Advanced Statistics – Tutorial 26/08/2021

Estimators – methods and desirable properties

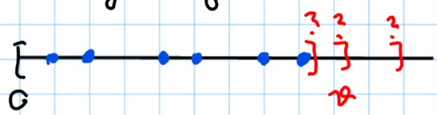
Setting:
- ϑ unknown parameter
- $X = (X_1, \dots, X_n)$ i.i.d. observations with density $f_{X;\vartheta}(x)$ under P_ϑ
- $u_n(x)$ function of observations

⇝ Then $T_n := u_n(X)$ is an estimator for ϑ (very general definition!)

Example:

Numbers $X_1, \dots, X_n$ are drawn independently and uniformly from $[0, \vartheta]$.

What is a good estimator for ϑ?



Density of $X = (X_1, \dots, X_n)$

$$f_{X;\vartheta}(x) = \begin{cases} \frac{1}{\vartheta^n} & \text{if } \forall i \in \{1, \dots, n\}: x_i \in [0, \vartheta] \\ 0 & \text{if } \exists i \in \{1, \dots, n\}: x_i \notin [0, \vartheta] \end{cases}$$

Classical ansatz: Maximum Likelihood Estimator (MLE)

→ choose $u_n(x)$ s.t.

$$f_{X;u_n(x)}(x) = \max_\vartheta f_{X;\vartheta}(x)$$

Example (ctd):

$T_n^{MLE} = \max_i X_i$ is MLE for ϑ

$\overline{T}_n = \frac{2}{n} \sum_i X_i$ is another estimator for ϑ

Desirable properties of estimators:

- Consistent: $\forall \varepsilon > 0: \lim_{n \to \infty} P_\vartheta(|T_n - \vartheta| > \varepsilon) = 0$

$$\left.\begin{array}{l} \cdot \text{ Unbiased: } E_\vartheta[T_n] = \vartheta \\ \cdot \text{ Effective: } Var_\vartheta(T_n) \text{ as small as possible} \end{array}\right\} \text{minimum variance unbiased estimator (MVUE)}$$

- Sufficient: conditional on $T_n = u_n(x)$, the distribution of X does not depend on ϑ, i.e.

$$\frac{f_{X;\vartheta}(x)}{f_{T_n;\vartheta}(u_n(x))} \text{ does not depend on } \vartheta.$$

Example (ctd):

Properties of T_n^{MLE}:

Let $c \in [0, \vartheta]$.

$\Rightarrow P_\vartheta(T_n^{MLE} \leq c) = P_\vartheta(\max_i X_i \leq c) = \prod_{i=1}^n P(X_i \leq c) = \left(\frac{c}{\vartheta}\right)^n$

$\Rightarrow f_{T_n^{MLE};\vartheta}(t) = \frac{n}{\vartheta^n} t^{n-1}, \quad t \in [0, \vartheta]$

Let $\varepsilon > 0 \Rightarrow P_\vartheta(|T_n^{MLE} - \vartheta| > \varepsilon) \leq P_\vartheta(T_n^{MLE} \leq \vartheta - \varepsilon)$

$$= \left(\frac{\vartheta - \varepsilon}{\vartheta}\right)^n \to 0 \text{ for } n \to \infty.$$

$\Rightarrow T_n^{MLE}$ is consistent.

$E[T_n^{MLE}] = \int_0^\vartheta \frac{n}{\vartheta^n} t^n \, dt = \frac{n}{n+1} \vartheta$

$\Rightarrow T_n^{MLE}$ is biased but $T_n^* := \frac{n+1}{n} T_n^{MLE}$ is unbiased.

Let $x = (x_1, \dots, x_n) \in [0, \vartheta]^n$.

$$\frac{f_{X;\vartheta}(x)}{f_{T_n^{MLE};\vartheta}(\max_i x_i)} = \frac{\frac{1}{\vartheta^n}}{\frac{n}{\vartheta^n}(\max_i x_i)^{n-1}} = \frac{1}{n(\max_i x_i)^{n-1}} \text{ does not depend on } \vartheta$$

$\Rightarrow T_n^{MLE}$ is sufficient.

Remark: T_n^* is consistent, sufficient and a MVUE for ϑ.

Properties of $\overline{T}_n$:

$E_\vartheta[\overline{T}_n] = \vartheta \Rightarrow \overline{T}_n$ is unbiased

For $\varepsilon > 0$, $P_\vartheta(|\overline{T}_n - \vartheta| > \varepsilon) \to 0$ for $n \to \infty$ by the weak law of large numbers $\Rightarrow \overline{T}_n$ is consistent

But $\overline{T}_n$ is not a MVUE for ϑ.