VII Asymptotic evaluation (let $N \to \infty$)

VII.1: Consistency, sufficiency, and robustness

We shall consider a family of estimators (= statistics) $\{W_n\}_{n\in\mathbb{N}}$ with $W_j(X_1, X_2, \cdots, X_j)$ $\sim f(\cdot|\theta)$

Example: $W_1 = X_1$, $W_2 = \frac{1}{2}(X_1+X_2)$, $W_3 = \frac{1}{3}(X_1+X_2+X_3) \cdots$

$$\overline{X}_n := W_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

Def. A sequence of estimators $\{W_n\}_{n\in\mathbb{N}}$, with $W_j = W_j(X_1, X_2, \cdots, X_j)$, is a consistent sequence for the parameter θ if

$$\forall \varepsilon > 0\ \forall \theta \in \Theta: \lim_{n\to\infty} P(|W_n - \theta| < \varepsilon) = 1$$
$$\Leftrightarrow \lim_{n\to\infty} P(|W_n - \theta| \geq \varepsilon) = 0$$

similar to "convergence in probability"

Example: $X_j \sim n(\theta, 1)$ and $W_n = \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$

Recall that $\frac{\overline{X}_n - \theta}{1/\sqrt{n}} \sim n(0,1)$ $\Rightarrow \overline{X}_n \sim n(\theta, \frac{1}{n})$

$$P(|\overline{X}_n - \theta| < \varepsilon) = \int_{\theta-\varepsilon}^{\theta+\varepsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}(s-\theta)^2} ds$$
$$\overset{s-\theta=t}{=\!=} \int_{-\varepsilon}^{\varepsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}t^2} dt$$
$$\overset{\sqrt{n}t=x}{=\!=} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{x^2}{2}} dx$$
$$\xrightarrow[\varepsilon \text{ fixed}]{n\to\infty} 1$$

Remark: We cannot do these computations $\overbrace{\ }^{\text{always}}$ so explicitly, but we're lucky! ☺

(Appendix 3)

$$P(|W_n - \theta| < \varepsilon) = P((W_n-\theta)^2 < \varepsilon^2)$$
$$\left(\text{Markov inequality}\right) \leq \frac{E((W_n-\theta)^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2}\left(\text{Bias}_\theta(W_n)^2 + \text{Var}_\theta(W_n)\right)$$

see IV.2

Thm. If for $\theta \in \Theta$ and if

$$\lim_{n\to\infty} \text{Bias}_\theta(W_n) = 0 \text{ and } \lim_{n\to\infty} \text{Var}_\theta(W_n) = 0$$

Then $\{W_n\}$ is a consistent family for θ.

Corollory: If $\underline{X} = (X_1, X_2, \cdots)$ with $E(X_j) = \theta < \infty$ and $\text{Var}(X_j) < \infty$, then $\overline{X}_n$ is a consistent sequence of estimator for θ.

(Based on §II.1)

What about efficiency? It is measured with variance.

Def. Let $\{W_n\}$ be a sequence of estimators for θ, and
let $\{k_n\}$ be a natural family of scaling parameters. Natural = coming from other reasons e.g. $k_n = \sqrt{n}$

1) If $\lim\limits_{n\to\infty} k_n \operatorname{Var}(W_n) = \tau^2$ then τ^2 is called the limiting variance.

2) If $k_n(W_n - \theta) \xrightarrow[\text{in distribution}]{n\to\infty} n(0,\sigma^2)$, then σ^2 is called the asymptotic variance.

⇕ Recall in App. 3

$$\forall x \in \mathbb{R}: F_n(x) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{s^2}{2\sigma^2}}\,ds$$

cdf of $k_n(W_n - \theta)$

Remark: $\tau^2 = \sigma^2$ in general but not always. (See example 10.1.10)

The asymptotic parameter σ will be a measurement of the efficiency.

Question: Is there a "best" σ^2?

Def. $\{W_n\}_{n\in\mathbb{N}}$ is asymptotically efficient to θ if

$$\sqrt{n}(W_n - \theta) \xrightarrow[\text{in distribution}]{n\to\infty} n(0, v(\theta)) \quad \text{with}$$

$$v(\theta) := \frac{1}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\ln(f(\cdot|\theta))\right)^2\right)}$$

related to Gramer-Rao bound (see 10.6.2)

Thm. If $X_j \sim f(\cdot|\theta)$ with $f(\cdot|\theta)$ satisfying some weak technical assumption and if $\hat{\theta}_n(X_1,\cdots,X_n)$ is the maximum likelihood estimator for θ, then $\{\hat{\theta}_n\}$ is a consistent and asymptotically efficient family of estimators.

Maximum likelihood estimator: see § IV.2

Likelihood function $\theta \mapsto L(\theta|\underline{x}) := f(\underline{x}|\theta)$ and look for its global maximum

Def. If $\{W_n\}$ and $\{V_n\}$ are 2 sequences of estimators for θ satisfying

$$\sqrt{n}(W_n - \theta) \xrightarrow[\text{in distr.}]{n\to\infty} n(0, \sigma_W^2) \quad \text{and}$$

$$\sqrt{n}(V_n - \theta) \xrightarrow[\text{in distr.}]{n\to\infty} n(0, \sigma_V^2)$$

then the asymptotic relative efficiency (ARE) of $\{V_n\}$ with respect to $\{W_n\}$ is

$$\operatorname{ARE}(\{V_n\},\{W_n\}) := \frac{\sigma_W^2}{\sigma_V^2}$$

Remark: $\frac{v(\theta)}{\sigma_V^2} \leq 1$

We should look for a sequence of estimators such that this ratio is close to 1.

Question: Is there always a sequence of estimators with the best σ^2?

Unfortunately no.

What about robustness? (See the Appendix 10)

Idea: What happens if $X_j \not\sim f(\cdot|\theta)$ for some j (rare events)?

Based on this idea, there should be a trade-off between efficiency & robostness.

$\rightsquigarrow$ many books.

Example: Consider the sample $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100\}$

Outlier: a point that differs a lot from the others

Then sample mean $\bar{x} \approx 14$ not robust

sample median $= 6$ more robust

If we compute the ARE

(If the pdf f is symetric $\Rightarrow$ median = mean)

$$\text{ARE}(\underset{\circledast}{\text{median}}, \underset{\text{based on } \bar{X}_n}{\text{mean}}) = \begin{cases} 0.64 & \text{normal} \\ 0.82 & \text{logistic} \\ 2 & \text{double exponential} \end{cases}$$

$\circledast$ define $M_n :=$ median for a sample $(X_1, \cdots, X_n)$

and then $\sqrt{n}(M_n - \theta) \xrightarrow[\text{in distr.}]{n \to \infty} n(0, \sigma^2_{\text{median}})$

Then for normal and logistic distributions, mean has a more efficient behavior;

for double exponential distribution, median is more efficient.

because of heavy tail

One way to take the best of both sequences of estimation is

to consider a mixture:

Consider $\sum_{j=1}^{n} \rho(X_j - a)$ and $\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq k \quad \text{related to mean} \\ k|x| - \frac{1}{2}k^2 & \text{if } |x| \geq k \quad \text{related to median} \end{cases}$

and set $\hat{\theta}_n(X_1, \cdots, X_n)$ for the

minimizer (as a function of a) of this expression.

k is a parameter that can be fixed freely.

With this new sequence of estimators we can compute (for $k = 1.5$)

	normal	logistic	d.exp
ARE (new, mean)	0.96	1.08	1.37 ← new is an improvement
ARE (new, median)	1.51	1.31	0.68

stable

Conclusion: $\{\hat{\theta}_n\}$ is a balance between mean and median, but is more robust.

$(n \to \infty)$