Chapter 7: Point Estimation

Wang Shujia

Contents

1 Properties of Point Estimators

Definitions

Estimator: Let $X_1,\ X_2,\ \ldots,\ X_n$ be an iid sample from a population $X \sim f(x|\theta)$, then a point *estimator* is any function $\hat{\theta} = T(X_1,\ X_2,\ \ldots,\ X_n)$ of a sample.

Estimate: An *estimate* is the specific realization of an estimator, i.e. is a function of the observed values $x_1, x_2, \ldots, x_n$.

- There are many potential estimators to estimate a parameter: e.g. sample mean, sample median, trimmed mean, mode,...

Methods of Evaluating Estimators

- The distribution of the estimator should be somehow centred with respect to θ (Accuracy, unbiased)
  - If it is not, the estimator will tend either to *under-estimate* or *overestimate*
- The *dispersion* of the distribution should be small (Precision)
- These two properties need to be considered together.
- The difference $\hat{\theta} - \theta$ is referred to as an *error*, and the *mean squared error* is a commonly used measure of performance of an estimator.

1.1 Methods of Evaluating Estimators

Unbiased Estimator

Bias: Let $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ be an estimator of the parameter θ, the bias in $\hat{\theta}$ is defined by

$$\mathrm{Bias}(\theta) = E(\hat{\theta}) - \theta$$

Unbias estimator: $\hat{\theta}$ is called an *unbiased estimator* of θ if $\mathrm{Bias}(\theta) = 0$.

- The sample mean $\overline{X}$ is an unbiased estimator of the population mean μ;
- The sample variance S^2 is an unbiased estimator of the population variance σ^2.

The sample variance S^2 is unbiased of σ^2

Let $X_1,\ X_2,\ \ldots,\ X_n$ be iid $N(\mu, \sigma^2)$, shown that the sample variance S^2 is an unbiased estimator of σ^2.

Proof:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Note that $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ and $\sigma^2 = E(X^2) - \mu^2$, we have

$$\begin{aligned} E[(n-1)S^2] &= E\big[\sum(X_i - \overline{X})^2\big] \\ &= E\big[\sum X_i^2 - n\overline{X}^2\big] \\ &= n(\sigma^2 - \mu^2) - n(\frac{\sigma^2}{n} - \mu^2) = (n-1)\sigma^2 \end{aligned}$$

Hence $E(S^2) = \sigma^2$

Mean Square Error

The *Mean Squared Error* (MSE) of an estimator $\hat{\theta}$ is

$$\text{MSE} = E(\hat{\theta} - \theta)^2$$

- MSE is a trade-off between the precison and accuracy

$$\text{MSE} = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2$$

- If $\hat{\theta}$ is unbiased, MSE = Var($\hat{\theta}$)
- We call $\hat{\theta}$ is a *consistent estimator* of θ if

$$\lim_{n\to\infty} E(\hat{\theta} - \theta)^2 = \lim_{n\to\infty} \text{MSE} = 0$$

Standard Error

Standard error:The *standard error* (SE) is defined as the standard deviation of the sampling distribution of the statistic.

If $\hat{\theta}$ is unbiased, SE $= \sqrt{\text{MSE}}$.

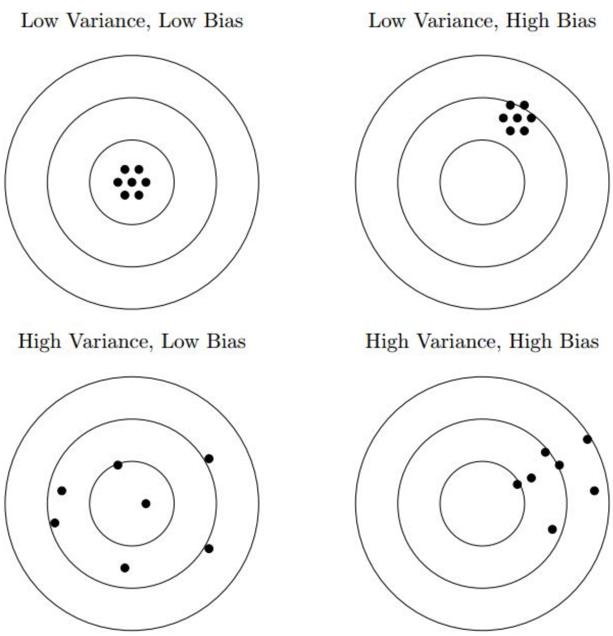
- Standard error of the sample mean ($\overline{X}$)

$$\text{SE}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

- An estimate of SE($\bar{x}$) is

$$\hat{\text{SE}}(\bar{x}) = \frac{s}{\sqrt{n}}$$

Visual representations of variance and bias



Example: Compare the MSE of S^2 and S_u^2

Let X_1, X_2, …, X_n be iid $N(\mu, \sigma^2)$, compare the MSE of S^2 and $S_u^2 = \frac{1}{n}\sum(X_i - \overline{X})^2$.

Note that $B = \frac{1}{\sigma^2}\sum(X_i - \overline{X})^2 = \frac{(n-1)}{\sigma^2}S^2 \sim \chi^2(n-1)$ and $\text{Var}(\chi_n^2) = 2(n-1)$, we have

$$\begin{aligned}
\text{E}(S^2) &= \sigma^2 \\
\text{Var}(S^2) &= \frac{2\sigma^4}{n-1} \\
\text{E}(S_u^2) &= \frac{n-1}{n}\sigma^2 \\
\text{Var}(S_u^2) &= \frac{2(n-1)\sigma^4}{n^2}
\end{aligned}$$

$$\text{MSE}(S^2) = \frac{2\sigma^4}{n-1} > \frac{(2n-1)\sigma^4}{n^2} = \text{MSE}(S_u^2)$$

Effciency

Efficiency: The efficiency of an estimator T is the inverse of its mean squared error, written as

$$\text{eff}(T) = \frac{1}{\text{MSE}(T)}$$

- An estimator T_1 is said to be more *precise* than the estimator T_2 if $\text{eff}[T_1]) \geq \text{eff}[T_2]$ or if $\text{MSE}[T_1] \leq \text{MSE}[T_2]$.
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two *unbiased* estimators of θ, based on the same sample. We said $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$.

Relative Efficiency

Relative Efficiency:

$$\text{eff}(T_1, T_2) = \frac{\text{eff}(T_1)}{\text{eff}(T_2)} = \frac{\text{MSE}(T_2)}{\text{MSE}(T_1)}$$

- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two *unbiased* estimators of θ, then

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{eff}(\hat{\theta}_1)}{\text{eff}(\hat{\theta}_2)} = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$$

Example 1. Let $X_1, X_2, \ldots, X_n$ denote a sample from $U(0, \theta)$, with $Y_1, Y_2, \ldots, Y_n$ be the corresponding ordered sample.

(i) Show that $T_1 = 2\overline{X}$ and $T_2 = \frac{n+1}{n}Y_n$ are both unbiased of θ.

(ii) Find eff(T_1,T_2).

(i) Now $E(X_i) = \theta/2,\ Var(X_i) = \theta^2/12$, so

$$E(T_1) = 2E(\overline{X}) = 2 \cdot \theta/2 = \theta$$

To find the mean of T_2, first note that the pdf of Y_n is

$$f_{Y_n}(y) = \begin{cases} \frac{n}{\theta^n}y^{n-1} & 0 < y < \theta \\ 0 & \text{Otherwise} \end{cases}$$

$$E(Y_n) = \int_0^\theta y \cdot \frac{n}{\theta^n} y^{n-1} dy = \frac{n}{n+1}\theta$$

Hence

$$E(T_2) = E(\frac{n+1}{n} Y_n) = \theta$$

So both T_1 and T_2 are unbiased.

(ii) Find eff(T_1,T_2).

$$Var(T_1) = Var(2\overline{X}) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

To find $Var(T_2)$, first we need to find $E(Y_n^2)$ from

$$E(Y_n^2) = \int_0^\theta y^2 \cdot \frac{n}{\theta^n} y^{n-1} dy = \frac{n}{n+2}\theta^2$$

$$Var(Y_n) = E(Y_n^2) - [E(Y_n)]^2 = \frac{n\theta^2}{n+2} - \big(\frac{n\theta}{n+1}\big)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

The variance of T_2 is

$$Var(T_2) = Var(\frac{n+1}{n} Y_n) = \frac{\theta^2}{n(n+2)}$$

The relative efficiency

$$\text{eff}(T_1, T_2) = \frac{Var(T_2)}{Var(T_1)} = \frac{3}{n+2}$$

This is less than 1 for $n > 1$ so T_2 is more efficient than T_1.

1.2 Cramer-Rao Lower Bound

Cramer-Rao Lower Bound

Theorem 2. *If $T = \hat{\theta}$ is an unbiased estimator of θ and a random sample of size n, $X_1, X_2, \ldots, X_n$, has pdf $f(x|\theta)$, then the variance of the unbiased estimator, $\hat{\theta}$, must satisfy the inequality*

$$\text{Var}(\hat{\theta}) \geq \frac{1}{nI_X(\theta)}$$

where $I_X(\theta)$ is the Fisher information *of X:*

$$I_X(\theta) = E\big[\big(\frac{\partial \ln f(X|\theta)}{\partial \theta}\big)^2\big]$$

Computation of Fisher Information

$$I_X(\theta) = E\big[\big(\frac{\partial \ln f(X|\theta)}{\partial \theta}\big)^2\big]$$

$$I_X(\theta) = -E\big[\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2}\big]$$

$$I_X(\theta) = Var\big(\frac{\partial \ln f(X|\theta)}{\partial \theta}\big)$$

Minimum variance unbiased estimator

Definition 3. If $\hat{\theta}$ is an unbiased estimator of θ and

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{nI_X(\theta)}$$

then $\hat{\theta}$ is called an ***efficient estimator*** or a ***minimum variance unbiased estimator*** of θ.

In general

Theorem 4. *If $T = \hat{\theta}$ is any estimator of θ and a random sample of size n, $X_1, X_2, \ldots, X_n$, has pdf $f(x|\theta)$, then the following inequality valid*

$$\mathrm{Var}(\hat{\theta}) \geq \frac{[1 + b_T^{'}(\theta)]^2}{nI_X(\theta)}$$

and

$$\mathrm{MSE} \geq \frac{[1 + b_T^{'}(\theta)]^2}{nI_X(\theta)} + b_T^2(\theta)$$

where $b_T(\theta)$ is the bias of θ, $b_T(\theta) = E(T) - \theta$.

Example

Show that $\overline{X}$ is a minimum variance unbiased estimator of the mean λ of a Poisson population.

Proof:

If $X \sim Pois(\lambda)$ then $E(X) = \mathrm{Var}(X) = \lambda$. So $E(\overline{X}) = \lambda$ and $\mathrm{Var}(\overline{X}) = \lambda/n$. We need to show the CRLB equals to λ/n.

Computing the Fisher information:

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

$$\ln f(x|\lambda) = x \ln \lambda - \lambda - \ln(x!)$$

$$\frac{\partial \ln f(x|\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1$$

$$\frac{\partial^2 \ln f(x|\lambda)}{\partial^2 \lambda} = -\frac{x}{\lambda^2}$$

$$I_X(\lambda) = -[\frac{\partial^2 \ln f(x|\lambda)}{\partial^2 \lambda}] = \frac{E(X)}{\lambda^2} = \frac{1}{\lambda}$$

Or

$$I_X(\lambda) = \mathrm{Var}[\frac{\partial \ln f(x|\lambda)}{\partial \lambda}] = \frac{1}{\lambda^2}\mathrm{Var}(X) = \frac{1}{\lambda}$$

the CRLB is

$$\frac{1}{nI_X(\lambda)} = \frac{\lambda}{n}$$

Hence $\mathrm{Var}(\overline{X}) = \lambda/n$ attains the CRLB.

2 Point Estimation Techniques

2.1 Method of Moments Estimators

Methods of Finding Estimators

1. Method of moments
   - Let sample moments equals to population monemts
   - The order of the moments equals to the number of unknown parameters
2. Maximum likelihood estimate (MLE)
   - The maximum point of the likelihood function

Example: method of moments estimators

Given a random sample of size n from a $Gamma(\alpha, \lambda)$ population, find the method of moments estimators of α and λ.

Solution: The population moments $E(X) = \alpha/\lambda$, and $Var(X) = \alpha/\lambda^2$. Let

$$\begin{aligned} E(X) = \alpha/\lambda &= m_1 = \overline{X} \\ E(X^2) = \frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2 &= m_2 = \frac{1}{n}\sum X_i^2 \end{aligned}$$

It can be solved that

$$\tilde{\alpha} = \frac{\overline{X}^2}{S_u^2} \text{ and } \tilde{\lambda} = \frac{\overline{X}}{S_u^2}$$

where $S_u^2 = \frac{1}{n}\sum(X_i - \overline{X})^2$.

2.2 Likelihood and Maximum Likelihood Estimators

Maximum Likelihood Estimate (MLE)

Definition 5. Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x|\theta)$ and $x = (x_1, x_2, \ldots, x_n)^T$ be the corresponding observed values. The likelihood function is

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

The value of θ that maximizes the likelihood function is called the *Maximum Likelihood Estimate* (MLE), i.e.

$$L(\hat{\theta}|x) = \max_{\theta\in\Theta} L(\theta)$$

Example: MLE of Poisson

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $Pois(\lambda)$ population. Compute the maximum likelihood estimator and the maximum likelihood estimate for the parameter l. Verify your answer with simulation by generating 20,000 random values from a $Pois(\lambda = 5)$ population.

Solution: The log-likelihood function is

$$\ln L(\lambda|x) = \ln\left(\prod \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right) = -n\lambda + \ln\lambda\sum x_i - \sum \ln(x_i!)$$

$$\frac{\partial \ln L(\lambda|x)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} \overset{set}{=} 0$$

The solution is $\hat{\lambda} = \bar{x}$, the second order derivative

$$\frac{\partial^2 \ln L(\lambda|x)}{\partial^2 \lambda} = -\frac{\sum x_i}{\lambda^2} < 0$$

The maximum likelihood estimate is $\hat{\lambda}(x) = \bar{x}$ and the maximum likelihood estimator is $\hat{\lambda}(X) = \overline{X}$.

MLE of Poisson: simulation

```
> set.seed(99)
> x <- rpois(20000, 5)
> lam <- c(mean(x),median(x))
> lam

## [1] 4.99415 5.00000
```

Example: MLE of Uniform

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $Unif(0,\theta)$ population. Find the maximum likelihood estimator of θ. Find the maximum likelihood estimate for a randomly generated sample of 1000 $Unif(0,2)$ random variables.

Solution: The pdf of a random variable $X \sim Unif(0,\theta)$ is $f(x|\theta) = 1/\theta$ $(0 \le x \le \theta)$. The likelihood function is

$$\boldsymbol{L(\theta|x) = \begin{cases} \frac{1}{\theta^n} & 0 \le x_1, x_2, \ldots, x_n \le \theta \\ 0 & \text{otherwise.} \end{cases}}$$

$L(\theta|x) = 1/\theta^n$ if and only if $\theta \ge x_{(n)} = max(x_i)$, hence it follows that the maximum likelihood estimator is $\hat{\theta}(X) = X_{(n)}$.

Properties of Maximum Likelihood Estimators

1. MLEs are not necessarily unbiased.
2. If T is a MLE of θ and g is any function, then $g(T)$ is the MLE of $g(\theta)$. This is known as the *invariance property* of MLEs.
3. Under certain regularity conditions on $f(x|\theta)$, as $n \to \infty$,

$$\hat{\theta}(X) \sim N(\theta, [I_X(\theta)]^{-1})$$

4. An efficient estimator (if exists) must be MLE, but MLE may not be efficient.

Example: MLE of Gamma

Given a random sample of size n from a population with pdf

$$f(x|\theta) = \frac{x}{\theta^2} e^{-x/\theta}, x \ge 0, \theta > 0$$

1. Find an estimator of θ using the method of moments.

2. Find an estimator of θ using the method of maximum likelihood.
3. Are the method of moments and maximum likelihood estimators of θ unbiased?
4. Compute the variance of the MLE of θ.
5. Is the MLE of θ efficient?