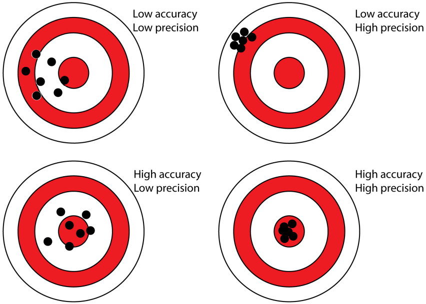
Matematisk Statistik

9. Forelæsning 02.03.2021

Estimation i klassisk statistik: punktestimatorernes egenskaber

- bias / centralitet,
- efficiens,
- MSE (mean square error),
- konsistens,
- omparametrisering, Jensens ulighed .



Definition

Bias af en estimator $\hat{\theta}$ til parameteren θ er givet ved

$$\mathrm{Bias}\,\hat{\theta} = \mathrm{E}\hat{\theta} - \theta.$$

Estimatoren $\hat{\theta}$ kaldes for **middelret** eller **central** (unbiased) hviss $\mathrm{E}\hat{\theta} = \theta$, dvs, $\mathrm{Bias}\,\hat{\theta} = 0$.

(se bogen: side 164)

Betragt en Bernoulli eksperiment, bestående af n uafhængige stokastiske variabler $X_1, \ldots, X_n$, der antager værdien 1 med sandsynlighed p, og er lig 0 ellers.

Andelen $\hat{p}$ af positive udfald i n uafængige Bernoulli eksperimenter,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

er en estimator til andelen p i populationen.

Proposition 6.3.1 If $X_1, X_2, \ldots, X_n$ are Bernoulli random variables with parameter p, then $\mathrm{E}[\hat{p}] = p$.

Proof. Let $X = \sum_{i=1}^{n} X_i$. Then

$$\mathrm{E}[\hat{p}] = \mathrm{E}\left[\frac{X}{n}\right] = \frac{np}{n} = p.$$

$\square$

Eksempel: estimator til varians

(se bogen: side 164)

Theorem 6.3.2 Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution with unknown $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then an unbiased estimator of σ^2 is $S^2 = (1/(n-1)) \sum_{i=1}^{n} (X_i - \overline{X})^2$.

Værktøj til beviset: (se side 164 MSRR)

- **Proposition A.2.1** $\mathrm{Var}[X] = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2$.
- **Theorem A.4.1** Let $X_1, X_2, \ldots, X_n$ be identically distributed random variables with mean μ and variance σ^2.
  Then $\mathrm{E}[\overline{X}] = \mu$.
  If, in addition, $X_1, X_2, \ldots, X_n$ are independent, then $\mathrm{Var}[\overline{X}] = \sigma^2/n$.
- samt identiteten

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \left(\sum_{i=1}^{n} X_i^2 \right) - n\overline{X}^2. \qquad (6.16)$$

(se også side 164 MSRR)

Definition

Lad $X_1, X_2, \ldots X_n$ være uafhængige, identisk fordelte stokastiske variabler fra en fordeling med varians $\sigma^2 < \infty$. Stikprøvefunktionen

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X)^2$$

kaldes **stikprøvevarians** (sample variance).

Stikprøvevariansen er en middelret estimator af σ^2.

Eksempel: den uniforme fordeling

Example 6.12 Let $X_1, X_2, \ldots, X_n$ be i.i.d. from Unif$[0, \beta]$. We have already seen that the MLE of β is $\hat{\beta}_{\text{mle}} = X_{\max}$.

With $f_{\max}(x) = (n/\beta^n)x^{n-1}$ (see Corollary 4.2.2), we have

$$\begin{aligned} \mathrm{E}[X_{\max}] &= \int_0^{\beta} x \, \frac{n}{\beta^n} x^{n-1} \, \mathrm{d}x \\ &= \frac{n}{n+1}\beta, \end{aligned}$$

derfor er $\hat{\beta}_{\text{mle}} = X_{\max}$ ikke middelret , men

$$\hat{\beta} = \frac{n+1}{n} X_{\max}$$

er middelret.

En anden middelret estimator er den fra moment metoden,

$$\hat{\beta}_{\text{mom}} = 2\bar{X}.$$

Definition

En estimator $\hat{\theta}$ kaldes **asymptotisk middelret**, hvis

$$\lim_{n\to\infty} \mathrm{E}\hat{\theta}(X_1,\ldots,X_n) = \theta,$$

for uafhængige $\mathbf{X}_1, \mathbf{X}_2, \cdots \sim F(.,\theta)$.

Eksempler

- ML estimatoren til $N(\mu,\sigma^2)$ normalfordelingens varians er asymptotisk middelret:

$$\mathrm{E}\hat{\sigma}^2 = \frac{n-1}{n}\sigma^2 \xrightarrow[n\to\infty]{} \sigma^2.$$

- ML estimator for parameter β i Unif$[0,\beta]$-fordelingen er $X_{\max}$. Den er asymptotisk middelret:

$$\lim_{n\to\infty} \mathrm{E}[X_{\max}] = \lim_{n\to\infty} \frac{n}{n+1}\beta = \beta$$

Motivation

Tit findes mange middelrette estimatorer for den samme parameter
— hvilken skal så foretrækkes?

Eksempel:

$\mathbf{X} = (X_1, X_2, X_3)$, uafhængige og identisk fordelte variabler med middelværdi μ og varians σ^2.

Både

$$\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$$

og

$$Y = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3$$

er middelrette estimatorer for μ.

Men

$$\operatorname{Var}\bar{X} = \frac{1}{3}\sigma^2 \quad \text{mens} \quad \operatorname{Var}Y = \frac{7}{18}\sigma^2.$$

Definition

Hvis $\hat{\theta}_1$ og $\hat{\theta}_2$ både er middelrette estimatorer til samme parameter θ, men $\operatorname{Var}\hat{\theta}_1 < \operatorname{Var}\hat{\theta}_2$, så kaldes $\hat{\theta}_1$ for **mere efficient** end $\hat{\theta}_2$.

Den **relative efficiens** af middelrette estimatorer $\hat{\theta}_1$ og $\hat{\theta}_2$ defineres som

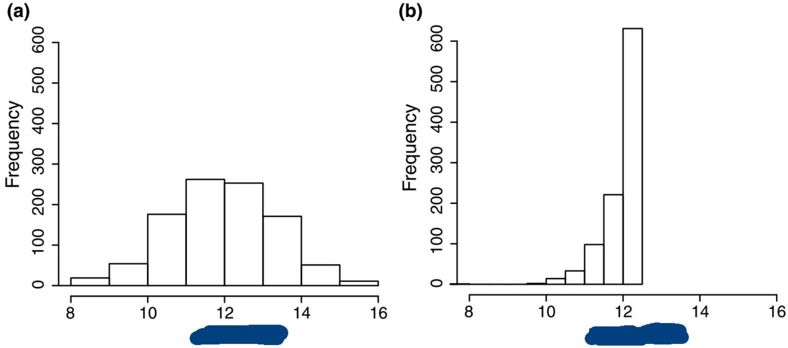
$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{\operatorname{Var}\hat{\theta}_1}{\operatorname{Var}\hat{\theta}_2}$$

Eksempel 6.13: Uniform fordeling

To middelrette estimatorer til parameteren af en Unif$[0,\beta]$-fordeling:

$$\hat{\beta}_{\text{mom}}(X_1,\ldots,X_n) = 2\bar{X}, \qquad \hat{\beta}_{\text{mle}}(X_1,\ldots,X_n) = \frac{n+1}{n} X_{\max}$$

Experiment med R: til $n = 25$, stikprøvefordelingen



☞ ? hvilket histogram stammer fra $\hat{\beta}_{\text{mom}}$, hvilket fra $\hat{\beta}_{\text{mle}}$?

? hvilken af de to estimatorer er mere efficient?

Cramér-Rao ulighed: nedre grænse på variansen

(se bogen: 6.3.2.1)

Flere teoretiske statistikere udledte i 1940erne en **nedre grænse for variansen** af unbiased estimators. Denne grænse afhænger af fordelingen af de X_i.

> If $X_1, X_2, \ldots, X_n$ are a random sample from a distribution with continuous pdf $f(x;\theta)$ and f satisfies certain smoothness criteria, then any unbiased estimator $\hat{\theta}$ of θ satisfies
>
> $$\mathrm{Var}[\hat{\theta}] \geq \frac{1}{n\,\mathrm{E}[(\partial/\partial\theta(\ln(f(X;\theta))))^2]}.$$

Udtrykket i nævneren kaldes for **Fisher information**,

$$\mathscr{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right] \qquad \text{(MSRR: 6.17)}$$

Det er variansen af score funktionen (fra ML estimationen).

Mere om Cramér-Rao uligheden: kandidatkurset "statistisk inferens".

Mean squared error kombinerer præcision og nøjagtighed i ét mål:

Definition (6.5)

Mean squared error af en estimator $\hat{\theta}$ er defineret som

$$\mathrm{MSE}\,\hat{\theta} = \mathrm{E}[(\hat{\theta} - \theta)^2].$$

Proposition 6.3.3:

$$\mathrm{MSE}\,\hat{\theta} = \mathrm{Var}\,\hat{\theta} + (\mathrm{Bias}\,\hat{\theta})^2.$$

☞ Når man sammenligner "gode"estimatorer for samme parameter, ser man tit at enten varians eller bias er lille.

Eksempel 6.15: Estimatorer til p i binomialfordelingen

Der betragtes antal af success i n uafhængige Bernoulli eksperimenter, altså en stokastisk variabel $X \sim \mathrm{Binom}(n, p)$, med n kendt og p ukendt.

ML estimator til successraten:

$$\hat{p}_1 = \frac{X}{n},$$

$$\mathrm{E}\,\hat{p}_1 = p, \quad \mathrm{Var}\,\hat{p}_1 = p(1-p)/n \implies \mathrm{MSE}\,\hat{p}_1 = \frac{p(1-p)}{n}.$$

har stor MSE når p ligger i nærheden af 0.5.

Alternativ estimator: tilføj en succes og en flop til data

$$\hat{p}_2 = \frac{X+1}{n+2}.$$

$$\mathrm{E}\,\hat{p}_2 = \frac{np+1}{n+2}, \quad \mathrm{Var}\,\hat{p}_2 = \frac{np(1-p)}{(n+2)^2} \implies \mathrm{MSE}\,\hat{p}_2 = \frac{np(1-p)+(1-2p)^2}{(n+2)^2}.$$

Den nye estimator er biased, men har mindre MSE for mellemstore værdier af p, som ikke ligger tæt ved 0 eller 1.

Eksempel 6.15, fortsat

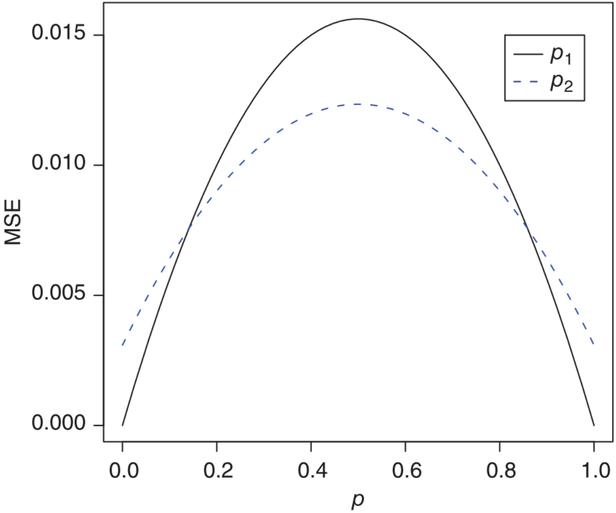


Figure 6.6 Mean square error against p, $n = 16$.

Konsistens

Vi betragter den samme estimator $\hat{\theta}$ anvendt på stikprøver med forskellige stikprøvestørrelser n:

Lad $X_1, X_2, \ldots$ være i.i.d. fra en fordeling med ukendt parameter θ. Vi skriver:

$$\hat{\theta}_n := \hat{\theta}(X_1, \ldots, X_n).$$

Hvad sker med voksende stikprøvestørrelse?

Definition 6.6 For a random sample of size n, let $\hat{\theta}_n$ denote an estimator of θ and let $\{\hat{\theta}_n\}_{n=0}^{\infty}$ be a sequence of estimators. The estimators are *consistent* for θ if and only if

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \tag{6.18}$$

for every $\varepsilon > 0$. ||

Denne type konvergens betegnes som konvergens i sandsynlighed.

Betragt $X_{\mathbf{1}}, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, 1)$, og $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Så er $\bar{X}_n \sim N(\mu; \frac{1}{n})$.

Beregn

$$\begin{aligned}
\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) &= \lim_{n\to\infty} P(\mu - \epsilon < \bar{X}_n < \mu + \epsilon) \\
&= \lim_{n\to\infty} P\Big(\frac{\mu - \epsilon - \mu}{\sqrt{1/n}} < \frac{\bar{X}_n - \mu}{\sqrt{1/n}} < \frac{\mu + \epsilon - \mu}{\sqrt{1/n}}\Big) \\
&= \lim_{n\to\infty} P\Big(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}\Big), \quad Z \sim N(0,1) \\
&= \lim_{n\to\infty} \big(\Phi(\epsilon\sqrt{n}) - \Phi(-\epsilon\sqrt{n})\big) \\
&= 1 - 0.
\end{aligned}$$

Markovs ulighed

(Pishro-Nik, afsn. 6.2.2)

Lad X være en stokastisk variabel, $X \geq 0$, og lad $c > 0$. Så gælder

$$P(X \geq c) \leq \frac{\mathrm{E}\, X}{c}.$$

Bevis: på tavlen.

 (se bogen: Thm. A.6.2, s. 484)

Lad X være en stokastisk variabel med endelig middelværdi μ og endelig varians σ^2. Så gælder for alle $k > 0$:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Bevis: på tavlen.

Korollar (Ex. 6.17):
$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ er en konsistent estimator for $\mathrm{E}\,X = \mu$, hvis både μ og $\mathrm{Var}\,X$ er endelige.

Tilstrækkelige betingelser for konsistens

(se bogen: Proposition 6.3.4, s.174)

Lad $\hat{\theta}_n$ være en følge af estimatorer for θ, som opfylder

$$\lim_{n\to\infty} \mathrm{E}\hat{\theta}_n = \theta \quad \text{og} \quad \lim_{n\to\infty} \mathrm{Var}\,\hat{\theta}_n = 0.$$

Så er $\hat{\theta}_n$ en konsistent følge af estimatorer for θ.

Eksempel 6.17: Eks 6.16 igen

Betragt $X_{\mathbf{1}}, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, 1)$, og $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Så er $\overline{X}_n \sim N(\mu; \frac{1}{n})$.

Så gælder $\lim_{n\to\infty} \mathrm{E}[\bar{X}_n] = \mu$ og $\lim_{n\to\infty} \mathrm{Var}[\overline{X}_n] = \lim_{n\to\infty} \sigma^2/n = 0$.

Derved danner $\overline{X}_1, \overline{X}_2, \ldots$ en konsistent følge af estimatorer for μ.

Standard Cauchy fordeling: lad $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} N(0,1)$. Så er $X = Y_1 / Y_2 \sim \text{Cauchy}(0,1)$.

Tæthed af Cauchy(θ, λ)-fordelingen:

$$f(x;\theta,\lambda) = \frac{1}{\pi}\frac{\lambda}{\lambda^2 + (x-\theta)^2}.$$

Fordelingsfunktion:

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \frac{1}{\pi}\arctan\left(\frac{x-\theta}{\lambda}\right) + \frac{1}{2}.$$

- Middelværdi og varians eksisterer ikke!
  (se opgave MSRR 2.11)
- Stikprøvegennemsnit $\overline{X}_n$ har samme fordeling som X_1
  $\implies \overline{X}_n$ er ikke en konsistent estimator for θ.

- Heldigvis er fordelingen symmetrisk om θ. Altså er medianen egnet som estimator for θ:
$$\hat{\theta}_n = \operatorname{med}(X_1, \ldots, X_n).$$

- Kvartiler:

$$F(x) = \frac{1}{4} \implies \arctan\left(\frac{x-\theta}{\lambda}\right) = \frac{\pi}{4} \implies x = \theta - \lambda \qquad F(x) = \frac{3}{4} \implies x = \theta + \lambda$$

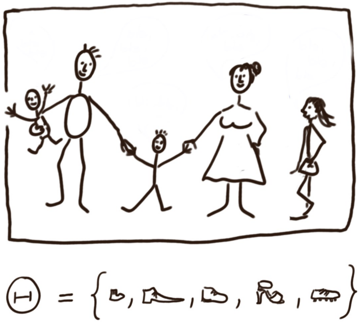
☞ Estimator for λ: half the interquartile range,

$$\hat{\lambda} = \operatorname{IQR}(X_1, \ldots, X_n)/2.$$

Vi undersøger lige, hvordan fordelingen af $\hat{\theta}_n$ afhænger af n.

Betragt en fordelingsfamilie med parameter θ.

I nogle situationer udtrykkes samme familien vha en anden parameter $\rho = h(\theta)$, hvor h er en **bijektiv** funktion.



$h(\Theta) = \{23, 45, 33, 39, 36\}$

Eksempel (R):

```
rgamma(n, shape, rate = 1, scale = 1/rate).
```

Parameteren "scale" kan bruges alternativt for "rate".

Proposition 6.3.5: Estimatorer kan genbruges i omparametriserede modeller.

$$\hat{\theta} \text{ er } \begin{cases} \text{ML-estimator} \\ \text{moment-estimator} \end{cases} \text{af } \theta \quad \Longrightarrow \quad \hat{\rho} = h(\hat{\theta}) \text{ er } \begin{cases} \text{ML-estimator} \\ \text{moment-estimator} \end{cases} \text{af } \rho$$

Normalfordelingen kan parametriseres ved (μ, σ) frem for (μ, σ^2) — se den måde R gør det på, fx `rnorm(n, mean = μ, sd = σ)`.

ML estimatorer til σ^2 og σ er givet ved

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \implies \hat{\sigma} = \sqrt{\widehat{\sigma^2}} = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}$$

Eksponentialfordelingen parametriseres nogle gange med skala $\beta = 1/\lambda$ frem for rate λ. Så skrives tætheden

$$f(x;\lambda) = \lambda \mathrm{e}^{-\lambda x} \quad \text{som} \quad \tilde{f}(x;\beta) = \frac{1}{\beta}\mathrm{e}^{-x/\beta}.$$

ML estimator til λ og $\beta = 1/\lambda$:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} \implies \hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Hvis $\hat{\theta}$ er ML-estimator for θ, så er $g(\hat{\theta})$ også ML estimator for $g(\theta)$.

Hvis $\hat{\theta}$ er moment-estimator for θ, så er $g(\hat{\theta})$ også moment estimator for $g(\theta)$.

MEN..

Hvis $\hat{\theta}$ er en middelret estimator for θ, så er $g(\hat{\theta})$ generelt set *ikke* middelret for $g(\theta)$.

Jensens ulighed[a]

Lad g være en streng konveks funktion og X en ikke degenereret stokastisk variabel[b]. Så er

$$g(\mathrm{E}\,X) < \mathrm{E}\,g(X).$$

[a] Johan Jensen, 1908

[b] en stok. variabel kaldes degenereret, hvis den kun antager 1 værdi

(se bogen: side 151)

Jensens ulighed betyder for en unbiased estimator $\hat{\theta}$, hvis g er streng konveks:

$$\mathrm{E}\, g(\hat{\theta}) > g(\theta).$$

Eksempel (MSRR, side 166): For Unif$[0,\beta]$ fordelingen er $\hat{\beta} = 2\bar{X}$ middelret. Men $(\hat{\beta})^2 = 4\bar{X}^2$ er ikke middelret for β^2:

$$\begin{aligned}\mathrm{E}[4\bar{X}^2] &= 4\,\mathrm{E}[\bar{X}^2] = 4(\mathrm{Var}[\bar{X}] + \mathrm{E}[\bar{X}]^2) \\ &= 4\left(\frac{\beta^2}{12n} + \left(\frac{\beta}{2}\right)^2\right) = \beta^2 + \frac{\beta^2}{3n}\end{aligned}$$

Altså $\mathrm{E}[\hat{\beta}^2] > \beta^2$.

Theorem 6.3.6:

maksimum likelihood estimatorer ofte er de bedste blandt alle centrale (= middelrette) estimatorer, i det de antager den minimale varians givet ved Cramér-Rao ulighed,

$$\mathrm{Var}(\hat{\theta}_n) \geq \frac{1}{n\mathscr{I}(\theta)}, \quad \mathscr{I}(\theta) = \mathrm{E}\,\frac{\partial(\ln(f(X;\theta)))}{\partial\theta}.$$

For store n må man desuden antage, at $\hat{\theta}_n$ er approksimativt normalfordelt.

Eksempel 6.19

Lad $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f$ med $f(x;\theta) = \theta x^{\theta-1}$, hvor $\theta > 0,\ 0 < x < 1$.

$$\log f(x;\theta) = \log\theta + (\theta - 1)\log x$$

MLE estimator er $\hat{\theta}_n = -\frac{n}{\sum_{i=1}^n \ln(x_i)}$.

For store n er $\hat{\theta}_n \sim N(\theta, \theta^2/n)$.

☞ mere om dette i Statistisk Teori