Chapter 7
Point Estimation

Best Estimators

Outline

Overview

Lehman-Scheffé Theorem

Rao-Blackwell Theorem

Overview

Our goal is to find situations where the notion of a best estimator as measured by mean square exists. Among unbiased estimators, we may have *nothing* to choose from.

Example. Let $X \sim Exp(\beta)$ and let d be a candidate unbiased estimator for β, then

$$\beta = E_\beta d(X) = \int_0^\infty d(x)\beta e^{-\beta x}\,dx.$$

Divide by β and differentiate the Riemann integral with respect to β. Then

$$\begin{aligned} 1 &= \int_0^\infty d(x) e^{-\beta x}\,dx. \\ 0 &= -\int_0^\infty x d(x) e^{-\beta x}\,dx = -\frac{1}{\beta} E_\beta X d(X). \end{aligned}$$

Because X is a complete and sufficient statistic,

$$d(X) = 0 \quad \text{a.s.} \quad P_\beta \text{ for all } \beta$$

and so there are *no* unbiased estimators for β.

Overview

Definition. An estimator T^* is called best unbiased estimator for $h(\theta)$ if

- $E_\theta T^*(X) = h(\theta)$ for all θ (unbiased)
- For any other unbiased estimator T,

$$\text{Var}_\theta(T^*(X)) \leq \text{Var}_\theta(T(X))$$

  for all θ.

Such an unbiased estimator T^* is called a uniformly minimum variance unbiased estimator (UMVUE)

Lehman-Scheffé Theorem

Theorem. Let $T(X)$ be a complete sufficient statistic. Then all unbiased estimators of $h(\theta)$ that are functions of $T(X)$ alone are equal a.s. P_θ for all $\theta \in \Theta$.

Proof. Let $g_1(T(X))$ and $g_2(T(X))$ be two unbiased estimators of $h(\theta)$, then

$$E_\theta[g_1(T(X)) - g_2(T(X))] = 0 \quad \text{for all } \theta.$$

Because T is complete

$$g_1(T(X)) = g_2(T(X)) \quad \text{a.s. } P_\theta \text{ for all } \theta \in \Theta.$$

Rao-Blackwell Theorem

Theorem. If an unbiased estimator is a function of a complete sufficient statistic, then it is UMVUE.

If $W(X)$ is any unbiased estimator with finite variance then so is

$$W^*(T(X)) = E_\theta[W(X)|T(X)].$$

Note that the sufficiency of $T(X)$ guarantees that W^* is not a function of θ and thus is a statistic.

The conditional variance formula states that $\text{Var}_\theta(W^*(T(X))) \leq \text{Var}_\theta(W(X))$ and by the Lehman-Scheffé Theorem, $W^*(T(X))$ is UMVUE.

Example. For $X_1, \ldots X_n \sim N(\mu, \sigma^2)$, $T(\mathbf{X}) = (\bar{X}, S^2)$ is a complete sufficient statistic whose components are unbiased estimators of μ and σ^2, respectively. Thus, they are UMVUE.

Rao-Blackwell Theorem

Example. For $X_1, \ldots X_n \sim Bin(k, p)$, the sum $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a complete sufficient statistic. Moreover, $T(\mathbf{X}) \sim Bin(nk, p)$. We will use this to find a UMVUE for

$$g(p) = kp(1-p)^{k-1},$$

the probability of one success. Start with the the simple minded estimator,

$$d(X_1) = I_{\{1\}}(X_1).$$

,

$$E_p d(X_1) = P\{X_1 = 1\} = g(p).$$

So, it is unbiased.

Unbiased Estimation

By the Rao-Blackwell Theorem, the unbiased estimator

$$d^*(T(\mathbf{X})) = E_p[h(X_1)|T(\mathbf{X})]$$

must be UMVUE. For $T(\mathbf{X}) = t$,

$$\begin{aligned} d^*(t) &= P_p\{X_1 = 1|T(\mathbf{X}) = t\} = \frac{P_p\{X_1 = 1, T(\mathbf{X}) = t\}}{P_p\{T(\mathbf{X}) = t\}} \\ &= \frac{P_p\{X_1 = 1, \sum_{i=2}^n X_i = t-1\}}{P_p\{T(\mathbf{X}) = t\}} = \frac{P_p\{X_1 = 1\}P_p\{\sum_{i=2}^n X_i = t-1\}}{P_p\{T(\mathbf{X}) = t\}} \\ &= \frac{\left(kp(1-p)^{k-1}\right)\left(\binom{k(n-1)}{t-1}p^{t-1}(1-p)^{n(k-1)-(t-1)}\right)}{\binom{kn}{t}p^t(1-p)^{kn-t}} = k\frac{\binom{k(n-1)}{t-1}}{\binom{kn}{t}} \\ d^*(T(\mathbf{X})) &= k\frac{\binom{k(n-1)}{T(\mathbf{X})-1}}{\binom{kn}{T(\mathbf{X})}} \end{aligned}$$

Rao-Blackwell Theorem

Example. For $X_1, \ldots X_n \sim Unif(0, \theta)$, $T(\mathbf{X}) = \max_{1 \le i \le n} X_i$ is a complete sufficient statistic, and

$$T^*(\mathbf{X}) = \frac{n+1}{n} \max_{1 \le i \le n} X_i$$

is unbiased and thus is UMVUE.

For $g(\theta)$, we must find d so that

$$\begin{aligned} g(\theta) &= \int_0^\theta d(t) \frac{nt^{n-1}}{\theta^n}\, dt \\ \theta^n g(\theta) &= \int_0^\theta d(t) nt^{n-1}\, dt \\ n\theta^{n-1} g(\theta) + \theta^n g'(\theta) &= d(\theta) n\theta^{n-1} \\ g(\theta) + \frac{\theta}{n} g'(\theta) &= d(\theta) \end{aligned}$$

Rao-Blackwell Theorem

$$d(t) = g(t) + \frac{t}{n} g'(t)$$

For $g(t) = ct^p$,

$$\begin{aligned} d(t) &= ct^p + \frac{t}{n} cpt^{p-1} = c\left(1 + \frac{p}{n}\right) t^p \\ d(T(\mathbf{X})) &= c\left(1 + \frac{p}{n}\right) T(\mathbf{X})^p \end{aligned}$$

is unbiased and is thus UMVUE.

For $P_\theta\{X \leq x\} = x/\theta = g(\theta)$, then $c = x, p = -1$ and

$$\left(1 - \frac{1}{n}\right) \frac{x}{T(\mathbf{X})} \quad \text{is UMVUE.}$$

Rao-Blackwell Theorem

Example. For $X_1, \dots X_n \sim Pois(\lambda)$, $T(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n X_i$ is a complete sufficient statistic. Moreover,

$$E_\lambda T(\mathbf{X}) = \lambda.$$

So, $T(\mathbf{X})$ is UMVUE for λ.

For λ^2,

$$\begin{aligned}
\mathrm{Var}_\lambda(T(\mathbf{X})) &= \frac{\lambda}{n}. \\
E_\lambda T(\mathbf{X})^2 &= \frac{\lambda}{n} + \lambda^2 \\
E_\lambda T(\mathbf{X})^2 &= \frac{1}{n} E_\lambda T(\mathbf{X}) + \lambda^2 \\
E_\lambda \left[T(\mathbf{X})^2 - \frac{1}{n} T(\mathbf{X}) \right] &= \lambda^2
\end{aligned}$$

and $T(\mathbf{X})^2 - T(\mathbf{X})/n$ is UMVUE for λ^2.

Rao-Blackwell Theorem

Example. Suppose that the random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \ldots, n$$

where $x_1, \ldots, x_n$ are *fixed* constants, and $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, \sigma^2)$ with σ^2 known. Then, the likelihood

$$\begin{aligned} \mathbf{L}(\beta|\mathbf{x}, \mathbf{y}) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2)\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{n}{2\sigma^2}(\overline{y^2} - 2\beta\overline{xy} + \beta^2\overline{x^2})\right) \end{aligned}$$

then, by the Fisher-Neyman factorization theorem $T(\mathbf{x}, \mathbf{y}) = (\overline{xy}, \overline{x^2})$ is a sufficient statistic. It is also complete.

Rao-Blackwell Theorem

The likelihood $\mathbf{L}(\beta|\mathbf{x},\mathbf{y})$ is maximized when

$$SS(\beta) = n(\overline{y^2} - 2\beta\overline{xy} + \beta^2\overline{x^2})$$

is minimized. So, take a derivative,

$$SS'(\beta) = n(-2\overline{xy} + 2\beta\overline{x^2}).$$

Thus, $SS'(\hat{\beta}) = 0$ when

$$\hat{\beta} = \frac{\overline{xy}}{\overline{x^2}}.$$

Its mean

$$E_\beta\hat{\beta}(\mathbf{X},\mathbf{Y}) = \frac{1}{n\overline{x^2}}\sum_{i=1}^{n} x_i E_\beta Y_i = \frac{1}{n\overline{x^2}}\sum_{i=1}^{n} x_i(\beta x_i) = \beta$$

The estimator is unbiased and thus is UMVUE.

Unbiased Estimation

Define the space of all unbiased estimators of zero.

$$\mathcal{U} = \{U : E_\theta[U(X)] = 0, \text{for all } \theta \in \Theta\}.$$

If $T(X)$ is an unbiased estimator of $g(\theta)$, then *every* unbiased estimator of $g(\theta)$ has the form

$$T(X) + U(X)$$

for some $U \in \mathcal{U}$.

Theorem. An unbiased estimator $T(X)$ is UMVUE for $g(\theta) = E_\theta T(X)$ if and only if, for every $U \in \mathcal{U}$,

$$\mathrm{Cov}_\theta(T(X), U(X)) = 0.$$

Unbiased Estimation

Proof. (sufficiency) Let $T(X)$ be an unbiased estimator of $g(\theta)$ that is uncorrelated with all $U \in \mathcal{U}$. For any other unbiased estimator $T^*(X)$, there exists $U \in \mathcal{U}$ so that

$$T^*(X) = T(X) + U(X).$$

Because $\mathrm{Cov}_\theta(T(X), U(X)) = 0$

$$\mathrm{Var}_\theta(T^*(X)) = \mathrm{Var}_\theta(T(X)) + \mathrm{Var}_\theta(U(X)) \geq \mathrm{Var}_\theta(T(X))$$

and $T(X)$ is UMVUE.

Unbiased Estimation

Proof. (necessity) Assume that $T(X)$ is UMVUE for $g(\theta)$. For $\alpha \in \mathbb{R}$ and $U \in \mathcal{U}$, define the unbiased estimator

$$T_\alpha(X) = T(X) + \alpha U(X).$$

Then

$$\begin{aligned}
\mathrm{Var}_\theta(T(X)) &\leq \mathrm{Var}_\theta(T_\alpha(X)) \\
&= \mathrm{Var}_\theta(T(X)) + 2\alpha \mathrm{Cov}_\theta(T(X), U(X)) + \alpha^2 \mathrm{Var}_\theta(U(X)) \\
0 &\leq 2\alpha \mathrm{Cov}_\theta(T(X), U(X)) + \alpha^2 \mathrm{Var}_\theta(U(X))
\end{aligned}$$

The right side is an equation in α for an upward facing parabola that contains the origin. Thus, the inequality holds for all α if and only if $\mathrm{Cov}_\theta(T(X), U(X)) = 0$.

Unbiased Estimation

Let $X \sim Unif(\theta - 1/2, \theta + 1/2)$. Then $E_\theta X = \theta$ and $\text{Var}_\theta(X) = 1/12$. Thus,

$$T(X) = X$$

is an unbiased estimator of θ. We will show that it is *not* UMVUE by finding a *correlated estimator* $U \in \mathcal{U}$.

To charaterize $U \in \mathcal{U}$, note that

$$\int_{\theta-1/2}^{\theta+1/2} U(x)\ dx = 0 \quad \text{for all } \theta \in \Theta.$$

and

$$0 = \frac{d}{d\theta}\int_{\theta-1/2}^{\theta+1/2} U(x)\ dx = U(\theta + 1/2) - U(\theta - 1/2).$$

Thus, $\mathcal{U}$ consists of all integrable functions that have period 1 and have average 0 on every interval of length 1.

Unbiased Estimation

So, take $U(x) = \sin(2\pi x)$, then because $E_\theta U(X) = 0$.

$$\mathrm{Cov}_\theta(T(X), U(X)) = \int_{\theta-1/2}^{\theta+1/2} T(x)U(x)\ dx = \int_{\theta-1/2}^{\theta+1/2} x\sin(2\pi x)\ dx.$$

Now integrate by parts

$$\begin{aligned}
\int_{\theta-1/2}^{\theta+1/2} x\sin(2\pi x)\ dx &= -\frac{1}{2\pi}x\cos(2\pi x)\Big|_{\theta-1/2}^{\theta+1/2} + \int_{\theta-1/2}^{\theta+1/2}\frac{1}{2\pi}\cos(2\pi x)\ dx \\
&= -\frac{1}{2\pi}(\theta+1/2)\cos(2\pi(\theta+1/2)) + \frac{1}{2\pi}(\theta-1/2)\cos(2\pi(\theta-1/2)) \\
&= -\frac{1}{2\pi}\cos(2\pi(\theta+1/2)) \neq 0.
\end{aligned}$$

$$\begin{aligned}
u(x) &= x & v(x) &= -\tfrac{1}{2\pi}\cos(2\pi x) \\
u'(x) &= 1 & v'(x) &= \sin(2\pi x)
\end{aligned}$$

and $T(\mathbf{X})$ is *not* UMVUE.

Constructing UMVUE

1. Compute the lowest variance bound (Cramér-Rao Bound), and find the estimator which achieves the bound.
2. Use the Rao-Blackwell Theorem to construct an unbiased estimator $d^*(T(\mathbf{X}))$ of $g(\theta)$, where $T(\mathbf{X})$ is a complete and sufficient statistic.
3. Construct $d^*(T(\mathbf{X}))$ using $d^*(T(\mathbf{X})) = E_\theta[d(\mathbf{X})|T(\mathbf{X})]$, where $d(\mathbf{X})$ is unbiased for $g(\theta)$, and $T(\mathbf{X})$ is a complete and sufficient statistic.