Lecture 6
Methods for Point Estimator

Fall 2013
Prof. Yao Xie, yao.xie@isye.gatech.edu
H. Milton Stewart School of Industrial Systems & Engineering
Georgia Tech
Methods of Point Estimation

• Method of Moments (MoM)
• Method of Maximum Likelihood
• Bayesian methods
• …
Methods of Moments

Population and samples moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The $k$th population moment (or distribution moment) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding $k$th sample moment is $(1/n) \sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Population moments

$$
\mu'_k = \begin{cases} 
\int x^k f(x) dx & \text{If } x \text{ is continuous} \\
\sum_x x^k f(x) & \text{If } x \text{ is discrete}
\end{cases}
$$

Sample moments

$$
m'_k = \frac{\sum_{i=1}^{n} X_i^k}{n}
$$
Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with $m$ unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The moment estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ are found by equating the first $m$ population moments to the first $m$ sample moments and solving the resulting equations for the unknown parameters.

\[ m_1' = \mu_1' \]
\[ m_2' = \mu_2' \]
\[ \vdots \]
\[ m_m' = \mu_m' \]

$m$ equations for $m$ parameters
Example

1) What is the point estimator of $\lambda$ in the exponential distribution?

2) What is the point estimator of $p$ in the Binomial distribution?

3) What is the point estimator for mean and variance in normal distribution?
Methods of Point Estimation

- Method of Moments (MoM)
- Method of Maximum Likelihood
- Bayesian methods
- …
Method of Maximum Likelihood

Suppose that $X$ is a random variable with probability distribution $f(x; \theta)$, where $\theta$ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size $n$. Then the likelihood function of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \quad (7-9)$$

Note that the likelihood function is now a function of only the unknown parameter $\theta$. The maximum likelihood estimator (MLE) of $\theta$ is the value of $\theta$ that maximizes the likelihood function $L(\theta)$.

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

$$l(\theta; x) = \sum_{i=1}^{n} \log[f(x_i; \theta)]$$

$$\hat{\Theta}(x) = \arg \max_{\theta} L(\theta; x) = \arg \max_{\theta} l(\theta; x)$$
Example

Let $X$ be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1 - p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where $p$ is the parameter to be estimated. The likelihood function of a random sample of size $n$ is

$$L(p) = p^{x_1}(1 - p)^{1-x_1}p^{x_2}(1 - p)^{1-x_2} \cdots p^{x_n}(1 - p)^{1-x_n}$$

$$= \prod_{i=1}^{n} p^{x_i}(1 - p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1 - p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(p) = \left( \sum_{i=1}^{n} x_i \right) \ln p + \left( n - \sum_{i=1}^{n} x_i \right) \ln (1 - p)$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left( n - \sum_{i=1}^{n} x_i \right)}{1 - p} \quad \rightarrow \quad \hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$
Example

Let $X$ be normally distributed with mean $\mu$ and variance $\sigma^2$, where both $\mu$ and $\sigma^2$ are unknown. The likelihood function for a random sample of size $n$ is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} e^{-(x_i - \mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2) \sum_{i=1}^{n} (x_i - \mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$
Example ( Continued )

\[
\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0
\]

\[
\frac{\partial \ln L(\mu, \sigma^2)}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0
\]

The solutions to the above equation yield the maximum likelihood estimators

\[
\hat{\mu} = \bar{X} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2
\]

Once again, the maximum likelihood estimators are equal to the moment estimators.
Exponential MLE

Let $X$ be a exponential random variable with parameter $\lambda$. The likelihood function of a random sample of size $n$ is:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{d \ln L(\lambda)}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\hat{\lambda} = n \left/ \sum_{i=1}^{n} x_i \right. = 1/\bar{X} \quad \text{(same as moment estimator)}$$
MLE Properties

Under very general and not restrictive conditions, when the sample size $n$ is large and if $\hat{\theta}$ is the maximum likelihood estimator of the parameter $\theta$,

1. $\hat{\theta}$ is an approximately unbiased estimator for $\theta$ [$E(\hat{\theta}) = \theta$],
2. the variance of $\hat{\theta}$ is nearly as small as the variance that could be obtained with any other estimator, and
3. $\hat{\theta}$ has an approximate normal distribution.

Example:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$
Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

**Example:**

In the normal distribution case, the maximum likelihood estimators of $\mu$ and $\sigma^2$ were $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function $h$, which yields

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{1/2}$$

Thus, the maximum likelihood estimator of the standard deviation $\sigma$ is *not* the sample standard deviation $S$. 
Complications in Using MLE

• It is not always easy to maximize the likelihood function because the equation(s) obtained from setting derivative to be 0 may be difficult to solve.

• It may not always be possible to use calculus methods directly to determine the maximum of the likelihood function.
Example: Uniform Distribution MLE

Let $X$ be uniformly distributed on the interval $0$ to $a$.

$$f(x) = \frac{1}{a} \text{ for } 0 \leq x \leq a$$

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n} = a^{-n} \text{ for } 0 \leq x_i \leq a$$

$$\frac{dL(a)}{da} = -n \frac{a^{-n-1}}{a^{n+1}} = -na^{-(n+1)}$$

$$\hat{a} = \max(x_i)$$

Calculus methods don’t work here because $L(a)$ is maximized at the discontinuity.

Clearly, $a$ cannot be smaller than $\max(x_i)$, thus the MLE is $\max(x_i)$. 

Figure 7-8 The likelihood function for this uniform distribution