

Sufficiency and unbiasedness

How we ~~assume~~ the concept of sufficiency helps us to find the UMVUE

Theorem 7.3.17 (Rao - Blackwell) Let W be any unbiased estimator of $\tau(\theta)$, and $T(x)$ be a sufficient statistic for θ . Define $\phi(T) = E[W|T]$

Then $\forall \theta \in \Theta$, $E[\phi(T)] = \tau(\theta)$ and $Var_{\theta}(\phi(T)) \leq Var_{\theta}(W)$

that is $\phi(T)$ is an uniformly better estimator than W ,

Proof $\tau(\theta) = E_{\theta}[W] = E_{\theta}[E[W|T]] = E[\phi(T)]$

$$Var_{\theta}(W) = Var_{\theta}(E[W|T]) + E[Var(W|T)]$$
$$\quad \quad \quad \downarrow$$
$$= Var_{\theta}(\phi(T)) + \text{positive quantity} \geq Var(\phi(T))$$

Moreover $\phi(T)$ is an estimator:
• is a function only of the data, because W is a function of only the data
• is $\perp \theta$ because T is sufficient

Note: if we look for the best estimator (UMVUE), we only need to consider functions of sufficient statistic
• we could ~~reduce~~ the variance by conditioning of any quantity, but we need sufficiency for the function to be an estimator

Theorem 7.3.19 If W is UMVUE of $\tau(\theta)$, then W is unique.

We still do not know if an unbiased estimator is the best unbiased one

IDEA: • check if we can improve an estimator

• let W be an unbiased estimator of $\tau(\theta)$, $E[W] = \tau(\theta)$

• let U be another estimator, s.t. $E[U] = 0$

• consider the unbiased estimator $\phi_a = W + aU$, where a is a constant

$$Var_{\theta}(\phi_a) = Var_{\theta}(W) + 2a Cov(W, U) + Var_{\theta}(U)$$

If for some θ_0 we have $Cov(W, U) \neq 0$, we can choose an a s.t.

$$2a Cov(W, U) + a^2 Var_{\theta}(U) < 0$$

so that $Var(\phi_a) < Var(W) \Rightarrow W$ is not the ~~best~~ ^{UMVUE} can be used

Therefore, the relationship between W and an unbiased estimator of 0 ^{to characterise the UMVUE}

Theorem 7.3.20. If $E[W] = \tau(\theta)$, W is the UMVUE if and only if W is uncorrelated with any unbiased estimators of θ

Note: an unbiased estimator of θ is basically random noise, if we manage to improve our W by adding random noise, it cannot be the best

- limited practical utility of this theorem (we cannot show that W is uncorrelated with all unbiased estimators of θ)
- we know that if a statistic is complete, there are no unbiased estimators of θ but θ itself

Theorem 7.3.23 (Lehmann-Schiffé) also 7.5.1. Let $T(x)$ be a complete sufficient statistic for θ and let $\phi(T)$ be an estimator based on $T(x)$ with $E[\phi(T)] = \tau(\theta)$

Then $\phi(T)$ is the unique UMVUE of $\tau(\theta)$

Combining Rao-Blackwell and Lehmann-Schiffé theorems, we obtain

Let $T(x)$ be a complete sufficient statistic for θ and let W be an unbiased estimator of $\tau(\theta)$. Then $\phi(T) = E[W | T]$ is the unique UMVUE of $\tau(\theta)$

Profile likelihood

Consider a density $f(y; \eta, \lambda)$ where η and λ are two parameters. Sometimes we are only interested in one of them, let say η , while λ is present only for modelling reasons.

η is called parameter of interest

λ is called nuisance parameter

Inference on η can be carried through the profile likelihood

def $L_{\text{prof}}(\eta) = L(\eta, \hat{\lambda}(\eta))$, where $\hat{\lambda}(\eta)$ is the maximiser of $L(\eta, \lambda)$ in λ given η , $\hat{\lambda}(\eta) = \sup_{\lambda} L(\eta, \lambda)$

Note:

- the profile likelihood is not a proper likelihood, because it is not based on any density (often it is referred to as pseudo-likelihood)
- it shares many of the properties of a proper likelihood

eg. $\sup_{\eta} L_{\text{prof}}(\eta) = \sup_{\eta, \lambda} L(\eta, \lambda)$ same MLE

similar asymptotic properties (we will see soon) Δ bounded

but not all of them, eg $E[L_{\text{prof}}(\eta)] \neq 0$, $O(1)$

Example

$Y_1, \dots, Y_n$ iid $Y_i \sim \text{Gamma}(\alpha, \beta)$ $f_Y(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$
 α shape, β rate parameters

$$\ell(\alpha, \beta) = n\alpha \log \beta + \alpha \sum \log y_i - \beta \sum y_i - \log \Gamma(\alpha)$$

we are only interested on the shape α

$$\ell_p(\alpha, \beta) = \frac{n\alpha}{\beta} - \sum y_i = 0 \quad \hat{\beta} = \frac{n\alpha}{\sum y_i} = \frac{\alpha}{\bar{y}} = \hat{\beta}(\alpha)$$

$$\begin{aligned} \ell_{\text{prof}}(\alpha) &= n\alpha \log \frac{\alpha}{\bar{y}} + \alpha \sum \log y_i - \frac{\alpha \sum y_i}{\bar{y}} - \log \Gamma(\alpha) \\ &= n\alpha \log \alpha - n\alpha \log \bar{y} + \alpha \sum \log y_i - n\alpha - n \log \Gamma(\alpha) \end{aligned}$$

$$\ell_{\alpha, \text{prof}}(\alpha) = n \log \alpha + \frac{n\alpha}{\alpha} - n \log \bar{y} + \sum \log y_i - n - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\ell_{\alpha, \text{prof}}(\alpha) = 0 \rightarrow \hat{\alpha}$$

$$\ell_{\alpha, \text{prof}}(\alpha) = \frac{\partial \ell_{\text{prof}}(\alpha)}{\partial \alpha^2} \dots$$

Asymptotic properties of the point estimator §10.1

• Consistency: the estimator converges to the true parameter for $n \rightarrow \infty$

Definition 10.1.1: A sequence of estimators $W_n(X_1, \dots, X_n)$ is a consistent sequence of estimators of the parameter θ if

$$\forall \varepsilon > 0, \forall \theta \in \Theta \quad \lim_{n \rightarrow \infty} P_\theta(|W_n - \theta| < \varepsilon) = 1$$

$$\text{or equivalently, } \lim_{n \rightarrow \infty} P_\theta(|W_n - \theta| \geq \varepsilon) = 0$$

Very similar to convergence in probability, but it needs to hold $\forall \theta \in \Theta$

To verify that an estimator is consistent:

(a) use the definition

(b) use theorem 10.1.3

Theorem 10.1.3 If W_n is a sequence of estimators of θ s.t.

(i) $\lim_{n \rightarrow \infty} \text{Var}_\theta(W_n) = 0$ $\forall \theta \in \Theta$ then W_n is a consistent sequence of estimators of θ

(ii) $\lim_{n \rightarrow \infty} \text{bias}(W_n) = 0$

Proof $P_\theta(|W_n - \theta| \geq \varepsilon) \leq \frac{E_\theta[(W_n - \theta)^2]}{\varepsilon^2} = \frac{\text{Var}(W_n) + \text{bias}^2(W_n)}{\varepsilon^2} = 0$

Chebyshev's inequality

Example: Let $X_1, \dots, X_n$ be iid $N(\mu, 1)$ and $\bar{X}_n = \frac{1}{n} \sum X_i$

We know that $\bar{X}_n \sim N(\mu, \frac{1}{n})$

$$\begin{aligned} (a) P_\theta(|\bar{X} - \mu| < \varepsilon) &= \int_{-\varepsilon}^{\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-\frac{n}{2}(\bar{x}_n - \mu)^2} d\bar{x}_n & y = \bar{x}_n - \mu \\ &= \int_{-\varepsilon}^{\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-\frac{ny^2}{2}} dy & t = \sqrt{n}y \\ &= \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{1/2} e^{-\frac{1}{2}t^2} dt \\ &= \Phi(\varepsilon\sqrt{n}) - \Phi(-\varepsilon\sqrt{n}) \xrightarrow{n \rightarrow \infty} 1 \end{aligned}$$

$$\begin{aligned} (ii) E[\bar{X}_n] &= \mu \Rightarrow \lim_{n \rightarrow \infty} \text{bias}(\bar{X}_n) = 0 \\ \text{Var}[\bar{X}_n] &= \frac{1}{n} \Rightarrow \lim_{n \rightarrow \infty} \text{var}(\bar{X}_n) = 0 \end{aligned} \Rightarrow \bar{X}_n \text{ is a consistent estimator of } \mu$$