Point Estimation

Feng-Chang Lin

Department of Biostatistics
University of North Carolina at Chapel Hill

flin@bios.unc.edu

(C&B §7)

Introduction

- Random sample $X_1, \cdots, X_n$ from $f(x|\theta)$, where θ is either a scalar or vector.
- We want to estimate θ or $\tau(\theta)$.
- **Example** If $X \sim N(\mu, \sigma^2)$, how do we estimate $\theta = (\mu, \sigma^2)$?
- **Example** If $X \sim N(\mu, \sigma^2)$, how do we estimate $\tau(\theta) = \mu/\sigma^2$?
- **Example** If $X \sim N(\mu, \sigma^2)$, how do we estimate $\tau(\theta) = P(X_1 > 100) = \Phi((100 - \mu)/\sigma)$?

Introduction (cont'd)

- *Point estimator*: Any function of the sample, a statistic, $W(X_1, \cdots, X_n)$, also simply called *estimator*. Specifically, an estimator can not be a function of θ. It must be a statistic.
- *Estimator*: The random variable $W(X_1, \cdots, X_n)$.
- *Estimate*: The realized value $W(x_1, \cdots, x_n)$.
- We want a good point estimator.
- How to find good estimators?
- What is a "good" estimator?

Method of Moments

- Match sample moments with population moments.
- Use as many sample moments as needed. Start with lower order moments first.
- The kth population moment: $\mu_k = EX_1^k$.
- The kth sample moment: $M_k = \frac{1}{n}\sum_{i=1}^n X_i^k$. What is M_1?
- Finding the moment estimator: Set $M_1 = \mu_1$, $M_2 = \mu_2$, $\cdots$, and solve for θ.
- The moment estimator will be denoted by $\hat{\theta}_{MM}$.
- **Example** $X_1, \cdots, X_n$ iid Bernoulli(θ), $\theta \in [0, 1]$. $M_1 = \mu_1$ gives $\hat{\theta}_{MM} = \bar{X}$.
- **Example** $X_1, \cdots, X_n$ iid $N(0, \theta)$, $M_1 = \mu_1 = 0$ is not usable. $M_2 = \mu_2 = \theta$ gives $\hat{\theta}_{MM} = \frac{1}{n}\sum_{i=1}^n X_i^2$.

Method of Moments (cont'd)

- **Example** $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$, both μ and σ^2 unknown.

$$M_1 = \mu, \text{ and } M_2 = \mu^2 + \sigma^2.$$

$$\hat{\mu}_{MM} = \bar{X}, \hat{\sigma}^2_{MM} = M_2 - M_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2.$$

- **Example** $X_1, \cdots, X_n$ iid binomial(m, p), both m and p unknown, $p \in [0, 1]$, $m \in \{0, 1, \cdots\}$.

$$M_1 = mp, M_2 = (mp)^2 + mp(1-p).$$

$$\frac{M_2}{M_1} - M_1 = 1 - p, \hat{p}_{MM} = 1 - \frac{M_2 - M_1^2}{M_1}, \hat{m}_{MM} = \frac{M_1}{\hat{p}_{MM}}.$$

- Negative $\hat{p}_{MM}$ and $\hat{m}_{MM}$ is possible. Out of range moment estimators are not rare in applications.

Maximum Likelihood

- The *likelihood function* is the joint pdf or pmf, but viewed as a function of θ with the sample x being fixed.
- If X is a random vector representing the observable data, then

$$L(\theta|x) = f(x|\theta).$$

- If $X_1, \cdots, X_n$ is a random sample from a pdf or pmf $f(x|\theta)$, then

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta),$$

with the log-likelihood function

$$\ell(\theta|x) = \log L(\theta|x) = \sum_{i=1}^{n} \log f(x_i|\theta).$$

Maximum Likelihood (cont'd)

- For a given sample x, the maximum likelihood estimator (MLE), denoted $\hat{\theta}(x)$ is a value of θ at which $L(\theta|x)$ attains its maximum over the parameter space.
- The abbreviation MLE is used for both maximum likelihood *estimator* and maximum likelihood *estimate*.
- If the range of x depends on θ, that dependence should be built into $L(\theta|x)$.
- **Example** $X_1, \cdots, X_n$ iid uniform on $[0, \theta]$.

$$L(\theta|x) = \theta^{-n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta) = \theta^{-n} I(x_{(n)} \leq \theta).$$

One has $\hat{\theta} = X_{(n)}$.

Maximum Likelihood (cont'd)

- Multiplied by a positive constant that does not involve the unknown parameters does not change the final answers.
- **Example** $X_1, \cdots, X_n$ iid Binomial(m, θ), with m known and $\theta \in [0, 1]$ unknown. The likelihood is

$$L(\theta|x) = \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i}(1-\theta)^{m-x_i} = C(x) \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{m-x_i},$$

  where $C(x)$ depends on x but not θ.
- Dropping $C(x)$ does not affect the maximization over θ.
- A value of θ that maximizes the log-likelihood $\ell(\theta|x)$ will also maximize the likelihood $L(\theta|x)$.

Maximum Likelihood (cont'd)

- There is no single simple procedure that is applicable to all types of problems for finding MLE.
- **Example** X is a single observation from the Binomial(m, θ), with unknown $\theta \in [0, 1]$, and known $m \geq 1$.

$$L(\theta|x) = \left(\begin{array}{c} m \\ x \end{array} \right) \theta^x (1-\theta)^{m-x}.$$

- If $x = 0$, the likelihood $L(\theta|0) = (1-\theta)^m$, which is monotone decreasing in θ. One would say $\hat{\theta} = 0$.
- If $x = m$, the likelihood $L(\theta|m) = \theta^m$, which is monotone increasing in θ. One would say $\hat{\theta} = 1$.
- If $0 < x < m$, the likelihood $L(\theta|x)$ is maximized at $\hat{\theta} = x/m$.
- In all cases, $\hat{\theta} = x/m$.

Maximum Likelihood (cont'd)

- **Example** Let $X_1, \cdots, X_n$ be iid random variables distributed as $N(\theta, 1)$, $\theta \in (-\infty, \infty)$.

$$L(\theta|x) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right\},$$

$$\ell(\theta|x) = (-n/2)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 - \frac{n}{2}(\bar{x} - \theta)^2,$$

- The log-likelihood is a quadratic function in θ that has a unique global maximum at $\theta = \bar{x}$, so $\hat{\theta} = \bar{x}$.

Maximum Likelihood (cont'd)

- **Example (restricted range)** Let $X_1, \cdots, X_n$ be iid random variables distributed as $N(\theta, 1)$, $\theta \in [0, \infty)$. If $\bar{x} \geq 0$, then $\bar{x}$ is the MLE. If $\bar{x} < 0$, the log-likelihood

$$\ell(\theta|x) = (-n/2)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 - \frac{n}{2}(\bar{x} - \theta)^2,$$

  will be monotone decreasing over $[0, \infty)$, hence its maximum will be at $\hat{\theta} = 0$.

- **Example (flat likelihood function)** $X_1, \cdots, X_n$ iid Uniform$(\theta - 1/2, \theta + 1/2)$.

$$L(\theta|x) = I(x_{(1)} > \theta - \frac{1}{2})I(x_{(n)} < \theta + \frac{1}{2})$$

  The likelihood $L(\theta|x) = 1$ over $\theta \in (x_{(n)} - 1/2, x_{(1)} + 1/2)$ and $L(\theta|x) = 0$ otherwise.

Maximum Likelihood (cont'd)

- **Discrete parameter, MLE of the binomial m with known p**
  Consider a single observation X from the Binomial(m, p), with p known and m unknown. We want to find the MLE of m.
- The parameter space is the set of integers $\{1, 2, \cdots\}$.
- Suppose that $p = 0.71$ and the observed value is $x = 7$. What is the MLE of m?
- Because $P(X = x|m) = 0$ if $x > m$, we get $L(m|x) = 0$ for $m < 7$ and $L(m|x) = \left(\begin{array}{c} m \\ x \end{array} \right) p^x(1-p)^{(m-x)}$ for integer $m \geq 7$.
- $L(m|x)$ is increasing for $7 \leq m \leq 9$ and decreasing for for $m \geq 9$. We can conclude that the MLE is $\hat{m} = 9$.
- Since $EX = mp$, the moment estimate is $\hat{m}_{MM} = x/p = 7/0.71 \approx 9.86$, which is not far from the MLE.

MLE for a 2-dimensional Parameter

- A two-dimensional parameter, and the likelihood is twice-differentiable.
- Use rules of calculus to find “local" maximum.
- The rules for a local maximum:
  a) Two first-order partial derivatives are zero.
  b) At least one second-order partial derivatives is negative.
  c) The Jacobian of the second-order partial derivatives is positive.
- **Example**: The $N(\mu, \sigma^2)$ model with both parameters unknown.

$$\ell(\mu, \sigma^2 | x) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 / \sigma^2.$$

Example (Normal Distribution)

$$\frac{\partial}{\partial \mu}\ell(\mu, \sigma^2|x) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{\partial}{\partial \sigma^2}\ell(\mu, \sigma^2|x) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial^2}{\partial \mu^2}\ell(\mu, \sigma^2|x) = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2}{\partial (\sigma^2)^2}\ell(\mu, \sigma^2|x) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial^2}{\partial \mu \partial \sigma^2}\ell(\mu, \sigma^2|x) = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu)$$

$$J(\hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}^6}\frac{n^2}{2} > 0.$$

Successive 1-dimensional Maximization

- To maximize $L(\alpha, \beta|x)$ over both α and β, we can proceed as follows.
- First, for a fixed α, we maximize $L(\alpha, \beta|x)$ over β.
- Let $\hat{\beta}(\alpha)$ be the value of β that maximizes $L(\alpha, \beta|x)$ for a fixed α.
- The function
$$H(\alpha|x) = L(\alpha, \hat{\beta}(\alpha)|x)$$
depends on α.
- We call this kind of function $H(\alpha|x)$ as the *profiled likelihood* for α.
- Then, the MLE of β is simply $\hat{\beta}(\hat{\alpha}_H)$, where $\hat{\alpha}_H$ is the maximizer of $H(\alpha|x)$.

Successive 1-dimensional Maximization (cont'd)

- **Example (MLE of the Weibull parameters)** Let $X_1, \cdots, X_n$ are iid Weibull(α, β) with density

$$f(x|\alpha,\beta) = \frac{\alpha}{\beta} x^{\alpha-1} \exp\left(-\frac{x^{\alpha}}{\beta}\right), x \geq 0, \alpha > 0, \beta > 0.$$

- The log-likelihood

$$\ell(\alpha,\beta|x) = n\log\alpha - n\log\beta + (\alpha-1)\sum_{i=1}^{n}\log x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i^{\alpha}.$$

- Maximized over β by setting the derivative

$$\frac{d}{d\beta}\ell(\alpha,\beta|x) = -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i^{\alpha} = 0.$$

- The solution is $\hat{\beta}(\alpha) = n^{-1}\sum_{i=1}^{n} x_i^{\alpha}$.

Successive 1-dimensional Maximization (cont'd)

- The solution is verified to be a maximum since

$$\frac{d^2}{d\beta^2}\ell(\alpha,\beta|x)\bigg|_{\beta=\hat{\beta}(\alpha)} = -\frac{n}{\hat{\beta}(\alpha)^2} < 0.$$

- The profile log-likelihood for α is

$$\begin{aligned} h(\alpha|x) &= \ell(\alpha,\hat{\beta}(\alpha)|x) \\ &= n\left\{\log\alpha - \log\frac{\sum_{i=1}^n x_i^\alpha}{n} + (\alpha-1)\frac{\sum_{i=1}^n \log x_i}{n} - 1\right\}. \end{aligned}$$

- There is no "closed" form for α. Maximization over α can be done either graphically or by numerical methods.

Invariance Property of MLE

Theorem (Theorem 7.2.10)

If $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

- If the mapping $\theta \to \tau(\theta)$ is one-to-one, then it is easy to see that the MLE of $\eta = \tau(\theta)$ is the same since

$$L^*(\eta|x) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) = L(\tau^{-1}(\eta)|x),$$

  and

$$\sup_\eta L^*(\eta|x) = \sup_\eta L(\tau^{-1}(\eta)|x) = \sup_\theta L(\theta|x).$$

- The proof is more complicated if the τ function is not one-to-one. Check p. 320 in C&B.

Invariance Property of MLE (cont'd)

- **Example** What is the MLE of θ^2 if $X_1, \cdots, X_n$ follows $N(\theta, \sigma^2)$?
- **Example** What is the MLE of $\sqrt{p(1-p)}$ if $X_1, \cdots, X_n$ follows Binomial(n, p)?

$$
\begin{aligned}
L(p|x) &= \prod_{i=1}^{n} \left(\begin{array}{c} n \\ x_i \end{array} \right) p^{x_i}(1-p)^{n-x_i} \\
&= C(x)p^{\sum_{i=1}^n x_i}(1-p)^{n^2-\sum_{i=1}^n x_i}.
\end{aligned}
$$

$$
\ell(p|x) \propto \sum_{i=1}^{n} x_i \log p + (n^2 - \sum_{i=1}^{n} x_i) \log(1-p).
$$

- How do we find the MLE of p?

Instability of MLE

- The MLE can be highly unstable if the likelihood function is very flat in the neighborhood of its maximum.
- **Example** $X_1, \cdots, X_5$ follows Binomial(n, p) with both n and p unknown.

$$\text{Sample 1: } (16, 18, 22, 25, 27) \Rightarrow \hat{n} = 99;$$
$$\text{Sample 2: } (16, 18, 22, 25, 28) \Rightarrow \hat{n} = 190.$$

- Even worse, there is no finite maximum. The MLE doesn't exist.

Method of Evaluating Estimators

- Bias: $\text{Bias}_\theta W(X) = E_\theta W(X) - \theta$.
- Variance: $Var_\theta W(X)$.
- Mean Squared Error (MSE): $E_\theta(W(X) - \theta)^2$=Bias^2 + Variance.
- Other: $E_\theta g(|W - \theta|)$.
- **Example**: Let $X_1, \cdots, X_n$ be iid N(μ, σ^2). Since,

$$E\bar{X} = \mu, \quad ES^2 = \sigma^2,$$

for all μ and σ^2. The MSE of these estimators are given by

$$E(\bar{X} - \mu)^2 = Var\bar{X} = \frac{\sigma^2}{n},$$

$$E(S^2 - \sigma^2)^2 = VarS^2 = \frac{2\sigma^4}{n-1}.$$

Method of Evaluating Estimators (cont'd)

- What is the MLE of μ and σ^2?

$$E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2,$$

$$Var(\hat{\sigma}^2) = Var\left(\frac{n-1}{n}S^2\right) = \frac{2(n-1)}{n^2}\sigma^4.$$

- The MSE of $\hat{\sigma}^2$ is given by

$$E(\hat{\sigma}^2 - \sigma^2)^2 = \frac{2(n-1)}{n^2}\sigma^4 + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4.$$

- We have

$$E(\hat{\sigma}^2 - \sigma^2)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = E(S^2 - \sigma^2)^2.$$

Best Unbiased Estimators

- Having a “biased" estimator may not be acceptable.
- Finding an estimator that minimizes MSE may not be reasonable for “scale" parameters.
- One can restrict their searching for the “best" estimator only form those “unbiased" estimators.
- **Uniformly Minimum Variance Unbiased Estimators (UMVUE)**: An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $E_\theta W = \tau(\theta)$, we have $Var_\theta W^* \leq Var_\theta W$ for all θ.
- W^* is called *uniformly minimum variance unbiased estimators* (UMVUE) of $\tau(\theta)$.
- “Uniformly" means the statement holds for all $\theta \in \Theta$.

Best Unbiased Estimators (cont'd)

- **Example** Let $X_1, \cdots, X_n$ be iid Poisson(λ), and let $\bar{X}$ and S^2 be the sample mean and variance, respectively. One has

$$E_\lambda \bar{X} = \lambda, \text{ and } E_\lambda S^2 = \lambda,$$

so both $\bar{X}$ and S^2 are unbiased estimators of λ. By linear combinations of $\bar{X}$ and S^2, we can create infinitely many unbiased estimators. Do we have the best one?

Cauchy-Schwarz Inequality

- For random variables X and Y,

$$Cov(X, Y) \leq \sqrt{Var(X)Var(Y)}.$$

or, equivalently,

$$VarX \geq \frac{\{Cov(X, Y)\}^2}{VarY}.$$

Cramér-Rao Lower Bound (CRLB)

- **Cramér-Rao Inequality** Let $X_1, \ldots, X_n$ be a sample with pdf $f(x|\theta)$, and let $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ be any unbiased estimator of $\tau(\theta)$ satisfying

$$\frac{d}{d\theta} E_\theta W(\boldsymbol{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}[W(\boldsymbol{x}) f(\boldsymbol{x}|\theta)] d\boldsymbol{x},$$

and

$$Var_\theta W(\boldsymbol{X}) < \infty.$$

Then

$$Var_\theta W(\boldsymbol{X}) \geq \frac{\{d\tau(\theta)/d\theta\}^2}{E_\theta \{U(\theta|\boldsymbol{x})\}^2},$$

where $U(\theta|\boldsymbol{x}) = \frac{\partial}{\partial\theta} \log f(\boldsymbol{x}|\theta)$ is called score function.

Cramér-Rao Lower Bound (cont'd)

- **Proof**: Note that,

$$
\begin{aligned}
\frac{d}{d\theta} E_\theta W(\boldsymbol{X}) &= \int_{\mathcal{X}} W(\boldsymbol{x}) \left\{ \frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta) \right\} d\boldsymbol{x} \\
&= E_\theta \left\{ W(\boldsymbol{X}) \frac{\frac{\partial}{\partial\theta} f(\boldsymbol{X}|\theta)}{f(\boldsymbol{X}|\theta)} \right\} \\
&= E_\theta \left\{ W(\boldsymbol{X}) \frac{\partial}{\partial\theta} \log f(\boldsymbol{X}|\theta) \right\}. \qquad (1)
\end{aligned}
$$

- If $W(\boldsymbol{X}) = 1$ in (1), one can have

$$
E_\theta \left\{ \frac{\partial}{\partial\theta} \log f(\boldsymbol{X}|\theta) \right\} = \frac{d}{d\theta} E_\theta(1) = 0.
$$

Cramér-Rao Lower Bound (cont'd)

- According to (1), we have

$$\begin{aligned} Cov_\theta\left\{W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right\} &= E_\theta\left\{W(\boldsymbol{X})\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right\} \\ &= \frac{d}{d\theta}E_\theta W(\boldsymbol{X}). \end{aligned}$$

- Also, we have

$$Var_\theta\left\{\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right\} = E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right\}^2\right].$$

- By the Cauchy-Schwarz Inequality, we have

$$Var_\theta W(\boldsymbol{X}) \geq \frac{\left\{\frac{d}{d\theta}E_\theta W(\boldsymbol{X})\right\}^2}{E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right\}^2\right]}.$$

Cramér-Rao Lower Bound (cont'd)

- If $X_1, \cdots, X_n$ are iid with pdf $f(x|\theta)$, then

$$Var_\theta W(\boldsymbol{X}) \geq \frac{\left\{\frac{d}{d\theta} E_\theta W(\boldsymbol{X})\right\}^2}{nE_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\}^2\right]}.$$

- If $f(x|\theta)$ satisfies

$$\frac{d}{d\theta} E_\theta\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\} = \int \frac{\partial}{\partial\theta}\left[\left\{\frac{\partial}{\partial\theta}\log f(x_1|\theta)\right\} f(x_1|\theta)\right] dx_1,$$

  then

$$E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\}^2\right] = -E_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log f(X_1|\theta)\right\}.$$

- Proof can be found in Exercise 7.39 of C&B.

Cramér-Rao Lower Bound (cont'd)

- In the Poisson example, $\tau(\lambda) = \lambda$ so $\tau'(\lambda) = 1$.
- One can show

$$
\begin{aligned}
E_\lambda\{U(\lambda|\boldsymbol{X})\}^2 &= -nE_\lambda\left\{\frac{\partial^2}{\partial\lambda^2}\log f(X_1|\lambda)\right\} \\
&= \frac{n}{\lambda}.
\end{aligned}
$$

- Hence for any unbiased estimator, W, of λ, we must have

$$
Var_\lambda W \geq \frac{\lambda}{n}.
$$

- Since $Var_\lambda \bar{X} = \lambda/n$, $\bar{X}$ is the best unbiased estimator of λ.

Violation of the Assumption in CRLB

- Let $X_1, \cdots, X_n$ be iid pdf $f(x|\theta) = 1/\theta$, $0 < x < \theta$. Since $\frac{\partial}{\partial\theta} \log f(x|\theta) = -1/\theta$. We have

$$E_\theta \left[\left\{ \frac{\partial}{\partial\theta} \log f(X_1|\theta) \right\}^2 \right] = \frac{1}{\theta^2}.$$

- The CRLB indicates $Var_\theta W \geq \theta^2/n$.
- However, $EX_{(n)} = \frac{n}{n+1}\theta$, and

$$Var_\theta \left(\frac{n+1}{n} X_{(n)} \right) = \frac{1}{n(n+2)}\theta^2 < \frac{1}{n}\theta^2.$$

- The problem is $\frac{d}{d\theta} \int_0^\theta h(x) f(x|\theta) dx \neq \int_0^\theta h(x) \frac{d}{d\theta} f(x|\theta) dx$.

Uniqueness of UMVUE

Theorem (7.3.19 in C&B)

If W is the best unbiased estimator of $\tau(\theta)$, then W is unique.

- Suppose that W' is another best unbiased estimator of $\tau(\theta)$, i.e., $Var(W') = Var(W)$.
- Take $W^* = (W + W')/2$; one can easily see $EW^* = \tau(\theta)$.
- Using covariance inequality, one can show $Var(W^*) \leq Var(W)$.
- However, since W is the best, $Var(W^*)$ can only equal $Var(W)$.
- When the equality stands, it implies that $W' = a + bW$.
- One can show that $a = 0$, $b = 1$, and $W' = W$.

Sufficiency and Unbiasedness

Theorem (Rao-Blackwell Theorem)

Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then $E_\theta \phi(T) = \tau(\theta)$ and $Var_\theta \phi(T) \leq Var_\theta W$ for all θ.

- **Proof** We have $\phi(T)$ as an unbiased estimator of $\tau(\theta)$ since

$$\tau(\theta) = E_\theta W = E_\theta\{E(W|T)\} = E_\theta \phi(T).$$

- Also,

$$\begin{aligned} Var_\theta W &= Var_\theta\{E(W|T)\} + E_\theta\{Var(W|T)\} \\ &= Var_\theta\{\phi(T)\} + E_\theta\{Var(W|T)\} \\ &\geq Var_\theta \phi(T). \end{aligned}$$

- We must show that $\phi(T) = E(W|T)$ is a function of only the sample and independent of θ (sufficiency!!).

Sufficiency/Completeness and Unbiasedness

Theorem (Lehmann-Sheffe Theorem)

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

- **Proof** Assume both W_1 and W_2 are unbiased estimator of $\tau(\theta)$.
- If we let $\phi_1(T) = E(W_1|T)$ and $\phi_2(T) = E(W_2|T)$, then

$$E\{\phi_1(T) - \phi_2(T)\} = E(W_1) - E(W_2) = 0.$$

- By the definition of completeness, $\phi_1 - \phi_2$ is a zero function.
- Hence $\phi_1(T) = \phi_2(T)$ (uniqueness).

Find UMVUE

- Method 1:
  - Find an unbiased estimator W for $\tau(\theta)$.
  - Look for a complete sufficient statistic for θ.
  - Derive $\phi(t) = E(W|T = t)$.
  - Then $\phi(T)$ is the UMVUE of $\tau(\theta)$.
- Method 2:
  - **Theorem 7.3.23** Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.
  - Adjusting a complete sufficient statistic to be unbiased gives the UMVUE.

Find UMVUE (cont'd)

Example Assume $X_1, \cdots, X_n$ are iid and follow Poisson(θ).

(1) Show that $I(X_1 = 0)$ is an unbiased estimator for $e^{-\theta}$.

(2) Find UMVUE for $e^{-\theta}$.

Solution

- Since $E\{I(X_1 = 0)\} = P(X_1 = 0) = e^{-\theta}$, $I(X_1 = 0)$ is an unbiased estimator for $e^{-\theta}$.
- Since the Poisson distribution belongs to an exponential family, $\sum_{i=1}^n X_i$ is a complete sufficient statistic.
- By the Lehmann-Scheffe Theorem, we know

$$\phi\left(\sum_{i=1}^n X_i\right) = E\left\{I(X_1 = 0) | \sum_{i=1}^n X_i\right\}$$

is the UMVUE for $e^{-\theta}$.

Find UMVUE (cont'd)

$$
\begin{aligned}
\phi(t) &= E\left\{ I(X_1 = 0) | \sum_{i=1}^{n} X_i = t \right\} = P\left(X_1 = 0 | \sum_{i=1}^{n} X_i = t \right) \\
&= \frac{P\left(X_1 = 0, \sum_{i=1}^{n} X_i = t\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} = \frac{P\left(X_1 = 0\right) P\left(\sum_{i=2}^{n} X_i = t\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \left(1 - \frac{1}{n}\right)^t .
\end{aligned}
$$

- One can conclude $\phi(\sum_{i=1}^{n} X_i) = (1 - 1/n)^{\sum_{i=1}^{n} X_i}$ is the UMVUE for $e^{-\theta}$.
- What is the MLE for $e^{-\theta}$?
- What does the $\phi(\sum_{i=1}^{n} X_i)$ converge to when $n \to \infty$?