Unbiasedness

Sudheesh Kumar Kattumannil

University of Hyderabad

1 Introduction

Assume that the data $X = (X_1, ..., X_n)$ comes from a probability distribution $f(x|\theta)$, with θ unknown. Once the data is available our task is to find (estimate) the value of θ. That is we need to construct good estimators for θ or its function $g(\theta)$. Then an important questions in this point, How to measure/evaluate the closeness of the estimator we obtained? How to find the best possible value? What is the best? We shall try to answer this questions by introducing some properties of the estimator. This include unbiasedness, consistency and efficiency.

2 Unbiasedness

After finding point estimator we are interested to develop criteria to compare different point estimators. One such measure is the mean square error (MSE) of an estimator. MSE of an estimator T of θ is defined as

$$MSE(T) = E(T - \theta)^2.$$

The MSE is a function of θ and has the interpretation in terms of variance and bias

$$MSE(T) = Var(T) + (B_T(\theta))^2,$$

where $B_T(\theta)$ is the bias given by $B_T(\theta) = E(T) - \theta$. Hence one may interested to minimize both the bias (inaccuracy) and variance (precision) of an estimator. When bias is is equal to zero we say that the estimator is unbiased. Then MSE become variance and we need to minimize the variance. Next we are dealing with unbiased estimator and our task is then to find the minimum variance unbiased estimator

Definition 1 *A statistic $T(X)$ is said to be an unbiased estimator of $g(\theta)$, a function of the parameter θ, if*

$$E(T(X)) = g(\theta) \quad \text{for} \quad \text{all} \quad \theta \in \Theta.$$

As mentioned above, Any estimator that not unbiased is called biased. Clearly the bias is the difference $Bias(\theta) = E(T(X)) - g(\theta)$.

When used repeatedly, an unbiased estimator in the long run will estimate the right value "on the average".

Example 1 Let $X_1, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$, then show that Show $\bar{X}$ is unbiased for θ and $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k = \bar{X})^2$ is unbiased for σ^2, and compute their MSEs. Show the estimator $S^{*2} = \frac{1}{n}\sum_{k=1}^{n}(X_k = \bar{X})^2$ is biased for σ^2, but it has a smaller MSE than S^2.

Note that unbiased estimators of $g(\theta)$ may not exist.

Example 2 Let X be a distributed according to the binomial distribution $B(n, p)$ and suppose that $g(p) = 1/p$. Then, unbiasedness of an estimator $T(x)$ of $g(\theta)$ should satisfy

$$\sum_{k=0}^{n} T(k)\binom{n}{k}p^k(1-p)^{(n-k)} \quad \text{for} \quad \text{all} \quad 0 < p < 1.$$

It can be seen that as $p \to 0$, the left side tends to $T(0)$ and the right side to ∞. Hence no such $T(x)$ exist.

If there exists an unbiased estimator of g, the estimand g will be called U-estimable.

Although unbiasedness is an attractive condition, after an unbiased estimator has been found, its performance should be investigated. That is, we need to find an estimator that has smaller variance among all unbiased estimator.

Definition 2 *An unbiased estimator $T(x)$ of $g(\theta)$ is the uniform minimum variance unbiased (UMVU) estimator of $g(\theta)$ if*

$$Var(T(X) \leq Var(T^{'}(X)) \quad \text{for} \quad \text{all} \quad \theta \in \Theta.$$

where $T^{'}(x)$ is any other unbiased estimator of $g(\theta)$. The estimator $T(x)$ is locally minimum variance unbiased (LMVU) at $\theta = \theta_0$ if $Var(T(X)) \leq Var(T^{'}(X))$ for any other unbiased estimator $T^{'}(x)$.

How To Find The Best Unbiased Estimator? There are different way in finding UMVU estimator. The relationship of unbiased estimators of $g(\theta)$ with unbiased estimators of zero can be helpful in characterizing and determining UMVU estimators when they exist. The following theorem is in that direction.

Theorem 1 *Let X have distribution $\{f_\theta(x), \theta \in \Theta$, let $T(x)$ be an unbiased estimator of $g(\theta)$, and let U denote the set of all unbiased estimators of zero. Then, a necessary and sufficient condition for T to be a UMVU estimator of its expectation $g(\theta)$ is that*

$$E(TU) = 0 \quad \text{for} \quad \text{all} \quad \mathrm{U} \in \mathrm{U} \quad \text{and} \quad \text{all} \quad \theta \in \Theta.$$

The proof was discussed in the class.

Theorem 2 *Let X be distributed according to a distribution $\{f_\theta(x), \theta \in \Theta$, suppose that $T(x)$ is a complete sufficient statistic for the same family, then for every U-estimable function $g(\theta)$ has one and only one unbiased estimator that is a function of T. Moreover it uniformly minimizes the variance; therefore, this estimator in particular is UMVU.*

Proof: If T_1 and T_2 are two unbiased estimators of $g(\theta)$, their difference $f(T) = T_1 - T_2$ satisfies

$$E(f(T)) = 0 \quad for \quad all \quad \theta \in \Theta,$$

and hence by the completeness of T , $T_1 = T_1$ a.e., hence the uniqueness. Now by Lehman-Sheffe' theorem (will discuss later), T is UMVU estimator for θ.

Next we will discuss the method for deriving UMVU estimators

If T is a complete sufficient statistic, the UMVU estimator of any U-estimable function $g(\theta)$ is uniquely determined by the set of equations

$$E(f(T)) = g(\theta) \quad for \quad all \quad \theta \in \Omega$$

Example 3 Suppose that T has the binomial distribution $b(p, n)$ and that $g(p) = p(1-p)$. Then, the above equation will becomes

$$\sum_{k=0}^{n} \binom{k}{n} T(k) p^k (1-p)^{n-k} = p(1-p) \quad for \quad all \quad 0 < p < 1.$$

If $\rho = p/q$ so that $p = \rho/(1+\rho)$ and $(1-p) = 1/(1+\rho)$, and the above equation can be rewritten as

$$\sum_{k=0}^{n} \binom{n}{k} T(k) \rho k = \rho(1+\rho)^{n-2} = \sum_{k=0}^{n-1} \binom{n-2}{k-1} (\rho)^k \quad 0 < \rho < \infty.$$

Comparing coefficient on the left and right sides leads to

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

The following theorem will help us to find UMVU estimator.

Theorem 3 *(Rao-Blackwell Theorem) Let* $\{F_\theta, \theta \in \Theta\}$ *be a family of probability distributions and* h *be any statistics in* U*, where* U *is the all unbiased estimate of* θ *with* $E(h^2 < \infty$*. Let* T *be a sufficient statistics for* $\{F_\theta, \theta \in \Theta\}$*. Then the conditional expectation* $E_\theta(h|T)$ *is independent* *of θ and is unbiased estimators of θ. Moreover*

$$E(E_\theta(h|T)-\theta)^2 = E_\theta(h-\theta)^2 \quad for \quad all \quad \theta \in \Theta.$$

The equality holds if and only if $E_\theta(h|T) = h$.

Theorem 4 *(Lehmann-Sheffe' Theorem) If T is a complete sufficient statistics and there exist an unbiased estimator h of θ, then there exist a unique UMVU estimator of θ, which is given by $E(h|T)$.*

Proof: If $h_1, h_2 \in U$, then $E_\theta(h_1|T)$ and $E_\theta(h_2|T)$ are both unbiased and

$$E(E_\theta(h_1|T) - E_\theta(h_2|T)) = 0 \quad for \quad all \quad \theta \in \Theta.$$

Since T is a complete sufficient statistics, we have

$$E_\theta(h_1|T) = E_\theta(h_2|T).$$

Hence by Theorem 3 $E(h|T)$ is the UMVU estimator.

Example 4 Suppose that $X_1, ..., X_n$ are iid according to the uniform distribution $U(0, \theta)$ and that $g(\theta) = \theta/2$. Then, $T = X_{(n)}$, the largest of the Xs, is a complete sufficient statistic. Since $E(X_1) = \theta/2$ and $T = X_{(n)}$ is a complete sufficient statistic, by Lehmann-Sheffe the UMVU estimator of $\theta/2$ is $t = E[X_1|X(n)]$. Hence we have to evaluate $E[X_1|X(n)]$.

If $X_{(n)} = t$, then $X_1 = t$ with probability $1/n$, and X is uniformly distributed on $(0, t)$ with the remaining probability $(n-1)/n$. Hence

$$E[X_1|t] = \frac{1}{n}t + \frac{n-1}{n}\frac{t}{2} = \frac{n+1}{n}\frac{t}{2}.$$

Thus, $[(n+1)/n]T/2$ and $[(n+1)/n]T$ are the UMVU estimators of $\theta/2$ and θ, respectively. Now we prove a theorem which was already discussed in connection with sufficiency.

Theorem 5 *A complete sufficient statistic is minimal sufficient.*

Proof: Let T be a complete sufficient statistics for the family $\{F_\theta\}$ and S be any sufficient statistics for which $E(h(X))$ is finite. Denote $h(T) = E(S|T)$, then by Lehmann-Sheffe Theorem h is UMVU estimator of $E(S)$. Consider any other sufficient statistics T_1. Then we show that $h(T)$ is a function of T_1. Suppose that $h(T)$ is not a function of T_1, then the function defined by $h_1(T_1) = E(h(T)|T_1)$ is unbiased for $E(S)$. And by Rao-Blackwell theorem,

$$V(h_1(T_1)) \leq V(h(T)),$$

which is contradiction to the fact that $h(T)$ is UMVU estimator for θ. Hence $h(T)$ is not a function of T_1. Since h and T_1 be arbitrary, T must be a function of every sufficient statistic, hence the proof.

Remark: In view of Lehmann-Sheffe' theorem once we have a complete sufficient statistics T which is unbiased for θ, then T will be UMVU for θ.

3 Lower bound to the Variance

In this section, we will discuss the attainment of a lower bound to the variance of an unbiased estimator. As a prerequisite we shall discuss about Fisher Information.

Definition 3 *Information(Fisher information) that X contains about the parameter θ denoted by $I(\theta)$ is defined as*

$$I(\theta) = E\Big(\frac{d}{d\theta} logp(X, \theta)\Big)^2,$$

where $s(X, \theta) = \frac{d}{d\theta} logp(X, \theta)$ is called the score or score function for X.

Clearly $I(\theta)$ is the average of the square of relative rate at which the density f_θ changes at x. It is plausible that the greater this expectation is at a given value θ_0, the easier it is to distinguish θ_0 from neighboring values θ, and, therefore, the more accurately θ can be estimated at $\theta = \theta_0$.

Remark: If $X_1, ..., X_n$ is a random sample with the pdf $f(x|\theta)$. Then the score for the entire sample $X_1, , X_n$ is

$$s_n(X, \theta) = \sum_{k=1}^{x} s(X_k, \theta).$$

And the Fisher information for the entire sample $X_1, ..., X_n$ is

$$I_n(\theta) = I(\theta).$$

It is important to realize that $I(\theta)$ depends on the particular parametrization chosen. In fact, if $\eta = h(\theta)$ and h is differentiable, the information that X contains about θ is

$$I(\theta) = I(h(\theta))(h^{'}(\theta))^2.$$

The proof of the alternative expressions of $I(\theta)$ given below will discuss in the class.

$$I(\theta) = E(-\frac{d^2}{d\theta^2} logp_\theta(X))$$

$$I(\theta) = E(\frac{d}{d\theta} logh_\theta(X))^2$$

$$I(\theta) = E(\frac{d}{d\theta} logr_\theta(X))^2,$$

where $h(x) = f(x)/(1 - F(x))$ and $r(x) = f(x)/F(x)$, the hazard rate and reversed hazard rate of X respectively.

Theorem 6 *Suppose that the distribution of X belongs to exponential family given by*

$$f(x|\theta) = h(x)exp\Big[T(x)s(\theta) - c(\theta)\Big],$$

and let $g(\theta) = E(T)$, then $T(x)$ satisfies

$$I(g(\theta)) = \frac{1}{V_\theta(T)}.$$

And for any differentiable function $h(\theta)$, we have

$$I(h(\theta)) = \left[\frac{s^{'}(\theta)}{h^{'}(\theta)}\right]^2 V_{\theta}(T).$$

Example 5 Let $X \sim Gamma(\alpha, \beta)$, where we assume that α is known. The density is given by

$$f(x) = \frac{1}{\Gamma\alpha\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} = e^{(\frac{-1}{\beta})x - \alpha\log(\beta)} h(x),$$

where $h(x) = x^{\alpha-1}/\Gamma\alpha$. We know that $E(T) = \alpha\beta$ and $V(T) = \alpha\beta^2$ and the information in X (using the Theorem 6) about $\alpha\beta$ is $I(\alpha\beta) = 1/\alpha\beta^2$.

If we are instead interested in the information in X about β, then we can reparameterize the density using $\eta(\beta) = -\alpha/\beta$ and $T(x) = x/\alpha$. And using

$$I(\theta) = I(h(\theta))(h^{'}(\theta))^2,$$

we have, quite generally, that $I[ch(\theta)] = \frac{1}{c2} I[h(\theta)]$, so the information in X about β is $I(\beta) = \alpha/\beta^2$.

Remark: Suppose that X is the whole data and $T = T(X)$ is some statistic. Then, $I_X(\theta) \geq I_T(\theta) \quad \text{for} \quad \text{all} \quad \theta \in \Theta$. The two information measures match with each other for all θ if and only if T is a sufficient statistic for θ.

Theorem 7 *(Cramer-Rao Lower bound) Let $X_1, ..., X_n$ be a sample with the joint pdf $f(x|\theta)$. Suppose $T(X)$ is an estimator satisfying (i) $E_{\theta}(T(X)) = g(\theta)$ for any $\theta \in \Theta$; and (ii) $Var_{\theta}(T(X)) < \infty$. If the following equation (interchangeability)*

$$\frac{d}{d\theta}\int_{\chi} h(x) f(x|\theta) dx = \int_{\chi} h(x) \frac{d}{d\theta} f(x|\theta) dx$$

holds for $h(x) = 1$ and $h(x) = T(x)$. Then

$$V(T(X)) \geq \frac{(g^{'}(\theta))^2}{I(\theta)}.$$

Theorem 8 *(Cramer-Rao Lower bound iid case) Let $X_1, ..., X_n$ be a iid with common pdf $f(x|\theta)$. Suppose $T(X)$ is an estimator satisfying all the conditions stated in Theorem 7, then*

$$V(T(X)) \geq \frac{(g^{'}(\theta))^2}{nI(\theta)}.$$

Theorem 9 *Let $X_1, ..., X_n$ be a iid with common pdf $f(x|\theta)$. Suppose that all the conditions stated in the Theorem 7 is satisfied, then the equality hold if and only if*

$$k(\theta)(T(x) - g(\theta)) = \frac{d}{d\theta} \log f_\theta,$$

for some function $k(\theta)$.

Another popular lower bound to the variance of an unbiased estimator is provided by the Chapman- Robbins inequality and it has better performance than the Cramer-rao inequality in the non-regular case as the later fails in that case. Next we discuss about it.

Theorem 10 *(Chapman-Robbins Inequality) Let $E(T(X)) = g(\theta)$, $\theta \in \Theta$ with $E(T(X))^2 < \infty$. If $\theta \neq \varphi$, assume that f_θ and f_φ are distinct and further assume that there exist $\theta \neq \varphi$ such that $\{f_\theta(x) > 0\} \supset \{f_\varphi(x) > 0\}$. Then*

$$V(T(X)) \geq \sup_{\varphi} \frac{(g(\varphi) - g(\theta))^2}{V(f_\varphi(X)/f_\theta(X))}.$$

The following example verifies the Theorem 10.

Example 6 Let X be distributed according to $U(0, \theta)$. Clearly the regularity condition of Cramer-Rao inequality does not hold in this case. We illustrate that the variance of the UMVU estimator for θ is greater that the bound provided by Chapman-Robbins inequality. Let $g(\theta) = \theta$. If $\varphi < \theta$, then $\{f_\theta(x) > 0\} \supset \{f_\varphi(x) > 0\}$. Consider

$$E(f_\varphi(X)/f_\theta(X))^2 = \int_0^\varphi (\frac{\theta}{\varphi})^2 \frac{1}{\theta} = \frac{\theta}{\varphi}.$$

and

$$E(f_\varphi(X)/f_\theta(X))^2 = \int_0^\varphi (\frac{\theta}{\varphi})\frac{1}{\theta} = 1.$$

Hence

$$V(f_\varphi(X)/f_\theta(X)) = \frac{\theta - \varphi}{\varphi}.$$

Thus by Chapmann-Robbins inequality

$$V(T(X)) \geq \sup_{\varphi;\varphi<\theta} \frac{(\theta-\varphi)^2}{(\theta-\varphi)\varphi} = \sup_{\varphi;\varphi<\theta} (\theta-\varphi)\varphi = \frac{\theta^2}{4},$$

for any unbiased estimator $T(X)$ of θ. Clearly $2X$ is unbiased for θ. Also it is UMVU estimator for θ as X is complete sufficient statistics. Consider

$$V(2X) = 4V(X) = \frac{\theta^2}{3},$$

which is greater than $\frac{\theta^2}{4}$, the bound provided by Chapmann-Robbins inequality.