20180507 Kom ihåg

Vi hade $X_1, X_2, \dots, X_n$ iid, med okänd parameter θ.

Vi (punkt-) skattar θ mha $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$. Här är $\hat{\theta}$ en s.v.

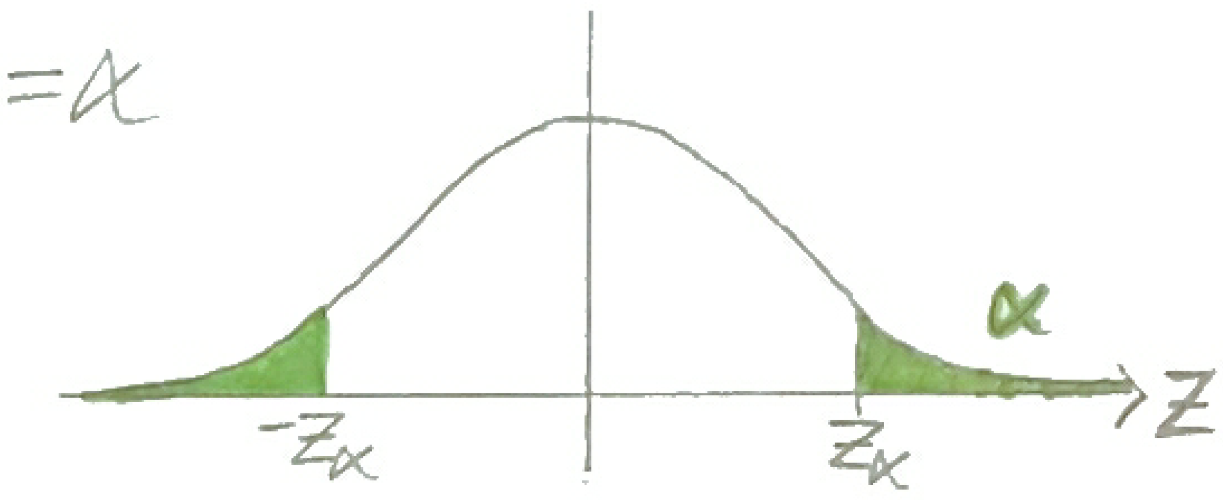
Efter mätning då data $x_1, \dots, x_n$ har blivit observerade blir den (numeriska) realiserade gissningen (av θ) $\hat{\theta}(x_1, \dots, x_n)$

Obs. ofta kallas både $\hat{\theta}(X_1, \dots, X_n)$ & $\hat{\theta}(x_1, \dots, x_n)$ för skattning (estimate). Vissa skiljer på de två (skattare/skattning alt. estimator/estimate).

Def. för $Z \sim N(0,1)$ & $0 \le \alpha \le 1$ def. vi z_α via $\mathbb{P}(Z \ge z_\alpha) = \alpha$.

Anm: symmetri ger $\mathbb{P}(Z \le z_\alpha) = \alpha$

$\Rightarrow \mathbb{P}(-z_\alpha \le Z \le z_\alpha) = 1 - 2\alpha$



Sats Låt $X_1 \sim N(\mu_1, \sigma_1^2)$ & $X_2 \sim N(\mu_2, \sigma_2^2)$ vara ober.

Då gäller att $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Bevis Vi har att $M_{X_i}(t) = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}$, $i = 1, 2$.

Då gäller $M_{X_1+X_2}(t) \overset{\text{ober.}}{=} M_{X_1}(t) M_{X_2}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t} =$

$= e^{(\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}$ $\blacksquare$

Vi har lärt oss hur vi gör en pktskattning, men hur bra är den?

Konfidensintervall (K.I.)

Fall 1 Antag $X_1, X_2, \dots$ iid $N(\mu, \sigma^2)$ med σ känd.

Vi skattar μ med $\hat{\mu}(X_1, \dots, X_n) = \bar{X}$. Vi vet $X_1 + \dots + X_n \sim N(n\mu, n\sigma^2)$ (enl. satsen ovan)

$$\Rightarrow \bar{X} = \frac{X_1+\ldots+X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$$

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

Kom ihåg: $Y \sim N(\mu, \sigma^2)$
$\Rightarrow \frac{Y-\mu}{\sigma} \sim N(0,1)$

Vi får

$$1-2\alpha = \mathbb{P}(-z_\alpha \le Z \le z_\alpha) = \mathbb{P}(-z_\alpha \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_\alpha) = \mathbb{P}(-z_\alpha \frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le z_\alpha \frac{\sigma}{\sqrt{n}}) =$$

$$= \mathbb{P}(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}).$$

Dvs. med sanno. $1-2\alpha$ tillhör μ det slumpmässiga intervallet $I_\mu = [\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}]$. I_μ är ett K.I. för μ med konfidens grad $1-2\alpha$.

Efter att data $x_1, x_2, \ldots, x_n$ har observerats bildas det numeriska K.I. $I_\mu = [\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}]$

Def låt $\underline{\theta} \le \bar{\theta}$ vara skattningar av θ. Om $\mathbb{P}(\underline{\theta}(X) \le \theta \le \bar{\theta}(X)) = 1-\alpha$ sägs intervallet $I_\theta = [\underline{\theta}(X), \bar{\theta}(X)]$ vara ett K.I. för θ med konfidensgraden $1-\alpha$.

Anm: i) $\underline{\theta}, \bar{\theta}$ är "under"- resp "överskattning" av θ

ii) om $\underline{\theta}$ är $-\infty$ (alt. $\bar{\theta} = \infty$) kallas intervallen $(-\infty, \bar{\theta}]$ (alt. $[\underline{\theta}, \infty)$) ensidiga K.I.

iii) Fall 1 fungerade för att referensvariabeln $R := \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ innehöll μ & hade känd fördelning.

Fall 2 Återigen är $X_1, X_2, \ldots$ iid. $N(\mu, \sigma^2)$ men här är både μ & σ^2 okända.

Vi skattar σ^2 med $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ & bildar $R = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim ?$

Def en s.v. X med tfkn $f(s) = c(1+\frac{s^2}{n})^{-\frac{n-1}{2}}$ för $s \in \mathbb{R}$, är t-fördelad med n frihetsgrader ($X \sim t(n)$).

Sats För $X_1, X_2, \dots, X_n$ iid $N(\mu, \sigma^2)$ gäller att $R = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$

Bevis Orgie i mgf, se kap 6 ∎

Anm: i) tfkn för $t(n)$ är symmetrisk.

ii) om $X_n \sim t(n) \Rightarrow X_n \xrightarrow{d} Z$, $n \to \infty$, där $Z \sim N(0,1)$ (dvs. $t(n) \to N(0,1)$).

iii) vi använder tabeller för $t_\alpha(n)$ där $\alpha = \mathbb{P}(T \geq t_\alpha(n))$ där $T \sim t(n)$.

åter till fall 2

$R = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$ får vi att $1-\alpha = \mathbb{P}(R \leq t_\alpha(n-1)) =$

$= \mathbb{P}\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} \leq t_\alpha(n-1)\right) = \mathbb{P}\left(\bar{X}-\mu \leq \frac{s}{\sqrt{n}} t_\alpha(n-1)\right) = \mathbb{P}\left(\mu \geq \bar{X} - \frac{s}{\sqrt{n}} t_\alpha(n-1)\right)$

s.a. $I_\mu = [\bar{X} - \frac{s}{\sqrt{n}} t_\alpha(n-1), \infty)$ är ett ensidigt K.I. för μ med konfidensgrad $1-\alpha$.

Anm: om $\mathbb{P}(\underline{\mu}(X) < \mu < \bar{\mu}(X)) = 1-\alpha$ & $\underline{\mu} > 0$

$\Rightarrow \mathbb{P}\left(\frac{1}{\bar{\mu}(X)} < \frac{1}{\mu} < \frac{1}{\underline{\mu}(X)}\right) = 1-\alpha$

s.a. $I_{\mu^{-1}} = \left[\frac{1}{\bar{\mu}(X)}, \frac{1}{\underline{\mu}(X)}\right]$ är ett K.I. för μ^{-1} med konf.grad $1-\alpha$.

Tal utfallen av en serie experiment anses komma från en $N(\mu, \sigma^2)$. Med $n = 13$ datapkter erhölls:

2.81, 5.14, 2.39, 2.63, 7.47, 7.23, 7.25, 5.01, -0.68, 5.15, 7.89, 4.47, 6.1

a) skatta μ & σ^2 b) hitta ett 99 % (tvåsidigt) K.I. för μ

2 a) då $X_1, X_2, \ldots, X_{13} \sim N(\mu, \sigma^2)$ (ober. !?) använder vi $\hat{\mu} = \bar{X}$ & $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ som är VVR.

Insättning av data ger de numeriska skattningarna $\hat{\mu}(x) = \bar{x} \approx 4.84$ & $s^2(x) \approx 6.21$

b) Vi har att $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) = t(12)$.

Tabell ger $1-0.01 = \mathbb{P}\left(-3.05 \le \frac{\bar{X}-\mu}{S\sqrt{n}} \le 3.05\right) = \mathbb{P}\left(\bar{X} - 3.05\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + 3.05\frac{S}{\sqrt{n}}\right)$

tvåsidigt, $2 \cdot 0.005 = 0.01$ totalt $\longrightarrow$ $t_{0.005}(12)$

S.a. med sanno. 99% har vi att $\mu \in \left[\bar{X} - \frac{S}{\sqrt{n}}3.05, \bar{X} + \frac{S}{\sqrt{n}}3.05\right]$.

Vårt 99%-numeriska K.I. blir

$I_\mu = 4.84 \pm 3.05\frac{\sqrt{6.21}}{\sqrt{13}} = [2.73, 6.95]$.

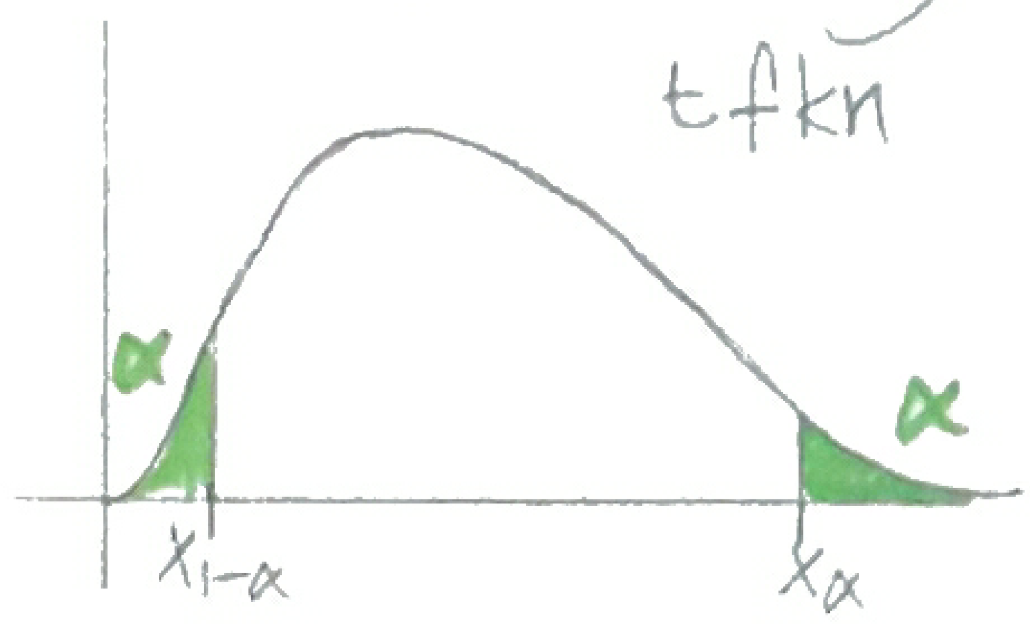
2018 05 08 Generella idén bakom K.I.

Steg 1 hitta en referensvariabel R där $X_1, \dots, X_n$ & θ ingår. R väljs s.a. dess fördelning är känd.

Steg 2 hitta lämpliga kvantiler $(z_\alpha, t_\alpha, \dots)$

Steg 3 lös ut θ ur uttrycket $1-2\alpha = \mathbb{P}(x_{1-\alpha} < R < x_\alpha)$ eller motsvar.

Anm: om fördelningen för R är symmetrisk $\Rightarrow x_{1-\alpha} = -x_\alpha$



Normalapproximation för K.I.

Vi betraktar proportions-skattning. Principen samma i andra fall.

Antag att i en population så har proportionen p en viss egenskap. Vi tar ett stickprov av storlek n & låter $X = \#$ med egenskapen. Vi antar att $X \sim Bin(n,p)$ (ok? ober.?!)

Vi pktskattar p med $\hat{p} = \frac{X}{n}$, men hur hittar vi ett K.I.?

Alt 1 använd exakt fördelning (dvs $X \sim Bin(n,p)$). (steg 1)

steg 2: hitta k_α s.a. $\mathbb{P}(|\frac{X}{n} - p| > k_\alpha) = 2\alpha$.

steg 3: $\mathbb{P}(|\frac{X}{n} - p| > k_\alpha) = \mathbb{P}(-k_\alpha < \frac{X}{n} - p < k_\alpha) = \mathbb{P}(\frac{X}{n} - k_\alpha < p < \frac{X}{n} + k_\alpha)$.

ok? Nej! Värdet på k_α kommer bero på p som är okänd. Dessutom är X ej symm., bör då K.I. vara det? $(k_{1-\alpha} = -k_\alpha)$

Alt 2 om n är "stort" ger CGS (centrala gränsv. satsen) att $X \approx N(\mathbb{E}[X], Var(X)) =$
$= N(np, np(1-p))$

$$\Rightarrow \frac{X-np}{\sqrt{np(1-p)}} \approx N(0,1) \quad (\text{steg } 1)$$

steg 2: kvantiler (z_α) är kända.

steg 3: lös ut p ur $\mathbb{P}\left(-z_\alpha \le \frac{X-np}{\sqrt{np(1-p)}} \le z_\alpha\right)$

steg 3': lös ut p ur $\mathbb{P}\left(-z_\alpha \le \frac{X-np}{\sqrt{n\hat{p}(1-\hat{p})}} \le z_\alpha\right) = \mathbb{P}\left(-z_\alpha\sqrt{n\hat{p}(1-\hat{p})} \le X - np \le \right.$

gör 3'

$$\left. \le z_\alpha\sqrt{n\hat{p}(1-\hat{p})}\right) = \mathbb{P}\left(\frac{X}{n} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \frac{X}{n} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Anm: principen densamma, men detaljer skiljer ngt fall till fall.

K.I. för σ^2 då $X_i \sim N(\mu, \sigma^2)$

Låt $X_1, \dots, X_n$ vara iid $N(\mu, \sigma^2)$ & låt $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Def en s.v. Y är chi-kvadrat fördelad med n frihetsgrad. om $Y = Z_1^2 + Z_2^2 + \dots + Z_n^2$, där Z_i är iid $N(0,1)$, $Y \sim \chi^2(n)$

Sats $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

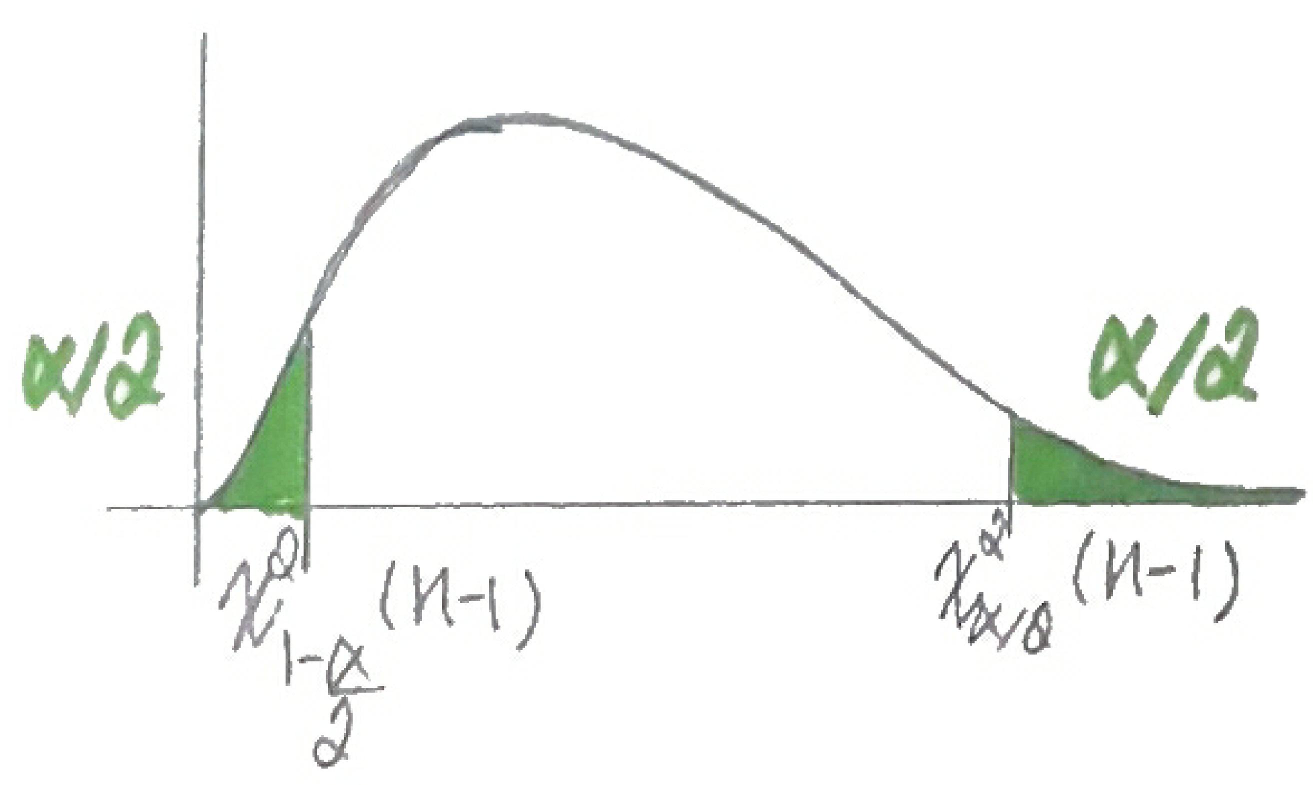
Bevis Lång räkning med mgf! ∎

Låt $\chi^2_\alpha(n-1)$ vara α-kvantilen för $\chi^2(n-1)$-förd.

Vi hittar ett K.I. med konfidensgrad $1-\alpha$ genom att

$$1-\alpha = \mathbb{P}\left(\chi^2_{1-\frac{\alpha}{2}}(n-1) \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}}(n-1)\right) = \mathbb{P}\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

Anm:



ej symmetrisk!

dvs. $\chi^2_{1-\frac{\alpha}{2}}(n-1) \neq \chi^2_{\frac{\alpha}{2}}$

Tal (forts från 7/5)

Data kom från en normalfördelning. ($X_i \sim N(\mu, \sigma^2)$)

& vi fick $\bar{x} = 4.84$, $s^2(x) \approx 6.81$ & $n = 13$.

c) hitta ett 95% K.I. för σ^2.

2 enl. ovan är $I_{\sigma^2} = \left[\frac{(n-1)s^2(X)}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2(X)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right]$ ett K.I. för σ^2 med konfidensgrad $1-\alpha$.

med $X = X_1, \ldots, X_n$

Här är $\alpha = 0{,}05$ & $n = 13$. Tabell (s. A8 i Rice) ger

$\chi^2_{0.025}(12) \approx 23.34$ & $\chi^2_{0.975}(12) \approx 4.40$

s.a.

$$I_{\sigma^2} = \left[\frac{12 \cdot 6.81}{23.34}, \frac{12 \cdot 6.81}{4.40}\right] \approx [3.19, 16.94]$$

är ett numeriskt 95% K.I. för σ^2.

Två stickprov

Vi har 2 ober. stickprov $X = (X_1, \ldots, X_n)$ resp. $Y = (Y_1, \ldots, Y_m)$

där $X_i \sim N(\mu_1, \sigma_1^2)$ & $Y_i \sim N(\mu_2, \sigma_2^2)$.

Vi vill analysera $\theta = \mu_1 - \mu_2$ som skattas med $\hat{\theta} = \hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$.

Fall 1 (σ_1^2, σ_2^2 kända)

Vi har att $\bar{X} \sim N(\mu_1, \sigma_1^2/n)$ & $\bar{Y} \sim N(\mu_2, \sigma_2^2/m)$ s.a.

$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$.

Låt $D^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ och bilda $R = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{D} \sim N(0,1)$

$$1-\alpha = P\left(-z_{\frac{\alpha}{2}} \leq R \leq z_{\frac{\alpha}{2}}\right) = P\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} D \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} D\right)$$

$$\Rightarrow I_{\mu_1 - \mu_2} = \bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} D$$

Tal I en undersökning uppmättes surheten i två olika sjöar. Den första kontrollerades på $n = 20$ platser & den andra på $m=29$. Resultaten av mätningarna anses vara ober. & $N(\mu_1, \sigma_1^2)$ resp. $N(\mu_2, \sigma_2^2)$ fördelade. Om $\sigma_1^2 = 1.24$ & $\sigma_2^2 = 1.9$, hitta ett 95% K.I. för $\mu_1 - \mu_2$ om data gav $\bar{x} = 5.3$ & $\bar{y} = 6.1$.

L Vi får att $D = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \approx 0.368$. Tabell ger $z_{0.025} \approx 1.96$, (N(0,1) parym. lös)

så ett 95% numeriskt K.I. blir $I_{\mu_1 - \mu_2} = \bar{x} - \bar{y} \pm 1.96 \cdot 0.368$
$\approx [-1.52, -0.079]$ //

Anm: har vi fog att hävda att surheterna skiljer sig mellan sjöarna? 0 inte med i $I_{\mu_1 - \mu_2}$, ok?

Antag nu att $X_i \sim N(\mu_1, \sigma^2)$ & $Y_i \sim N(\mu_2, \sigma^2)$

$\Rightarrow \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2(n^{-1} + m^{-1}))$.

Hur ska σ^2 skattas?

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2.$$

Man kan visa att $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$ är VVR & har minst varians av alla skattare på formen $\beta S_X^2 + (1-\beta) S_Y^2$.

20180509 Två stickprov forts.

Antag att $X_1, \dots, X_n$ & $Y_1, \dots, Y_m$ är ober. med $X_i \sim N(\mu_1, \sigma^2)$ medans $Y_i \sim N(\mu_2, \sigma^2)$ (samma varians!)

Vi får att $\bar{X} \sim N(\mu_1, \sigma^2/n)$ & $\bar{Y} \sim N(\mu_2, \sigma^2/m)$ s.a. $\theta = \mu_1 - \mu_2$ skattas av $\hat{\theta} = \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2(n^{-1}+m^{-1}))$

Obs. $Var(\bar{X}-\bar{Y}) = Var(\bar{X}) + Var(-\bar{Y}) + 2Cov(\bar{X}, -\bar{Y}) =$
$= Var(\bar{X}) + Var(\bar{Y}) - 2Cov(\bar{X}, \bar{Y}) \overset{ober.}{=} Var(\bar{X}) + Var(\bar{Y})$

Här är σ^2 okänd & skattas mha $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$.

Vi använder $R = \frac{\hat{\theta}(X, Y) - \theta}{S_p(X, Y)\sqrt{n^{-1}+m^{-1}}}$

Sats Om R är som ovan gäller att $R \sim t(n+m-2)$

Vi får att $1-2\alpha = \mathbb{P}(|\hat{\theta}-\theta| \leq t_\alpha(m+n-2) S_p(X, Y)\sqrt{n^{-1}+m^{-1}})$
s.a. ett K.I. med konf. grad. $1-2\alpha$ för θ ges av
$I_\theta = \bar{x} - \bar{y} \pm t_\alpha(m+n-2) S_p(x, y)\sqrt{n^{-1}+m^{-1}}$

Tal (som sura sjöar) 2 sjöar, $n=20$, $m=29$, $\hat{\mu}_1 = \bar{x} = 5.3$, $\hat{\mu}_2 = \bar{y} = 6.1$
$s_x^2 = 1.52$, $s_y^2 = 1.83$.

Hitta ett 95% K.I. för $\mu_1 - \mu_2$.

L Vi slår ihop skattningarna för σ^2

$$s_p^2 = \frac{19}{47} s_x^2 + \frac{28}{47} s_y^2 = \frac{19}{47} 1.52 + \frac{28}{47} 1.83 \approx 1.705$$

Enl. tabell är $t_{0.025}(47) \approx 2.01$

Obs. Tabell anger $t_{0.025}(40)$ & $t_{0.025}(60)$, vi interpolerar.
(Vi ger inte upp eller ngt annat åt hellskotta)

interpolera på känsla!

Så ett 95% numeriskt K.I. blir

$$I_{\mu_1-\mu_2} = \bar{x}-\bar{y} \pm t_{0.025}(47)\, S_p(\underline{x},\underline{y})\sqrt{\frac{1}{20}+\frac{1}{29}} \approx [-1,563, -0,037]$$ //

Vi har (minst) fallet $\sigma_1 \neq \sigma_2$ okända. Vi skippar detta!

Man kan förståss tänka sig två stickprov där data ej kommer från en normal fördelning.

Tal två väljarundersökningar genomfördes med fyra månaders mellanrum. Resultatet för parti Q blev

1) 143 av $n_1 = 1217$ tillfrågade stöttade Q.

2) 139 av $n_2 = 1078$ —//—

Skapa ett 99% K.I. för $p_2 - p_1$ där p_i är Q's väljarantal vid undersökning i.

L låt $\hat{p}_1 = \frac{X_1}{n_1}$ & $\hat{p}_2 = \frac{X_2}{n_2}$ där X_i = # anhängare vid unders. i

Obs. X_i är en s.v. dvs. "innanperspektiv".

Vi har att $X_i \sim Bin(n_i, p_i)$ s.a. $\hat{p}_i \underset{CGS}{\approx} N(\mathbb{E}[\hat{p}_i], Var(\hat{p}_i)) =$
$= N\left(p_i, \frac{p_i(1-p_i)}{n_i}\right)$

$\Rightarrow \hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$.

Vi ersätter variansen av skattningen $\hat{p}_1 - \hat{p}_2$ med skattn. av denna varians $S^2_{\hat{p}_1-\hat{p}_2} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$.

Vi låter $R = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{S_{\hat{p}_1-\hat{p}_2}} \approx N(0,1)$

Tabell ger $0.99 = \mathbb{P}(-2.575 \le Z \le 2.575),\ Z \sim N(0,1)$

$$\approx \mathbb{P}\left(-2.575 \le \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{S_{\hat{p}_1 - \hat{p}_2}} \le 2.575\right)$$

$$= \mathbb{P}(\hat{p}_1 - \hat{p}_2 - 2.575\, S_{\hat{p}_1 - \hat{p}_2} \le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + 2.575\, S_{\hat{p}_1 - \hat{p}_2})$$

Ett 90% numeriskt K.I blir $I_{p_1 - p_2} = \hat{p}_1(x_1) - \hat{p}_2(x_2) \pm 2.575\, S_{\hat{p}_1 - \hat{p}_2}(x_1, x_2)$

$\approx [-0.047, 0.024]$ //

Anm: har vi fog att hävda att sympatierna har förändrats? ($p_1 \neq p_2$)? Antagligen nej, då $0 \in I_{p_1 - p_2}$.

Stickprov i par

Dyker upp naturligt när vi har ett tydligt "före-/efter-perspektiv".

Ta! halten bly uppmäts i 20 vattendrag som ligger i anslutning till en nyetablerad fabrik. Mätningen genomfördes innan & sedan igen ett år efter att fabriken startades.

Resultatet av mätningarna blev:

Drag nr.	1	2	3	4	......	19	20
Innan	1.7	1.6	2.7	4.3		0.7	2.9
Efter	1.9	1.5	3.1	3.9		0.8	3.4

Under antagandet att mätningarna kommer från normalfördelningar, bestäm ett 99% K.I. för skillnaden i halten bly.

Anm: låt μ_k = halten bly i vattendrag k innan mätning.

Det är rimligt att förvänta sig skillnader mellan vattendragen, dvs. $\mu_1 \neq \mu_2$ etc. Vi antar dock att skillnaden Δ innan/efter är densamma för alla vattendrag, dvs

$\mu_1 \to \mu_1 + \Delta,\ \mu_2 \to \mu_2 + \Delta, \ldots,\ \mu_{20} \to \mu_{20} + \Delta.$

Vi vill skatta Δ!

Vad innebär normalantagandet?

Jo: X_k = uppmätt halt i drag k innan

Y_k = —//— efter

$\Rightarrow X_k \sim N(\mu_k, \sigma_1^2)$ & $Y_k \sim N(\mu_k + \Delta, \sigma_2^2)$,

och därmed blir $Z_k := Y_k - X_k \sim N(\Delta, \underbrace{\sigma_1^2 + \sigma_2^2}_{\text{lika}})$

$\mathcal{L}$ Vi skattar Δ med $\hat{\Delta} = \bar{Z}$ där $Z_k \sim N(\Delta, \underline{\sigma^2})$.

Vi får att $R = \dfrac{\hat{\Delta} - \Delta}{S(Z)/\sqrt{n}} \sim t(n-1)$.

Detta ger oss $\bar{Z} = 0.42$ & $s_Z = 0.32$ s.a.

$$I_\Delta = \bar{Z} \pm t_{0.005}(19)\frac{s_Z}{\sqrt{20}} = 0.42 \pm 2.86\,\frac{0.32}{\sqrt{20}} = 0.42 \pm 0.205 \quad //$$

Anm: parat stickprov reduceras till ett stickprov!

20180509 Räkneövning 6
Parameterskattning & inferens

8.8 Antal hopp för fåglar mellan flygningar ges av

# hopp	1	2	3	...	10	11	12
frekvens	48	31	20	...	1	2	1

Anta att X_i = antal hopp för flygning $i \sim \mathrm{Geom}(p)$

a) Skatta p.

För $X_i \sim \mathrm{Geom}(p)$ gäller att $E(X_i) = p^{-1}$.

Momentmetoden: $p = \frac{1}{E(X_i)} \Rightarrow \hat{p} = \frac{1}{\bar{X}}$, $\bar{X} = \frac{1}{n}\sum_{k=1}^{12} n_k k$,

där n_k är frekvensen av k hopp & $n = \sum_{k=1}^{12} n_k$

$\Rightarrow \hat{p} \approx 0{,}36$

b) Approx. ett 95% K.I. för p.

Vi har att $\bar{X} \overset{approx.}{\sim} N(E(\bar{X}), Var(\bar{X}))$

$$E(\bar{X}) = n^{-1}\sum_{k=1}^{n} E(x_k) = p^{-1}$$

$$Var(\bar{X}) = Var\left(n^{-1}\sum_{k=1}^{n} X_k\right) = n^{-2}\sum Var(X_k) = \frac{1}{n^2}\, n\, \frac{1-p}{p^2} = \frac{1-p}{np^2}$$

$$\approx \frac{1-\hat{p}}{n\hat{p}^2} =: s_{\bar{X}}^2$$

$$\Rightarrow \bar{X} \overset{approx.}{\sim} N\left(\frac{1}{p}, \frac{1-\hat{p}}{n\hat{p}^2}\right).$$

Låt $Z = \frac{\bar{x} - p^{-1}}{s_{\bar{x}}} \Rightarrow Z \overset{appr.}{\sim} N(0,1)$.

tabell

$$0{,}95 = P(-1.96 \le Z \le 1.96) = P(-1.96 s_{\bar{x}} \le \bar{x} - p^{-1} \le 1.96 s_{\bar{x}}) =$$

$$= P(\bar{x} - 1.96 s_{\bar{x}} \le p^{-1} \le \bar{x} + 1.96 s_{\bar{x}}) = P\left(\frac{1}{\bar{x}+1.96 s_{\bar{x}}} \le p \le \frac{1}{\bar{x}-1.96 s_{\bar{x}}}\right)$$

$\bar{x} \approx 2.79$ (räkna ur tabellen $\in$ uppg.)

$s_{\bar{x}} \approx 0.196$ —"—

$\Rightarrow I_p \approx [0.32, 0.42]$

8.18 Anta att $X_1, \dots, X_n$ iid på $[0,1]$ med täthet

$$f(x|\alpha) = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x^{\alpha-1}(1-x)^{2\alpha-1}, \quad E(X) = \frac{1}{3}$$

& $Var(X) = \frac{2}{9(3\alpha+1)}$

a) Hitta MM-skattaren för α.

$Var(X) = E(X^2) - E(X)^2 = \frac{2}{9(3\alpha+1)}$. MM = skatta moment med medelvärde.

$$\overline{X^2} - 1/9 = \frac{2}{9(3\hat{\alpha}+1)} \iff 3\hat{\alpha} = \frac{2}{9(\overline{X^2}-1/9)} - 1 \iff \hat{\alpha} = \frac{1}{3}\left(\frac{2}{9(\overline{X^2}-1/9)} - 1\right)$$

b) Vilken ekv. gäller för ML-skattaren av α?

Derivatan för (log-)likelihood är 0 i $\hat{\alpha}$.

$$L(\alpha) = \prod_{i=1}^{n} f(X_i|\alpha) = \prod_{i=1}^{n} \underbrace{\frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)}}_{\gamma(\alpha)} X_i^{\alpha-1}(1-X_i)^{2\alpha-1} =$$

$$= \gamma(\alpha)^n \prod_{i=1}^{n}(X_i^{\alpha-1}) \prod_{i=1}^{n}((1-X_i)^{2\alpha-1}).$$

$$l(\alpha) = \log L(\alpha) = n \log \gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n} \log X_i + (2\alpha-1)\sum_{i=1}^{n}(1-X_i)$$

$$l'(\alpha) = \frac{n\gamma'(\alpha)}{\gamma(\alpha)} + \sum_{i=1}^{n} \log X_i + 2\sum_{i=1}^{n}(1-X_i) = 0$$

8.27 Några elektriska komponenter har livslängder som är $Exp(\tau^{-1})$-fördelade.

Fem komp. testas samtidigt. Den första går sönder efter 100 dagar, varefter testet avbryts.

a) Vad är likelihood-funktionen för τ?

T_k = livslängd för komp. $k \sim Exp(\tau^{-1})$ med täthet

$f(t|\tau) = \frac{1}{\tau} e^{-t/\tau}, \quad t > 0.$

Låt $T = \min(T_1, \dots, T_5)$, hint $T \sim Exp(5/\tau)$

$\Rightarrow L(\tau) = f_T(100 \mid \tau) = \frac{5}{\tau} e^{-\frac{5 \cdot 100}{\tau}}$

b) Vad är MLE för τ?

$l(\tau) = \log L(\tau) = \log 5 - \log \tau - 500/\tau$

$l'(\tau) = -\tau^{-1} + 500/\tau^2 = 0 \quad \Rightarrow \hat{\tau} = 500$

$l''(\tau) = +\tau^{-2} - 1000/\tau^3 \overset{\tau=500}{=} \frac{1}{2.5 \cdot 10^5} - \frac{1}{1.25 \cdot 10^8} < 0 \Rightarrow \hat{\tau} = 500$ maximum

c) Vilken fördelning har $\hat{\tau}$?

$T \sim Exp(5/\tau)$, $\hat{\tau} = 500 = 5T$

$P(5T \leq t) = P(T \leq t/5) = 1 - e^{-\frac{5}{\tau}\frac{t}{5}} = 1 - e^{-t/\tau}$

$\Rightarrow \hat{\tau} \sim Exp(1/\tau)$

d) Vad är standardfelet för $\hat{\tau}$?

$\sigma_{\hat{\tau}} = \sqrt{Var(\hat{\tau})} = \sqrt{(\tau^{-1})^{-2}} = \tau$

8.57 Skattning av varians för normalfördelning med okänt väntevärde.

$X_1, \ldots, X_n$ är iid normal. Använd $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ & att $E(\chi^2_r) = r$, $Var(\chi^2_r) = 2r$.

a) Vilken av skattarna är VVR?

$s^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad \sigma^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$$\left.\begin{aligned} E\left(\frac{(n-1)s^2}{\sigma^2}\right) &= \frac{n-1}{\sigma^2} E(s^2) \\ E\left(\frac{(n-1)s^2}{\sigma^2}\right) &= n-1 \end{aligned}\right\} \Rightarrow \frac{n-1}{\sigma^2} E(s^2) = n-1 \Leftrightarrow E(s^2) = \sigma^2$$

så VVR!

$E(\sigma^2) = E\left(\frac{n-1}{n} s^2\right) = \frac{n-1}{n} \underset{E(s^2)}{\sigma^2} \overset{?}{\neq} s^2$ så ej VVR!

b) Vilken har minst MSE (mean square error)

MSE för s^2: $E((s^2 - \sigma^2)^2) = E(\sigma^4 (s^2/\sigma^2 - 1)^2) =$

$$= E\left(\frac{\sigma^4}{(n-1)^2}\left(\frac{(n-1)s^2}{\sigma^2} - (n-1)\right)^2\right) = \frac{\sigma^4}{(n-1)^2} E\left(\left(\frac{(n-1)s^2}{\sigma^2} - E\left(\frac{(n-1)s^2}{\sigma^2}\right)\right)^2\right)$$

$$= \frac{\sigma^4}{(n-1)^2} Var\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

MSE för $\hat{\sigma}^2$: $E((\hat{\sigma}^2 - \sigma^2)^2) = E\left(\left(\frac{n-1}{n}s^2 - \sigma^2\right)^2\right) = \left(\frac{n-1}{n}\right)^2 E\left(\left(s^2 - \frac{n}{n-1}\sigma^2\right)^2\right)$

$$= \left(\frac{n-1}{n}\right)^2 E\left(\left(s^2 - \sigma^2 + \sigma^2 - \frac{n}{n-1}\sigma^2\right)^2\right) = \left(\frac{n-1}{n}\right)^2 \left(E\left((s^2-\sigma^2)^2\right) + E\left(\left(\sigma^2 - \frac{n}{n-1}\sigma^2\right)^2\right) +\right.$$

$$\left. \underbrace{+ 2E\left(\underbrace{(s^2-\sigma^2)}_{\text{VVR} \Rightarrow E=0}\underbrace{\left(\sigma^2 - \frac{n}{n-1}\sigma^2\right)}_{\text{flyttakt}}\right)}_{=0}\right) = \left(\frac{n-1}{n}\right)^2\left(\frac{2\sigma^4}{n-1} + \left(\sigma^2 - \frac{n}{n-1}\sigma^2\right)^2\right) =$$

$$= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} + \left(\frac{n-1}{n}\right)^2 \sigma^4\left(1 - \frac{n}{n-1}\right)^2 = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} =$$

$$= \frac{2\sigma^4}{n} - \frac{\sigma^4}{n} < \frac{2\sigma^4}{n} < \frac{2\sigma^4}{n-1}$$

$\Rightarrow$ MSE för $\hat{\sigma}^2$ är minst (= bäst?)