Konf. int. , forts.

$X_1, X_2, \dots, X_n$ uavh $N(\mu, \sigma^2)$!

$T = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim T_{n-1}$

$\to \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$ = eksakt $(1-\alpha)100\%$ KI for μ.

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

$\to \left(\frac{S^2}{\chi_{\alpha/2}}(n-1), \frac{S^2}{\chi_{1-\alpha/2}}(n-1)\right)$ er et eksakt $(1-\alpha)100\%$ KI for σ^2

$P(\chi^2_{n-1} > \chi_\alpha) = \alpha$

T-intervallet er robust for normalitetsantagelsen når n er stor

Kji-int. krever normalfordeling også ved stor n.

Men når n er liten og/eller normalfordeling ikke holder er det ikke sikkert at

Dekningsgrad = Coverage $= P(\theta \in (\theta_L, \theta_U)) = 1-\alpha$

sep 12-12:13

Hvis antagelsene svikter kan vi istedet beregne KI ved bootstrapping.

Bootstrap Data $X_1, \dots, X_n \sim f(x;\theta)$

(0) Estimat $\hat{\theta}$ fra $X_1, \dots, X_n$

(1) Generer pseudodata $X_1^*, \dots, X_n^*$

(2) Beregner θ^* fra pseudo dataene

(3) Gjentar (1) - (2) B ganger, $\theta_1^*, \dots, \theta_B^*$

Hvis fordelingen til θ_b^*-ene tiln. normal får et tiln KI ved

$\hat{\theta} \pm t_{\alpha/2} S_{\theta^*}$ Boot-t intervall

der $S^2_{\theta^*} = \frac{1}{B}\sum_{b=1}^{B}(\theta_b^* - \bar{\theta}^*)^2$

sep 12-12:44

Alternativt kan vi benytte persentilintervallet

Sorterer θ_b^*-ene: $\theta^*_{(1)} < \theta^*_{(2)} < \dots < \theta^*_{(B)}$

La f.eks. $b_L = [0.025B]$ (og $[x]$ = heltallsverdi av x)

og $b_U = [0.975B]$

Da er $(\theta^*_{(b_L)}, \theta^*_{(b_U)})$ et 95% persentil intervall KI for θ.

Parametrisk bootstrap:

Pseudo data $x_1^*, \dots, x_n^*$ trekkes fra $f(x;\hat{\theta})$

Ikke parametrisk bootstrap:

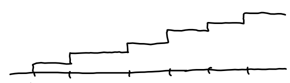
Generer $X_1^*, \dots, X_n^*$ ved å trekke med tilbakelegging fra $x_1, \dots, x_n$

sep 12-12:51

Ikke-parametrisk bootstrap er å trekke fra fordeling $P(X^* = x_i) = \frac{1}{n}$ observerte $x_1, \dots, x_n$. (hvis alle ulike)

Generelt: Trekk fra empiriske kumulative

$\hat{F}_n(x) = \frac{1}{n}\#\{x_i \le x\}$



Har at $\hat{F}_n(x) \to F(x) = P(X^* \le x)$ $n \to \infty$

$E\hat{F}_n(x) = F(x)$

$\mathrm{Var}\,\hat{F}(x) = \frac{1}{n}F(x)(1-F(x))$

Med boot-t interv. trenger vi bare est. se$(\hat{\theta})$, kan holde med $B = 100$

Men for persentilintervallet bør $B = 1000$ (eller større)

sep 12-12:58

Teoretisk median M er def. ved

$P(X \le M) = \frac{1}{2} = P(X \ge M)$

(når X er kont. fordelt)

Median: utvalget $m = \begin{cases} X_{m+1} & \text{når } n = 2m+1 \\ \frac{X_m + X_{m+1}}{2} & \text{når } n = 2m \end{cases}$

Har da $\sqrt{n}(m - M) \to N\left(0, \frac{1}{4f'(M)^2}\right)$

der X har tetthet $f(x)$

Men $f(x)$ er vanskelig å estimere og $f'(x)$ enda vanskeligere.

sep 12-13:26

Hvorfor ikke alltid bootstrap?

- Numerisk tungt
- Litt ubehagelig at vi ikke får samme resultat - eksakt - ved gjentatt bootstrap
- For mange problemstillinger fungerer "klassiske" metoder godt nok.
- Ikke alltid opplagt hvordan man skal bootstrappe, f.eks. med avhengige data.

R har innebygget en automatisk funksjon for bootstrapping: boot

Man må skrive library(boot) først.

sep 12-13:49