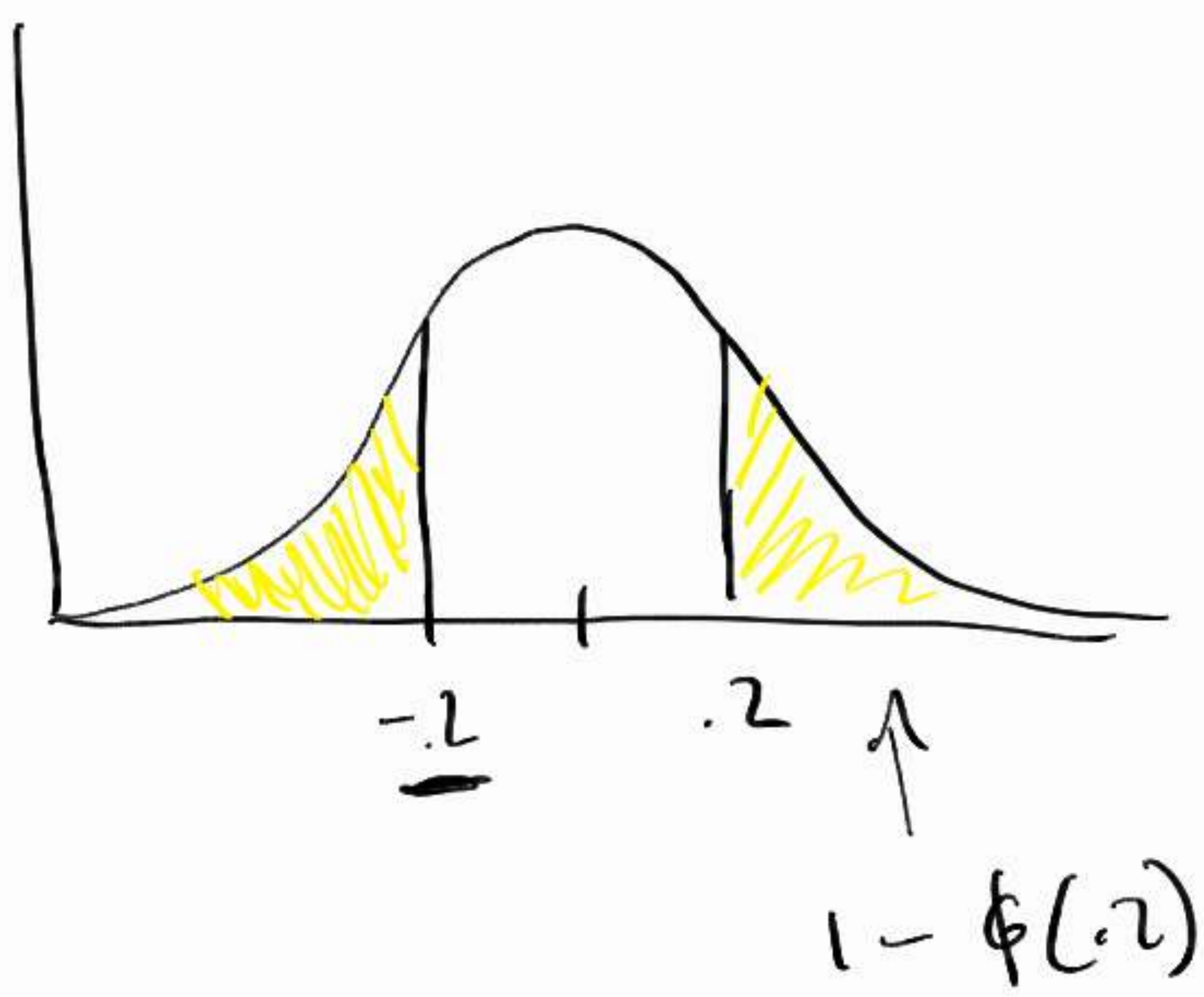
$$p(a < X \leq b) = F_X(b) - F_X(a) = \phi(b) - \phi(a)$$

$b = .95 \quad a = .2$

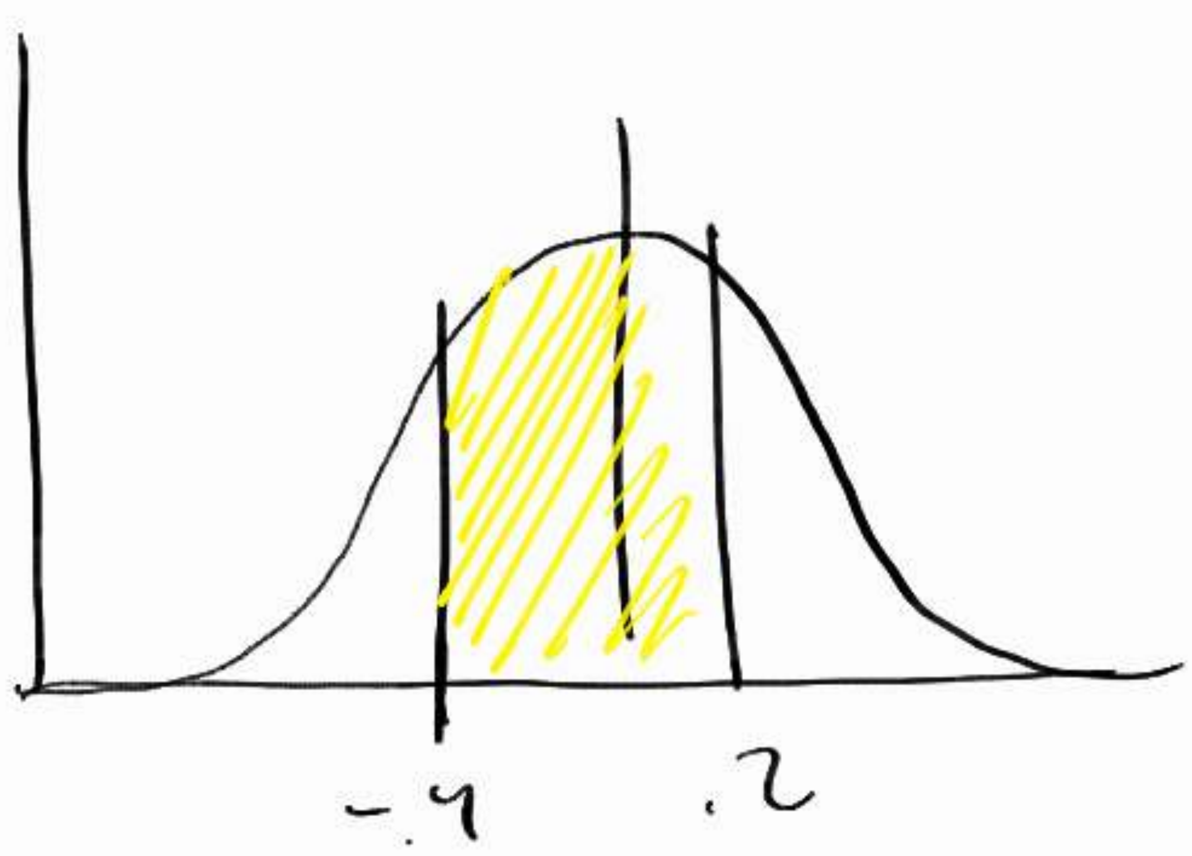
$\phi(.95) = .8289$

$\phi(.2) = .5793$

$\phi(b) - \phi(a) = .2496$



$\phi(-b) = 1 - \phi(b)$



$$\left.\begin{array}{l}\phi(.2) - \phi(-.4) \\ \qquad\quad = \\ \qquad (1 - \phi(.4))\end{array}\right\} \phi(.2) - [1 - \phi(.4)]$$

$=$

$\phi(.2) - [1 - \phi(.4)]$

$=$

$.57 - [1 - .65]$

$=$

$\sim .22$

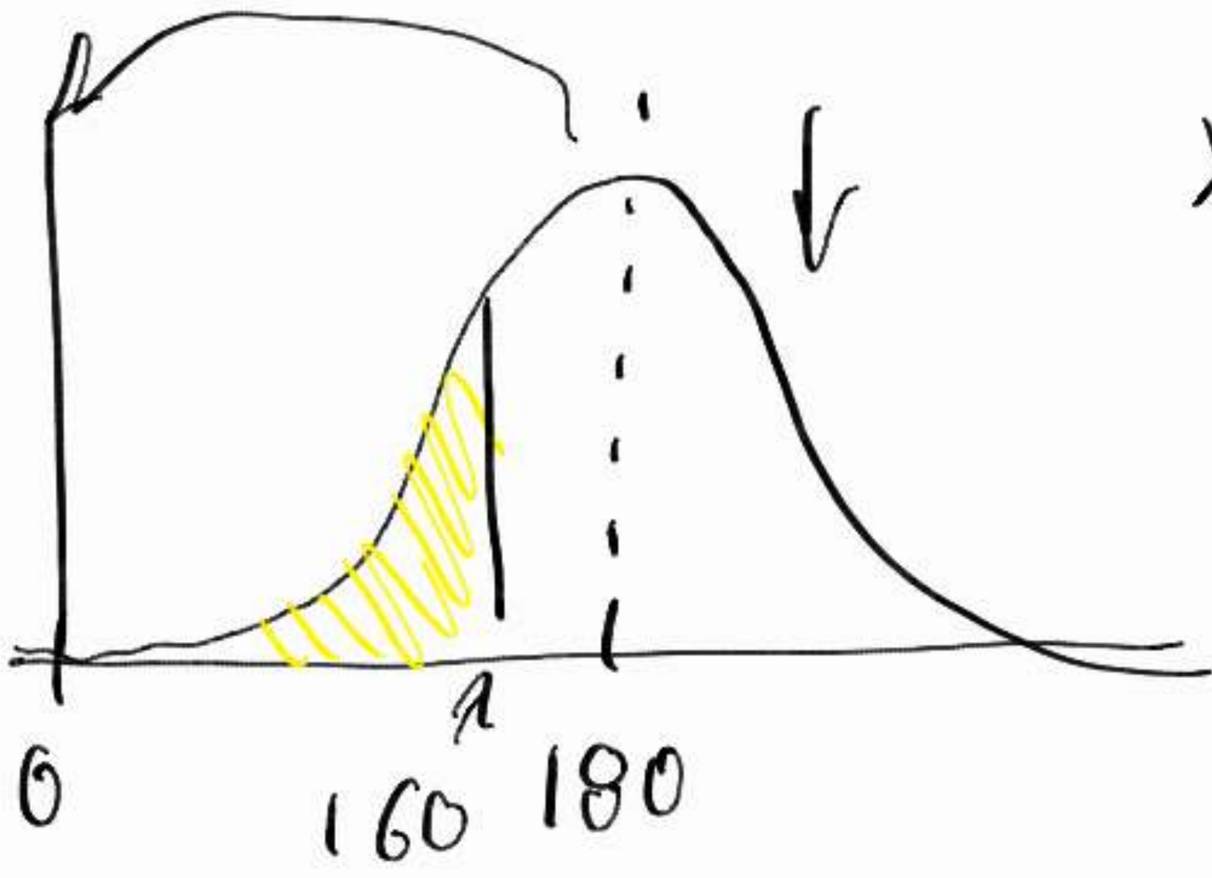
OM X ÄR NORMALFÖRDELAD

$X \in N(\mu; \sigma)$

OM $X \in$ NORM $\quad E(X) = \mu$

$D(X) = \sigma$

$Y = \dfrac{X - \mu_X}{\sigma_X} \Rightarrow Y \in N(0;1) =$ Y ÄR NORMALISERAD :)



$X \in N(180; 5)$

VÄNTE-VÄRDE, STANDARD AVVIKELSE

$Y = \frac{X-180}{5} = Y \in N(0;1)$

$F_X(160)$

$P(X<160) = P(X-180 < 160-180) = P(X-180 < -20)$

$\Rightarrow P\left(\boxed{\frac{X-180}{5}} < \frac{-20}{5}\right) \Rightarrow P\left(\underbrace{\frac{X-180}{5}}_{=Y} < -4\right) = P(Y<-4)$

$= \phi(-4) = 1 - \phi(4) = 1 - .99997$

5 VA FÖR LITEN STANDARDAVVIKELSE

RAST 16:15 :)

6.4 BOK: BESTÄM a) $P(X \le 6)$

$X \in N(5;2)$ b) $P(1.8 < X < 7.2)$

a) $P(X \le 6) = P(X-5 \le 6-5) = P\left(\frac{X-5}{2} \le \frac{6-5}{2}\right) = \phi(.5)$

$= .6915$:)

b) $P(1.8 < X < 7.2) = P\left(\frac{1.8-5}{2} < X \le \frac{7.2-5}{2}\right)$ 0.05

$= \phi\left(\frac{7.2-5}{2}\right) - \phi\left(\frac{1.8-5}{2}\right)$ .8643 $1 - .9452$

$\phi(1.1) - \phi(-1.6) = \phi(1.1) - [1 - \phi(1.6)] = .80953$:)

$$\boxed{\begin{aligned} E(aX+b) &= aE(X)+b \\ V(aX+b) &= a^2V(X) \end{aligned}}$$

$$D(aX+b) = |a|D(X)$$

$X \in N(\mu;\sigma)$

$Y = aX+b \in N(a\mu+b;\ |a|D(X))$

$X \in N(\mu_X;\sigma_X);\ Y \in N(\mu_Y;\sigma_Y) \Rightarrow$

$X \pm Y \in N(\mu_X \pm \mu_Y;\ \sqrt{\sigma_X^2+\sigma_X^2})$

$$V(X+Y) = V(X)+V(Y)$$

$$\sqrt{D(X+Y)^2} = \sqrt{D(X)^2+D(Y)^2}$$

$X_1, X_2 \dots X_n \in N(\mu,\sigma)$, OBEROENDE

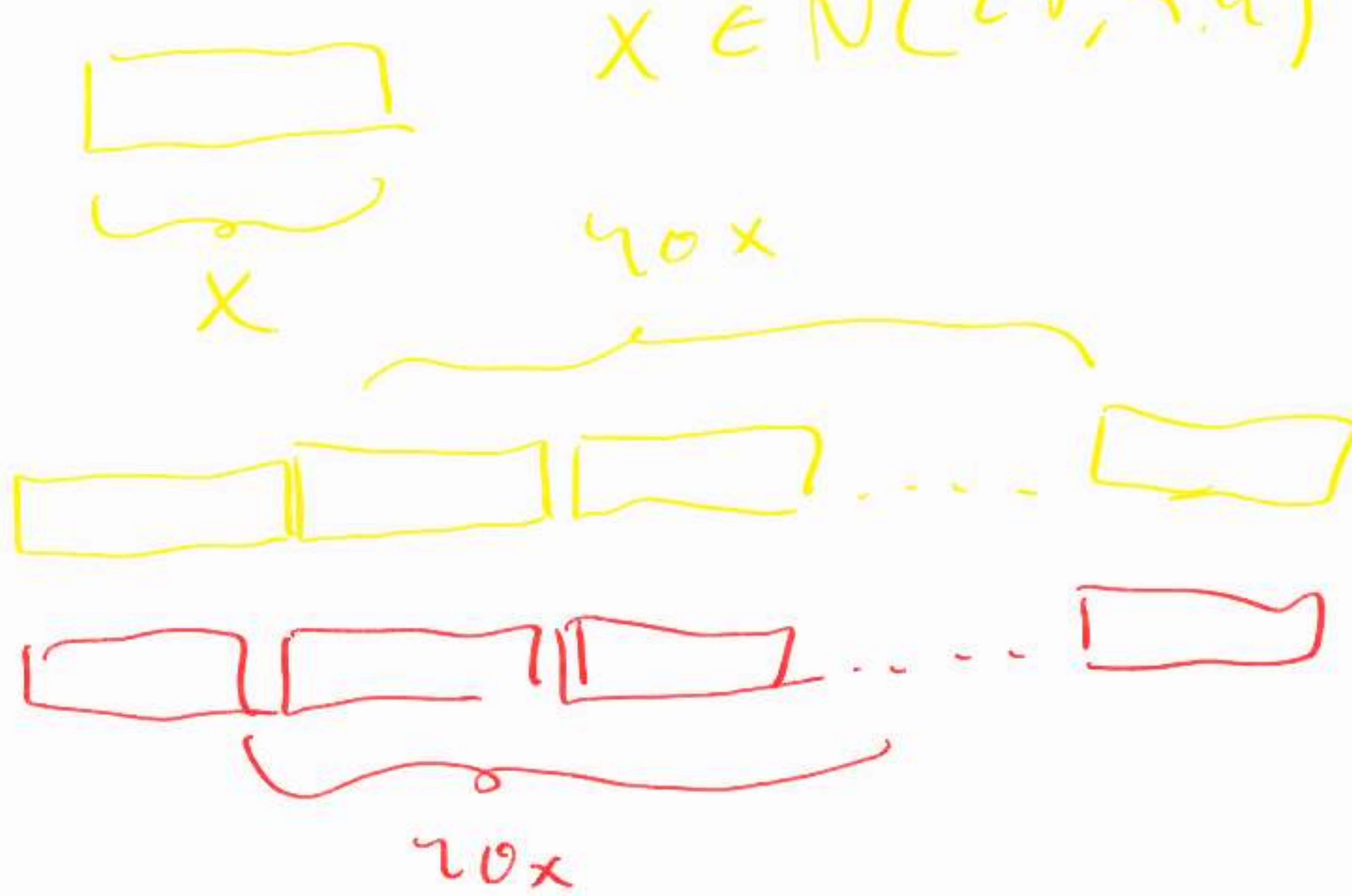
$$\bar{X} = \frac{\sum_{1}^{n} X_i}{n} = \frac{X_1+X_2\dots X_n}{n}$$

$$\Downarrow$$

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

Ex 6.3

$X \in N(25;\ \sqrt{.4})$



VAD ÄR SANNOLIKHET

ATT $|\ \_\_\_ - \_\_\_\ | < 10$

$X_1, X_2 \dots X_{40} = 40 \cdot \bar{X}$ $\qquad X_i, Y_i \in N(25, \sqrt{.04})$

$Y_1, Y_2 \dots Y_{40} = 40 \cdot \bar{Y}$

$$\bar{X} = \frac{\sum_{1}^{n} x_i}{n} = \frac{x_1 + x_2 \dots x_n}{n}$$

$$\Downarrow$$

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

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

$$\bar{X} = N\left(25; \frac{\sqrt{.04}}{\sqrt{40}}\right) = N\left(25; \sqrt{\frac{.04}{40}}\right)$$

$$40 \cdot \bar{X} = N\left(40 \cdot 25; 40 \cdot \frac{\sqrt{.04}}{\sqrt{40}}\right) = N(1000; \sqrt{40 \cdot .04})$$

$$40 \cdot \bar{Y} = \text{---}\,''\,\text{---} \qquad \text{---}\,''\,\text{---} = N(1000; \sqrt{40 \cdot 0.4})$$

$$X \pm Y \in N\left(\mu_X \pm \mu_Y; \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

$$Z = 40\bar{X} - 40\bar{Y} \in N\left(1000 - 1000; \sqrt{\sqrt{40 \cdot .04}^2 + \sqrt{40 \cdot 0.4}^2}\right)$$

$$\parallel$$

$$Z = N(0; \sqrt{80 \cdot 0.4})$$

$$P(-10 < Z < 10)$$

$$= P\left(\frac{-10}{\sqrt{80 \cdot 0.4}} < \frac{Z}{\sqrt{80 \cdot 0.4}} < \frac{10}{\sqrt{80 \cdot 0.4}}\right) = \phi(1.77) - \phi(-1.77)$$

$$\frac{10}{\sqrt{80 \cdot 0.4}} = 1.77$$

$$= \phi(1.77) - [1 - \phi(1.77)]$$

$$= 2\phi(1.77) - 1 = 0.9616$$

:-)

CENTRALA GRÄNSVÄRDESSATSEN

$X_1, X_2 \dots \to \infty$: OBEROENDE LIKAFÖRDELADE S.V

MED $E(X_i) = \mu$; $D(X_i) = \sigma$

$$Y_n = X_1 + X_2 \dots + X_n$$

$$P\left(a < \frac{Y_n - n\cdot\mu}{\sigma\cdot\sqrt{n}} \le b\right) \to \phi(b) - \phi(a) \quad \text{när } n \to \infty$$

VAD ÄR SANNOLIKHETEN ATT FÅ 320–380 SOM SUMMA NÄR MAN SLÅR 100 TÄRNINGAR

X = ANTALET 6:OR I EN TÄRNING

$$E(X) = \frac{1+2+3 \ldots +6}{6} = 3.5 = \frac{7}{2}$$

$$E(X^2) = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{70}{24}$$

$$D(X) = \sqrt{\frac{70}{24}} \leftarrow \sigma$$

$$P\left(a < \frac{Y_n - n\cdot\mu}{\sigma\cdot\sqrt{n}} \le b\right) \to \phi(b) - \phi($$

100

$$Y = N\left(100\cdot E(X); \sqrt{\frac{70}{24}}\cdot\sqrt{100}\right)$$

$$Y = N\left(350; \sqrt{\frac{70\cdot 100}{24}}\right) \simeq N(350; 17)$$

$$P(320 < Y < 380) = P\left(\frac{320-350}{17} < \frac{Y-350}{17} < \frac{380-350}{17}\right)$$

$$= \phi\left(\frac{30}{17}\right) - \phi\left(\frac{-30}{17}\right)$$

$$= \phi(1,76) - (1 - \phi(1.76)) = 2\cdot\phi(1,76) - 1 = 2\cdot .9608 - 1 = \boxed{\sim 0,9216}$$

Crash course: Statistik

Sannolikhet vs Statistik

- Sannolikhet är en teori och statistik är det som vi får i praktiken

Sannolikhet

- Väntevärde μ:

= det man får i genomsnitt om man gör oändligt många försök

Statistik

- Oändligt många försök är inte realistiskt i praktiken

- Medelvärde: $\bar{x}$

= det man får i genomsnitt om man gör ändligt många försök

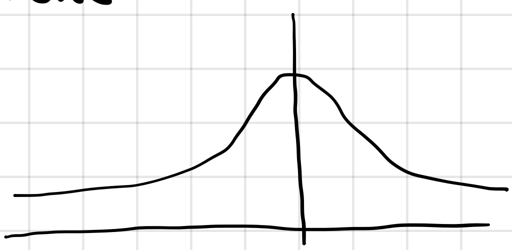
miniräknare: (TI-84)

stat → edit → L1 → sätt in alla värde

(2nd) → list → math → mean() → välj L1 (via stat)

- Varians σ^2

-avvikelse



- Stickprovsvariansen s^2

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

→

$n-1$ istället för n för att vi ska få det väntevärdesriktigt ska bli σ^2 i genomsnitt

-väljer ut några slumpmässiga stickprov, representerar alla resultat (praktiken)

• Stickprovsvarians på miniräknare:

stat → edit → skapa lista tex L1 → sätt in värde

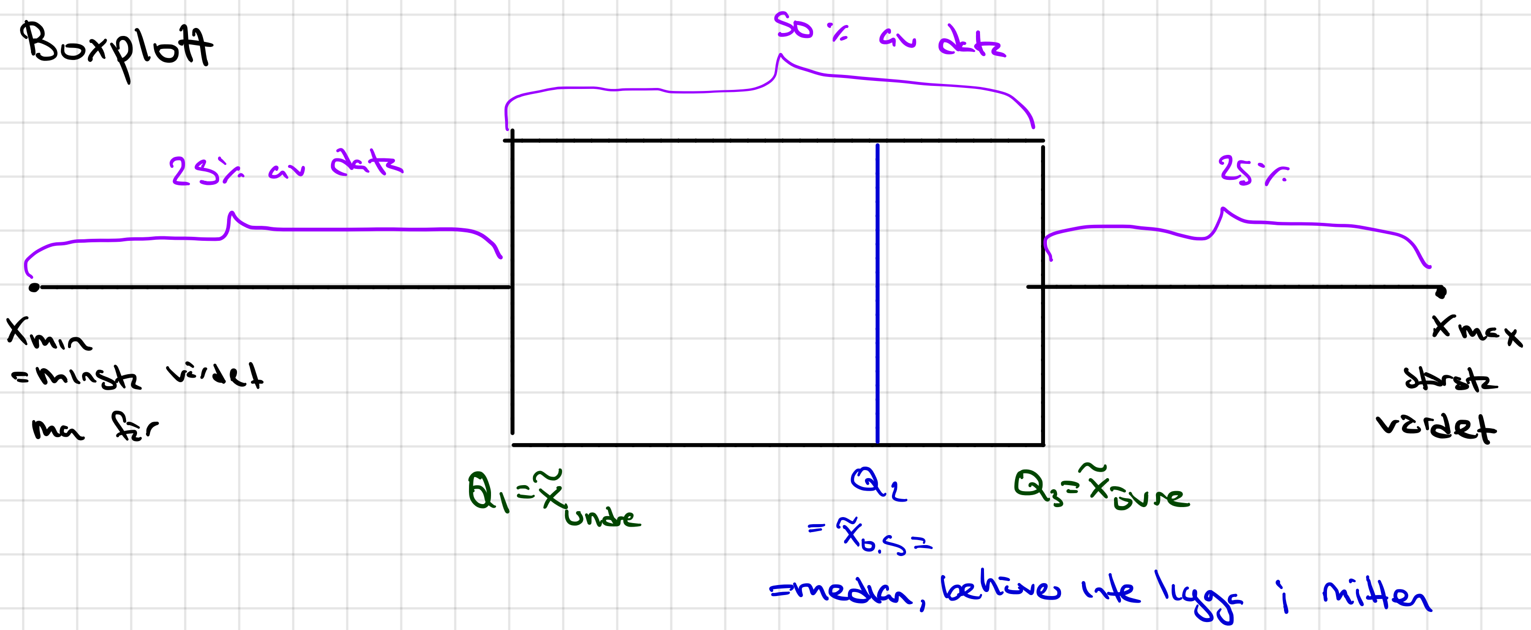
2nd → list - math → varians (för s^2) / std dev (s) $\}$ ⇒ välj rätt lista

• Statistik: populationsvarians: $= \frac{1}{n} \sum_{i=1}^{n} (x_i - x)^2$

- Ingen osäkerhet, tex i ett samhälle så kollar man alla individer

• I alla koncept vi ska kolla på i denna kurs använd s^2

Boxplott



• $Q_1 =$ första kvantilen, kallas även för 25%-percentilen, 25% av all data ligger till vänster

• $Q_2 = \tilde{x}_{0.5}$ = median, andra kvantilen, 50%-percentilen, 50% av data ligger till vänster

• Q_3 = tredje kvantilet, 75%-percentilen, 75% av data ligger till vänster

- (Q_1, Q_3) är kvantilintervall (50 % av data)
- $Q_3 - Q_1$ = kvantilavstånd
- (X_{min}, X_{max}) är variationsintervall

$X_{max} - X_{min}$ = variationsbredd

Hur tar man fram Q_x?

- Anta att vi söker Q_2 dvs medianen

- om vi har n st mätdata borde det vara mätdata nummer k där $\frac{k}{n} = 0.50$

väljer k till det heltal som uppfyller olikheten

$$\boxed{0.50n \le k \le 0.50n + 1}$$

- fungerar på alla sätt för alla Q_x

exempel

Anta att $n=11$ dvs vi har 11 st data. Vad är Q_1 och Q_3?

Q_1: $0.25n \le k \le 0.25n + 1$

$0.25n = 0.25 \times 11 = \frac{11}{4} = 2.75$

$0.25n + 1 = 3.75$

$2.75 \le k \le 3.75 \implies k=3$ är det heltal som uppfyller olikheten $\implies$

$Q_1 = X_3$

i ~~växande~~ ordning

Q_1 Q_3

$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9 \quad X_{10} \quad X_{11}$

$Q_3: \; 0.75n \le k \le 0.75n + 1$

$$\left.\begin{array}{l} 0.75 \times 11 = 8.25 \\ 0.75 \times 11 + 1 = 9.25 \end{array}\right\} \; k = 9 \quad \Rightarrow \; Q_3 = x_9$$

Vad händer om vi får flera k som uppfyller olikheten?

$n = 12$, Q_1: $\left.\begin{array}{l} 0.25 \times 12 = 3 \\ 0.25 \times 12 + 1 = 4 \end{array}\right\}$ $k=3$ och $k=4$ båda uppfyller olikheten

x_3 och x_4 ⇒ tar medelvärdet

$$Q_1 = \frac{x_3 + x_4}{2}$$

tenta 2020-05-26 # 6

Ange kvantilavstånd för följande mätdata

7 10 10 11 12 12 12 12 13 15 19 20 21

21 21 22 23 26 26 26 27 30 36

Lösning: kvantilavstånd: $Q_3 - Q_1$ (leja i FS)

$n = 24$

Q_1: $0.25 \times 24 \le k \le 0.25 \times 24 + 1$

$6 \le k \le 7$

$$Q_1: \; \frac{x_6 + x_7}{2} = \frac{12+12}{2} = 12$$

Q_3: $0.75 \times 24 \le k \le 0.75 \times 24 + 1$

$18 \le k \le 19$

$$Q_3: \; \frac{x_{18} + x_{19}}{2} = \frac{23+26}{2} = 24.5$$

k.a: $24.5 - 12 = \underline{\underline{12.5}}$

Punkt skattning

- En skattning av θ kallas θ^*_{obs} (observerade skattningen av θ)

- θ^*_{obs} är ett utfall av den stokastiska variabeln θ^*

- θ^* - stickprovs variabel : $\theta^* = \theta(X_1, X_2 \dots X_n)$

Skattningar i vanliga fördelningar

- viktigt för konfidens intervall

- Binomial-fördelning: $X \in Bin(n, p)$: n = antal försök vi gör, känd
  x - antal gånger vi lyckas.
  p - vill skatta

$$p^*_{obs} = \frac{x}{n}$$

- Hypergeometrisk fördelning: $X \in Hyp(N, n, p)$, väldigt lik Bin,

$$p^*_{obs} = \frac{x}{n}$$

- Poisson-fördelning: $X \in Po(\mu)$. $\mu = E(X)$, väntevärde, det som behövs för att kunna räkna ut sannolikheter

väntevärde är medelvärde i statistik:

$$\mu^*_{obs} = \bar{x}$$

- Exponential-fördelning: $X \in Exp(\lambda)$, λ okänd:

$$E(X_i) = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{E(X_i)} \quad \lambda^*_{obs} = \frac{1}{\bar{x}}$$

• ffg-fördelning: $X \in ffg(p)$, vill skatta p. $p = \frac{1}{E(X_i)}$ $p_{obs}^* = \frac{1}{\bar{x}}$

• Normalfördelning: $X \in N(\mu, \sigma)$ $\mu^*_{obs} = \bar{x}$

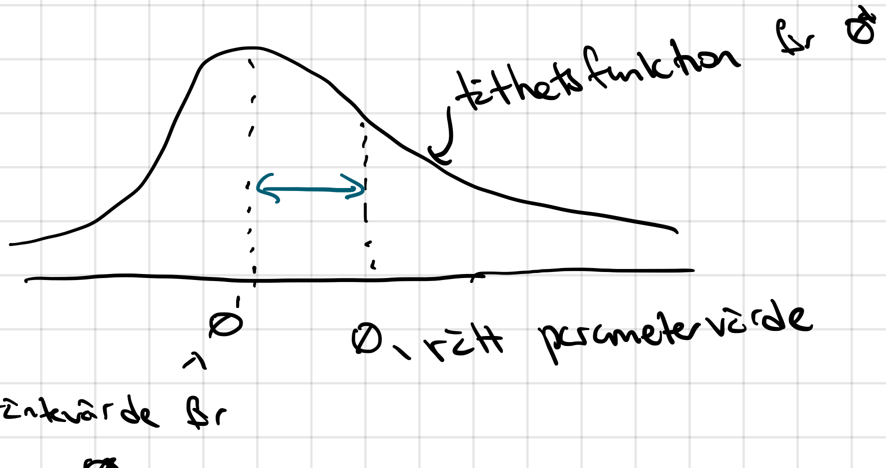
$\sigma^*_{obs} = s$

Väntevärdesriktighet

• Vill att praktiken ska vara lik teorin.

• En punktskattning θ^*_{obs} sägs vara väntevärdesriktig om tillhörande stickprovsvariabel θ^* har väntevärdet θ

$$E(\theta^*) = \theta$$



- detta är t.ex. en icke-väntevärdesriktig skattning då det är ett stort systematiskt fel

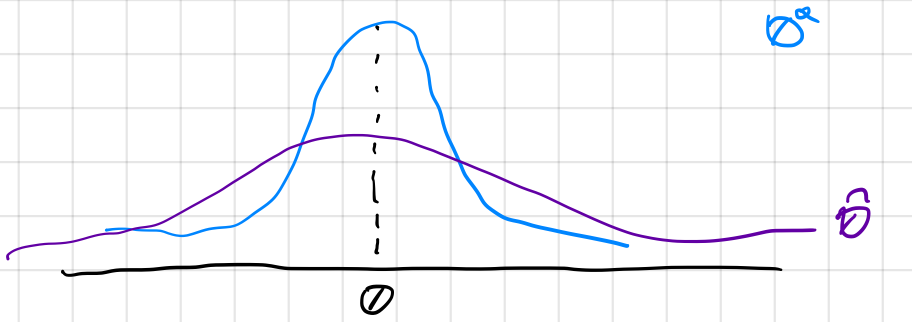
- här skiljer sig skattningens väntevärde $\theta' = E(\theta^*)$ kraftigt från θ.

Effektivitet

• Om två skattningar θ^*_{obs} och $\hat{\theta}_{obs}$ är väntevärdesriktiga och motsvarande stickprovsvariabler uppfyller

$V(\theta^*) < V(\hat{\theta})$ för alla $\theta \in \Omega_\theta$ så är θ^*_{obs} effektivare än $\hat{\theta}_{obs}$

- en väntevärdesriktig skattning med mindre varians är bättre



effektivitet är en jämförelse! kan säga att en skattning är mer effektiv än en annan, säger inte om skattningen är bra!

tenta 2019-12-19 # 7

Antag att X och Y är oberoende s.v. så att $E(X) = E(Y) = \theta$ och $V(X) = 3\sigma^2$ och $V(Y) = 2\sigma^2$. Låt x och y är observationer av X. resp. Y. Två skattningar av θ har föreslagits

$$\theta^*_{obs} = \frac{1}{2}(x+y) \quad \text{och} \quad \hat{\theta}_{obs} = \frac{1}{5}(2x+3y)$$

Väntevärdesriktighet? Effektivitet?

Lösning: Väntevärdesriktiga?

$$E(\theta^*) = \{= \theta ?\} = E\left[\frac{1}{2}X + \frac{1}{2}Y\right] = \{E(aX) = aE(X)\} =$$

$$= \frac{1}{2}E(X) + \frac{1}{2}E(Y) = \frac{1}{2}\theta + \frac{1}{2}\theta = \theta \quad \text{ok! vvr!}$$

$$E(\hat{\theta}) = E\left[\frac{2}{5}X + \frac{3}{5}Y\right] = \frac{2}{5}E(X) + \frac{3}{5}E(Y) = \frac{2}{5}\theta + \frac{3}{5}\theta = \theta \quad \text{ok!}$$

- Båda skattningar är väntevärdesriktiga. kollar variansen för att avgöra vilken skattning som är mer effektiv:

$$\underline{\underline{V(\theta^*)}} = V\left[\frac{1}{2}X + \frac{1}{2}Y\right] = \{V(aX) = a^2V(X)\} = \frac{1}{4}V(X) + \frac{1}{4}V(Y) =$$

$$= \frac{1}{4}3\sigma^2 + \frac{1}{4}2\sigma^2 = \frac{5}{4}\sigma^2 = \underline{\underline{1.25\sigma^2}}$$

$$V(\hat{\theta}) = V\left[\frac{2}{5}X + \frac{3}{5}Y\right] = \frac{4}{25}V(X) + \frac{9}{25}V(Y) =$$

$$= \frac{4}{25} \times 3\sigma^2 + \frac{9}{25} \times 2\sigma^2 = \frac{30}{25}\sigma^2 = 1.2\sigma^2$$

$V(\hat{\theta}) < V(\theta^*) \Rightarrow \hat{\theta}_{obs}$ är mer effektiv än θ^*_{obs}

Maximum likelihood-metoden (ML)

- bygger på att vi kan (eller antar att vi kan) sannolikhetsfunktionen eller täthetsfunktionen.

- eftersom vi har fått det mätdata man fick så bör sannolikheten vara stor att få just de värdena

$\Rightarrow$ maximera sannolikheten.

. Det värde $\theta_{obs\,ML}$ för vilket $L(\theta)$ antar sitt största värde inom Ω_θ kallas för ML-skattningen av θ.

om oberoende: diskret: $L(\theta) = p_{X_1}(x_1;\theta) \cdot p_{X_2}(x_2;\theta) \cdots p_{X_n}(x_n;\theta)$

kontinuerliga fallet: $L(\theta) = f_{X_1}(x_1;\theta) f_{X_2}(x_2;\theta) \cdots f_{X_n}(x_n;\theta)$

går ut på att hitta maximum till funktionen:

$$\frac{dL(\theta)}{d\theta} = 0$$

- om en jobbig funktion $\Rightarrow$ logaritmera först

Minsta-kvadrat metoden (MK)

- Går ut på att minimera skillnaden mellan praktiska och teoretiska värdena.

$$Q = \sum_{i=1}^{n} (x_i - \mu_i(\theta_1 \theta_2 \dots \theta_n))^n$$

$x_1 \dots x_n$ är oberoende och $\theta_1 \dots \theta_n$ är okända parametrar.

går ut på att hitta minimum till funktionen

$$\frac{dQ}{d\theta} = 0$$

- om Q beror på flera okända parametrar så kollar man på partiella derivator dvs tar gradienten till funktionen.

tenta 2020-03-10 #7

Låt X_1, X_2 och X_3 vara tre oberoende lika fördelade s.v. sådana att $X_i \in Po(\mu)$. Beräkna ML-skattningen av μ då $x_1 = 4$, $x_2 = 10$ och $x_3 = 1$

Lösning: $X_i \in Po(\mu)$ $\begin{cases} x_1 = 4 = k_1 \\ x_2 = 10 = k_2 \\ x_3 = 1 = k_3 \end{cases}$

ML: $L(\mu) = \{\text{oberoende}\} = p_{X}(x_1; \mu)\, p_{X_2}(x_2; \mu)\, p_{X_3}(x_3; \mu)$

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}$$

$$L(\mu) = \frac{\mu^4}{4!} e^{-\mu} \frac{\mu^{10}}{10!} e^{-\mu} \frac{\mu^1}{1!} e^{-\mu} = \frac{e^{-3\mu} \cdot \mu^{15}}{4! \cdot 10! \cdot 1!}$$

-hitta max genom att kolla på kritiska punkter till $L(\mu)$.

-derivatan kräver kedje och produkt regeln, har stora konstanter, stora potenser, enklare om vi logaritmerar först då max av $L(\mu)$ är samma som max av $\ln[L(\mu)]$.

$\ln(ab) = \ln(a) + \ln(b)$

$$\ln[L(\mu)] = \ln\left[\frac{\mu^{15} e^{-3\mu}}{10!\,4!}\right] = \ln(\mu^{15}) + \ln(e^{-3\mu}) - \ln(10!\,4!) =$$

$$= 15\ln(\mu) - 3\mu - \ln(10!\,4!)$$

max $\dfrac{d\ln[L(\mu)]}{d\mu} = 0$

$$\ln[L(\mu)]' = \frac{15}{\mu} - 3 = 0 \Leftrightarrow \frac{15}{\mu} = 3 \Rightarrow \mu_{obs,ML} = 5$$

tenta 2019-06-05

(9) Antag att vi har två oberoende observationer av x_1 och x_2 av $X_i \in Bin(n_i, p)$, $i = 1, 2$. Ange ML skattningen av p.

$$L(p) = \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{x_2} p^{x_2} (1-p)^{n_2-x_2}$$

eftersom vi har oberoende observationer och $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$L(p) = \frac{n_1!\, n_2!}{(n_1-x_1)!\, x_1!\, (n_2-x_2)!\, x_2!}\, p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2}$$

n_1, n_2, x_1, x_2 är konstanter

-jobbig funktion, tar ln

$$L(p) = c\, p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2} \quad \text{där } c = \frac{n_1!\, n_2!}{(n_1-x_1)!\, x_1!\, (n_2-x_2)!\, x_2!}$$

$$\ln[L(p)] = \ln\left[c\, p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2}\right] =$$

$$= \ln(c) + \ln(p^{x_1+x_2}) + \ln\left[(1-p)^{n_1+n_2-x_1-x_2}\right] =$$

$$= \ln(c) + (x_1+x_2)\ln p + (n_1+n_2-x_1-x_2)\ln(1-p)$$

$$\ln[L(p)]' = \frac{x_1+x_2}{p} + \frac{n_1+n_2-x_1-x_2}{1-p}(-1) = 0$$

$$\frac{x_1+x_2}{p} = \frac{n_1+n_2-x_1-x_2}{1-p} \Rightarrow (x_1+x_2)(1-p) = p(n_1+n_2-x_1-x_2)$$

$$x_1 + x_2 - x_1 p - x_2 p = p n_1 + p n_2 - p x_1 - x_2 p$$

$$p^*_{\text{obs m.}} = \frac{x_1+x_2}{n_1+n_2}$$

(10) Antag att vi har N st s.v. X_i som är oberoende och tillhör Γ-fördelningen $\Gamma(k, \theta)$

$$E(X_i) = \frac{k}{\theta} \quad \text{och} \quad V(X_i) = \frac{k}{\theta^2}$$

ML skattningen av θ är $\theta^*_{obs\,ML} = \frac{\sum^N x_i}{kN}$. Vad är variansen av denna skattning?

Lösning: $V(\theta^*) = V\left[\frac{\sum X_i}{kN}\right] = \frac{1}{k^2N^2} V(\sum X_i) = \frac{1}{k^2N^2} N V(X_i) =$

$$= \frac{1}{k^2 N} \cdot \frac{k}{\theta^2} = \frac{1}{kN\theta^2}$$

tenta 2020-01-08 #7

Låt x_i, $i = 1,2,3$ vara observationer av de oberoende s.v. X_i, $i=1,2,3$. De är gamma fördelade $\Gamma(2, \lambda)$ dvs har täthets funktionen:

$$f_{X_i}(x) \begin{cases} f_{X_i}(x) = \frac{\lambda^2}{1!} x e^{-\lambda x} & \text{om } x > 0, i=1,2,3 \\ 0 & \text{om } x \leq 0 \end{cases}$$

för $\lambda > 0$. Detta medför att $E(X_i) = \frac{2}{\lambda}$ och $V(X_i) = \frac{2}{\lambda^2}$ $i = 1,2,3$

Antag att följande värden har erhållits: $x_1 = 1.4075$
$x_2 = 0.7142$
$x_3 = 0.4004$

Bestäm MK-skattningen av λ

Lösning: $Q = \sum_{i=1}^{n} (x_i - \mu_i(\lambda))^2$

$\mu_i(\lambda) = \frac{2}{\lambda}$

$$Q = \left(1.4075 - \frac{2}{\lambda}\right)^2 + \left(0.7142 - \frac{2}{\lambda}\right)^2 + \left(0.4004 - \frac{2}{\lambda}\right)^2$$

$\frac{dQ}{d\lambda} = 0$ ger min:

$$Q' = 2\left(1.4075 - \frac{2}{\lambda}\right)\frac{2}{\lambda^2} + 2\left(0.7142 - \frac{2}{\lambda}\right)\left(\frac{2}{\lambda^2}\right) +$$

$$+ 2\left(0.4004 - \frac{2}{\lambda}\right)\left(\frac{2}{\lambda^2}\right) =$$

$$= \frac{4}{\lambda^2}\left[1.4075 - \frac{2}{\lambda} + 0.7142 - \frac{2}{\lambda} + 0.4004 - \frac{2}{\lambda}\right] = 0$$

$\lambda > 0$, ka förkorta:

$$1.4075 + 0.7142 + 0.4004 = \frac{2}{\lambda} + \frac{2}{\lambda} + \frac{2}{\lambda}$$

$$2.37897 = \frac{6}{\lambda} \Rightarrow \lambda^*_{obs\,mk} = \frac{6}{2.37897} = \underline{\underline{2.379}}$$

tenta 2020-05-26 # 13 (10p)

För att undersöka en bilmodellens bromseffekt mättes bromssträckan efter inbromsning vid hastigheterna 20, 40, 60, 80 och 100 km/h.

Resultat

hastighet (v)	20	40	60	80	100
bromssträcka (x) (m)	12	25	52	78	119

Rent teoretiskt skulle bromssträckan vara en kvadratisk funktion av hastigheten plus en slumpmässig normalfördelad avvikelse och därför antas bromssträckan vid inbromsning vid hastigheten v vara $N(\gamma(\frac{v}{20})^2, \sigma^2)$ där γ är okänd.

Skatta γ med Minsta-kvadrat metoden.

Lösning: MK : $Q = \sum_{i=1}^{n} (x_i - \mu_i(\theta))^2$

vårt fall: $\mu(\gamma) = E(X) = \gamma(\frac{v}{20})^2$

x_i är olika bromssträckor:

$x_1 = 12$ $\qquad \mu_1 = \gamma(\frac{20}{20})^2 = \gamma$

$x_2 = 25$ $\qquad \mu_2 = \gamma(\frac{40}{20})^2 = 4\gamma$

$x_3 = 52$ $\qquad \mu_3 = \gamma(\frac{60}{20})^2 = 9\gamma$

$x_4 = 78$ $\qquad \mu_4 = \gamma(\frac{80}{20})^2 = 16\gamma$

$x_5 = 119$ $\qquad \mu_5 = \gamma(\frac{100}{20})^2 = 25\gamma$

$$Q = (12-\gamma)^2 + (25-4\gamma)^2 + (52-9\gamma)^2 + (78-16\gamma)^2 + (119-25\gamma)^2$$

min av Q: $\dfrac{dQ}{d\gamma} = 0$

$$Q' = 2(12-\gamma)(-1) + 2(25-4\gamma)(-4) + 2(52-9\gamma)(-9) + 2(78-16\gamma)(-16) + 2(119-25\gamma)(-25) =$$

$$= -2(12-\gamma) - 8(25-4\gamma) - 18(52-9\gamma) - 32(78-16\gamma) - 50(119-25\gamma) \overset{!}{=} 0$$

$$-24 + 2\gamma - 200 + 32\gamma - 936 + 162\gamma - 2496 + 512\gamma$$

$$-5950 + 1250\gamma = 0$$

$$-9606 + 1958\gamma = 0 \quad \Rightarrow \gamma = 4.91$$

$$\gamma^*_{obs,\,mk} = 4.91$$

Medelfel

- En skattning av $D(\theta^*)$ kallas medelfelet för θ^* och betecknas $d(\theta^*)$ (en siffra)
- Medelfelet för skattningen θ är alltså $D^*(\theta^*)_{obs}$

observerade skattningen av standard avvikelsen för skattningen av θ.

exempel

Anta att $X \in N(\mu, \sigma)$ och vi har $\mu^*_{obs} = \bar{x}$. Vad är medelfelet för denna skattning?

- vi söker är $D^*(\mu^*)_{obs}$

$$\mu^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$D(\mu^*) = D(\bar{X}) = \sqrt{V(\bar{X})}$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\, nV(X_i) = \frac{V(X_i)}{n} = \frac{\sigma^2}{n}$$

$$D(\bar{X}) = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

medelfel

$$D^*(\mu^*)_{obs} = \left(\frac{\sigma}{\sqrt{n}}\right)^*_{obs} = \frac{\sigma^*_{obs}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

exempel

Anta att $X \in Bin(n, p)$ $\quad p^*_{obs} = \frac{x}{n}$. Vad är $d(p^*)$?

$$d(p^*) = D^*(p^*)_{obs}$$

$$D(p^*) = D\left(\frac{X}{n}\right) = \sqrt{V\left(\frac{X}{n}\right)}$$

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \{Bin\} = \frac{1}{n^2}\, np(1-p) = \frac{p(1-p)}{n}$$

$$D(p^*) = \sqrt{\frac{p(1-p)}{n}}$$

medelfelet: $D^*(p^*)_{obs} = \left(\sqrt{\frac{p(1-p)}{n}}\right)^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} =$

$= \left\{ p^*_{obs} = \frac{x}{n} \right\} = \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$

exempel

$X \in Po(\mu)$. Vad är medelfelet av μ^*?

vi vet att $\mu^*_{obs} = \bar{X}$

$D(\mu^*) = \sqrt{V(\mu^*)}$

$V(\mu^*) = V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} nV(X_i) = \frac{\mu}{n}$

$\Rightarrow D(\mu^*) = \sqrt{\frac{\mu}{n}}$

$D^*(\mu^*)_{obs} = \left(\sqrt{\frac{\mu}{n}}\right)^*_{obs} = \sqrt{\frac{\mu^*_{obs}}{n}} = \sqrt{\frac{\bar{X}}{n}}$

tenta 2018-10-24 #8

Låt $x = 38$ vara ett utfall av en s.v. $X \in Bin(600, p)$ där p är okänd. Bestäm medelfelet för skattningen $p^* = \frac{x}{n}$

$p^*_{obs} = \frac{x}{n} = \frac{38}{600}$

$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$

$D(p^*) = \sqrt{\frac{p(1-p)}{n}} \Rightarrow D^*(p^*)_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$

$\frac{X}{n} = p^*_{obs} = \frac{38}{600} \Rightarrow D^*_{obs} = \sqrt{\frac{\frac{38}{600}\left(1-\frac{38}{600}\right)}{600}} = 0.00994$

tenta 2013-12-20 # 11

Låt X_i, $i=1,2,3$ vara oberoende s.v. med väntevärde μ och standardavvikelse σ. Anta att μ skattas med

$$\mu^*_{obs} = \frac{(x_1 + x_2 + 2x_3)}{4}$$

och σ skattas med s. Låt utfallen på X_i vara $\begin{cases} x_1 = 5 \\ x_2 = 9 \\ x_3 = 7 \end{cases}$ Bestäm medelfelet för skattningen av μ.

Lösning: $V(\mu^*) = V\left(\frac{X_1+X_2+2X_3}{4}\right) = \frac{1}{16}\left[V(X_1) + V(X_2) + 4V(X_3)\right] =$

$= \frac{6}{16} V(X_i) = \frac{6}{16}\sigma^2 = \frac{3}{8}\sigma^2$

$D(\mu^*) = \sqrt{\frac{3}{8}\sigma^2} = \sqrt{\frac{3}{8}}\,\sigma \Rightarrow D^*_{obs} = \sqrt{\frac{3}{2}}\, s.$

måste hitta s:

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

måste hitta $\bar{x}$ först. $\bar{x} = \frac{x_1+x_2+x_3}{3} = \frac{5+9+7}{3} = 7$

$s^2 = \frac{1}{2}\left[(5-7)^2 + (9-7)^2 + (7-7)^2\right] = \frac{1}{2}(4+4+0) = 4$

$s = \sqrt{4} = 2. \rightarrow D^*_{obs} = \sqrt{\frac{3}{2}}\,2 = 1.22.$

tenta 2019-03-12 #7

Antag att $X \in N(\mu, \sigma)$ och $Y \in N(2\mu, 2\sigma)$ och de s.v. X och Y är oberoende. Vi skattar μ med

$$\mu^*_{obs} = \frac{x}{2} + \frac{y}{4} \quad \text{och} \quad \sigma \text{ med}$$

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + \left(\frac{y}{2} - \mu^*_{obs}\right)^2}$$

Vi har vidare fått de observerade värdena $x=7$ och $y=15$. Beräkna medelfelet av μ^* utgående från denna information.

$$\mu^*_{obs} = \frac{x}{2} + \frac{y}{4} = \frac{7}{2} + \frac{15}{4} = \frac{29}{4}$$

$$V(\mu^*) = V\left[\frac{X}{2} + \frac{Y}{4}\right] = \frac{1}{4}V(X) + \frac{1}{16}V(Y) = \begin{Bmatrix} V(X) = \sigma^2 \\ V(Y) = (2\sigma)^2 \end{Bmatrix} =$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{16}(2\sigma)^2 = \frac{\sigma^2}{4} + \frac{4\sigma^2}{16} = \frac{\sigma^2}{2} \Rightarrow D(\mu^*) = \frac{\sigma}{\sqrt{2}}$$

medelfel: $D^*(\mu^*)_{obs} = \dfrac{\sigma^*_{obs}}{\sqrt{2}}$

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + \left(\frac{y}{2} - \mu^*_{obs}\right)^2} = \sqrt{\left(7 - \frac{29}{4}\right)^2 + \left(\frac{15}{2} - \frac{29}{4}\right)^2} =$$

$$= 0.3536 \quad \Rightarrow D^*_{obs} = \frac{0.3536}{\sqrt{2}} = \frac{1}{4} = 0.25$$

tenta 2019-01-08 #11

Två undersökningar gjordes på två grupper av patienter, som hade blivit vaccinerade respektive inte vaccinerade. 500 vaccinerade patienter undersöktes där 49 blev sjuka. I den icke-vaccinerade gruppen var 600 undersökta där 58 blev sjuka. Låt p_1 vara andelen sjuka och vaccinerade och p_2 andelen sjuka och inte vaccinerade. Vad är medelfelet för skattningen av $p_1 - p_2$?

X_1 - vaccinerade och sjuka: $X_1 \sim Bin(500, p_1)$

$X_2 \sim Bin(600, p_2)$

$p^*_{1\,obs} = \frac{x_1}{n_1} = \frac{49}{500}$, $p^*_{2\,obs} = \frac{x_2}{n_2} = \frac{58}{600}$

$$V(p_1^* - p_2^*) = V\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = \frac{1}{n_1^2}V(X_1) + \frac{1}{n_2^2}V(X_2) = \{\text{Bin-fördelning}\} =$$

$$= \frac{1}{n_1^2} n_1 p_1(1-p_1) + \frac{1}{n_2^2} n_2 p_2 (1-p_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$D(p_1^* - p_2^*) = \sqrt{V(p_1^* - p_2^*)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$D^*_{obs} = \sqrt{\frac{p^*_{1\,obs}(1-p^*_{1\,obs})}{n_1} + \frac{p^*_{2\,obs}(1-p^*_{2\,obs})}{n_2}} =$$

$$= \sqrt{\frac{\frac{49}{500}\left(1-\frac{49}{500}\right)}{500} + \frac{\frac{58}{600}\left(1-\frac{58}{600}\right)}{600}} = 0.018$$

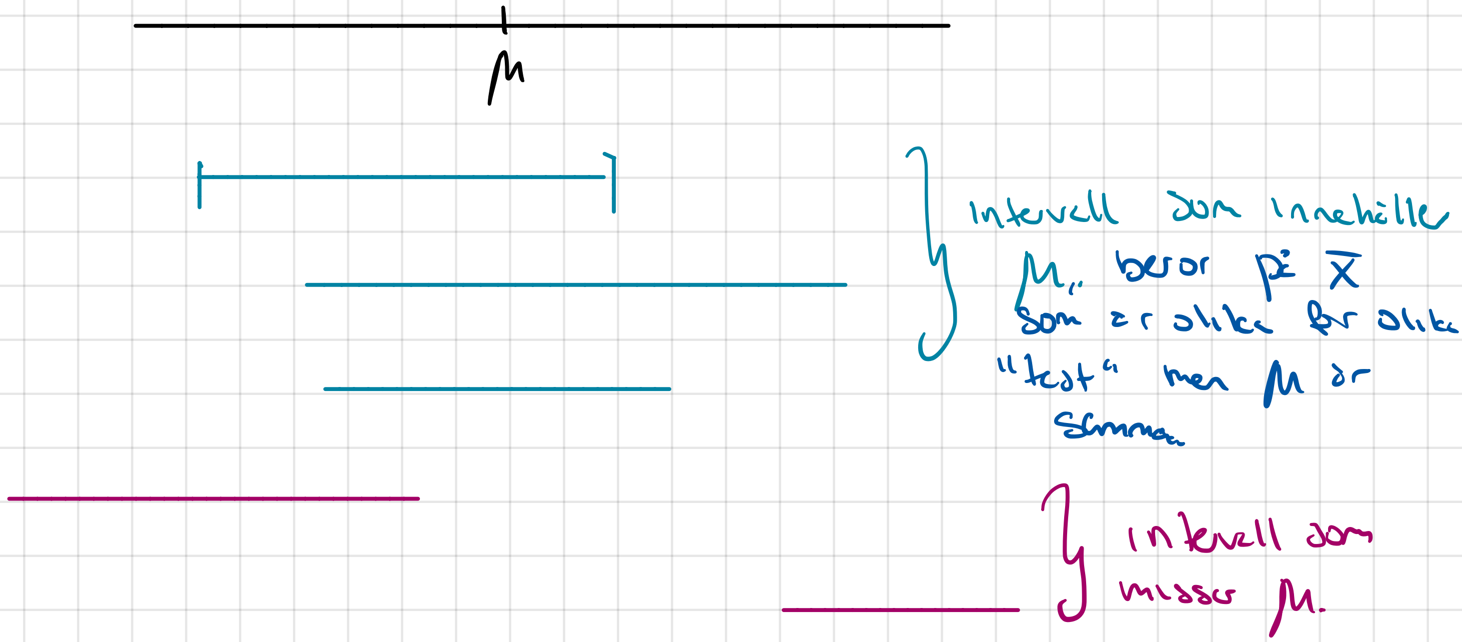
Sammanfattning för medelfel

1. Hitta $V(\theta^*)$
2. Hitta $D(\theta^*) = \sqrt{V(\theta^*)}$
3. Skatta $D(\theta^*) \Rightarrow D^*(\theta^*)_{obs}$

Konfidensintervall (konf-int)

- Definition: Ett intervall I_θ som med sannolikheten $1-\alpha$ täcker över θ kallas för ett konfidens-intervall för θ med konfidensgraden $1-\alpha$

$$P(\theta \in I_\theta) = 1-\alpha$$



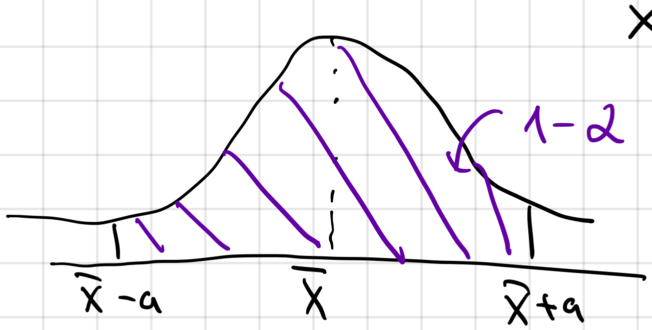
Förklaring till formulan i formelbladet §12.1

· Anta att X_i:na är oberoende och $X_i \in N(\mu, \sigma)$ med σ känd. Gör n st mätningar och vill bilda ett konf-int av graden $1-\alpha$ för μ.

- Skattar μ med $\bar{x}$ $\rightarrow$ $\hat{\mu}_{obs} = \bar{x}$

$X_i \in N(\mu, \sigma) \Rightarrow$ $\bar{X}$ också normalfördelad:

$\bar{X} \in N(\mu, \frac{\sigma}{\sqrt{n}})$



vi vill hitta a.

$$P(\mu \text{ finns i intervallet}) = 1-\alpha \Leftrightarrow P(|\bar{X}-\mu| < a) = 1-\alpha$$

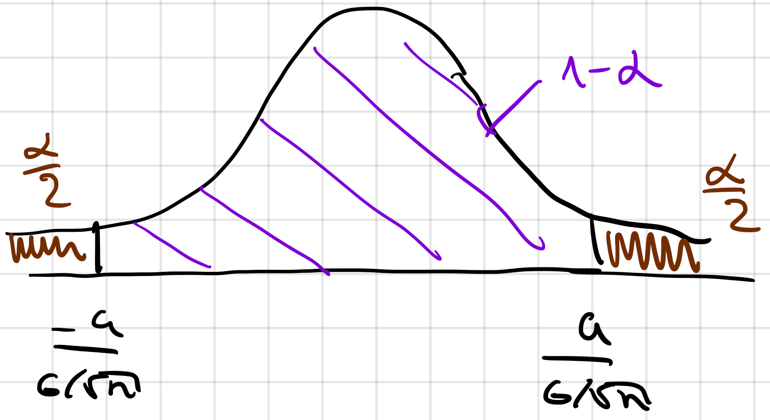
$$P(-a < \bar{X}-\mu < a) = 1-\alpha$$

$$P(\mu-a < \bar{X} < \mu+a) = 1-\alpha$$

gör om till $N(0,1)$ för att kunna använda tabell 2:

$$P\left[\frac{\mu-a-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\mu+a-\mu}{\sigma/\sqrt{n}}\right] = 1-\alpha$$

$$P\left(\frac{-a}{\sigma/\sqrt{n}} < Y < \frac{a}{\sigma/\sqrt{n}}\right) = 1-\alpha \quad , Y \in N(0,1)$$



Vill använda tabell 2: $P\left(Y > \frac{a}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2}$

$$\frac{a}{\sigma/\sqrt{n}} = \lambda_{\frac{\alpha}{2}} \Rightarrow a = \frac{\sigma}{\sqrt{n}} \lambda_{\frac{\alpha}{2}}$$

$$I_\mu = \left(\bar{x} - \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

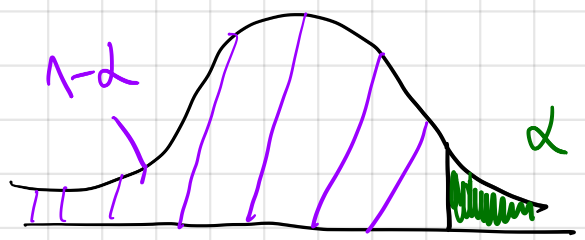
$$\Rightarrow \boxed{I_\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \lambda_{\frac{\alpha}{2}}}$$

ett stickprov, 2sidigt konf-int för väntevärdet

FS: §12.1 används när man söker konf-int för väntevärde när σ är känd.

Om vi har ensidigt intervall då?

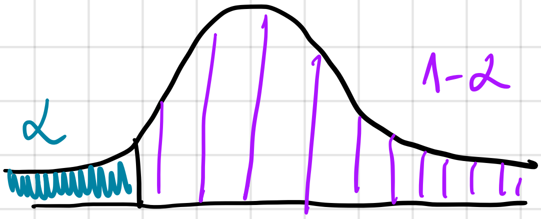
① Ensidigt uppåt begränsad konfidensintervall



$$I_\mu = \left(-\infty, \bar{x} + \frac{\sigma}{\sqrt{n}} \lambda_{\alpha}\right)$$

ensidigt!!!!!

② Ensidigt nedåt begränsad konfidensintervall:



$$I_\mu = \left(\bar{x} - \frac{\sigma}{\sqrt{n}}\lambda_\alpha \, , \, +\infty\right)$$

notera: $\boxed{\begin{array}{ll} \lambda_\alpha & \text{i ensidiga intervall} \\ \lambda_{\frac{\alpha}{2}} & \text{i tvåsidiga intervall} \end{array}}$

Konfidensintervallets bredd beror på:

1. En stor spridning hos mätdata (stor σ) ger bredare intervall

2. Ju fler observationer (stort n), ju smalare intervall

3. Ju större säkerhet vi vill ha, ju bredare intervall ($\lambda_{\frac{\alpha}{2}}$)

Konfidensintervall med σ okänd

- Måste skatta σ med s, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sigma^*_{obs}$

har då: $P\left(\frac{-a}{S/\sqrt{n}} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{a}{S/\sqrt{n}}\right) = 1-\alpha$

- S är inte konstant, utan beror på flera s.v.

⇒ använder t-fördelningen

precis som $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \in N(0,1)$ så gäller att $\frac{\bar{X} - \mu}{S/\sqrt{n}} \in t(n-1)$

§ 11.1 d

t-fördelningen finns i tabell 3. tar n-1 istället för n, fördelningen är annars ganska lik normalfördelningen:

$$\boxed{I_\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)}$$

→ §12.2 - används när σ okänd och vill ha konfidensintervall för μ.

$f = n-1$, frihetsgrader.

tenta 2020-04-16 #8

Antag att vi vill skatta väntevärdet μ för en normalfördelning med okänd standardavvikelse och tar därför ett slumpmässigt stickprov $x_1\, x_2 \dots x_n$ av storlek $n=15$ ur denna fördelning. Har $\bar{x} = 20$ och $s = 5$. Ange ett 99% konfidensintervall för μ.

- både μ och σ är okända ⇒ §12.2 i FS

$$I_\mu = \bar{x} \pm D^*_{obs}\, t_{\frac{\alpha}{2}}(n-1)$$

$$f = n-1 = 15-1 = 14$$

$$D^*_{obs} = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{15}} = 1.291$$

tvåsidigt intervall: $\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$

$$\Rightarrow t_{0.005}(14) = 2.98$$

$$I_\mu = 20 \pm 1.291 \times 2.98 = (16.15,\ 23.85)$$

tenta 2019-03-12 # 9

Vi mäter syrekoncentrationen i vatten och resultatet anses vara ett utfall av oberoende s.v. som följer $N(\mu, \sigma)$ där σ betecknar standardavvikelsen. Summan av de 31 mätvärden är 69.7 och stickprovsvariansen är 1.2772. Ange den övre gränsen till det ensidigt uppåt begränsade 95% - konfidensintervallet för μ.

μ och σ är okända $\Rightarrow$ §12.2

ensidigt (uppåt begränsad) $\Rightarrow$ $I_\mu = (-\infty, \bar{x} + D^*_{obs}\, t_{\alpha}(n-1))$

ensidigt. Inte $\frac{\alpha}{2}$!

$$\bar{x} = \frac{69.7}{31} = 2.248$$

$$D^*_{obs} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1.2772}{31}} = 0.203$$

$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

$t_{0.05}(31-1) = t_{0.05}(30) = 1.7$

$$I_\mu = (-\infty, \overbrace{2.248 + 0.203 \times 1.7}^{\text{övre gräns}})$$

övre gränsen: 2.59

Summanfattning (I_μ)

- Om σ känd $\Rightarrow$ §12.1

ensidigt: α

- Om σ okänd $\Rightarrow$ §12.2

om tvåsidigt: $\frac{\alpha}{2}$

Skillnad mellan 2 stickprov

- Anta att vi har X_i:na som är oberoende och $N(\mu_x, \sigma_x)$ fördelade, samt vi har Y_i:na som är oberoende och $N(\mu_y, \sigma_y)$ fördelade

- Vi vill ta fram $I_{\mu_x-\mu_y}$, har vi då tre fall:

(1) σ_x och σ_y är kända

- använder §12.1 och §11.3

$$D(\bar{X}-\bar{Y}) = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

$$\Rightarrow I_{\mu_x-\mu_y} = (\bar{x}-\bar{y}) \pm \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\, \lambda_{\frac{\alpha}{2}}$$

dvs enligt §12.1: $\theta^*_{obs} = \bar{x}-\bar{y}$

$\theta = \mu_x - \mu_y$

$D = D(\bar{x}-\bar{y})$ vilket finns i §11.3

(2) σ_x och σ_y är okända och olika

- väldigt stor osäkerhet, använder §12.3 approximativa metoden,

$\sigma^*_{x\,obs} = s_x$ och $\sigma^*_{y\,obs} = s_y$. har en approximativ konfidensgrad $1-\alpha$

$$I_{\mu_x-\mu_y} = (\bar{x}-\bar{y}) \pm \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\, \lambda_{\frac{\alpha}{2}}$$

③ σ_X och σ_Y är okända men lika

- har $\sigma_X = \sigma_Y = \sigma$ och är okända

§ 11.2 b): vikter ihop s_X och s_Y och får

$$s^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2}$$

· Eftersom σ är okänd ⟹ § 12.2 för att ta fram konf-int:

$$I_{\mu_X-\mu_Y} = (\bar{x}-\bar{y}) \pm s\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}\, t_{\frac{\alpha}{2}}(n_x+n_y-2)$$

$$s = \sqrt{\frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2}}$$

Stickprov i par

· Om X och Y är beroende av varandra och vi vill bilda ett konfidens intervall för $\mu_Y - \mu_X$ så måste vi först skapa en ny s.v. Z som för varje i: $Z_i = Y_i - X_i$

$Z_i \in N(\mu_z, \sigma_z)$

tenta 2019-01-08 # 10

Två stickprov från två populationer och varje stickprov uppfattas som observationer på $N(\mu_i, \sigma_i)$ där vi antar att $\sigma_1 = \sigma_2$. Har fått följande sammanfattande mått:

Stickprov 1: $n_1 = 4$ $\bar{x}_1 = 1007.25$ $s_1 = 143.66$

Stickprov 2: $n_2 = 4$ $\bar{x}_2 = 817.75$ $s_2 = 73.627$

Beräkna den undre gränsen i ett 95%-igt tvåsidigt konfidens intervall för $\mu_1 - \mu_2$

Lösning:

- Har 2 populationer som inte beror på varandra ⇒ skillnad mellan 2 stickprov med samma varians som är okänd. (fall 3)

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{3(143.66)^2 + 3(73.627)^2}{4+4-2} =$$

$$= 13029.56 \Rightarrow s = \sqrt{13029.56} = 114.147$$

$$D^*_{obs} = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 114.147\sqrt{\frac{1}{4}+\frac{1}{4}} = 80.714$$

$$§\ 12.2: \quad I_{\mu_x - \mu_y} = (\bar{x}_1 - \bar{x}_2) \pm D^*_{obs}\, t_{\underset{\alpha/2}{0.025}}(n_1+n_2-2)$$

$$\text{undre gränsen: } \bar{x}_1 - \bar{x}_2 - D^*_{obs}\, t_{0.025}(6) = -8.25$$

tenta 2020-01-08 #8

Kolesterolhalten i blodet mäts hos 5 personer före och efter en period med förändrad kost och motion. Antag att mätvärdena är normal fördelade. Ange med 95% säkerhet ett konfidens intervall för vilken genomsnittlig förändring av kolesterolhalten som kan påvisas om följande data finns

	1	2	3	4	5
före	262	272	284	298	294
efter	252	262	278	282	278

Lösning: mäter före och efter ⇒ beroende av varandra ⇒ stickprov i par!

Skapa: $z_i = x_i - y_i$ där x_i är före och y_i är efter

$z_1 = x_1 - y_1 = 262 - 252 = 10$

$z_2 = x_2 - y_2 = 272 - 262 = 10$

$z_3 = x_3 - y_3 = 284 - 278 = 6$

$z_4 = 298 - 282 = 16$

$z_5 = x_5 - y_5 = 294 - 278 = 16$

Skapar en lista: stat-edit-L1 och sätter in värdena för z_i

räkna ut: list-math - mean(L1) ⇒ $\bar{z} = 11.6$

-stdDev(L1) = $s_z = 4.3359$

$Z \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \sim \mu = \bar{x} = 11.6$

$\frac{s}{\sqrt{n}} = 1.9391$

både μ och σ okända ⇒ §12.2

har 5 st data: $n-1 = 4$

$t_{0.025}(4) = 2.78$

$I_z = \bar{z} \pm \frac{s_z}{\sqrt{n}} t_{0.025}(4) = 11.6 \pm 1.9391 \times 2.78 = [6.21, 16.99]$

Approximativa konfidens-intervall

- Baseras på användning av normalapproximation
- I FS: §12.3 och den används när mätdata inte är normalfördelad men vårt θ^* är approximativt normalfördelad

  - får då ett konfidens intervall med approximativa konfidensgraden $1-\alpha$

- Vi kan använda §5 och §6 för att veta när vi kan approximera

$$I_\theta = \theta^*_{obs} \pm d(\theta^*)\lambda_{\frac{\alpha}{2}}$$

$d(\theta^*)$ är medelfelet, samma uträkning som innan!

Binomial-fördelning

- används tex vid kvalitetskontroll

- Vi har $X \sim Bin(n,p)$ och vi skattar p med $p^*_{obs} = \frac{x}{n}$.
- p^*_{obs} är en skattning $\Rightarrow$ vi vill bilda ett konfidens intervall för p.
- Vi har inte en Normal fördelning $\Rightarrow$ använder §12.3

θ^* i vårt fall är $\theta^*_{obs} = p^*_{obs} = \frac{x}{n} = \frac{\text{antal lyckade försök}}{\text{total antal försök}}$

$\theta^* = \frac{X}{n}$. om X är approximativt N-fördelad så är även θ^* N-fördelad.

$X \sim N$-fördelad om $np^*_{obs}(1-p^*_{obs}) \geq 10$ då kan §12.3 användas.

$$I_\theta = \theta^*_{obs} \pm D^*_{obs} \frac{\lambda}{2}$$

$\theta = p$, $\theta^*_{obs} = p^*_{obs} = \frac{x}{n}$

medelfel: $V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$

$$D\left(\frac{X}{n}\right) = \sqrt{\frac{p(1-p)}{n}} \Rightarrow D^*_{obs} = \left(\sqrt{\frac{p(1-p)}{n}}\right)^*_{obs} =$$

$$= \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$$

$$\boxed{I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} \lambda_{\frac{\alpha}{2}}}$$

Konfidens intervall för skillnad mellan andelar

- Anta att vi har $X \in Bin(n_x, p_x)$ och $Y \in Bin(n_y, p_y)$ och vi vill bilda ett konfidensintervall för $p_x - p_y$

- Om $n_x p^*_x(1-p^*_x) \geq 10$ och $n_y p^*_y(1-p^*_y) \geq 10$ så kan vi använda §12.3

$$I_\theta = \theta^*_{obs} \pm D^*_{obs} \lambda_{\frac{\alpha}{2}} \quad : \quad \theta = p_x - p_y \text{ och } \theta^*_{obs} = \frac{x}{n_x} - \frac{y}{n_y}$$

$$V\left(\frac{X}{n_x} - \frac{Y}{n_y}\right) = V\left(\frac{X}{n_x}\right) + V\left(\frac{Y}{n_y}\right) = \frac{1}{n_x^2} V(X) + \frac{1}{n_y^2} V(Y) =$$

$$= \frac{1}{n_x^2} n_x p_x(1-p_x) + \frac{1}{n_y^2} n_y p_y(1-p_y) = \frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}$$

$$D\left(\frac{X}{n_x} - \frac{Y}{n_y}\right) = \sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}} \Rightarrow D^*_{obs} = \sqrt{\frac{p^*_{x\,obs}(1-p^*_{x\,obs})}{n_x} + \frac{p^*_{y\,obs}(1-p^*_{y\,obs})}{n_y}}$$

$$I_{p_x - p_y} = p^*_{x\,obs} - p^*_{y\,obs} \pm \sqrt{\frac{p^*_{x\,obs}(1-p^*_{x\,obs})}{n_x} + \frac{p^*_{y\,obs}(1-p^*_{y\,obs})}{n_y}} \lambda_{\frac{\alpha}{2}}$$

tenta 201 - 10-24 #10

Låt 20 personer utföra mätningar på en $N(\mu, \sigma)$-fördelad s.v. Därefter gör de vart och ett 95% konfidens-intervall för μ. Vilket är det troligaste värdet på det antal intervall som kommer att missa μ?

$$P(\text{missa}) = 1 - 0.95 = 0.05 \Rightarrow X \in Bin(20, 0.05)$$

$$E(X) = np = 20 \times 0.05 = 1$$

testtenta 2019 #10

Antag att $X \in Bin(200, p)$. Vi gör ett försök och får $x = 23$. Ange den övre gränsen för det ensidigt uppåt begränsade konf-int för p med approximativ konfidensgrad 95%

$$p^*_{obs} = \frac{x}{n} = \frac{23}{200}$$

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

$$D^*_{obs} = \sqrt{\frac{p^*_{obs}(1 - p^*_{obs})}{n}} = \sqrt{\frac{\frac{23}{200}\left(1 - \frac{23}{200}\right)}{200}} = 0.02256$$

ensidigt uppåt: $\lambda_\alpha = \lambda_{0.05} = 1.6449$

$$I_p^{(\text{övre})} = p^*_{obs} + D^*_{obs} \lambda_\alpha = \frac{23}{200} + 0.02256 \times 1.6449 = 0.152$$

tenta 2019-03-11 #9

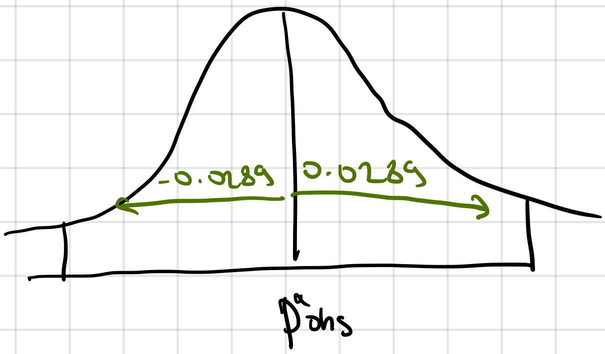
Antag att $X \in Bin(n, p)$ och vi vill bilda ett approximativt konf-int I_p för p. Vi gör $n = 500$ försök och får $x = 62$. Ange den tvåsidiga intervallets **bredd** på den approximativa graden 95%.

$$p^*_{obs} = \frac{x}{n} = \frac{62}{500}$$

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{p(1-p)}{n} \Rightarrow D^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$$

$$\lambda_{0.025} = 1.96$$

$$I_p = \frac{62}{500} \pm \sqrt{\frac{\frac{62}{500}\left(1 - \frac{62}{500}\right)}{500}} \times 1.96 = \frac{62}{500} \pm 0.0289$$



bredd: $2 \times 0.0289 = 0.0578$

Poisson-fördelning

- Vi antar att $X_i \in Po(\mu)$ och skattar μ med $\bar{X}$
- Vi vill bilda ett konf-int för μ och vill använda §12.3

- måste ha Θ^* är approximativt N-fördelad:

$$\Theta = \mu \Rightarrow \Theta_{obs} = \mu^*_{obs} = \bar{x} \Rightarrow \Theta^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

§6 i FS: $Po(\mu) \sim N(\mu, \sqrt{\mu})$ om $\mu \geq 15$

enligt satsen: Om $X_1 \in Po(\mu_1)$ och $X_2 \in Po(\mu_2)$ och X_1 och X_2 är oberoende så är

$$X_1 + X_2 \in Po(\mu_1 + \mu_2)$$

$$\sum X_i \in Po(\mu, \sqrt{\mu}) \Rightarrow \sum X_i \sim N(n\mu, \sqrt{n\mu})$$

om $n\mu \geq 15$

$\theta^*_{obs} = \bar{x}$, vad är $D^*_{obs} = ?$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} V\left(\sum X_i\right) = \frac{1}{n^2} n \mu = \frac{\mu}{n}$$

$$D^*_{obs} = \left(\sqrt{\frac{\mu}{n}}\right)^*_{obs} = \sqrt{\frac{\bar{x}}{n}}$$

$$\Rightarrow \boxed{I_\mu = \bar{x} \pm \sqrt{\frac{\bar{x}}{n}} \lambda_{\frac{\alpha}{2}}}$$

tenta 2019-12-19 #9

Antag att $X \in Po(\mu)$ där μ är okänt. Ange den övre gränsen för det ensidigt uppåt begränsade konfidens-intervallet för μ som har approximativa konfidensgraden 90%. Utgå från följande värden som är framtagna från fyra oberoende försök med $x_1 = 12$, $x_2 = 16$, $x_3 = 19$, $x_4 = 15$

$$\mu^*_{obs} = \bar{x} = \frac{12+16+19+15}{4} = 15.5 \qquad \mu \geq 15 \Rightarrow \sim N(\mu, \sqrt{\mu})$$

kan använda §12.3: $X \sim N(15.5, 3.937)$

$$I_\mu = (-\infty, \bar{x} + D^*_{obs} \lambda_{0.1})$$

tar 0.1 då vi har ensidigt intervall!

$\lambda_{0.1} = 1.2816$

$D^*_{obs} = \sqrt{\frac{\bar{X}}{n}} = \frac{3.937}{\sqrt{4}} = \frac{3.937}{2} = 1.9685$

övre gränsen: $\bar{x} + D^*_{obs}\lambda_{0.1} = 15.5 + 1.9685 \times 1.2816 =$

$= \underline{\underline{18.02}}$

tenta 2018-12-18 #12

Låt X_i, $i = 1, 2 \ldots 5$ vara oberoende Po-fördelade s.v. med väntevärde μ. Man får $\bar{x} = 16$ Ange den nedre gränsen för det ensidigt nedåt begränsade konfidensintervallet för μ med den approximativa konfidensgraden 95%.

$I_\mu = (\bar{x} - D^*_{obs}\lambda_{0.05}, +\infty)$

undre

$V(\bar{X}) = V\left(\frac{1}{5}\sum X_i\right) = \frac{1}{25} 5V(x) = \frac{1}{5}\mu$

$D^*_{obs} = \sqrt{\frac{\bar{x}}{5}} = \sqrt{\frac{16}{5}} = 1.789$

$\lambda_{0.05} = 1.6449$

$I_\mu = (\underbrace{16 - 1.789 \times 1.6449}, +\infty)$

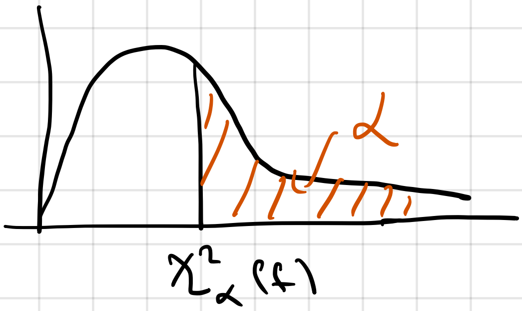
undre gräns: 13.06

Konfidens intervall för σ och σ^2

- Använder oss av §12.4 och χ^2 - fördelning.
  - Om $X_1, X_2 \dots X_f$ är $\in N(0,1)$ -fördelade och oberoende så gäller det att

$$\sum_{i=1}^{f} X_i^2 \in \chi^2(f) \qquad f = n-1$$

  - tabell 4:



$$P(X > \chi^2_\alpha(f)) = \alpha$$

- Anta att X_i:na är oberoende och $N(\mu, \sigma)$ -fördelade

$$\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \in \chi^2(n-1)$$

§11.2 b: $$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n-1}{\sigma^2} \underbrace{\frac{1}{n-1}\sum (x_i - \bar{x})^2}_{s^2} =$$

$$= \frac{n-1}{\sigma^2} s^2 \in \chi^2(n-1)$$

tabell 4: $$P\left(\frac{(n-1)s^2}{\sigma^2} > \chi^2_\alpha(n-1)\right) = \alpha$$

$$P\left(\sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_\alpha(n-1)}}\right) = \alpha$$

vänstra gränsen: $$P\left(\sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}\right) = \frac{\alpha}{2}$$

högra gränsen: $$P\left(\sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right) = 1-\frac{\alpha}{2}$$

$$\Rightarrow \boxed{I_\sigma = \left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right)}$$ §12.4

$$\boxed{I_{\sigma^2} = \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)}$$

tenta 2019-01-03 #9

Antag att $X_1 \ldots X_n$ utgör ett stickprov på $N(\mu, \sigma)$. Från 20 observationer erhålls följande värden: $\bar{x} = 0.46$ och $s = 0.43$. Ange den nedre gränsen för den tvåsidiga konfidens intervallet för σ med konfidensgraden 99%.

$\sigma^*_{obs} = s \Rightarrow$ använder §12.4

nedre gränsen: $$\theta^*_{obs}\sqrt{\frac{f}{\chi^2_{\frac{\alpha}{2}}(f)}}$$

$\sigma = \theta$

$\theta^*_{obs} = s$

$$\frac{\alpha}{2} = \frac{1-0.99}{2} = 0.005$$

$f = n-1 = 20-1 = 19$

$$\chi^2_{0.005}(19) = 38.6$$

$\Rightarrow \quad \sigma^*_{obs} \sqrt{\frac{f}{\chi^2_{\frac{\alpha}{2}}(f)}} = s\sqrt{\frac{19}{\chi^2_{0.005}(19)}} = 0.43\sqrt{\frac{19}{38.6}} = 0.302$

tenta 2019-05-28 #7

Antag att $X_1 \dots X_n$ utgör ett stickprov på $N(\mu, \sigma)$. Från 10 observationer erhålls följande värden: $\bar{x} = 2.13$ och $s = 1.03$. Ange den övre gränsen för den tvåsidiga konfidens-intervallet för σ^2 med konfidensgraden 95%.

Skattar σ med $\sigma^*_{obs} = s$, använder §12.4

övre gränsen ges av $s\sqrt{\frac{f}{\chi^2_{1-\frac{\alpha}{2}}(f)}}$ för σ då vi skattar sigma med s.

$s = 1.03, \quad 1-\frac{\alpha}{2} = 0.975, \quad f = n-1 = 10-1 = 9$

$\sigma: \quad 1.03\sqrt{\frac{9}{\chi^2_{0.975}(9)}} = 1.88$

för σ^2: $1.88^2 = 3.54$

Hypotesprövning

- Den hypotes som man vill testa om man ska förkasta eller inte kallas för nollhypotesen, H_0, ska alltid kunna förkastas.

- Mothypotesen till H_0 kallas för H_1

- Risknivå = signifikansnivå = α, högsta risken vi kan ta att förkasta H_0 om H_0 är sann.

$P(\text{förkasta } H_0)$ om H_0 sann $< \alpha$

- p-värde: Sannolikheten att förkasta H_0 under förutsättningen att H_0 är sann givet att vi har fått ett visst värde på test-variabeln = observationen

$$p\text{-värdet} < \alpha \Rightarrow \text{förkasta } H_0 \text{, annars inte.}$$

- Styrkefunktionen: $h(\theta) = P(\text{förkasta } H_0)$ om θ är rätt parametervärde

Styrkan hos testet för $\theta = \theta_1$: $P(\text{förkasta } H_0 : \theta = \theta_0)$ om $H_1 : \theta = \theta_1$ är sann.

Konfidensintervall metoden

- Om $H_0 : \mu = a$, $H_1 : \mu \neq a$ } tvåsidigt konf-int

  - bilder ett konfidens intervall för μ och om a ligger i konf-int & så kan vi inte förkasta H_0 på den risknivån

vid ensidig test: $H_0 : \mu = a$

$H_1 : \mu < a$ begränsas uppåt.

tenta 2019-10-23 # 9

Följande konfidens intervall med konfidensgraden 99%. för skillnaden mellan 2 stickprovs väntevärden är given:

$$I_{\mu_X - \mu_Y} = (0.12, 0.32)$$

Antag att stickproven kommer från $X_i \in N(\mu_X, \sigma_X)$ och $Y_i \in N(\mu_Y, \sigma_Y)$ samt alla s.v. är oberoende. Antag vidare att σ_X och σ_Y är kända. Man önskar testa $H_0 : \mu_X = \mu_Y$ mot $H_1 : \mu_X \neq \mu_Y$. Vilken slutsats kan vi dra om att förkasta H_0?

$H_0: \mu_X = \mu_Y$

$\mu_X - \mu_Y = 0$. Måste kolla om 0 finns i intervallet:

$I_{\mu_X - \mu_Y} = (0.12, 0.32)$ 99%

$0 \notin I_{\mu_X - \mu_Y} \Rightarrow$ kan förkasta H_0 på risknivån 1% och även 5%.

tenta 2019-08-12 #11

Vi antar att $X \in Exp(\lambda)$. Nollhypotesen är $H_0: \mu = 4$ där $\mu = E(X)$ och mothypotesen är $H_1: \mu > 4$. Har en observation $x = 8$. Bestäm testets p-värde.

p-värde = p (förkasta H_0) om H_0 sann. $= P(X \geq 8)$ om $\mu = 4$

$$P = \int_8^\infty \lambda e^{-\lambda x} \, dx$$

$$\lambda = \frac{1}{E(X)} = \frac{1}{4}$$

$$\text{p-värde} = \int_8^\infty \frac{1}{4} e^{-\frac{1}{4}x} \, dx = \left[-e^{-\frac{1}{4}x} \right]_8^\infty = e^{-\frac{1}{4} \cdot 8} = 0.135$$

tenta 2019-12-19 #1

Antag att $X \in ffg(p)$ och låt $H_0: p = 0.25$. Vill testa H_0 mot $H_1: p < 0.25$ och vill förkasta H_0 om vi observerar ett stort värde x på X. Antag att vi har observerat $x = 5$. Bestäm testets p-värde.

Stort X: $P(X \geq 5) = 1 - P(X \leq 4)$

$p = 1 - \left[p + p(1-p) + p(1-p)^2 + p(1-p)^3 \right] = \{ p = 0.25 \} =$

$= 0.3164$

tenta 2019-01-03 #12

En forskare har gjort 10 försök som anses vara oberoende av varandra där sannolikheten för lyckat försök är p. Låt X stå för antalet lyckade försök. Forskaren önskar pröva $H_0: p = \frac{1}{2}$ mot $H_1: p > \frac{1}{2}$. Resultatet av de 10 försök var 8 lyckade. Beräkna p-värdet.

$n = 10, \quad p = \frac{1}{2} \Rightarrow X \sim Bin(10, 0.5)$ enligt H_0

$P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7) = 0.0547$

tenta 2018-10-24 #12

Givet är 7 oberoende observationer av en s.v. X som är $N(\mu, 0.4)$. Medelvärdet beräknat på 7 observationer är 0.719. Vill testa $H_0: \mu = 1$ mot $H_1: \mu < 1$. Vad är testets p-värde?

p-värde $= P(X > \mu) = \{$ gör om till $N(0,1) \} = P\left(\frac{X - \bar{x}}{\sigma/\sqrt{n}} , \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \right) =$

$= P\left(Y > \frac{1 - 0.719}{0.4/\sqrt{7}}\right) = P(Y > 1.86) = 1 - P(Y \leq 1.86) = 0.0314$

χ^2 - test: test av en given fördelning

- Används när nollhypotesen är att vi har en viss sannolikhets funktion

$H_0: p(A_1) = p_1 \quad p(A_2) = p_2 \ldots$ summan av alla är 1.

§14.3 : $Q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$ villkor för test: alla $np_i > 5$

r - antal resulteler.

om $Q > \chi^2_\alpha(f)$, $f = r-1$ förkasta H_0 på risknivån α

om $Q \leq \chi^2_\alpha(f)$, förkasta ej H_0 på risknivån α

- tabell 4 för χ^2 - fördelnings värdena.

tenta 2019-01-08 #15 (10p)

Man misstänkte att ett roulettbord på ett kasino var manipulerat och genomförde ett test med 8000 försök. Om rouletten är korrekt ska röd, svart och grön komma i proportionerna 18 : 18 : 1. Testresultatet gav röd: 3751, svart: 4018, grön: 231. Avgör med felrisken 1%. om rouletten är korrekt.

- H_0 nollhypotesen att vi har en viss sannolikhetsfunktion.

$\Rightarrow$ test av en given fördelning.

$n = 8000$

$A_1 =$ röd $\quad x_1 = 3751$
$A_2 =$ svart $\quad x_2 = 4018$
$A_3 =$ grön $\quad x_3 = 231$
$r = 3$

$18+18+1=37$

H_0: $p(A_1) = \frac{18}{37}$ $\quad p(A_2) = p_2 = \frac{18}{37}$ $\quad p(A_3) = p_3 = \frac{1}{37}$

Villkor för χ^2-test: alla $np_i \geq 5$

np_3 (grön) har lägst sannolikhetsfunktion:

$8000 \times \frac{1}{37} > 5 \Rightarrow$ χ^2-test ok att använda.

$$Q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{\left(3751 - \frac{18}{37} \times 8000\right)^2}{\frac{18}{37} \times 8000} +$$

$$+ \frac{\left(4018 - \frac{18}{37} \times 8000\right)^2}{\frac{18}{37} \times 8000} + \frac{\left(231 - \frac{1}{37} \times 8000\right)^2}{\frac{1}{37} \times 8000} = 10.20$$

- Ska vi förkasta H_0?

$$\chi^2_\alpha(r-1) = \chi^2_{0.01}(2) = 9.21$$

$Q > \chi^2_\alpha(r-1) \Rightarrow Q =$ stort, förkasta H_0 på risknivå 1%.

χ^2-test: test av en given fördelning med skattade parametrar

- Vill testa $H_0: p(A_1) = p_1(\theta)$ ... för en okänd parameter θ.

- Skatta θ först, helst använda ML

$$Q' = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*}$$

approximativt $\chi^2(r-k-1)$ fördelad där k är antalet skattade parametrar.

villkor $np_i^* \geq 5$

$Q > \chi^2_\alpha(r-k-1) \rightarrow$ förkasta H_0 på risknivå α

$Q \leq \chi^2_\alpha(r-k-1) \Rightarrow$ förkasta inte H_0 på risknivå α.

tenta 2020-01-08 #11

Man vill undersöka om följande data kommer från en Bin-fördelning med parametern $n=3$ för något värde på p.

	0	1	2	3	totalt
antal observationer	15	54	28	3	100

Om $p^*_{obs} = 0.3967$ används så får man $Q = 6.5329$.
Vad kan man säga om att förkasta H_0?

1% och 5%

Lösning: χ^2-test med en skattad variabel

$r = 4$, $k = 1$ (endast p^*_{obs} är skattad)

$\chi^2_{0.01}(r-k-1) = \chi^2_{0.01}(4-1-1) = \chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.05}(2) = 5.99$

$Q > \chi^2_{0.05}(2) \Rightarrow$ kan förkasta H_0 på 5% risknivå

$Q < \chi^2_{0.01}(2) \Rightarrow$ kan inte förkasta på 1% risknivå.

Tenta 2019-06-05 #12

Man misstänker att ett roulettbord på ett kasino var manipulerat och genom förde ett test med 8000 försök. Om korrekt: 18:18:1. Testet gav: röd = 3771, svart = 4002, grön = 227. H_0 att vi teorier sannolikheter, $Q = 7.4$. P-värdet?

test av en given fördelning: $\chi^2(r-1) = \chi^2(2)$

för p-värde kolla tabell 4

$\chi^2_{0.025}(2) = 7.38$

$\chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.025} < Q < \chi^2_{0.01}$ $\quad$ Q närmare $\chi^2_{0.025}(2)$

p-värde mellan 1% och 2.5%, lite närmare till 2.5%.

χ^2-test: homogenitetstest:

- Används när nollhypotesen är att vi har samma sannolikhetsfunktion i alla grupper

$$Q = \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\left(x_{ij} - \frac{n_i m_j}{N}\right)^2}{\frac{n_i m_j}{N}}$$

villkor: alla $\frac{n_i m_j}{N} \geq 5$

om $Q > \chi^2_\alpha[(r-1)(s-1)]$ förkasta Ho

tenta 2020-03-10 #11

studerar 570 personer

	typ 1	typ 2	placebo
~~biverkning~~	27	17	10
~~ingen biverkning~~	373	73	70

För att testa hypotesen Ho: Det finns ingen skillnad mellan grupperna, fick $Q = 13.62$. Slutsats?

- här homogenitetstest: $s = 2$, $r = 3$ $= \chi^2[(3-1)(2-1)] = \chi^2(2)$

$\chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.001}(2) = 13.8$

$Q = 13.62 > \chi^2_{0.01}$ ⇒ förkasta Ho på risknivå 1%.

$Q < \chi^2_{0.001}$ ⇒ kan inte förkasta på risknivå 0.1%.

χ^2-test: oberoendetest

. Om vi vill undersöka om två egenskaper A och B är oberoende

- går till som homogenitetstest

Ho: antar att oberoende.