Cramér - Rao nedre grænse (regulært estimationsproblem)

sætn. $T(Y)$ estimator for θ $(k=1)$

$$a(\theta) = E[T(Y);\theta]$$

$$\left.\begin{aligned} a'(\theta) &= \frac{\partial}{\partial\theta}\int_y T(y) f(y;\theta)\, d\nu(y) \\ &= \int_y T(y) \frac{\partial}{\partial\theta} f(y;\theta)\, d\nu(y) \end{aligned}\right\} \text{eksi-stens}$$

$$0 < I(\theta) < \infty$$

$$\Rightarrow \operatorname{Var}[T(Y);\theta] \geq \frac{(a'(\theta))^2}{I(\theta)}$$

bevis:

$$\begin{aligned} a'(\theta) &= \int_y T(y) \frac{\partial}{\partial\theta} f(y;\theta)\, d\nu(y) \\ &= \int_y T(y) \frac{\partial}{\partial\theta}(\ln f(y;\theta))\, f(y;\theta)\, d\nu(y) \\ &= E[T(Y)\, u(\theta;Y)] \\ &= \operatorname{Cov}[T(Y), u(\theta;Y)] \end{aligned}$$

Af $-1 \leq \rho \leq 1 \Leftrightarrow \rho^2 \leq 1$ ser vi

$$(\operatorname{Cov}[T(Y), u(\theta;Y)])^2 \leq \operatorname{Var}[T(Y)]\operatorname{Var}[u(\theta;Y)]$$

$$\Leftrightarrow (a'(\theta))^2 \leq \operatorname{Var}[T(Y)]\, I(\theta)$$

$$\Leftrightarrow \operatorname{Var}[T(Y)] \geq \frac{(a'(\theta))^2}{I(\theta)}$$

spec. $T(Y)$ central $\Rightarrow a(\theta) = \theta \Rightarrow a'(\theta) = 1$

dvs. $\operatorname{Var}[T(Y)] \geq \dfrac{1}{I(\theta)}$

$k > 1$: $\operatorname{Var}[T(Y)] \geq (I(\theta))^{-1}$

(dvs. $\operatorname{Var}[T(Y)] - (I(\theta))^{-1}$ er pos. semidef.)

bemærk $\operatorname{Var}[T(Y)] \geq (I(\theta))^{-1}$

$\Rightarrow c^T \operatorname{Var}[T(Y)]\, c \geq c^T (I(\theta))^{-1} c$

$\Leftrightarrow \operatorname{Var}[c^T T(Y)] \geq c^T (I(\theta))^{-1} c$

eks. $y = (y_1, \dots, y_n)$, $Y_i \sim N(\theta, 1)$ uafh.

$l(\theta; y) = -\frac{1}{2} \sum_i (y_i - \theta)^2$

$u(\theta; y) = \frac{\partial l}{\partial \theta} = \sum (y_i - \theta) = n(\bar{y} - \theta)$

$u'(\theta; y) = -n$, $I(\theta) = -E[-n] = n$

$\operatorname{Var}[T(Y); \theta] \geq \frac{1}{n}$, når $T(y)$ central

tidl.: $\hat{\theta} = \bar{y}$, $\operatorname{Var}(\bar{Y}) = \frac{1}{n}$

dvs $\bar{Y}$ antager C-R nedre grænse.

Der findes altså ing. bedre centr. estimator.

Findes der flere lige så gode? Nej (w/ bevis)

□

eks. $y = (y_1, \dots, y_n)$, $Y_i \sim N(\mu, \sigma^2)$ uafh.

tidl. $\frac{d^2 l}{d\theta\, d\theta^T} = \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^4}(\bar{y} - \mu) \\ -\frac{n}{\sigma^4}(\bar{y} - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i (y_i - \mu)^2 \end{bmatrix}$

dvs $I(\theta) = I(\mu, \sigma^2) = n \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$ *

□

eks. $y = (y_1, \dots, y_n)$, $Y_i \sim e(\exp(\alpha + \beta x_i))$ uafh.

tidl. $\frac{d^2 l}{d\theta\, d\theta^T} = \begin{bmatrix} -e^{\alpha} \sum_i y_i e^{\beta x_i} & -e^{\alpha} \sum_i x_i y_i e^{\beta x_i} \\ -e^{\alpha} \sum_i x_i y_i e^{\beta x_i} & -e^{\alpha} \sum_i x_i^2 y_i e^{\beta x_i} \end{bmatrix}$

$I(\theta) = I(\alpha, \beta) = \begin{bmatrix} n & 0 \\ 0 & \sum x_i^2 \end{bmatrix}$, idet

$E[Y_i] = \frac{1}{\exp(\alpha + \beta x_i)}$

---

* $(I(\theta))^{-1} = \frac{1}{n} \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} = \begin{bmatrix} \operatorname{Var} \bar{Y} & 0 \\ 0 & \frac{n}{n-1} \operatorname{Var} \hat{\sigma}^2 \end{bmatrix}$

$k \geq 2$: $\quad Y_i \sim e(\exp x_i^T\beta)$, $\quad x_i^T\beta = \sum_{j=1}^{k} x_{ij}\beta_j$

$$\ell(\beta; y) = \sum_i x_i^T\beta - \sum_i y_i \exp x_i^T\beta$$

$$\frac{d\ell}{d\beta} = \sum_i x_i - \sum_i y_i (\exp x_i^T\beta)\, x_i$$

$$\frac{d^2\ell}{d\beta\, d\beta^T} = -\sum_i y_i (\exp x_i^T\beta)\, x_i x_i^T ,$$

dvs. $I(\beta) = \sum_i x_i x_i^T$, idet $EY_i = \dfrac{1}{\exp x_i^T\beta}$

$= X^TX$, hvor $X^T = [x_1\, x_2 \dots x_n]$

(X $n \times k$, X^TX $k \times k$) $\square$

Efficiens

$T(y)$ centr. estimator

efficiens af $T(y)$: $\dfrac{1}{Var[T(Y)]\, I(\theta)}$ $\quad (\leq 1)$

$T(y)$ ikke centr.:

$$MSE = E[(T(Y)-\theta)^2] = Var[T(Y)] + \left(E[(T(Y)-\theta)]\right)^2$$

ofte: $\sim O(n^{-1})$ $\qquad \sim O(n^{-2})$

dvs. dominerende

eks. $y = (y_1, \dots, y_n)$, $Y_i \sim U[0;\theta]$ uafh.

tidl.: $\hat{\theta} = y_{(n)}$

$P(Y_{(n)} \leq t) = \frac{1}{\theta^n} t^n$, $0 < t < \theta$, sæt $t := \frac{t}{\theta}$

$P(\frac{1}{\theta} Y_{(n)} \leq t) = t^n$, $0 < t < 1$

$f(\frac{1}{\theta} Y_{(n)}; \theta) = n\, t^{n-1}$

$\frac{1}{\theta} E[Y_{(n)}] = \int_0^1 t\, n t^{n-1}\, dt = n \left[\frac{t^{n+1}}{n+1}\right]_0^1 = \frac{n}{n+1}$,

dvs. $E[\hat{\theta}] = \frac{n}{n+1}\theta$

$$\frac{1}{\theta^2} E[Y_{(n)}^2] = \int_0^1 t^2 n t^{n-1} dt = n \left[\frac{t^{n+2}}{n+2}\right]_0^1 = \frac{n}{n+2}$$

dvs. $E[\hat{\theta}^2] = \frac{n}{n+2}\theta^2$

$$\mathrm{Var}[\hat{\theta}] = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2$$

korrektion for centralitet : $T(y) := \frac{n+1}{n}\hat{\theta}$

$\mathrm{Var}[T(Y)] \sim O(n^{-2})$, men C-R forventer $O(n^{-1})$?

(estimationsproblemet er ikke regulært !) □

Egenskaber ved MLE for $n \to \infty$

Asymptotisk teori handler om at studere stikprøve-variable $T_n(y)$ for $n \to \infty$, hvor n er stikprøvestørrelsen.

I praksis kan resultaterne ofte bruges på moderate stikprøvestørrelser, normalt ok for $n > 30$.

Konsistente estimatorer

$\mathcal{F}_n$ indiceret mgd. af stat. modeller med parameterrum Θ

$T_n(Y)$ følge af estimatorer for $\theta \in \Theta$

$T_n(Y)$ konsistent, når $T_n(Y) \xrightarrow{P} \theta$, dvs.

$$\lim_{n\to\infty} P(|T_n(Y) - \theta| > \varepsilon; \theta) = 0$$

(svag konsistens)

stærk konsistens, når $T_n(Y) \to \theta$ a.s.

eks. $y = (y_1, \dots, y_n)$, $Y_i \sim N(\theta, 1)$ uafh.

tidl. $\hat{\theta} = \bar{y}$

$\bar{Y} \to \theta$ a.s. ifølge store tals stærke lov

dvs. $\bar{Y}$ er en stærk konsistent estimator

□

sætn. $\left.\begin{array}{l} E[T_n(Y)] \to \theta_* \\ Var[T_n(Y)] \to 0 \end{array}\right\} \Rightarrow T_n(Y) \xrightarrow{r} \theta_*$

sætn. $g(\cdot)$ kont. fkt.

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n \xrightarrow{r} c \Rightarrow g(X_n) \xrightarrow{r} g(c)$$

$y = (y_1, \dots, y_n)$, Y_i uafh. og identisk fordelte, tæthed $g(y_i;\theta)$, regulært estimationsproblem

$l_n(\theta; y) = \sum_i \ln g(y_i;\theta)$

Lad θ_* betegne den sande parameter og $\theta \neq \theta_*$

$$\frac{1}{n}\left(l_n(\theta;y) - l_n(\theta_*;y)\right) = \frac{1}{n}\sum_i \ln \frac{g(Y_i;\theta)}{g(Y_i;\theta_*)}$$

$$\to E\left[\ln \frac{g(Y_i;\theta)}{g(Y_i;\theta_*)};\theta_*\right] \text{ a.s. (store tals stærke lov)}$$

Jensens ulighed* giver, når $\theta \neq \theta_*$ (ln er konkav)

$$E\left[\ln \frac{g(Y_i;\theta)}{g(Y_i;\theta_*)};\theta_*\right] \leq \ln E\left[\frac{g(Y_i;\theta)}{g(Y_i;\theta_*)};\theta_*\right] = 0^{**}$$

heraf $l_n(\theta;y) - l_n(\theta_*;y) \to -\infty$ for $n \to \infty$, dvs.

det må forventes, at $L_n(\theta_*;y) >> L_n(\theta;y)$ for n passende stor i alle pkt.'er med $\theta \neq \theta_*$.

Kan vi på denne baggrund forvente, at $\hat{\theta} \to \theta_*$ a.s.?

Specialtilfælde: $\Theta = \{\theta_*, \theta_1, \theta_2, \dots, \theta_m\}$

sæt $\forall \varepsilon > 0$: $A_j = \{y \mid l_n(\theta_*) - l_n(\theta_j) > \varepsilon \text{ for } n > n_0\}$,

$j = 1, \dots, m$

n_0 vælges, så $P(A_j) > 1 - \delta$, valg af δ fastlægger n_0

$$P\left(\bigcap_j A_j\right) = 1 - P\left(\bigcup_j \bar{A}_j\right) \geq 1 - \sum_j P(\bar{A}_j) \geq 1 - m\delta$$

* $E[g(X)] \geq g(E[X])$, når g er konveks

** Bemærk, at $E\left[\frac{g(Y_i;\theta)}{g(Y_i;\theta_*)};\theta_*\right] = \int_y \frac{g(y_i;\theta)}{g(y_i;\theta_*)} g(y_i;\theta_*)\, d\nu(y) = \int_y g(y_i;\theta)\, d\nu(y) = 1$

der $P(\{y \mid l_n(\theta) - l_n(\theta_*) > \varepsilon$ for alle $\theta_j \neq \theta_*$ og
for alle $n > n_0\}) \geq 1 - m\delta$

altså $\hat{\theta} \to \theta_*$ a.s., idet δ er arbitrær

bevis, når Θ er et interval, er komplicenet!

Asymptotisk ford. af MLE

ford. af $\hat{\theta} - \theta_*$ ~~søges~~

Forudsætninger

- estimationsproblemet er regulært
- Y_i, $i = 1, \dots, n$ uafh. og identisk fordelte, tæth. $g(y_i; \theta)$
- $0 < i(\theta) < \infty$
- $\hat{\theta}$ konsistent
- $\exists M(y;\theta) : \left| \frac{\partial^3}{\partial \theta^3} \ln g(y_i;\theta) \right| < M(y;\theta)$
- $\exists M_0 : E[M(Y_i;\theta);\theta] < M_0 < \infty$

Betragt $l'(\hat{\theta}; y)$, husk $l'(\hat{\theta}; y) = 0$

Taylorudvikling af $l'(\hat{\theta})$ omkr. θ_*:

$$l'(\theta_*) + l''(\theta_*)(\hat{\theta} - \theta_*) + \tfrac{1}{2} l'''(\tilde{\theta})(\hat{\theta} - \theta_*)^2 = 0,$$

$\tilde{\theta}$ ligger ml. $\hat{\theta}$ og θ_*

$$\Rightarrow \sqrt{n}(\hat{\theta} - \theta_*) = \frac{-\frac{1}{\sqrt{n}} l'(\theta_*)}{\frac{1}{n} l''(\theta_*) + \frac{1}{2n} l'''(\tilde{\theta})(\hat{\theta} - \theta_*)}$$

bemærk

$$-\frac{1}{\sqrt{n}} l'(\theta_*; Y) = -\frac{1}{\sqrt{n}} \sum_i \frac{\partial}{\partial \theta} \ln g(Y_i; \theta) \Big|_{\theta = \theta_*}$$

$$= -\frac{1}{\sqrt{n}} \sum_i u(\theta; Y_i) \Big|_{\theta = \theta_*} \xrightarrow{d} N(0, \tfrac{1}{n} n\, i(\theta))$$

med $\theta = \theta_*$

ifølge den centr. grænseværdisætn.

Altså $-\frac{1}{\sqrt{n}}\, l'(\theta_*; Y) \xrightarrow{d} N(0, i(\theta_*))$

$\frac{1}{n} l''(\theta_*) = \frac{1}{n}\sum_i \frac{\partial u}{\partial \theta}(\theta_*; Y_i) \to E\left[\frac{\partial u}{\partial \theta}(\theta_*; Y_i)\right]$ a.s.,

iflg store tals stærke lov, altså $\frac{1}{n} l''(\theta_*) \to -i(\theta_*)$

$$\frac{1}{n}\left|l'''(\tilde\theta; Y)\right| \le \frac{1}{n}\sum_i \left|\frac{\partial^3}{\partial\theta^3}\log g(Y_i;\theta)\right|\Bigg|_{\theta=\tilde\theta}$$

$$< \frac{1}{n}\sum_i M(Y_i;\theta) \sim O_p(1)$$

$\hat\theta - \theta_* \xrightarrow{P} 0$, dvs. $\hat\theta - \theta_* \sim o_p(1)$

dvs $\frac{1}{2n} l'''(\tilde\theta; Y)(\hat\theta - \theta_*) \sim o_p(1)$

Sammenfatning:

$$\sqrt{n}(\hat\theta - \theta_*) \xrightarrow{d} N\left(0, \frac{i(\theta_*)}{(-i(\theta_*))^2}\right) = N\left(0, \frac{1}{i(\theta_*)}\right)$$

approksimativt:

$$\hat\theta \sim N\left(\theta_*, \frac{1}{n\, i(\theta_*)}\right), \quad n > 30$$

$k > 1$:

$$\sqrt{n}(\hat\theta - \theta_*) \xrightarrow{d} N_k(0, (I(\theta_*))^{-1})$$

appr.: $\hat\theta \sim N_k(\theta_*, (n I(\theta_*))^{-1}))$, $n > 30$

Efficiens af MLE

den tidl. def. kræver bestemmelse af $E[T(Y)]$ og $Var[T(Y)]$

ny def:

$T(Y)$ er bedste asymptotiske normal, når

$$\sqrt{n}(T(Y) - \theta) \xrightarrow{d} N\left(0, \frac{1}{i(\theta)}\right)$$

MLE opfylder kravet, men også mange andre estimatorer for stor klasse af estimatorer:

$$E[T(Y)] = \theta + \frac{b(\theta)}{n} + O(n^{-2})$$

$$Var[T(Y)] = \frac{1}{n\,i(\theta)} + O(n^{-2})$$

dvs. MSE er domineret af $\frac{1}{n\,i(\theta)}$

Antag $E[(T(Y)-\theta)^2] = \frac{1}{n\,i(\theta)} + \frac{a_2(\theta)}{n^2} + o(n^{-2})$

estimator med 'mindste' $a_2(\theta)$ foretrækkes

(anden ordens efficiens)

MLE har høj anden ordens efficiens, når

bias er fjernet (eller næsten fjernet)

dvs. $\hat{\hat{\theta}} = \hat{\theta} - \frac{b(\theta)}{n}$ med ny bias $\sim O(n^{-2})$

Reparametrisering

$\psi = \psi(\theta)$ bijektiv, $\hat{\psi} = \psi(\hat{\theta})$, $k=1$, ψ diff.

$$i(\psi) = \left(\frac{d\theta}{d\psi}\right)^2 i(\theta)\Big|_{\theta=\theta(\psi)}$$

$$\sqrt{n}(\hat{\psi}-\psi) \xrightarrow{d} N\left(0, \frac{1}{\left(\frac{d\theta}{d\psi}\right)^2 i(\theta)}\right)\Big|_{\theta=\theta(\psi)}$$

eks $y = (y_1, \dots, y_n)$, $Y_i \sim N(\mu, \sigma^2)$ u.a.f.h.

tidl. $I(\mu, \sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$

betragt $\hat{\sigma}^2$

approksimativ fordeling:

$$\left.\begin{array}{l} \hat{\sigma}^2 \sim N\left(\sigma^2, \frac{2\sigma^4}{n}\right) \\ \text{eksakt: } \hat{\sigma}^2 \sim \frac{\sigma^2}{n}\chi^2(n-1) \end{array}\right\}$$ sammenl., se AA s. 87

parameterskift : $\psi = \sqrt{\sigma^2} \Leftrightarrow \psi^2 = \sigma^2, \quad \psi > 0$

$$\frac{d\sigma^2}{d\psi} = 2\psi$$

$$I(\psi) = 2\psi \, \frac{n}{2\sigma^4}\bigg|_{\sigma^2 = \psi^2} = \frac{2n}{\psi^2}$$

approksimativ fordeling of $\hat{\psi}$:

$$\hat{\psi} \sim N\left(\psi, \frac{\psi^2}{2n}\right)$$

sammenl. m. eksakt ford, se AA s. 87

$k > 1$: $\psi = (\psi_1(\theta), \psi_2(\theta), \ldots, \psi_k(\theta))$ bijektiv

$\psi_j(\theta), \quad j = 1, \ldots, k$, diff.

approksimativ fordeling of $\hat{\psi}$:

$$\hat{\psi} \sim N_k(\psi, \Delta^T I(\theta) \Delta^{-1})\bigg|_{\theta = \theta(\psi)}, \quad \Delta = \left[\frac{\partial \theta_i}{\partial \psi_j}\right]$$