KU LEUVEN

Chapter 1
Statistical models and estimators

Stefan Van Aelst
KU Leuven
2022-2023

0 Outline I

1 (Mathematical) statistics

- Study stochastic phenomena to gain information about underlying population
- First tool: **Probability theory**
- Second tool: **Random samples**

1 Probability theory

Setting: A probability space $(\Omega, \mathcal{A}, \mathrm{P})$

- Ω: **Universum**= the set of all possible outcomes of the stochastic phenomenon
- $\mathcal{A}$: **$\boldsymbol{\sigma}$-algebra** = set of **events**, i.e. the subsets of Ω which are measurable
- P: Probability measure

Example: $(\mathbb{R}, \mathcal{B}, \mathrm{N}(\mu, \sigma^2))$

- $\mathcal{B}$ is the Borel σ-algebra generated by the collection

$$\mathcal{C} = \{[a, b] \mid -\infty < a \leq b < +\infty\},$$

1 Random variable

A function $X : \Omega \to \mathbb{R}$ is a **random variable** (r.v.) if
$X : (\Omega, \mathcal{A}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B})$ is a measurable function which means

$$\forall B \in \mathcal{B} : X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$$

Then, P induces a probability measure P_X on $(\mathbb{R}, \mathcal{B})$:

$$\mathrm{P}_X(B) = \mathrm{P}(X \in B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

So, $(\Omega, \mathcal{A}, \mathrm{P}) \overset{X}{\to} (\mathbb{R}, \mathcal{B}, \mathrm{P}_X)$ with

- (Cumulative) distribution function: $F_X(x) = \mathrm{P}(X \leqslant x)$
- Density function $f_X = \frac{dF_X(x)}{dx}$ (if r.v. X is absolutely continuous)
- Moment Generating function: $M_X(t) = \mathrm{E}[e^{tX}]$ if moments exist
- Characteristic function:
  $\phi_X(t) = \mathrm{E}[e^{itX}] = \mathrm{E}[\cos(tX)] + i\ \mathrm{E}[\sin(tX)]$

1 Random vector

A function $\boldsymbol{X} = (X_1, \ldots, X_p) : \Omega \to \mathbb{R}^p$ is a **random vector** if $\boldsymbol{X} : (\Omega, \mathcal{A}, \mathrm{P}) \to (\mathbb{R}^p, \mathcal{B}^p)$ is a measurable function
Then, P induces a probability measure $\mathrm{P}_{\boldsymbol{X}}$ on $(\mathbb{R}^p, \mathcal{B}^p)$:

$$\mathrm{P}_{\boldsymbol{X}}(B_1 \times \cdots \times B_p) = P(X_1^{-1}(B_1) \cap \cdots \cap X_P^{-1}(B_p))$$

- (Cumulative) distribution function: $F_{\boldsymbol{X}}(x_1, \ldots, x_p) = \mathrm{P}(X_1 \leqslant x_1, \ldots, X_p \leqslant x_p)$
- Density function $f_{\boldsymbol{X}} = \frac{\partial F_{X_1, \ldots, X_p}}{\partial x_1 \ldots \partial x_p}(x_1, \ldots, x_p)$ (if all X_j are absolutely continuous)
- Moment Generating function: $M_{\boldsymbol{X}}(t_1, \ldots, t_p) = \mathrm{E}[e^{t_1 X_1 + \cdots + t_p X_p}]$ if moments exist
- Characteristic function: $\phi_{\boldsymbol{X}}(t_1, \ldots, t_p) = \mathrm{E}[e^{i(t_1 X_1 + \cdots + t_p X_p)}]$.

2 Probability theory

- Assume a probability space $(\Omega, \mathcal{A}, \mathrm{P})$
- P satisfies certain properties
- Study properties of the probability space(s)

KU LEUVEN

2 (Mathematical) statistics

- The probability measure P is unknown
- Assume a statistical model $(\Omega, \mathcal{A}, \{\mathrm{P}_\theta; \theta \in \Theta\})$
- $\{\mathrm{P}_\theta; \theta \in \Theta\}$ is an assumed family of probability measures on $(\Omega, \mathcal{A})$
- Θ is the **parameter space**
- Equivalently, a statistical model $(\Omega, \mathcal{A}, \{F_\theta; \theta \in \Theta\})$ with $\{F_\theta; \theta \in \Theta\}$ an assumed family of distributions

2 Types of statistical models

- **Parametric statistics**: Θ has a finite dimension k, i.e. $\Theta \subseteq \mathbb{R}^k$
  Then, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$.
  Statistical model: $(\Omega, \mathcal{A}, \{F_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k\})$
  Example: $\{\mathbb{R}, \mathcal{B}, \mathrm{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\}$
- **Nonparametric statistics**: Θ is infinite dimensional.
  Index θ is usually dropped in this case
  Statistical model: $(\Omega, \mathcal{A}, \{\mathrm{P}; \mathrm{P} \in \mathcal{P}\})$
  Example: $\{\mathbb{R}, \mathcal{B}, \{\mathrm{P}; \mathrm{P} \in \mathcal{P}\})$ with
  $\mathcal{P} = \{$all probability measures with continuous density$\}$
- **Semiparametric statistics**: $\Theta = \Theta_1 \times \Theta_2$ with Θ_1 finite dimensional and Θ_2 infinite dimensional
  Example: single index model: $Y = g(\boldsymbol{X}^\top \boldsymbol{\beta}) + \epsilon$ with both $\boldsymbol{\beta} \in \mathbb{R}^p$ and the smooth function g unknown

2 Statistical inference

Gain information about the unknown θ that generated the data

- **Point estimation**: Find a good 'approximation' of the unknown θ.
- **Confidence interval/region**: Determine a subset of Θ which contains the unknown θ with high 'confidence'.
- **Hypothesis test**: 'Decide' whether the unknown θ belongs to $\Theta_0 \subset \Theta$ or to $\Theta_1 = \Theta \setminus \Theta_0$.

2 Random sample

A **random sample** $(X_1, \ldots, X_n)$ is a collection of independent random variables that all have the same distribution as X, i.e. P_X. These are called **i.i.d. random variables**
The realization of a random sample $(X_1, \ldots, X_n)$ is denoted by $(x_1, \ldots, x_n)$.

A **statistic** is a measurable function
$T : (\Omega^n, \mathcal{A}^{\otimes n}) \to (\mathbb{R}^p, \mathcal{B}^p) : (X_1, \ldots, X_n) \to T(X_1, \ldots, X_n)$.

- If $p = 1$ then $T_n = T(X_1, \ldots, X_n)$ is a random variable
- If $p > 1$ then $T_n = T(X_1, \ldots, X_n)$ is a random vector whose components $T_1(X_1, \ldots, X_n), \ldots, T_p(X_1, \ldots, X_n)$ are one-dimensional statistics

3 Point estimator

A statistic $T_n = T(X_1, \ldots, X_n)$ which for a random sample $(X_1, \ldots, X_n)$ provides an approximation for θ is an **estimator** of θ
The value $t_n = T(x_1, \ldots, x_n)$ is an **estimate** of θ based on the available sample data.
Examples

- $\theta = \mathrm{E}[X]$
  Estimator: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, the sample mean
- $\theta = \mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$
  Estimator: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, the sample variance
- $\theta = \mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^r]$
  Estimator: $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^r$
- $\theta = F_X(x)$
  Estimator: $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq y)$, the empirical distribution function

4 Example

Nonparametric statistical model $\{\mathbb{R}, \mathcal{B}, \{\mathrm{P} \in \mathcal{P}\}\}$ with
$\mathcal{P} = \{\mathrm{P}; \mathrm{P}\text{ has finite variance } \sigma^2 > 0\}$
Assume that $\mu = \mathrm{E}[X]$ is known and we want to estimate the unknown parameter σ^2.
Many estimators are possible, we could use for example

- $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$
- $\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
- $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$

Is there a preferable estimator?

4 Performance measures

- **Bias**: $b_\theta(T_n) = \mathrm{E}_\theta[T_n(\boldsymbol{X})] - \theta$
  Unbiased estimator: $b_\theta(T_n) = 0$ $(\forall\, \theta \in \Theta)$ i.e.
  $\mathrm{E}_\theta[T_n(\boldsymbol{X})] = \theta$ $(\forall\, \theta \in \Theta)$
- **Mean squared error**: $\mathrm{MSE}_\theta(T_n) = \mathrm{E}_\theta\left[(T_n(\boldsymbol{X}) - \theta)^2\right]$
  If $\mathrm{Var}_\theta[T_n] < \infty$ $(\forall\, \theta \in \Theta)$: $\mathrm{MSE}_\theta(T_n) = (b_\theta(T_n))^2 + \mathrm{Var}_\theta[T_n]$
- **Mean absolute deviation error**: $\mathrm{ABS}_\theta(T_n) = \mathrm{E}_\theta\left[|T_n(\boldsymbol{X}) - \theta|\right]$
- **General expected loss (risk)**: $R_\theta(T_n) = \mathrm{E}_\theta\left[L(T_n(\boldsymbol{X}), \theta)\right]$

Examples

- L_1-loss: $L(x, \theta) = |x - \theta|$: Mean absolute deviation
- L_2-loss: $L(x, \theta) = (x - \theta)^2$: Mean squared error
- L_p-loss: $L(x, \theta) = |x - \theta|^p$ for $p > 0$
- Large deviation loss: $L(x, \theta) = I(|x - \theta| > c)$

4 Asymptotic properties

Consider a sequence of statistics $\{T_n = T(X_1, \ldots, X_n); n \geq n_0\}$
Asymptotic properties of T_n are obtained as $n \to \infty$

- T_n is an **asymptotically unbiased estimator** if

$$\lim_{n\to\infty}(b_\theta(T_n)) = 0 \text{ i.e. } \mathrm{E}_\theta[T_n(\boldsymbol{X})] \xrightarrow{n\to\infty} \theta \quad (\forall\, \theta \in \Theta)$$

- T_n is **(weakly) consistent** if

$$\forall\, \theta \in \Theta : T_n \xrightarrow{P} \theta \text{ if } n \to \infty$$

- T_n is **strongly consistent** if

$$\forall\, \theta \in \Theta : T_n \xrightarrow{a.s.} \theta \text{ if } n \to \infty$$

- T_n is **mean square consistent** if

$$\forall\, \theta \in \Theta : \mathrm{MSE}_\theta(T_n) \to 0 \text{ if } n \to \infty$$

4 Asymptotic properties

- Strong consistency $\Rightarrow$ (weak) consistency
- Mean square consistency $\Rightarrow$ (weak) consistency
- An (asymptotically) unbiased estimator T_n is mean square consistent if $\mathrm{Var}_\theta[T_n] \xrightarrow{n\to\infty} 0 \; (\forall\, \theta \in \Theta)$

4 Asymptotic normality

- A univariate estimator T_n for $\theta \in \Theta \subseteq \mathbb{R}$ is **asymptotically normal (distributed)** if $\forall\, \theta \in \Theta$ there exists a $\mathrm{V}_\theta > 0$ such that

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathrm{N}(0, \mathrm{V}_\theta) \text{ if } n \to \infty$$

  $\mathrm{V}_\theta\, 0$ is called the asymptotic variance of the estimator T_n and $T_n \approx \mathrm{N}\left(\theta, \frac{\mathrm{V}_\theta}{n}\right)$

- A multivariate estimator T_n for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ is **asymptotically normal (distributed)** if $\forall\, \theta \in \Theta$ there exists a positive definite symmetric matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ such that

$$\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \mathrm{N}_k(0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \text{ if } n \to \infty$$

  $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is called the asymptotic variance-covariance matrix of the estimator T_n and $T_n \approx \mathrm{N}_k\left(\boldsymbol{\theta}, \frac{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{n}\right)$

4 Asymptotic normality

If T_n is asymptotically normal, then $\sqrt{n}(T_n - \theta)$ is **bounded in probability**, denoted by

$$\sqrt{n}(T_n - \theta) = O_{\mathrm{P}}(1)$$

That is,

$$\forall \epsilon > 0, \exists M_\epsilon, n_\epsilon : \forall n \geq n_\epsilon : \mathrm{P}(|\sqrt{n}(T_n - \theta)| \leq M_\epsilon) > 1 - \epsilon$$

4 Functions of estimators

If T_n is an estimator of θ, then $g(T_n)$ is an estimator of $g(\theta)$

Delta method

If T_n is an asymptotically normal estimator for θ:

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathrm{N}(0, \mathrm{V}_\theta)$$

and the function $g : \mathbb{R} \to \mathbb{R}$ is differentiable at θ with $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} \mathrm{N}(0, g'(\theta)^2\, \mathrm{V}_\theta)$$

4 Delta method: proof

Rewrite $$g(T_n) - g(\theta) = (T_n - \theta)\frac{g(T_n) - g(\theta)}{T_n - \theta}$$
$$= (T_n - \theta)g'(\theta) + (T_n - \theta)\left(\frac{g(T_n) - g(\theta)}{T_n - \theta} - g'(\theta)\right)$$
$$= (T_n - \theta)g'(\theta) + (T_n - \theta)h(T_n)$$

with $h(t) = \begin{cases} 0 & \text{if } t = \theta \\ \frac{g(T_n)-g(\theta)}{T_n-\theta} - g'(\theta) & \text{if } t \neq \theta \end{cases}$

Then $$\sqrt{n}\frac{(g(T_n) - g(\theta))}{g'(\theta)\sqrt{V_\theta}} = \sqrt{n}\frac{T_n - \theta}{\sqrt{V_\theta}} + \sqrt{n}\frac{T_n - \theta}{\sqrt{V_\theta}}h(T_n)\frac{1}{g'(\theta)}$$

4 Delta method: proof

$$\sqrt{n}\frac{(g(T_n) - g(\theta))}{g'(\theta)\sqrt{\mathrm{V}_\theta}} = \sqrt{n}\frac{T_n - \theta}{\sqrt{\mathrm{V}_\theta}} + \sqrt{n}\frac{T_n - \theta}{\sqrt{\mathrm{V}_\theta}} h(T_n)\frac{1}{g'(\theta)}$$

- $\sqrt{n}\frac{T_n - \theta}{\sqrt{\mathrm{V}_\theta}} = O_{\mathrm{P}}(1)$
- $h(T_n) \xrightarrow{P} h(\theta) = 0$ if $n \to \infty$
- $g'(\theta) \neq 0$ so $\frac{1}{g'(\theta)} < \infty$

Hence, $\sqrt{n}\frac{T_n - \theta}{\sqrt{\mathrm{V}_\theta}} h(T_n)\frac{1}{g'(\theta)} \xrightarrow{P} 0$ if $n \to \infty$

Since, $\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathrm{N}(0, \mathrm{V}_\theta)$ Slutsky's lemma yields the result.

4 Delta method

Example: Consider a r.v. X with $\mathrm{E}[X^4] < \infty$
Assume that $\mu = \mathrm{E}[X]$ is known and we want to estimate $\sigma^2 = \mathrm{Var}[X]$
Set $\tau^2 = \mathrm{E}[(X-\mu)^4] - \sigma^4$ and assume $0 < \tau^2 < \infty$
Based on a random sample $X_1, \ldots, X_n$, estimate σ^2 by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

- $\hat{\sigma}_n^2$ is an unbiased estimator of σ^2
- Since $\mathrm{Var}[(X-\mu)^2] = \mathrm{E}[((X-\mu)^2 - \sigma^2)^2] = \tau^2$, the central limit theorem (CLT) yields

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(X_i - \mu)^2 - \sigma^2] \xrightarrow{D} \mathrm{N}(0, \tau^2)$$

4 Delta method

The standard deviation σ can now be estimated by $\hat{\sigma}_n$
Apply the delta method with $g(x) = \sqrt{x}$ for which $g'(x) = \frac{1}{2\sqrt{x}}$

- $g'(\sigma^2) = \frac{1}{2\sigma} \neq 0$ for $\sigma^2 > 0$
- $g'(\theta)^2 \, \mathrm{V}_\theta = \frac{\tau^2}{4\sigma^2}$

which yields

$$\sqrt{n}(\hat{\sigma}_n - \sigma) \xrightarrow{D} \mathrm{N}\left(0, \mathrm{V}_\sigma\right) \text{ with } \mathrm{V}_\sigma = \frac{\tau^2}{4\sigma^2}$$

4 Variance stabilizing transformation

Often, the asymptotic variance V_θ of an asymptotically normal estimator depends on θ:

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathrm{N}(0, \mathrm{V}_\theta) \text{ if } n \to \infty$$

Can we find a transformation such that the variance does not depend on θ anymore?
That is, a function g such that

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} \mathrm{N}(0, c^2) \text{ if } n \to \infty$$

for some constant $c > 0$ independent of θ
The delta method yields that g needs to satisfy

$$g'(\theta)^2\,\mathrm{V}_\theta = c^2$$

so, g needs to solve the differential equation

$$g'(\theta) = \frac{c}{\sqrt{\mathrm{V}_\theta}}$$

4 Variance stabilizing transformation

Example 1: square root transformation

- T_n is a r.v. following a Poisson distribution with parameter $n\lambda$ for some $\lambda > 0$
- Then, $\sqrt{n}(\frac{T_n}{n} - \lambda) \xrightarrow{D} \mathrm{N}(0, \lambda)$
- Set $g(x) = \sqrt{x}$ then $g'(\lambda) = \frac{c}{\sqrt{\lambda}}$

We obtain that $\sqrt{n}\left(\sqrt{\frac{T_n}{n}} - \sqrt{\lambda}\right) \xrightarrow{D} \mathrm{N}(0, \frac{1}{4})$

Hence, if $X \sim$ Poisson(λ), then $\sqrt{X}$ behaves like $\mathrm{N}\left(\sqrt{\lambda}, \frac{1}{4}\right)$

4 Variance stabilizing transformation

Example 2: arcsin transformation

- $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter $\theta \in]0, 1[$
- CLT yields $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{D} \mathrm{N}(0, \theta(1-\theta))$
- g needs to satisfy the equation $g'(\theta) = \frac{c}{\sqrt{\theta(1-\theta)}}$
- With $c = 1/2$ this becomes $g'(\theta) = \frac{1}{2\sqrt{\theta(1-\theta)}}$ with solution

$$g(\theta) = \arcsin(\sqrt{\theta})$$

We obtain that $\sqrt{n}(\arcsin(\sqrt{\overline{X}_n}) - \arcsin(\sqrt{\theta})) \xrightarrow{D} \mathrm{N}\left(0, \frac{1}{4}\right)$

4 Multivariate Delta method

If T_n is an **asymptotically normal (distributed)** estimator for $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$:

$$\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \mathrm{N}_k(0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \text{ if } n \to \infty$$

and the function $g : \mathbb{R}^k \to \mathbb{R}$ is differentiable at $\boldsymbol{\theta}$ with gradient $\nabla g(\boldsymbol{\theta}) = (\frac{\partial g}{\partial t_1}\big|_{\boldsymbol{t}=\boldsymbol{\theta}}, \ldots, \frac{\partial g}{\partial t_k}\big|_{\boldsymbol{t}=\boldsymbol{\theta}})^\top \neq \mathbf{0}$, then

$$\sqrt{n}(g(T_n) - g(\boldsymbol{\theta})) \xrightarrow{D} \mathrm{N}(0, \nabla g(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \nabla g(\boldsymbol{\theta})) \text{ if } n \to \infty$$

5 Optimal estimators

For a parametric model $X : (\Omega, \mathcal{A}, \{F_\theta; \theta \in \Theta \subseteq \mathbb{R}\})$ can we find the best possible estimator of θ based on a random sample?

What is the **optimal estimator** depends on the performance criterion

In general we can try to find the estimator T_n which minimizes the risk $R_\theta(T_n) = \mathrm{E}_\theta\left[L(T_n(\boldsymbol{X}), \theta\right]$ for all possible values of $\theta \in \Theta$.

However, this is not feasible in general because the class of estimators is too large.
Restricted classes of estimators can be considered

5 Uniform Minimum Variance Unbiased Estimators

Consider the class of unbiased estimators with finite variance $\mathrm{Var}_\theta[T_n] < \infty$ $(\forall \theta \in \Theta)$, then
an estimator $T_n(\boldsymbol{X})$ is a **Uniform Minimum Variance Unbiased Estimator** (UMVUE) of θ if

1. $\mathrm{E}_\theta[T_n(\boldsymbol{X})] = \theta \quad \forall \theta \in \Theta$
2. For any other unbiased estimator $S_n(\boldsymbol{X})$ of θ:
   $\mathrm{Var}_\theta[T_n(\boldsymbol{X})] \leq \mathrm{Var}_\theta[S_n(\boldsymbol{X})] \quad \forall \theta \in \Theta$

5 Uniform Minimum Variance Unbiased Estimators

The following characterisation of a UMVUE was obtained by C.R. Rao

Consider a random sample $X_1, \dots, X_n$ from a statistical model with $\theta \in \Theta \subseteq \mathbb{R}$.

If T_n is an unbiased estimator of θ with variance $\mathrm{Var}_\theta[T_n(\boldsymbol{X})] < \infty (\forall \theta \in \Theta)$, then

T_n is a UMVUE of $\theta \Leftrightarrow \; \mathrm{E}_\theta[T_n U_n] = 0$ for all $\theta \in \Theta$ and for all U_n which is an unbiased estimator of 0 with finite variance.

5 Proof of the UMVUE characterisation

$\boxed{\Rightarrow}$ Suppose there exists a U_n such that $\mathrm{E}_\theta[T_n U_n] \neq 0$ for some $\theta \in \Theta$.

Then, $0 < |\mathrm{E}_\theta[T_n U_n]|^2 \leq \mathrm{E}_\theta[T_n^2]\,\mathrm{E}_\theta[U_n^2]$ (Cauchy-Schwarz), so $\mathrm{E}_\theta[U_n^2] > 0$

Set $a = -\dfrac{\mathrm{E}_\theta[T_n U_n]}{\mathrm{E}_\theta[U_n^2]} \neq 0$ then

- $\mathrm{E}_\theta[T_n + aU_n] = \theta$
- T_n is UMVUE, so $\mathrm{Var}_\theta[T_n] \leq \mathrm{Var}_\theta[T_n + aU_n]$

Hence, $\mathrm{E}_\theta[T_n^2] - \mathrm{E}_\theta[T_n]^2 \leq \mathrm{E}_\theta[(T_n + aU_n)^2] - (\mathrm{E}_\theta[T_n] + a\,\mathrm{E}_\theta[U_n])^2$

Since $\mathrm{E}_\theta[U_n] = 0$ this reduces to

$$\mathrm{E}_\theta[T_n^2] \leq \mathrm{E}_\theta[(T_n + aU_n)^2]$$

5 Proof of the UMVUE characterisation

On the other hand,

$$\begin{aligned}
\mathrm{E}_\theta[(T_n + aU_n)^2] &= \mathrm{E}_\theta[T_n^2] + 2a\,\mathrm{E}_\theta[T_nU_n] + a^2\,\mathrm{E}_\theta[U_n^2] \\
&= \mathrm{E}_\theta[T_n^2] - 2\frac{(\mathrm{E}_\theta[T_nU_n])^2}{\mathrm{E}_\theta[U_n^2]} + \frac{(\mathrm{E}[T_nU_n])^2}{(\mathrm{E}_\theta[U_n^2])^2}\,\mathrm{E}_\theta[U_n^2] \\
&= \mathrm{E}_\theta[T_n^2] - \frac{(\mathrm{E}_\theta[T_nU_n])^2}{\mathrm{E}_\theta[U_n^2]} \\
&< \mathrm{E}_\theta[T_n^2]
\end{aligned}$$

which contradicts the result on the previous slide, so we can conclude that such a U_n does not exist.

5 Proof of the UMVUE characterisation

$\boxed{\Leftarrow}$ Take an arbitrary unbiased estimator S_n of θ with finite variance

Then, $\mathrm{E}_\theta[T_n - S_n] = 0$ and $\mathrm{Var}_\theta[T_n - S_n] < \infty$
Hence, $\mathrm{E}_\theta[T_n(T_n - S_n)] = 0 \quad \forall\, \theta \in \Theta$ which yields

$$\mathrm{E}_\theta[T_n^2] = \mathrm{E}_\theta[T_n S_n] \leq \sqrt{\mathrm{E}_\theta[T_n^2]\,\mathrm{E}_\theta[{S_n}^2]} \quad \forall\, \theta \in \Theta \quad \text{(Cauchy-Schwarz)}$$

$$\begin{aligned}
\left(\mathrm{E}_\theta[T_n^2]\right)^2 &\leq \mathrm{E}_\theta[T_n^2]\,\mathrm{E}_\theta[S_n^2] \quad \forall\, \theta \in \Theta \\
\mathrm{E}_\theta[T_n^2] &\leq \mathrm{E}_\theta[S_n^2] \quad \forall\, \theta \in \Theta \\
\mathrm{Var}_\theta[T_n] &\leq \mathrm{Var}_\theta[S_n] \quad (\mathrm{E}_\theta[T_n]^2 = \mathrm{E}_\theta[S_n]^2) \quad \forall\, \theta \in \Theta
\end{aligned}$$

This holds for any unbiased estimator S_n of θ with finite variance, so T_n is a UMVUE

5 Uniqueness of UMVUE

Consider a random sample $X_1, \ldots, X_n$ from a statistical model with $\theta \in \Theta \subseteq \mathbb{R}$. It holds that

$$T_n \text{ and } S_n \text{ are both an UMVUE of } \theta \;\Rightarrow\; T_n \overset{\text{a.s.}}{=} S_n \quad (\forall\, \theta \in \Theta)$$

Proof

- $\mathrm{E}_\theta[T_n - S_n] = 0$ and $\mathrm{Var}_\theta[T_n - S_n] < \infty$
- T_n is UMVUE so $\mathrm{E}_\theta[T_n(T_n - S_n)] = 0 \Rightarrow \mathrm{E}_\theta[T_n^2] = \mathrm{E}_\theta[T_n S_n]$
- S_n is UMVUE so $\mathrm{E}_\theta[S_n(T_n - S_n)] = 0 \Rightarrow \mathrm{E}_\theta[S_n^2] = \mathrm{E}_\theta[T_n S_n]$

Hence, $\mathrm{E}_\theta[(T_n - S_n)]^2] = \mathrm{E}_\theta[T_n^2] + \mathrm{E}_\theta[S_n^2] - 2\,\mathrm{E}_\theta[T_n S_n] = 0$

Since $(T_n - S_n)^2 \geq 0 \Rightarrow (T_n - S_n)^2 \overset{\text{a.s.}}{=} 0$ so $T_n \overset{\text{a.s.}}{=} S_n$

5 Best Linear Unbiased Estimators

Sometimes the class can be restricted even further to unbiased estimators that are a linear combination of the observations $X_1, \ldots, X_n$

An estimator $T_n(\boldsymbol{X})$ is the **Best Linear Unbiased Estimator** (BLUE) of θ if

1. $\mathrm{E}_\theta[T_n(\boldsymbol{X})] = \theta \quad \forall\theta \in \Theta$
2. $T_n(\boldsymbol{X})$ is a linear estimator, i.e. it can be written as $T_n(\boldsymbol{X}) = \sum_{i=1}^n c_i X_i$ for some $c_i \in \mathbb{R}$
3. For any other linear unbiased estimator $S_n(\boldsymbol{X})$ of θ: $\mathrm{Var}_\theta[T_n(\boldsymbol{X})] \leq \mathrm{Var}_\theta[S_n(\boldsymbol{X})] \quad \forall\theta \in \Theta$

5 Best Linear Unbiased Estimators

- Linear unbiased estimators are a subset of unbiased estimators, so if the UMVUE is a linear estimator, then it is also the BLUE.
- For many parameters there do not exist linear unbiased estimators. For example, for $X \sim \mathrm{N}(\mu, \sigma^2)$ there is no linear unbiased estimator for σ^2.

5 Best Linear Unbiased Estimators

Example: If $\mathrm{Var}_\theta[X] < \infty$, then the sample mean is the BLUE of $\mu = \mathrm{E}[X]$

Gauss-Markov theorem: If X is a r.v. with mean $\mu = \mathrm{E}[X]$ and$\mathrm{Var}_\theta[X] = \sigma^2 < \infty$ and $X_1, \ldots, X_n$ is a random sample from X, then the sample mean $\overline{X}_n$ is the BLUE of μ

5 Proof of Gauss-Markov theorem

- Clearly, $\overline{X}_n$ is an unbiased estimator of μ
- $\overline{X}_n$ is a linear estimator with $c_i = \frac{1}{n}$ for $i = 1, \dots, n$
- To show that $\overline{X}_n$ has minimal variance, consider an arbitrary unbiased linear estimator $T_n(\boldsymbol{X}) = \sum_{i=1}^n c_i X_i$ of μ. Then
  - $\mathrm{E}[T_n(\boldsymbol{X})] = \sum_{i=1}^n c_i \,\mathrm{E}[X_i] = \mu \sum_{i=1}^n c_i$, so $\sum_{i=1}^n c_i = 1$
  - $\mathrm{Var}[T_n(\boldsymbol{X})] = \sum_{i=1}^n c_i^2 \,\mathrm{Var}[X_i] = \sigma^2 \sum_{i=1}^n c_i^2$
  - The BLUE thus minimizes $\sum_{i=1}^n c_i^2$ under the condition $\sum_{i=1}^n c_i = 1$
  - Under this condition it holds that
$$\sum_{i=1}^n \left(c_i - \frac{1}{n}\right)^2 = \sum_{i=1}^n c_i^2 - \frac{2}{n}\sum_{i=1}^n c_i + \frac{1}{n} = \sum_{i=1}^n c_i^2 - \frac{1}{n}$$
    so it suffices to minimize the left hand side
  - Since $\sum_{i=1}^n \left(c_i - \frac{1}{n}\right)^2 \geq 0$, the minimum is reached for $c_i = \frac{1}{n}$ for $i = 1, \dots, n$

5 Minimax estimators

Search for estimators for which the maximal possible risk (taken over θ) is minimal

An estimator T_n is called a **minimax estimator** of θ if

$$\sup_\theta R_\theta(T_n) \leq \sup_\theta R_\theta(S_n)$$

for any other estimator S_n of θ

6 Bayesian estimators

- In the Bayesian framework, the parameters are also random variables/vectors with a distribution over the parameter space Θ
- A **prior distribution** is assumed for the parameters based on past experience or current believes
- The choice of the prior distribution is an important (subjective) choice
- The information of the random sample is combined with the prior distribution to estimate θ
- For example, minimize the **Bayes risk** with respect to the prior distribution of θ

6 Bayes estimators

The Bayes risk of the estimator $T_n(\boldsymbol{X})$ with respect to the prior distribution with density $\pi(\theta)$ is defined as

$$\mathrm{BR}_\pi(T_n) = \begin{cases} \sum_{\theta\in\Theta} R_\theta(T_n)\pi(\theta) & \text{if } \pi \text{ discrete} \\ \int_{\theta\in\Theta} R_\theta(T_n)\pi(\theta)\,d\theta & \text{if } \pi \text{ continuous} \end{cases}$$

An estimator $T_n(\boldsymbol{X})$ is called the **Bayes estimator** of θ with respect to the prior density $\pi(\theta)$ if

$$\mathrm{BR}_\pi(T_n) \leq \mathrm{BR}_\pi(S_n)$$

for any other estimator $S_n(\boldsymbol{X})$ of θ.