Theory of Statistical Inference - Lecture IV.4
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

Definition IV.4.1 A $\delta \in \mathcal{D}(I^{dec})$ is *unbiased* if

$$R(\theta, \delta) \leq \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\theta', \psi)\, \delta(x, d\psi)\, P_\theta(dx)$$

for every $\theta, \theta' \in \Theta$. Let $\mathcal{D}_U(I^{dec}) \subset \mathcal{D}(I^{dec})$ denote the subclass of unbiased decision procedures. ■

- so, if closeness of ψ to $\Psi(\theta')$ is measured by $L(\theta', \psi)$, then an unbiased decision procedure chooses a value that on average is closer to the true value than any false value.

- so now look for optimal unbiased $\delta \in \mathcal{D}_U(I^{dec})$

Example IV.4.1 *0-1 loss*

suppose $\Theta = \cup_{i=1}^{k} C_i$ where $C_i \cap C_j = \phi$ whenever $i \neq j$

- so $\Psi(\Theta) = \{C_1, \ldots, C_k\}$ and 0-1 loss

$$Loss(\theta, \psi) = \begin{cases} 0 & \text{when } \theta \in C_i, \psi = C_i \\ 1 & \text{otherwise} \end{cases}$$

- $\delta(x, \cdot)$ is a probability measure on $\{C_1, \ldots, C_k\}$ and

$$R(\theta, \delta) = \sum_{i=1}^{k} I_{C_i}(\theta) E_\theta(\delta(x, \{C_i\}^c)$$

- for δ to be unbiased for every θ, θ'

$$\begin{aligned} \sum_{i=1}^{k} I_{C_i}(\theta) E_\theta(\delta(x, \{C_i\}^c) &\leq \sum_{i=1}^{k} I_{C_i}(\theta') E_\theta(\delta(x, \{C_i\}^c) \\ \text{iff when } \theta &\in C_i, \theta' \in C_j \\ I_{C_i}(\theta) E_\theta(\delta(x, \{C_i\}^c) &\leq I_{C_j}(\theta') E_\theta(\delta(x, \{C_j\}^c) \\ \text{prob. of rejecting } C_i \text{ when true} &\leq \text{prob. of rejecting } C_j \text{ when it is false} \end{aligned}$$

- this applies to hypothesis testing as well $(k = 2)$ ■

Example IV.4.2 *absolute error loss*

- $\Psi(\Theta)$ convex subset of $\mathbb{R}^1$ so wlog restrict to estimators d and for unbiasedness want, for all θ, θ'

$$R(\theta, d) = E_\theta|d - \Psi(\theta)| \le E_\theta|d - \Psi(\theta')|$$

- the following can be generalized to case $\Psi(\Theta) \subset \mathbb{R}^k$ (coordinate-wise)

Theorem IV.1 Estimator d is (abs. error) unbiased when $\Psi(\theta)$ is a median of the distribution of d when θ is true.
Proof: This follows from the following identity (**Exercise IV.4.1**) when m is a median of the distribution of r.v X with m a median of the distribution of X (stated here for the case $c \ge m$).

$$\begin{aligned} E|X - c| - E|X - m| &= (c - m)(P(X \le m) - P(X > m)) + \\ &\quad 2\int_m^c (c - x)P(dx) \ge 0 \end{aligned}$$

and equals 0 when $c = m$ and a similar identity applies when $c \le m$. ■

Example IV.4.3 *quadratic loss*

Theorem IV.2 For a convex decision function $\mathcal{D}(I^{dec})$ with quadratic loss an estimator d is unbiased iff $E_\theta(d) = \Psi(\theta)$ for every $\theta \in \Theta$.
Proof: $\Longleftarrow]$

$$\begin{aligned} & E_\theta((d - \Psi(\theta'))^t A((d - \Psi(\theta'))) \\ = \; & E_\theta((d - E_\theta(d))^t A((d - E_\theta(d))) + (E_\theta(d) - \Psi(\theta'))^t A(E_\theta(d) - \Psi(\theta') \\ \geq \; & E_\theta((d - E_\theta(d))^t A((d - E_\theta(d))) = R(\theta, d) \end{aligned}$$

and so d is unbiased.
$\Longrightarrow]$

$$\begin{aligned} & R(\theta, d) = E_\theta((d - \Psi(\theta))^t A((d - \Psi(\theta))) \\ = \; & E_\theta((d - E_\theta(d))^t A((d - E_\theta(d))) + (E_\theta(d) - \Psi(\theta))^t A(E_\theta(d) - \Psi(\theta)) \\ \leq \; & E_\theta((d - \Psi(\theta'))^t A((d - \Psi(\theta')) \end{aligned}$$

and in particular when $\Psi(\theta') = E_\theta(d)$ and so

$$(E_\theta(d) - \Psi(\theta))^t A(E_\theta(d) - \Psi(\theta)) = 0$$

which implies $E_\theta(d) = \Psi(\theta)$. ■

Example IV.4.4 *Not all parameters have (quadratic error) unbiased estimators*

- suppose the model is $\{\text{binomial}(n,\theta) : \theta \in [0,1]\}$

- then

$$E_\theta(T) = \sum_{k=0}^{n} T(k)\binom{n}{k}\theta^k(1-\theta)^{n-k}$$

which is always a polynomial of degree at most n in θ

- so the only parameters that have (squared error) unbiased estimators for this model are such polynomials, e.g. $\psi = \Psi(\theta) = \theta/(1-\theta)$ does not have an (quadratic error) unbiased estimator ■

- an optimal unbiased estimator minimizes the mean-squared error

$$MSE(\theta, d) = E_\theta((d - \Psi(\theta))^t A((d - \Psi(\theta))) = trace\,(Var_\theta(d))$$

but even if a (quadratic error) unbiased estimator exists there may not be an optimal one

- sometimes there is an optimal unbiased estimator

Example IV.4.5

- $x = (x_1, \ldots, x_n)$ iid $N_k(\mu, \Sigma_0)$ with $\Theta = \mathbb{R}^k, \Sigma_0 \in \mathbb{R}^{k \times k}$ p.d. known, and $\Psi(\mu) = \mu$

- then $\bar{x} \sim N_k(\mu, \Sigma_0 / n)$ so $\bar{x}$ is unbiased and is minimal sufficient

- the family $\{N_k(\mu, \Sigma_0) : \mu \in \mathbb{R}^k\}$ is complete, namely, if $E_{(\mu,\Sigma_0)}(d) = 0$ for every μ, then $d = 0$ with $P_{(\mu,\Sigma_0)}$ probability 1 (Lehmann-Scheffe)

- so, if $d(\bar{x})$ is unbiased for μ then $d(\bar{x}) - \bar{x}$ has mean 0 and so $d(\bar{x}) = \bar{x}$ and there is an optimal unbiased estimator ■

- note - that minimizing the variance doesn't mean that you have minimized the MSE so no guarantee a UMVU estimator is admissible

Example IV.4.6

- $x = (x_1, \ldots, x_n)$ iid $N(0, \sigma^2)$ with $\Theta = [0, \infty)$, and $\Psi(\sigma^2) = \sigma^2$, then $s^2 = \sum_{i=1}^n x_i^2 / n$ is a complete minimal sufficient statistic with $E_{\sigma^2}(s^2) = \sigma^2$ so s^2 is optimal and using $X = ns^2/\sigma^2 \sim$ chi-squared(n)

$$\begin{aligned}
& Var_{\sigma^2}(s^2) = MSE(\sigma^2, s^2) = E_{\sigma^2}(s^4) - \sigma^4 \\
= \; & \frac{\sigma^4}{n^2} E(X^2) - \sigma^4 = \frac{\sigma^4}{n^2} n(n+2) - \sigma^4 = \frac{2\sigma^4}{n} \text{ and} \\
& MSE\left(\sigma^2, \frac{n}{n+1} s^2\right) = E_{\sigma^2}\left(\frac{n}{n+1} s^2 - \sigma^2\right)^2 \\
= \; & \left(\frac{n}{n+1}\right)^2 \frac{\sigma^4}{n^2} n(n+2) - 2\left(\frac{n}{n+1}\right) \sigma^4 + \sigma^4 \\
= \; & \frac{\sigma^4}{n+1} \left\{ \frac{n(n+2)}{n+1} - 2n + n + 1 \right\} = \frac{\sigma^4}{n+1} \left\{ \frac{n(n+2)}{n+1} - n + 1 \right\} \\
= \; & \frac{(2n+1)\sigma^4}{(n+1)^2} = \left(\frac{2}{n+1} - \frac{1}{(n+1)^2}\right) \sigma^4 < \frac{2\sigma^4}{n}
\end{aligned}$$

and so the UMVU estimator is not admissible ■

Example IV.4.7 *hypothesis testing*

- $\Theta = H_0 \cup H_a, H_0 \cap H_a = \phi$

- we saw in Example IV.4.1 that δ is unbiased iff

$$
\begin{aligned}
&I_{H_0}(\theta)E_\theta(\delta(x, H_a)) \le I_{H_a}(\theta')E_\theta(\delta(x, H_0)) \text{ or} \\
&I_{H_0}(\theta)E_\theta(\varphi(x)) \le I_{H_a}(\theta')E_\theta(1-\varphi(x)) \text{ or} \\
&\text{prob. of rejecting } H_0 \text{ when it is true } \le \\
&\qquad\qquad\qquad\qquad \text{prob. of accepting } H_0 \text{ when it is true}
\end{aligned}
$$

- actually this just says that $E_\theta(\varphi(x)) \le 1/2$ but if you allow for different losses for type I error (c_0) and type II error (c_1) then unbiased φ must satisfy

$$
\begin{aligned}
E_\theta(\varphi(x)) &\le c_1/(c_0+c_1) \text{ when } \theta \in H_0 \\
E_\theta(\varphi(x)) &\ge c_1/(c_0+c_1) \text{ when } \theta \in H_a
\end{aligned}
$$

- for size α choose (c_0, c_1) s.t. $\alpha = c_1/(c_0+c_1)$

- there is a theory of unbiased testing for exponential families which generally works (depends on parameter)

Example IV.4.8 *binomial point null*

- $x = (x_1, \ldots, x_n)$ iid Bernoulli$(\theta), \Theta = [0,1], H_0 = \{\theta_0\}$ versus $H_a = \{\theta_0\}^c$

- the UMPU (uniformly most powerful unbiased) size α test is given by

$$\varphi_{\theta_0}(\bar{x}) = \begin{cases} 1 & \text{if } \bar{x} < c_1 \text{ or } \bar{x} > c_2 \\ \gamma_1 & \text{if } \bar{x} = c_1 \\ \gamma_2 & \text{if } \bar{x} = c_2 \\ 0 & \text{otherwise} \end{cases}$$

where the constants c_1, c_2 and $\gamma_1, \gamma_2 \in [0,1]$ are determined by requiring the test be exact size α and unbiased

- this leads to the randomized confidence region

$$C(u, x) = \{\theta : \varphi_\theta(\bar{x}) \le u\}$$

■

- what does unbiasedness mean for the associated confidence regions?

- suppose φ_{ψ_0} is unbiased size α for $H_0 = \{\theta : \Psi(\theta) = \psi_0\}$ versus $H_a = H_0^c$ and put $C(u, x) = \{\psi : \varphi_\psi(x) \leq u\}$

- note

$$\begin{aligned} & P_\theta^*(\Psi(\theta) \in C(u, x)) = 1 - \alpha \\ & P_\theta^*(\Psi(\theta') \in C(u, x)) = E_\theta(P\left(\varphi_{\Psi(\theta')}(x) \leq u)\right) \\ = \quad & E_\theta(1 - \varphi_{\Psi(\theta')}(x)) \\ = \quad & 1 - E_\theta(\varphi_{\Psi(\theta')}(x)) \leq 1 - \alpha \text{ whenever } \theta' \text{ is s.t. } \Psi(\theta') \neq \Psi(\theta) \\ & \text{by the unbiasedness of the test} \end{aligned}$$

Definition IV.4.2 A confidence region C for $\psi = \Psi(\theta)$ is *unbiased* whenever the probability of covering the true value is greater than or equal to the probability of covering a false value for every $\theta \in \Theta$.

- so for the binomial problem the confidence region obtained from the optimal unbiased size α tests is a uniformly most accurate unbiased $(1 - \alpha)$-confidence region for θ