Ch4. Unbiased Estimation

- A clean comparison between two estimators is not always possible
- Can we compare two estimators when both are unbiased?
- An estimator δ is called unbiased for $g(\theta)$ if

$$E_\theta[\delta(X)] = g(\theta) \qquad \forall\ \theta \in \Omega.$$

if an estimator exists, g is called <u>U-estimable</u>

ex. $X \sim \mathrm{Unif}(0, \theta)$ $\qquad$ δ is unbiased if

$$\int_0^\theta \delta(x) \frac{1}{\theta} dx = g(\theta) \qquad \int_0^\theta \delta(x) dx = \theta\, g(\theta)$$

g cannot be U-estimatable unless $\theta g(\theta) \to 0$ as $\theta \downarrow 0$.

if g' exists $\qquad \delta(x) = \frac{d}{dx}(x\, g(x)) = g(x) + x g'(x)$

if $g(\theta) = \theta$ $\qquad \delta(X) = 2X$.

ex. $X \sim \text{Bin}(n, \theta)$. $\qquad g(\theta) = \sin\theta$.

$$\sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sin\theta \qquad \forall \theta \in (0,1)$$

Left hand side = polynomial in θ with degree at most n

right hand side = cannot be written as a polynomial.

$\sin\theta$ is not U-estimable.

- squared error loss. $L(\theta, d) = (d - g(\theta))^2$.

if δ is unbiased,

$$R(\theta, \delta) = E\, L(\theta, \delta(X)) = Var_\theta(\delta(X)).$$

$\Rightarrow$ minimize the variance.

- UMVUE: Uniformly minimum variance unbiased estimator.

$Var_\theta(\delta) \le Var_\theta(\delta^*)$ $\forall \theta \in \Omega$) for any unbiased estimator δ^*.

Thm: Suppose g is U-estimable and T is complete sufficient. Then there is an essentially unique unbiased estimator based on T that is UMVUE.

pf. $\delta = \delta(X)$ — unbiased. $\eta(T) = E(\delta \mid T)$.

$g(\theta) = E_\theta \delta = E_\theta E_\theta(\delta \mid T) = E_\theta \eta(T)$ $\quad$ $\eta(T)$ — unbiased

if $E_\theta(\eta(T) - \eta^*(T)) = 0 \quad \forall \theta \in \Omega. \implies \eta(T) = \eta^*(T)$ a.e. P.

if δ^* is any unbiased estimator. $\eta^*(T) = E_\theta(\delta^* / T)$.

$Var_\theta(\delta^*) \geq Var_\theta(\eta^*(T)) = Var(\eta(T)) \quad \forall \theta \in \Omega)$

- T — complete suff. $\quad$ $E(\delta \mid T) \leftrightarrow$ UMVUE.
  δ — unbiased $\quad \Longrightarrow \quad \eta(T)$

ex. $\quad X_1 \cdots X_n \overset{iid}{\sim} Unif(0, \theta)$

$T = \max\{X_1 \cdots X_n\}$ — complete suff.

$\eta(T)$ is unbiased $\quad \int_0^\theta \eta(t) \frac{n t^{n-1}}{\theta^n} dt = g(\theta). \quad \theta > 0$

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

---

if $g(\theta) = \theta$. $\quad \frac{(n+1)}{n} T$ — UMVUE.

another unbiased estimator: $\delta = 2\overline{X}$.

$$E_\theta T^2 = \int_0^\theta t^2 \frac{n t^{n-1}}{\theta^n}\, dt = \frac{n}{n+2}\theta^2$$

$$E_\theta \eta^2(T) = \left(\frac{n+1}{n}\right)^2 E_\theta T^2 = \frac{(n+1)^2}{n(n+2)}\theta^2$$

$$Var(\eta(T)) = \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 = \frac{\theta^2}{n(n+2)}$$

$$Var(2\bar{X}) = \frac{\theta^2}{3n}.$$

$$\frac{Var(\eta(T))}{Var(\delta)} = \frac{3}{n+2}$$

Ex: $X_1 \cdots X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$

$T(x) = X_1 + \cdots + X_n$ —— complete and suff.

$T = T(X) \sim Bin(n, \theta)$

$g(\theta) = \theta^2$. $\quad E(X_1 X_2) = EX_1 EX_2 = \theta^2$

$$\eta(T) = E_\theta(X_1 X_2 | T) = P_\theta(X_1 = 1, X_2 = 1 | T) = \frac{P(X_1=1, X_2=1, T=t)}{P(T=t)}$$

$$= \frac{P(X_1=1)P(X_2=1)P\left(\sum_{i=3}^n X_i = t-2\right)}{P(T=t)} = \frac{\theta^2 \binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}$$

$$= \frac{t(t-1)}{n(n-1)}.$$

$\frac{T(T-1)}{n(n-1)}$ —— UMVUE.

- Is it reasonable that we only consider unbiased estimators?

ex. $X_1 \cdots X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$. $T = \max\{X_1 \cdots X_n\}$ complete suff.

$\frac{(n+1)}{n} T$ —— UMVUE.

$E_\theta T = \frac{n\theta}{n+1}$. $E T^2 = \frac{n\theta^2}{n+2}$.

$\delta_a = aT$. $R(\theta, \delta_a) = E_\theta (aT - \theta)^2 = a^2 E T^2 - 2a\theta E T + \theta^2$

$= \theta^2 \left(\frac{n}{n+2} a^2 - \frac{2n}{n+1} a + 1 \right)$

minimized when $a = \frac{n+2}{n+1}$. $R(\theta, \delta_a) = \frac{\theta^2}{(n+1)^2}$.

but $R(\theta, \frac{n+1}{n} T) = \frac{\theta^2}{n^2 + 2n}$.

$\frac{n+1}{n} T$ vs. $\frac{n+2}{n+1} T$

$$R(\theta, \delta) = E_\theta (\delta - g(\theta))^2 = \underbrace{\text{Var}_\theta(\delta)}_{\text{Var.}} + \underbrace{b^2(\theta, \delta)}_{\text{bias}}$$

$\frac{n+2}{n+1} T \longrightarrow$ smaller Var. larger bias.

ex. $$P(X=x) = \frac{\theta^x e^{-\theta}}{x!\,(1-e^{-\theta})} \qquad x = 1, 2, \cdots$$

Truncated Poisson to $\{1, 2, \cdots\}$. X — complete suff.

$g(\theta) = e^{-\theta}$ (the proportion lost through truncation).

$$e^{-\theta} = \sum_{k=1}^{\infty} \frac{\delta(k)\,\theta^k e^{-\theta}}{k!\,(1-e^{-\theta})}$$

$$\sum_{k=1}^{\infty} \frac{\delta(k)}{k!}\theta^k = 1 - e^{-\theta} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!}\theta^k. \qquad \theta > 0$$

$\delta(k) = (-1)^{k+1}$ UMVUE: $(-1)^{X+1}$ ← absurd!

Normal Theory.

$X \sim N(\mu, \sigma^2)$ $\quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$aX + b \sim N(a\mu + b,\ a^2\sigma^2)$ $\quad M_Z(t) = e^{\frac{t^2}{2}}$. $\quad M_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$

$X_1 \perp X_2 \Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.

$X_1 \cdots X_n \overset{iid}{\sim} N(\mu\ \sigma^2)$.

$\theta = (\mu, \sigma^2)$. $\quad T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$. complete suff.

$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ $\quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum X_i^2 - n\bar{X}^2 \right)$

$\bar{X} = \frac{T_1}{n}$. $\quad S^2 = \frac{T_2 - T_1^2/n}{n-1}$ $\quad$ or $\quad T_1 = n\bar{X}$. $\quad T_2 = (n-1)S^2 + n\bar{X}^2$.

$(\bar{X}, S^2)$ — complete suff.

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Gamma distribution: $\quad f_{\alpha,\beta}(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}$ $\quad x > 0$.

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

$\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$. $\quad \alpha > 0$. $\quad \Gamma(n+1) = n!$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

If $X \sim \Gamma(\alpha, 1)$. $\quad \beta X \sim \Gamma(\alpha, \beta)$.

β: scale parameter. $\quad \alpha$: shape parameter.

$X \sim \Gamma(\alpha, \beta)$ $\quad M_X(u) = Ee^{tu} = \int_0^\infty e^{ux} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx$

$= \frac{1}{(1-u\beta)^\alpha} \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dy = \frac{1}{(1-u\beta)^\alpha}$

$$X \sim \Gamma(\alpha_x, \beta), \quad Y \sim \Gamma(\alpha_y, \beta) \quad \text{independent} \implies X+Y \sim \Gamma(\alpha_x+\alpha_y, \beta)$$

The Chi-square distribution is a special case of Gamma.

$$M_{Z^2}(u) = \int e^{uz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{1-2u}}$$

$$Z^2 \sim \chi^2(1) := \Gamma(\tfrac{1}{2}, 2)$$

$$Z_1^2 + \cdots + Z_p^2 \sim \chi^2(p) = \Gamma(\tfrac{p}{2}, 2). \quad \text{mgf. } \frac{1}{(1-2u)^{p/2}} \quad u < \frac{1}{2}$$

$$Z_i = \frac{X_i - \mu}{\sigma} \qquad \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i-\mu}{\sigma} = \frac{\bar{X}-\mu}{\sigma} \sim N(0, \tfrac{1}{n})$$

$$n\bar{Z}^2 \sim \chi_1^2.$$

$$V \triangleq \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma} - \frac{\bar{X}-\mu}{\sigma}\right)^2$$

$$= \sum_{i=1}^{n}(Z_i - \bar{Z})^2 = \sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2$$

$$V + n\bar{Z}^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2.$$

$$\bar{X} \perp S^2. \qquad n\bar{Z}^2 \perp n\bar{Z}^2.$$

$$M_V(u)\, M_{n\bar{Z}}(u) = \frac{1}{(1-2u)^{n/2}}. \qquad M_V(u) = \frac{1}{(1-2u)^{\frac{n-1}{2}}}.$$

$$V \sim \chi_{n-1}^2.$$

Normal Estimation

$n+r>1$ $$Es^r = E\left(\frac{\sigma^r}{(n-1)^{\frac{r}{2}}} V^{\frac{r}{2}}\right) = \frac{\sigma^r}{(n-1)^{\frac{r}{2}}} \int_0^\infty \frac{x^{(r+n-3)/2} e^{-\frac{x}{2}}}{2^{(n-1)/2}\, \Gamma(\frac{n-1}{2})} dx$$

$$= \frac{\sigma^r 2^{r/2}\, \Gamma(\frac{r+n-1}{2})}{(n-1)^{r/2}\, \Gamma(\frac{n-1}{2})}.$$

$$\frac{(n-1)^{\frac{r}{2}}\, \Gamma(\frac{n-1}{2})}{2^{r/2}\, \Gamma(\frac{n+r-1}{2})} s^r$$ — unbiased estimate of σ^r.

UMVUE. function of $(\bar{x}, s^2)$

$r=2$: UMVUE S^2

$r=1$: UMVUE is not S.

By Stirling's formula,

$$\frac{(n-1)^{\frac{r}{2}}\, \Gamma(\frac{n-1}{2})}{2^{\frac{r}{2}}\, \Gamma(\frac{n+r-1}{2})} = 1 - \frac{r(r-2)}{4n} + O\left(\frac{1}{n^2}\right)$$

For large n, s^r has small bias.

$E\bar{X} = \mu$ $\bar{X}$ UMVUE for μ.

$E\bar{X}^2 = \mu^2 + \frac{\sigma^2}{n}$. $\bar{X}^2 - \frac{S^2}{n}$ — UMVUE for μ^2.

- $\frac{\mu}{\sigma}$: signal-to-noise ratio.

σ^{-1}: UMVUE: $$\frac{\sqrt{2}\,\Gamma(\frac{n-1}{2})}{S\sqrt{n-1}\,\Gamma(\frac{n-2}{2})}.$$

μ: UMVUE. $\bar{X}$.

independent: $\frac{\mu}{\sigma}$: UMVUE: $$\frac{\sqrt{2}\,\Gamma(\frac{n-1}{2})}{\sqrt{n-1}\,\Gamma(\frac{n-2}{2})}\frac{\bar{X}}{S}.$$

- pth quantile: $P(X_i \le x) = p$ $\quad \underline{\mu + \sigma\Phi^{-1}(p)}$.

UMVUE: $$\bar{X} + \frac{\sqrt{n-1}\,\Gamma(\frac{n-1}{2})}{\sqrt{2}\,\Gamma(\frac{n}{2})} S\,\Phi^{-1}(p).$$

Variance Bounds and Information.

- covariance inequality: $\mathrm{cov}^2(X,Y) \le \mathrm{Var}(X)\,\mathrm{Var}(Y)$

$$E\left((X-EX)\sigma_Y \pm (Y-EY)\sigma_X\right]^2 = 2\sigma_X\sigma_Y\left(\sigma_X\sigma_Y \pm \mathrm{cov}(X,Y)\right) \ge 0$$

δ — unbiased for $g(\theta)$.

$$\mathrm{Var}_\theta(\delta) \geq \frac{\mathrm{cov}_\theta^2(\delta, \psi)}{\mathrm{Var}_\theta(\psi)}.$$

choose ψ cleverly!

Thm (Cramér-Rao lower bound). $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$.

δ — unbiased for $g(\theta)$.

$$\mathrm{Var}_\theta(\delta) \geq \frac{g'(\theta)^2}{I(\theta)}.$$

$I(\theta)$. fisher information.

$$I(\theta) = E_\theta\left(\frac{\partial \log p_\theta(x)}{\partial \theta}\right)^2 = -E\left[\frac{\partial^2 \log p_\theta(x)}{\partial \theta^2}\right]$$

$$= \mathrm{Var}\left(\frac{\partial \log p_\theta(x)}{\partial \theta}\right).$$

$$0 = \frac{\partial}{\partial \theta} 1 = \frac{\partial}{\partial \theta} \int p_\theta(x)\, d\mu(x) = \int \frac{\partial}{\partial \theta} p_\theta(x)\, d\mu(x)$$

$$= \int \frac{\partial \log p_\theta(x)}{\partial \theta}\, p_\theta(x)\, d\mu(x) = E_\theta \frac{\partial \log p_\theta(x)}{\partial \theta}.$$

ex. exponential family:

$$P_\eta(x) = e^{\eta T(x) - A(\eta)} h(x).$$

$$\frac{\partial \log P_\eta(x)}{\partial \eta} = T - A'(\eta). \qquad I(\eta) = Var_\eta(T) = A''(\eta)$$

$$\frac{\partial^2 \log P_\eta(x)}{\partial \eta^2} = -A''(\eta).$$

ex. Location family: $P_\theta(x) = f(x-\theta)$.

$$I(\theta) = E_\theta\left(\frac{\partial \log f(x-\theta)}{\partial \theta}\right)^2 = E_\theta\left(-\frac{f'(x-\theta)}{f(x-\theta)}\right)^2 = \int \frac{f'^2}{f}$$

$$I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta). \qquad X, Y \text{ independent}$$

- Variance bounds in higher dimensions

Fisher information matrix:

$$(I(\theta))_{ij} = E_\theta\left(\frac{\partial \log P_\theta(x)}{\partial \theta_i} \frac{\partial \log P_\theta(x)}{\partial \theta_j}\right)$$

$$= Cov_\theta\left(\frac{\partial \log P_\theta(x)}{\partial \theta_i}, \frac{\partial \log P_\theta(x)}{\partial \theta_j}\right)$$

$$= -E_\theta\left(\frac{\partial^2 \log P_\theta(x)}{\partial \theta_i \partial \theta_j}\right)$$

$g(\theta)$: $\quad g: \Omega \to \mathbb{R}$.

$$\mathrm{Var}(\delta) \geq \nabla g(\theta)' \, I^{-1}(\theta) \, \nabla g(\theta)$$

ex: s-parameter exponential family:

$$p_\eta(x) = e^{\eta \cdot T(x) - A(\eta)} h(x)$$

$$\frac{\partial^2 \log p_\eta(x)}{\partial \eta_i \partial \eta_j} = - \frac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j} \qquad I(\eta) = \nabla^2 A(\eta).$$