Kræsjkurs stat 110

Innhold:

1. Mengdelære og sannsynlighetsbegrepet
2. Tilfeldige variabler og sannsynlighetsfunksjoner mm.
3. Viktige sannsynlighetsfordelinger
4. Simultanfordelinger
5. Grenseteoremer og utvalgsfordelinger
6. Punktestimatorer og intervallestimatorer
7. Hypotesetesting

Mengdelære

Mengde: En samling av matematiske objekter / elementer

Eksempel:

$S_1 = \{1, 2, 3\}$ $S_2 = \{(a,b), 7, 9\}$

Eksperiment:

En prosedyre der en tilhørende mengde representerer alle mulige utfall av prosedyren. Et eksperiment må kunne repeteres.

Utfallsrom:

mengden som tilhører et eksperiment. Denne mengden inneholder alle mulige utfall til eksperimentet.

Notasjon: Ω eller S

Eksempel: $\Omega = \{1,2,3,4,5,6\}$

Hendelse:

En delmengde av et utfallsrom tilhørende et eksperiment.

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

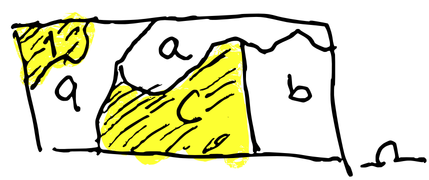
$A = \{1, 3, 5\} \subseteq S$ er en hendelse.

Er $B = \{1, 2, 3, 4, 5, 6\}$ en hendelse? Ja

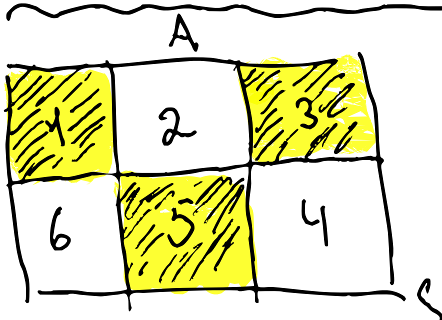
Hvis $\Omega = \{1, 9, a, b, c\}$.

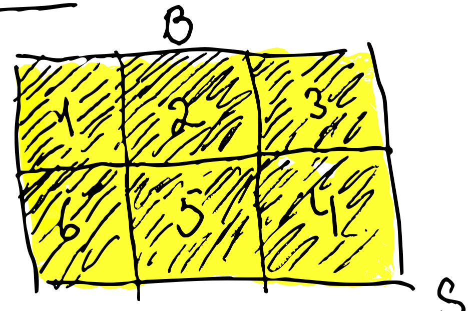
Er $G = \{1, d, c\}$ en hendelse?

Nei



Venndiagram





Mengdeoperatorer

Komplement:

Komplementet til en mengde A er hele utfallsrommet uten A.
↳ "S"

Notasjon: $\bar{A}$, A' eller A^c
↳ bok

Union:

Union til to (eller flere) hendelser A og B er hendelsen der A og/eller B inntreffer.

Notasjon: "$\cup$" ($A \cup B$)

Snitt:

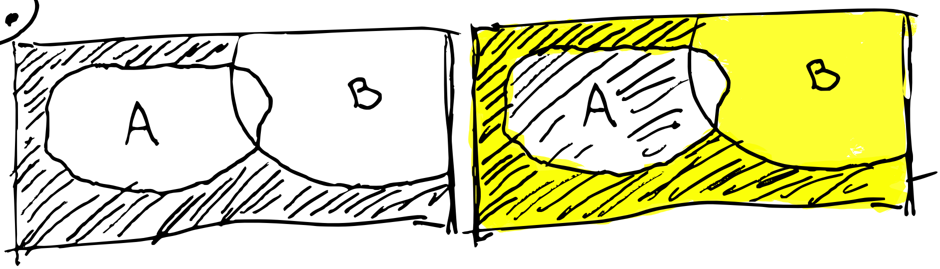
Snitt til to (eller flere) hendelser A og B er hendelsen der A og B inntreffer samtidig. **Notasjon:** "$\cap$" ($A \cap B$)

De' Morganes lover:

1. $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$

2. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

1.



Eksempel der vi bruker De Morganes lover:

$\bar{A} = S \setminus A$ (Def. komplement)

1) $S = A \cup \bar{A} \Rightarrow \bar{A} \cap A = \emptyset$

Bevis 1):

$S = A \cup \bar{A} \Rightarrow \bar{S} = \overline{(A \cup \bar{A})}$

$\Rightarrow \bar{S} = \bar{A} \cap \overline{(\bar{A})} = \bar{A} \cap A$

$\bar{S} = \emptyset \Rightarrow \bar{A} \cap A = \emptyset$

$\blacksquare$

Sannsynlighet

Def. Sannsynlighet til en hendelse A er størrelsen / målet til A. Notasjon: $P(A)$

Aksiomer.

1. $P(A) \in [0,1]$ for vilkårlig A
2. $P(S) = 1$, $P(\emptyset) = 0$
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ $\forall$ $A_i \cap A_j = \emptyset$ der $i \neq j$

b) $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ (generelt)

Eksempel:

La $P(A), P(B) > 0$ der $P(A \cap B) > 0$.

$$P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{>0}$$

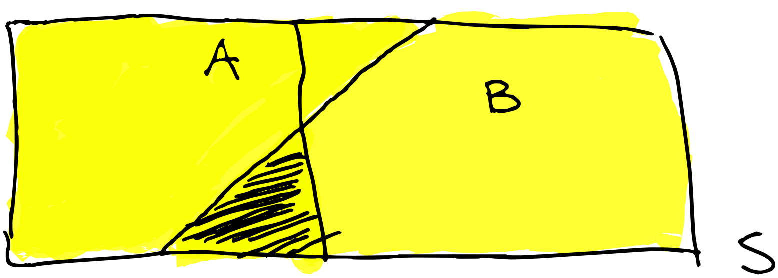
$$P(A) + P(B) = P(A) + P(B)$$

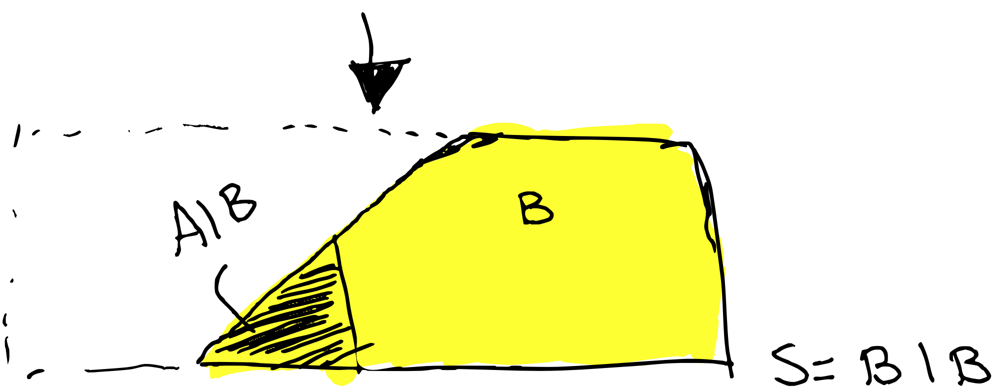
$$\Rightarrow P(A \cup B) < P(A) + P(B).$$

Betinget sannsynlighet

Def. A og B er hendelser der $P(B) > 0$.
Betinget sannsynlighet for A gitt B $(A|B)$ er

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





"Nytt" utfallsrom fordi der $P(B|B) = 1$

uavhengighet

Def. To hendelser A og B er uavhengige hvis B og A ikke gir ytterlig informasjon om A og B - henholsvis.

Viktig:

A og B er uavhengige $\iff$

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

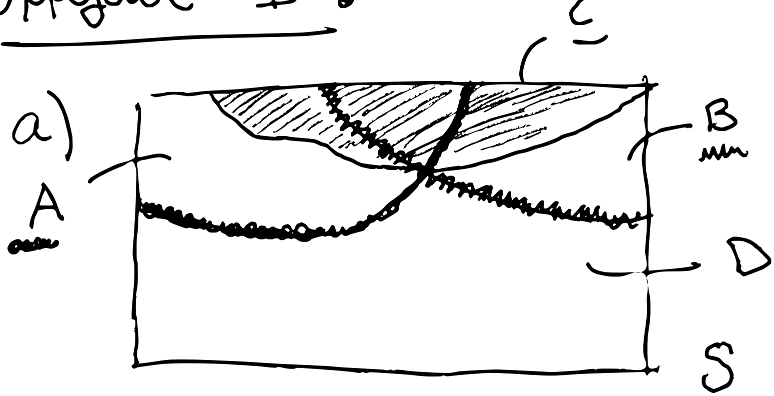
Konsekvens:

A og B er uavhengige $\iff$

1. $P(A \mid B) = P(A)$
2. $P(B \mid A) = P(B)$

oppgave 1:

a)



D og E er disjunkte fordi $E \cap D = \phi$ som er def. på disjunkte hendelser.

b) $P(C) = P(A \cap C) + P(B \cap C) - P(A \cap B)$

$= P(C|A)P(A) + P(C|B)P(B) - P(A \cap B)$

$= P(C|A)P(A) + P(C|A)P(B) - P(A \cap B)$

$= P(C|A)\{P(A) + P(B)\} - P(A \cap B)$

$= P(C|A)\{P(A \cup B) + P(A \cap B)\} - P(A \cap B)$

$= P(C|A)P(A \cup B) + P(A \cap B)\{P(C|A) - 1\}$

$= P(C|A)P(A \cup B) - P(A \cap B)\{1 - P(C|A)\}$

$= P(C|A)P(A \cup B) - P(C'|A)P(A \cap B)$

c)

Anta at c er uavhengig av A og B. Da

$$P(C) = P(A \cup B)P(C) - P(A \cap B)(1 - P(C))$$

$$P(C) - P(A \cup B)P(C) - P(A \cap B)P(C)$$

$$= -P(A \cap B) \Rightarrow$$

$$P(C)\{1 - (P(A \cup B) + P(A \cap B))\} = -P(A \cap B)$$

$$\Rightarrow$$

$$P(C) = -\frac{\overbrace{P(A \cap B)}^{>0}}{\underbrace{1 - \{P(A) + P(B)\}}_{>0}} < 0$$

d) $P(C) =$

$$P(A \cup B)P(C|A) - P(A \cap B)P(C'|A) =$$

$$\frac{1}{2}P(A \cup B) - \frac{1}{2}P(A \cap B) = \frac{1}{2}\{P(A \cup B) - P(A \cap B)\}$$

$$P(D) = 1 - P(A \cup B \cup C) = 1 - P(A \cup B).$$

$$P(D \cup C) = 1 - \frac{1}{2}\{P(A \cup B) + P(A \cap B)\}$$

$$= 1 - \frac{1}{2}\{P(A) + P(B)\} = 1 - 0{,}5 \cdot 0{,}45 = \underline{\underline{0{,}775}}$$

PAUSE

Tilfeldige variabler:

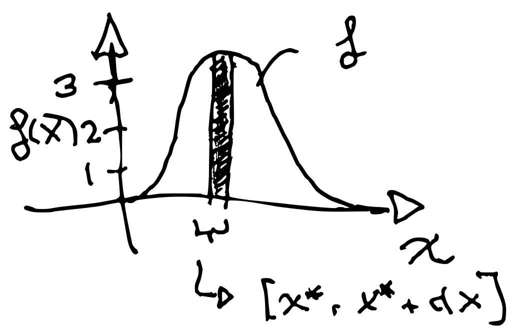
Def. En tilfeldig variabel X er definert ved

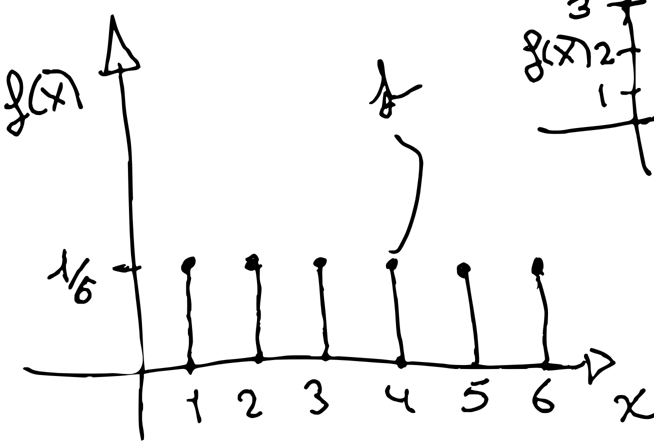
$$X : S \to \mathcal{M}$$

Der $\mathcal{M}$ er et målrom (ofte $\mathbb{R}$)

Sannsynlighetsfunksjoner:

Def. Gitt en T.V. X, er en sannsynlighetsfunksjon definert som en funksjon som beskriver sannsynligheten til alle utfall til X.





Forventning

Def. Forventning til en T.V. X er et vektet gjennomsnitt av alle mulige utfall tilhørende X.

Altså:

$$E[X] = \sum_{i \in S_X} f_X(i) \cdot i \qquad \text{eller}$$

$$E[X] = \int_{i \in S_X} f_X(i) \cdot i \, di$$

Forventning til en funksjon y

$$E[y(X)] = \sum_{i \in S_X} f_X(i)\, y(i) \qquad \text{eller}$$

$$E[y(X)] = \int_{i \in S_X} f_X(i)\, y(i)\, di$$

Resultat

$$E[aX] = a\,E[X] \quad \text{for alle} \quad a \in F$$

fordi

$$E[aX] = \sum_{i \in S_X} f_X(i) \cdot a \cdot x = a \overbrace{\sum_{i \in S_X} f_X(i)\, x}^{E[X]}$$

Varians:

Def. Varians til en T.F. X er et mål på spredningen til X, der denne spesifikke spredningen er definert ved

$$Var(X) = E\left[(X - E[X])^2\right].$$

Resultat.

$$Var(X) = E[X^2] - (E[X])^2$$

Varians til en funksjon:

$$Var(y(X)) = \sum_{i \in S_X} f_X(i) \cdot (y(i) - E[y(X)])^2$$

$$= \sum_{i \in S_X} f_X(i) \{ y(i)^2 - 2y(i)E[y(X)] + (E[y(X)])^2 \}$$

$$= E[\{y(X)\}^2] - 2E[y(X)] \underbrace{\sum_{i \in S_X} f_X(i) y(i)}_{= E[y(X)]} + E[y(X)]^2$$

$$= E[\{y(X)\}^2] - \{E[y(X)]\}^2$$

Varians til $y(X) = a + bX$

$$\mathrm{var}(a+bX) = E[\{a+bX\}^2] - \{E[a+bX]\}^2$$

$$= E[a^2 + 2abX + b^2X^2] - \{E[a] + bE[X]\}^2$$

$$= a^2 + 2abE[X] + b^2E[X^2] - a^2 - b^2E[X]^2 - 2abE[X]$$

$$= b^2E[X^2] - b^2(E[X])^2 = b^2\{\underbrace{E[X^2] - \{E[X]\}^2}_{\mathrm{var}(X)}\}$$

$$= b^2\,\mathrm{var}(X).$$

Standardavik:

$$SD(X) = \sqrt{\mathrm{var}(X)}$$

$$SD(y(X)) = \sqrt{\mathrm{var}(y(X))}$$

$$SD(a+bX) = \sqrt{b^2\,\mathrm{var}(X)} = b \cdot SD(X)$$

Momenter

m'te momentet til X er

$E[X^m]$.

Kjente diskréte fordelinger:

Binomisk fordeling:

$X \sim \text{Bin}(n, p)$

$f_X(i) = \binom{n}{i} p^i (1-p)^{n-i}$

$\Omega_X = \{0, 1, \dots, n\}$

$E[X] = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$

$= \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$ $\qquad \binom{n}{i} = \frac{n}{i}\binom{n-1}{i-1}$

$= \sum_{i=1}^{n} n \binom{n-1}{i-1} p^i (1-p)^{n-i}$

$= np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i}$

$= np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)}$

La $s = i-1$ og $r = n-1$

$= np \underbrace{\sum_{s=0}^{r} \binom{r}{s} p^s (1-p)^{r-s}}_{=1} = np.$ $\square$

$\text{Var}[X] = np(1-p)$

oppgave 2

$X \sim Bin(30, 0{,}60)$, $f(X) = X^2$

a) $P(X-1=30) = P(X=31) = \underline{0}$

$P(29 \geq X) = P(X \leq 29) = 1 - P(X=30)$

$= 1 - \binom{30}{30} p^{30} (1-p)^{30-30} = 1 - 0{,}6^{30} \underline{\approx 1}$

b) $E[X] = np = 30 \cdot 0{,}60 = \underline{18}$

$Var[X] = np(1-p) = 18 \cdot 0{,}4 = \underline{7{,}2}$

$E[f(X)] = E[X^2] = Var[X] + (E[X])^2$

$= 7.2 + 18^2 = \underline{331.2}$

$Var[f(X)] = E[f(X)^2] - (E[f(X)])^2$

$= 1{,}19 \cdot 10^5 - 331{,}2^2$

$= \underline{9.3 \cdot 10^3}$

$SD[2X] = 2 SD[X]$

$= 2 \cdot \sqrt{7.2}$

$\approx \underline{5.37}$

c)

$P\left(X + \sqrt{f(X)} \in \left(E[X+\sqrt{f(X)}] - SD[X+\sqrt{f(X)}], E[X+\sqrt{f(X)}] + SD[X+\sqrt{f(X)}]\right)\right)$

$P\left(2X \in (36 - 5{,}37, 36 + 5{,}37)\right) =$

$P\left(X \in \left(18 - \frac{5.37}{2}, 18 + \frac{5.37}{2}\right)\right)$

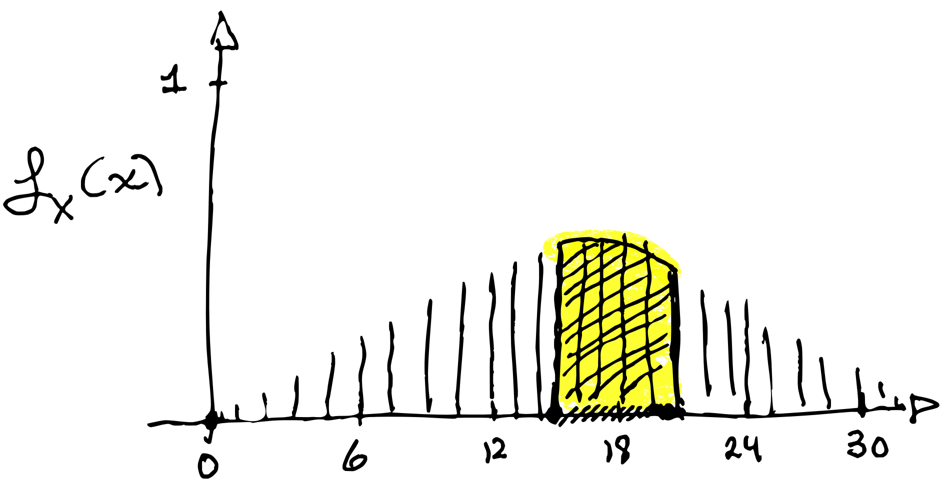
$= P(X < 20.685) - P(X < 15.315)$

$= P(X \leq 20) - P(X \leq 15)$.

$= F_X(20) - F_X(15) = 0{,}8237 - 0{,}1754$

$= \underline{\underline{0{,}6483}}$

Figur



d) Normalapproksimasjon. Best når $p \approx 0{,}5$. Så det optimale hadde vært å endre $p = 0{,}6$ til $p = 0{,}5$.

Poisson-fordeling

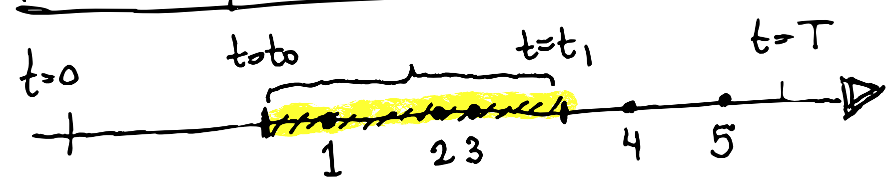
$X \sim \text{Poisson}(\mu)$

$$f_X(i) = \frac{\exp(-\mu)\mu^i}{i!}$$

$E[X] = \mu$

$\text{var}[X] = \mu$

Poisson-prosess



λ : hendelsefrekvens

$\Delta t = t_1 - t_0$, der $t_1 > t_0$

sannsynligheten for nøyaktig i hendelser finner sted i $[t_0, t_1]$ er dermed

$$f_{X(\Delta t)}(i) = \frac{\exp(-\lambda \Delta t)(\lambda \Delta t)^i}{i!}$$

$\Leftrightarrow X(\Delta t) \sim \text{Poisson}(\lambda \cdot \Delta t)$

for hver Δt.

Oppgave 3:

$\lambda = 20$; T: Antall dager (intervall)

$X(T) \sim \text{Poisson}(20T)$ hvis $T = t_2 - t_1$

$$f_{X(T)}(i) = \frac{\exp(-20T)(20T)^i}{i!}$$

a) $E[X_i(1/2)] = 20 \cdot \frac{1}{2} = \underline{\underline{10}}$ for $i = 1, 2$.

$P(X_1(1/2) < 10) = P(X_1(1/2) \leq 9)$

$= F_{X_1(\frac{1}{2})}(9) = \underline{0{,}458} \Rightarrow$

$P(X_2(1/2) < 10) = \underline{0{,}458}$

$P(X_1(1/2) < 10, X_2(1/2) < 10) \overset{\text{uavh.}}{=}$

$(0{,}458)^2 \approx \underline{\underline{0{,}2098}}$

$$\underbrace{X(t_2) - X(t_1)}_{X(T) - X(0) = X(T)} \sim \text{Poisson}(\overbrace{(t_2 - t_1)}^{T}\lambda)$$

PAUSE

uniform fordeling:

$X \sim \text{unif}(a,b]$ la $a^* < a$

$$f_X(i) = \frac{1}{b-a} \cdot \mathbb{1}(i \in [a,b])$$

$$F_X(i) = \int_{-\infty}^{i} \frac{1}{b-a} \mathbb{1}(j \in [a,b]) \, dj$$

$$= \int_{a^*}^{i} \frac{1}{b-a} \mathbb{1}(j \in [a,b]) \, dj$$

$$= \left[\frac{j}{b-a} \mathbb{1}(j \in [a,b]) \right]_{a^*}^{i}$$

$$= \left[\frac{i}{b-a} \mathbb{1}(i \in [a,b]) - \frac{a^*}{b-a} \underbrace{\mathbb{1}(a^* \in [a,b])}_{=0} \right]$$

$$= \frac{i}{b-a} \mathbb{1}(i \in [a,b])$$

$$E[X] = \int_{-\infty}^{\infty} i \cdot \frac{1}{b-a} \mathbb{1}(i \in [a,b]) \, di$$

$$= \int_{a}^{b} i \cdot \frac{1}{b-a} \, di = \frac{1}{b-a} \left[\frac{1}{2} i^2 \right]_a^b$$

$$= \frac{1}{2(b-a)} [b^2 - a^2] = \frac{a+b}{2}$$

$$\text{Var}[X] = \frac{1}{12} \{b-a\}^2$$

Normalfordeling:

$X \sim N(\mu, \sigma^2)$ eller $X \sim N(\mu, \sigma)$

$$f_X(i) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left(-\frac{1}{2\sigma^2}[i-\mu]^2\right)$$

$E[X] = \mu$ (Kontrollerer sentrum)

$Var[X] = \sigma^2$ (Kontrollerer spisshet)

Standard Normalfordeling:

$Z \sim N(0, 1)$

Standardisering av vilkårlig X

$$X \sim N(\mu, \sigma^2) \iff Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$$

Gammafordeling:

$X \sim \text{Gamma}(\alpha, \beta)$

α : Form

β : Skala

$E[X] = \alpha\beta$

$Var[X] = \alpha\beta^2$

$E[X^m] = \beta^m \dfrac{\Gamma(m+\alpha)}{\Gamma(\alpha)}$

Gamma function

$\Gamma(k) = \int_0^\infty s^{k-1}\exp(-s)\,ds \quad \forall\ k > 0$

$\Gamma(k) = (k-1)! \quad \forall\ k \in \mathbb{N}$

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Eksponentiell fordeling

$X \sim \text{Ekspo}(\lambda)$. $\quad \lambda = \frac{1}{\beta}$ og $\alpha = 1$

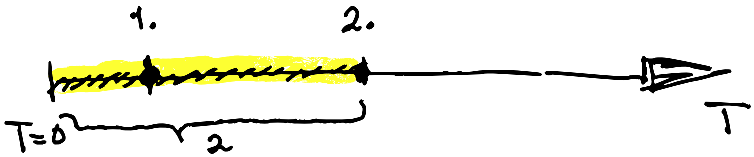
$f_X(i) = \lambda\exp(-i\lambda) \cdot \mathbb{1}(i > 0)$

$E[X] = \alpha\beta = \alpha\beta = \frac{1}{\lambda}$

$Var[X] = \alpha\beta^2 = 1 \cdot \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$

Oppgave 3:

b) Vi kaller de to første ventetidene for W_1 og W_2.



$$W = \sum_{i=1}^{2} W$$

$$P(W_1 \leq w) = 1 - P(W_1 \geq w) =$$

$$1 - P(X(w\lambda) = 0) = 1 - \frac{\exp(-w\lambda)(w\lambda)^0}{0!}$$

$$= 1 - \exp(-w\lambda)$$

$$\Rightarrow f_{W_1}(w) = \frac{dP(W_1 \leq w)}{dw} = \underline{\lambda \exp(-w\lambda)}$$

$$P(W_2 \leq w_1 + w) = 1 - P(W_2 \geq w_1 + w)$$

$$\overset{\text{uavh.}}{=} 1 - P(X((w + w_1)\lambda) - X(w_1\lambda) = 0)$$

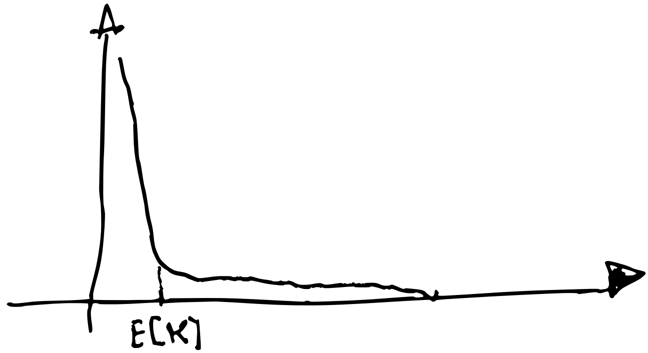
$$= 1 - \exp(-w\lambda) \Rightarrow f_{W_2}(w) = \lambda \exp(-w\lambda)$$

W_1 og W_2 er uavhengige og like fordelte $\Rightarrow$

$$W_1 + W_2 = \underline{W \sim \text{Gamma}(2, \lambda)}$$

Oppgave 4:

a)



b) $f_K(k) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} k^{\alpha-1} \exp(-k\beta^{-1}) \quad k > 0$

$= \frac{1}{\Gamma(1)\beta^{1}} k^{0} \exp\left(-\frac{k}{\beta}\right) \quad k > 0$

La $\lambda = \frac{1}{\beta}$

$= \underline{\lambda \exp(-\lambda k)} \quad k > 0$

c) $E[K] = \frac{1}{\lambda} = \frac{1}{2/5} = \frac{5}{2} = \underline{\underline{2{,}5}}$

$\mathrm{Var}[K] = \frac{1}{\lambda^2} = \frac{1}{4/25} = \frac{25}{4} = \underline{\underline{6{,}25}}$

$P(K > 0{,}05) = 1 - P(K < 0{,}05) =$

$1 - [1 - \exp(-0{,}05 \cdot \lambda)] = \exp\left(-0{,}05 \cdot \frac{5}{2}\right) = \underline{0{,}8825}$

Simultanfordelinger

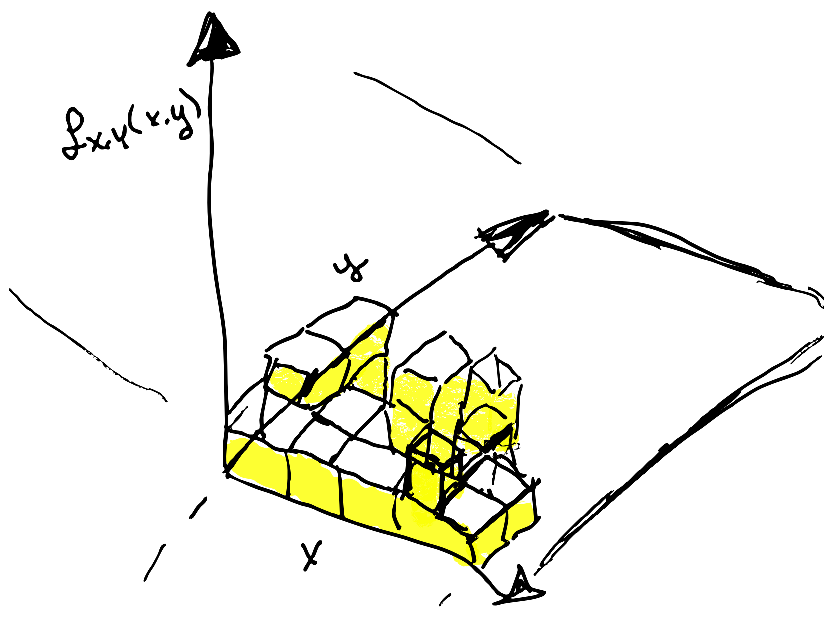
La X og Y være to T.V. som kan være avhengige med henholdsvise utfallsrom S_X og S_Y.

Simultan sannsynlighetsfunksjon.

$f_{X,Y}(x,y)$ er en sannsynlighetsfunksjon for X og Y hvis den beskriver sannsynlighetene

$P(X=x, Y=y) \; \forall \; (x,y) \in S_X \times S_Y$

Figur



Samlet variasjon og Korrelasjon

$$\mathrm{cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

$$\mathrm{cov}(aX,bY) = ab\,E[XY] - ab\,E[X]E[Y]$$
$$\hookrightarrow ab\,\mathrm{cov}(X,Y)$$

Avhengige X og Y:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{cov}(X,Y)$$
$$\mathrm{Var}[X-Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] - 2\mathrm{cov}(X,Y)$$

Korrelasjon:

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{SD(X)SD(Y)}$$

$$\mathrm{corr}(aX,bY) = \frac{ab\,\mathrm{cov}(X,Y)}{ab\,SD(X)SD(Y)}$$

$$= \frac{\mathrm{cov}(X,Y)}{SD(X)SD(Y)} = \mathrm{corr}(X,Y).$$

Forventning til XY:

$$E[XY] = \sum_{x\in S_X}\sum_{y\in S_Y} xy \cdot f_{X,Y}(x,y)$$

Marginale sannsynlighets-funksjoner:

La $f_{X,Y}(x,y)$ være punkt-sannsynlighetsfunksjonen til X og Y.

$$f_X(x) = \sum_{y \in S_Y} f_{X,Y}(x,y) \quad \forall\ x \in S_X$$

$$f_Y(y) = \sum_{x \in S_X} f_{X,Y}(x,y) \quad \forall\ y \in S_Y$$

Betingete sannsynlighets-funksjoner:

$$f_{X|Y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \forall\ x \in S_X$$

$$f_{Y|X}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad \forall\ y \in S_Y$$

Bayes Teorem

Konsekvens:

$$f_{X|Y}(x) = \frac{f_{Y|X}(y)\, f_X(x)}{f_Y(y)}$$

$$f_{Y|X}(y) = \frac{f_{X|Y}(x)\, f_Y(y)}{f_X(x)}$$

Betinget forventning:

$$E[\underbrace{X|Y}_{\text{funksjon av } x}] = \sum_{x \in S_X} x \cdot f_{X|Y}(x,y)$$

$$E[Y|X] = \sum_{y \in S_Y} y \cdot f_{Y|X}(x,y)$$

$$E[G(X)|Y] = \sum_{x \in S_X} G(x) f_{X|Y}(x,y)$$

$$E[G(Y)|X] = \sum_{y \in S_Y} G(y) f_{Y|X}(x,y)$$

$$G(X) = (X - E[X|Y])^2$$

$$\Rightarrow E[G(X)|Y] = Var[X|Y].$$

$$E[G(X)|Y] = \sum_{x \in S_X} G(x) f_{X|Y}(x,y) =$$

$$\sum_{x \in S_X} \{x - E[X|Y]\}^2 f_{X|Y}(x,y) =$$

$$\sum_{x \in S_X} \{x^2 - 2xE[X|Y] + (E[X|Y])^2\} f_{X|Y}(x,y) =$$

$$\sum_{x \in S_X} x^2 f_{X|Y}(x,y) - 2E[X|Y] \sum_{x \in S_X} x f_{X|Y}(x,y)$$

$$+ (E[X|Y])^2 \sum f_{X|Y}(x,y)$$

$= E[X^2|Y] - (E[X|Y])^2 \Rightarrow$

$Var[X|Y] = E[X^2|Y] - (E[X|Y])^2$

Statistikk:

Def. En statistikk er en størrelse en kan beregne fra data.

Før data er innsamlet vil en statistikk være en **t.v.**

Data:

Notasjon: $\{X_1, \dots, X_n\}$ eller $\{x_i\}_{i=1}^{n}$

hvis **alle** X'ene er **uavhengige og like fordelte** sier vi at de er **"iid."**

Fordelingen til en statistikk kaller vi en **utvalgsfordeling.**

Utvalgsfordelinger til noen statistikker basert på normalfordeling:

La $\{X_i\}_{i=1}^{n}$ være et t.u (tilfeldig utvalg) og iid. normalfordelte $\sim N(\mu, \sigma^2)$

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$	$\frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$
$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

Sentral Grenseteoremet:

Korrekt å bruke for summer og skalerte summer.

Eksempler:

$$\bar{X} \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\sum_{i=1}^{n} X_i \to N(n\mu, n\sigma^2)$$

$$\hat{p} = \frac{\sum_i \mathbb{1}(x_i = 1)}{n} \to N\left(p, \frac{p(1-p)}{n}\right)$$

Loven om store tall:

$\bar{X}$ konvergerer mot μ når $n \to \infty$.

$P(|\bar{x} - \mu| > \varepsilon) \to 0$ når $n \to 0$ for alle $\varepsilon > 0$.

"Konvergens i sannsynlighet"

Punktestimatorer

Populasjons-parameter

Def. Gitt en ukjent pop. parameter vil en punktestimator være en T.V. kun avhengig av data.

"punkt" fordi en observert punkt. estimator (estimat) er ett tall.

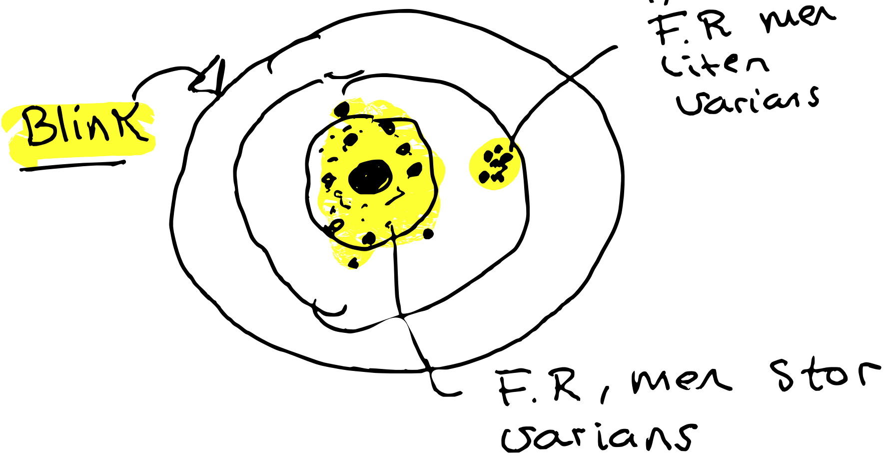
Mer:

En punktestimator for en pop-parameter er en statistikk.

Måle kvalitet hos P. estimator:

* Forventningsrett (F.R)
* varians (usikkerhet)

Bilde:



Det er viktigst å sjekke F.R.

Hvis F.R er ok, så ser vi på variansen til estimatoren.

F.R: $E[\hat{\Theta}] = \Theta$

Varians: $Var[\hat{\Theta}] = \underbrace{E[(\hat{\Theta} - E(\Theta))^2]}_{MSE.}$

Hvordan finne p. estimatorer:

* Intuisjon } livsfarlig!
* Gjette
* sannsynlighetsmaksimering (MLE)
* Minste kvadraters metode
* Momentmetoden

Sannsynlighetsmaksimeringsprinsippet:

For $\{x_i\}_{i=1}^n$ der disse er iid. med S.F $f_X(x_i)$, definerer vi

$$\mathcal{L}(\underline{\theta} \mid \{X_i\}_{i=1}^n) = \prod_{i=1}^n f_X(x_i \mid \underline{\theta})$$

Den $\underline{\theta}$ som maksimerer $\mathcal{L}$ vil være MLE-estimatoren for $\underline{\theta}$.

log-MLE:

$$\ell(\theta \mid \{x_i\}_{i=1}^n) = \sum_{i=1}^n \log f_X(x_i \mid \theta).$$

ℓ og $\mathcal{L}$ er maksimert ved samme $\underline{\theta}$.

Oppgave 4:

d) vi har $\{x_i\}_{i=1}^{n}$, $x_i \sim \exp(\frac{1}{\beta})$

$$\mathcal{L}(\tfrac{1}{\beta} \mid \{x_i\}_{i=1}^{n}) = \prod_{i=1}^{n} f_X(x_i) \quad (1)$$

La $\lambda = \frac{1}{\beta}$, og ta log på (1).

$$\ell(\lambda \mid \{x_i\}_{i=1}^{n}) = \sum_{i=1}^{n} \log f_X(x_i) \quad (2).$$

$$\sum_{i=1}^{n} \log f_X(x_i) = \sum_{i=1}^{n} \log(\lambda \exp\{-x_i\lambda\})$$

$$= \sum_{i=1}^{n} \{\log(\lambda) + \log\exp(-\lambda x_i)\} =$$

$$n\log(\lambda) - \lambda \sum x_i.$$

vi finner λ som maksimerer ℓ.

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0$$

$$\Rightarrow \frac{n}{\lambda} = \sum x_i \Rightarrow \underline{\frac{1}{\lambda} = \bar{x}}$$

$$\Rightarrow \hat{\lambda} = \underline{\underline{\frac{1}{\bar{X}}}}.$$

$$E[\hat{\lambda}] = E\left[\frac{1}{\bar{X}}\right] \neq \frac{1}{E[\bar{X}]} = \lambda.$$ **Nei!**

PAUSE

Hvordan vet vi om vi har et toppunkt og ikke bunnpunkt?

$$\frac{d^2 L}{d\underset{\sim}{\theta}^2}(\hat{\underset{\sim}{\theta}}) < 0$$ (neg. 2. derivert i ekstr. punktet)

Intervallestimatorer:

Vi holder oss til ett-utvalg-intervall-estimatorer!

Def.

Samme som punktestimator, men kombinert med usikkerheten til p. estimatoren.

Konfidens-intervall

Bredden til intervallet bestemmes av et konfidensnivå $100(1-\alpha)\%$

der α : signifikansnivå.

Bredden bestemmes også av usikkerheten til p. estimatoren

OBS

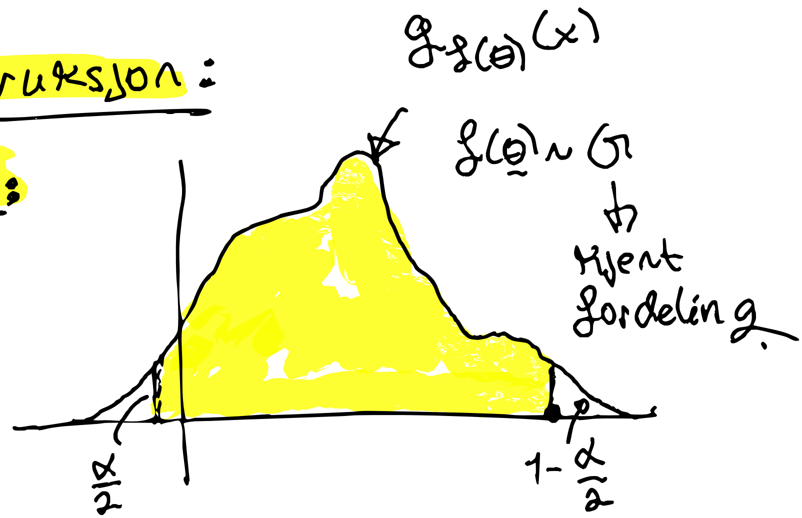
Parametriske Konfidens. Int.:

Konfidens. Int. laget ved å anta at $f(\underline{\theta})$ følger en kjent fordeling.

Huis utvalgsfordeling til $f(\underline{\theta})$ ikke har en kjent fordeling kan vi ikke konstruere paramet. Konf. Int.

Konstruksjon:

Bilde:



Matematisk:

$$P\left(G_{\alpha/2} \leq f(\underline{\theta}) \leq G_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

↳ Løse ulikhet m.h.p $\underline{\theta}$

Vi kan bruke visse statistikker for å bestemme ulike konf. int.

Eksempler

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

$$\frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

$f(\mu)$

$$P\left(z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(t_{n-1,\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{n-1,1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{S^2(n-1)}{\sigma^2} \leq \chi^2_{1-\alpha/2,\,n-1}\right) = 1-\alpha$$

$f(\sigma^2)$

mer:

Vi kan **justere** intervallbredden til ønsket bredde ved å endre **n**.

Bredde $(KI_{\mu|\sigma^2}) = 2 \cdot z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Oppgave 5:

$D_i = X_{i+1} - X_i \quad i \in \{1, \dots, 20\}$

a) Hvis X_{i+1} og X_i er positivt korrelerte $\Rightarrow$

$\text{corr}(X_{i+1}, X_i) > 0 \Rightarrow \text{cov}(X_{i+1}, X_i) > 0$

videre har vi

$\text{Var}[D_i] = \text{Var}[X_{i+1} - X_{i+1}] =$

$\text{Var}[X_{i+1}] + \text{Var}[X_i] - 2\underbrace{\text{cov}(X_{i+1}, X_i)}_{>0}$

$\Rightarrow \overbrace{\text{Var}[X_{i+1}] + \text{Var}[X_i]}^{2\sigma^2} > \text{Var}[D_i]$

vanskelig å finne forventning fordi forventningen er en funksjon av en T.V.

b) $T = \dfrac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim t_{n-1}$

$KI_{\mu_D} = \left[\bar{d} + t_{\alpha/2, n-1} \cdot \dfrac{S_D}{\sqrt{n}}, \bar{d} + t_{1-\alpha/2, n-1} \cdot \dfrac{S_D}{\sqrt{n}}\right]$

$= \left[3.75 + (-2.09) \cdot \dfrac{27.67}{\sqrt{20}}, 3.75 + (2.09) \cdot \dfrac{27.67}{\sqrt{20}}\right]$

$= [-9.18, 16.68]$

Hypotesetesting:

Def.

Statistisk hypotesetesting er en prosedyre ment for å trekke slutninger for utfallsrommet til en parameter.

Hvorfor:

vi ønsker trekke slutninger for intersante parametre.
En estimator i seg selv er ofte lite interresant!

Prosedyre:

Sign. nivå

H_0: ---
H_1: ---

1. Velg α
2. Sett opp to komplementære hypoteser basert på det vi lurer på.
3. bestem testobservator, som er en statistikk med kjent ford. under H_0
4. Bestem forkastningsområdet til testen (avhengig av fordeling til testobservatoren)
5. Hvis observert test observator er i forkastningsområdet

$\Rightarrow$ Forkast H_0 og godkjenn H_1 basert på α og data.

Potensielle feil i hypotesetesting:

Type I feil:

$P(\text{Forkast } h_0 \mid h_0 \text{ sann}) = \alpha$

Type II feil:

$P(\text{Behold } h_0 \mid h_1 \text{ sann}) = \beta$

Styrke:

$P(\text{Forkast } h_0 \mid h_1 \text{ sann}) = 1 - \beta$

P-verdi:

P(T.O. minst like ekstem som O.T.O.)

T.O.: Test-observator under H_0

minst like ekstem: Betydning avhengig av H_1

O.T.O.: observert test-observator.

Oppgave 5:

c)

μ_{D_0}

$H_0: \mu_D \leq 0$

$H_1: \mu_D > 0$

d) $\alpha = 0{,}05$ og $n = 20$

Testobservator:

$$T = \frac{\bar{D} - \mu_{D_0}}{S_D/\sqrt{n}} \sim t_{n-1}$$

Forkastningsområde:

$T \geq t_{\alpha, n-1} = -t_{0{,}05, 19}$ eller

$$R = \{T : T \in [\underbrace{-t_{0{,}05, 19}}_{=1{,}73}, \infty)$$

observert testobservator:

$$\tau = \frac{\bar{d} - 0}{s_D/\sqrt{n}} = \frac{\bar{d}}{s_D/\sqrt{n}} = \sqrt{n}\,\frac{\bar{d}}{s_D} = \sqrt{20} \cdot \frac{3{,}75}{27{,}67}$$

$= \underline{0{,}61}$ $\Rightarrow$ <u>ikke forkast H_0</u>

e) For å forkaste H_0 må

$$T = \frac{\bar{d} - \mu_{D_0}}{S_D / \sqrt{n}} \geq -t_{n-1,\alpha} \Rightarrow$$

$$\frac{\sqrt{n}(\bar{d} - \mu_{D_0})}{S_D} \geq t_{n-1,1-\alpha} \Rightarrow$$

$$S_D \leq \frac{\sqrt{n}(\bar{d} - \mu_{D_0})}{t_{n-1,1-\alpha}}$$

Maksimal S_D som forkaster H_0

$$S_D = \frac{\sqrt{n}(\bar{d})}{t_{n-1,1-\alpha}} = \frac{\sqrt{20} \cdot 3{,}75}{1{,}73} \approx \underline{\underline{9{,}69}}$$

f)

$$T = \frac{\bar{X} - \mu_{X_0}}{S_X / \sqrt{n+1}} \sim t_n$$

$$t = \frac{\bar{x} - a}{S_X / \sqrt{n+1}} = \frac{171.67 - 100}{34.25 / \sqrt{21}} = \underline{9.59}$$

$$\text{P-verdi} = P(T \geq 9.59) = \underline{3.18 \cdot 10^{-9}}$$

<u>Forkast H_0!</u>

fordi P-verdi < $\alpha = 0{,}05$.

G) **Bonus**

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma}}{\frac{S}{\sigma\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}}$$

$$= \frac{Z}{\frac{S\cdot\sqrt{n-1}}{\sigma\ \sqrt{n-1}}} = \frac{Z}{\frac{1}{\sqrt{n-1}}\cdot\underbrace{\left(\frac{S^2(n-1)}{\sigma^2}\right)}_{V\sim\chi^2_{n-1}}^{1/2}}$$

$$= \frac{Z}{\frac{(V)^{1/2}}{\sqrt{n-1}}} = \frac{Z}{\underbrace{\sqrt{V/_{n-1}}}_{}}$$

$v = n-1 \Rightarrow T \sim \underline{t_{n-1}}$

$\boxtimes$

$\boxed{\text{PAUSE}}$