Lecture Note 8 *
Point Estimators and Point Estimation Methods

MIT 14.30 Spring 2006

Herman Bennett

Given a parameter with unknown value, the goal of point estimation is to use a sample to compute a number that represents in some sense a good guess for the true value of the parameter.

19 Definitions

19.1 Parameter

A pmf/pdf can be equivalently written as $f_X(x)$ or $f_X(x|\theta)$, where θ represents the constants that fully define the distribution. For example, if X is a normally distributed RV, the constants μ and σ will fully define the distribution. These constants are called parameters and are generally denoted by the Greek letter θ. [1]

Example 19.1.

– Normal distribution: $f(x|\theta) \equiv f(x|\mu,\sigma)$, 2 parameters: $\theta_1 = \mu$ and $\theta_2 = \sigma$;

– Binomial distribution: $f(x|\theta) \equiv f(x|n,p)$, 2 parameters: $\theta_1 = n$ and $\theta_2 = p$;

– Poisson distribution: $f(x|\theta) \equiv f(x|\lambda)$, 1 parameter: $\theta = \lambda$;

– Gamma distribution: $f(x|\theta) \equiv f(x|\alpha,\beta)$, 2 parameters: $\theta_1 = \alpha$ and $\theta_2 = \beta$.

19.2 (Point) Estimator

A (point) estimator of θ, denoted by $\hat{\theta}$, is a statistic (a function of the random sample):

$$\hat{\theta} = r(X_1, X_2, ..., X_n). \tag{63}$$

*Caution: These notes are not necessarily self-explanatory notes. They are to be used as a complement to (and not as a substitute for) the lectures.

[1] Another way of saying this: parameters are constants that index a family of distributions.

- Note that $\hat{\theta}$ does not depend directly on θ, but only indirectly through the random process of each X_i.

A (point) estimate of θ is a realization of the estimator $\hat{\theta}$ (*i.e.*: a function of the realization of the random sample):

$$\hat{\theta} = r(x_1, x_2, ..., x_n). \tag{64}$$

Example 19.2. Assume we have a random sample $X_1, ..., X_{10}$ from a normal distribution $N(\mu, \sigma^2)$ and we want to have an estimate of the parameter μ (which is unknown). There is an infinite number of estimators of μ that we could construct. In fact, any function of the random sample could classify as an estimator of μ, for example:

$$\hat{\theta} = r(X_1, X_2, ..., X_{10}) = \begin{cases} X_{10} \\ 2 \\ \frac{X_{10}+X_1}{2} \\ \bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \\ 1.5X_2 \\ \text{etc.} \end{cases}$$

Example 19.3. Assume a random sample $X_1, ..., X_n$ from a U$[0, \theta]$ distribution, where θ is unknown. Propose 3 different estimators.

20 Evaluating (Point) Estimators

Since there are many possible estimators, we need to define some characteristic in order to evaluate and rank them.

20.1 Unbiasedness

An estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ, if for every possible value of θ:

$$E(\hat{\theta}) = \theta \tag{65}$$

If $\hat{\theta}$ is not unbiased, it is said to be a biased estimator, where the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$.

- If the random sample $X_1, ..., X_n$ is *iid* with $E(X_i) = \theta$, then the sample mean estimator

$$\hat{\theta} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{66}$$

is an unbiased estimator of the population mean: $E(\bar{X}_n) = \theta$ (see Example 18.1 in Lecture Note 7).

- If the random sample $X_1, ..., X_n$ is *iid* with $E(X_i) = \mu$ and $Var(X_i) = \theta$, then the sample variance estimator

$$\hat{\theta} = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \tag{67}$$

is an unbiased estimator of the population variance: $E(S^2) = \theta$ (see Example 18.1 in Lecture Note 7).

Example 20.1. Let $X_i \sim$ U $[0,\ \theta]$. Assume a random sample of size n and define an estimator of θ as:

$$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i. \qquad \text{Is } \hat{\theta} \text{ biased?}$$

20.2 Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$, if for a given sample size n,

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2) \tag{68}$$

Where $Var(\hat{\theta}_i)$ is the variance of the estimator.

Let $\hat{\theta}_1$ be an unbiased estimator of θ. $\hat{\theta}_1$ is said to be efficient, or minimum variance unbiased estimator, if for any unbiased estimators of θ, $\hat{\theta}_k$,

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_k) \tag{69}$$

- Do not confuse the variance of the estimator $\hat{\theta}$, $Var(\hat{\theta})$, with the sample variance estimator S^2, which is an unbiased estimator of the population variance σ^2(!).

Example 20.2. How would you compare the efficiency of the estimators in Example 19.2? Which of these estimators is unbiased?

20.3 Mean Squared Error

Why restrict ourselves to the class of unbiased estimators? The mean square error (MSE) specifies for each estimator $\hat{\theta}$ a trade off between bias and efficiency.

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + (bias(\hat{\theta}))^2 \tag{70}$$

$\hat{\theta}$ is the minimum mean square error estimator of θ if, among all possible estimators of θ, it has the smallest MSE for a given sample size n.

Example 20.3. Graph the pdf of two estimators such that the bias of the first estimator is less of a problem than inefficiency (and vice versa for the other estimator).

20.4 Asymptotic Criteria

20.4.1 Consistency

Let $\hat{\theta}$ be an estimator of θ. $\hat{\theta}$ is is said to be consistent if $\hat{\theta} \xrightarrow{p} \theta$ (Law of Large Numbers; Lecture Note 7).

Example 20.4. Assume a random sample of size n from a population $f(x)$, where $E(X) = \mu$ (unknown). Which of the following estimators is consistent?

$$\hat{\mu_1} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\mu_2} = \frac{1}{n-5}\sum_{i=1}^{n} X_i \qquad \hat{\mu_3} = \frac{1}{n-5}\sum_{i=1}^{n-5} X_i$$

- MSE $\to 0$ as $n \to \infty \Longrightarrow$ consistency.

20.4.2 Asymptotic Efficiency

Let $\hat{\theta}_1$ be an estimator of θ. $\hat{\theta}_1$ is is said to asymptotically efficient if it satisfies the definition of an efficient estimator, equation (69), when $n \to \infty$.

21 Point Estimation Methods

The following are two standard methods used to construct (point) estimators.

21.1 Method of Moments (MM)

Let $X_1, X_2, ..., X_n$ be a random sample from a population with pmf/pdf $f(x|\theta_1, ..., \theta_k)$, where $\theta_1, ..., \theta_k$ are unknown parameters. One way to estimate these parameters is by equating the first k population moments to the corresponding k sample moments. The resulting k estimators are called <u>method of moments (MM) estimators</u> of the parameters $\theta_1, ..., \theta_k$. [2]

The procedure is summarized as follows:

System of equations:
k equations and
k unknowns
$\Longrightarrow$ yields $\hat{\theta}_1, ..., \hat{\theta}_k$

Note that
i) $E(X_i^j) = g_j(\theta_1, ..., \theta_k)$
ii) the realization of the sample moment is a scalar.

$$\left\{ \begin{array}{cccc}
\textit{Sample moment} & & \begin{array}{c}\textit{Population moment} \\ (\textit{theoretical})\end{array} & \\
\frac{1}{n}\sum_{i=1}^{n} X_i & = & E(X_i^1) & \textit{First moment} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 & = & E(X_i^2) & \textit{Second moment} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^3 & = & E(X_i^3) & \textit{Third moment} \\
\vdots & \vdots & \vdots & \\
\frac{1}{n}\sum_{i=1}^{n} X_i^k & = & E(X_i^k) & k^{th} \textit{ moment}
\end{array} \right.$$

[2] This estimation method was introduced by Karl Pearson in 1894.

Example 21.1. Assume a random sample of size n from a $N(\mu, \sigma^2)$ population, where μ and σ^2 are unknown parameters. Find the MM estimator of both parameters.

Example 21.2. Assume a random sample of size n from a gamma distribution: $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ for $0 < x < \infty$. Assume also that the random sample realization is characterized by $\frac{1}{n}\sum_{i=1}^{n} x_i = 7.29$ and $\frac{1}{n}\sum_{i=1}^{n} x_i^2 = 85.59$. Find the MM estimators of α and β (parameters). Remember that $E(X_i) = \alpha\beta$ and $\text{Var}(X_i) = \alpha\beta^2$.

21.2 Maximum Likelihood Estimation (MLE)

Let $X_1, X_2, ..., X_n$ be a random sample from a population with pmf/pdf $f(x|\theta_1, ..., \theta_k)$, where $\theta_1, ..., \theta_k$ are unknown parameters. Another way to estimate these parameters is finding the values of $\hat{\theta_1}, ..., \hat{\theta_k}$ that maximize the likelihood that the observed sample is generated by $f(x|\hat{\theta_1}, ..., \hat{\theta_k})$.

The joint pmf/pdf of the random sample, $f(x_1, x_2, ..., x_n|\theta_1, ..., \theta_k)$, is called the likelihood function, and it is denoted by $L(\theta|\mathbf{x})$.

$$L(\theta|\mathbf{x}) = L(\theta_1, ..., \theta_k|x_1, ..., x_n) = \begin{cases} f(x_1, ..., x_n|\theta_1, ..., \theta_k) & \text{generally} \\ \\ \prod_{i=1}^{n} f(x_i|\theta_1, ..., \theta_k) & \text{random sample } (iid) \end{cases}$$

Where θ and $\mathbf{x}$ are vectors such that $\theta = (\theta_1, ..., \theta_k)$ and $\mathbf{x} = (x_1, ..., x_n)$.

For a given sample vector $\mathbf{x}$, denote $\hat{\theta}_{MLE}(\mathbf{x})$ the parameter value of θ at which $L(\theta|\mathbf{x})$ attains its **maximum**. Then, $\hat{\theta}_{MLE}(\mathbf{x})$ is the maximum likelihood estimator (MLE) of the (unknown) parameters $\theta_1, ..., \theta_k$. [3]

- Intuition: Discrete case.

[3] This estimation method was introduced by R.A. Fisher in the 1912.

- Does a global maximum exist? Can we find it? Unique?...Back to Calculus 101:

$$foc: \ \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_i} = 0, \quad i = 1, ..k \qquad \text{(for a well behaved function.)}$$

You also need to check that it is really a maximum and not a minimum (2^{nd} derivative).

- Many times it is easier to look for the maximum of $LnL(\theta|\mathbf{x})$ (same maximum since is a monotonic transformation...back to Calculus 101).

- Invariance property of MLE: $\hat{\tau}_{MLE}(\theta) = \tau(\hat{\theta}_{MLE})$.

- For large samples MLE yields an excellent estimator of θ (consistent and asymptotically efficient). No wonder it is widely used.

- But...

i) Numerical sensitivity (robustness).
ii) Not necessary unbiased.
iii) Might be hard to compute.

Example 21.3. Assume a random sample from a $N(\mu, \sigma^2)$ population, where the parameters are unknown. Find the ML estimator of both parameters.

Example 21.4. Assume a random sample from a $\mathrm{U}(0, \theta)$ population, where θ is unknown. Find $\hat{\theta}_{MLE}$.