Lecture 12
Estimation Methods: Maximum Likelihood Estimators

General methods for parameter estimation

- We saw that MVUE is the most desirable estimator. We discussed two approaches to find the MVUE:
  - MVUE through CRLB
  - MVUE through a complete sufficient statistic
- The above approaches may not be feasible for practical models of random data.
- In a practical situation, we have to apply some general techniques to construct a good estimator. The same technique applied on different probability models may produce different estimation rules. The goodness is measured in terms of the desired properties of an estimators like unbiasedness, consistence and efficiency.

General Estimation Techniques …

- Most of these methods are based on some optimality criteria. An optimality criterion tries to optimize some functions of the random samples with respect to the unknown parameter to be estimated.

- Some of the most popular estimation techniques are:
  - Method of moments
  - Maximum likelihood method
  - Bayesian methods.
  - Least squares methods

We will discuss these methods.

Method of Moments

- The method of moments (MM) is a simple criterion for parameter estimation. When other methods are mathematically intractable, an MM estimator is a simple alternative.
- Suppose $X_1, X_2, ..., X_N$ are iid random samples with the joint probability density function $f(x_1, x_2, ..., x_N; \theta_1, \theta_2, ..., \theta_K)$ which depends on unknown parameters $\theta_1, \theta_2, ..., \theta_K$.
- The r-th moment of each X_i is given by

$EX_1^r \quad r = 1, 2, ...$ Thus,

For $r = 1$ $\quad EX_1 = \text{Mean of } X_1$

For $r = 2$ $\quad EX_1^2 = \text{Mean-square value of } X_1$

and so on

Sample Moments

- The sample moments are given by

$$\hat{\mu}_r = \frac{1}{N}\sum_{i=1}^{N} X_i^r, \quad r = 1, 2, ..$$

For example,

$$\hat{\mu}_1 = \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad \text{and} \quad \hat{\mu}_2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2$$

- Note that

$$E\hat{\mu}_r = \frac{1}{N}\sum_{i=1}^{N} EX_i^r = EX_1^r \text{ and}$$

$$\operatorname{var}(\hat{\mu}_r) = \frac{1}{N^2}\sum_{i=1}^{N} \operatorname{var} X_i^r = \frac{\operatorname{var} X_i^r}{N}$$

So that $\lim_{N\to\infty} \operatorname{var}(\hat{\mu}_r) = \lim_{N\to\infty} \frac{\operatorname{var} X_i^r}{N} = 0$

$\therefore \hat{\mu}_r$ is an unbiased and a consistent estimator.

Method of Moments

- Based on the assumption that the observed data have the sample moments same as the population moments:

$\hat{\mu}_r = EX_1^r$

- MM estimation finds *k* equations relating the first k moments $\mu_1, \mu_2, \ldots \mu_k$ with the parameters $\theta_1, \theta_2, \ldots, \theta_K$ and then substitute the moments by the corresponding sample moments $\hat{\mu}_1, \hat{\mu}_2, \ldots \hat{\mu}_k$ in these equations. The solution of the equations give the estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$

- The MM method is also known as the method of substitution

Steps in MM estimation

❖ The following are the steps:

(1) Express $EX_1^r,\ r=1,2,\ldots,k$ as functions of $\theta_1,\theta_2,\ldots,\theta_K$ to get the equations

$$EX_1 = h_1(\theta_1,\theta_2,\ldots,\theta_K)$$
$$EX_1^2 = h_2(\theta_1,\theta_2,\ldots,\theta_K)$$
$$\vdots \quad \vdots$$
$$EX_1^k = h_k(\theta_1,\theta_2,\ldots,\theta_K)$$

(2) Substitute $EX_1^r,\ r=1,2,\ldots,k$ by the corresponding sample moments $\hat{\mu}_r$ to get the modified set of equations

$$\hat{\mu}_1 = h_1(\theta_1,\theta_2,\ldots,\theta_K)$$
$$\hat{\mu}_2 = h_2(\theta_1,\theta_2,\ldots,\theta_K)$$
$$\vdots \quad \vdots$$
$$\hat{\mu}_k = h_k(\theta_1,\theta_2,\ldots,\theta_K)$$

Solve the modified set of equations to get the MM estimators $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_K$

Example: Let $X_1, X_2, \ldots, X_N$ are iid with $X_i \sim N(\mu,\ \sigma^2)$. Find MM estimators for μ and σ^2.

❖ We have the first two moments,

$$EX_1 = \mu$$

$$EX_1^2 = \sigma^2 + \mu^2$$

❖ Substituting the moments by the sample moments,

$$\frac{1}{N}\sum_{i=1}^{N} x_i == \mu$$

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2 = \sigma^2 + \mu^2$$

❖ Solving we get,

$$\hat{\mu}_{MM} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and } \hat{\sigma}_{MM}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{\mu}_{MM}\right)^2$$

Properties of MM estimators

- The properties of the MM estimators depend on the nature of $f(x_1, x_2, ..., x_N; \theta_1, \theta_2, ..., \theta_K)$.
- As $\hat{\mu}_r$ is an unbiased and consistent estimator of EX_1^r, the MM estimators will be unbiased and consistent if $\hat{\theta}_r$ is a linear combination of sample moments.
- Properties of MM estimators are empirically studied through simulations.

We have to look for better estimation rules!

Maximum Likelihood Estimator (MLE)

- Suppose $X_1, X_2, \ldots, X_N$ are random samples with the joint probability density function $f(x_1, x_2, \ldots, x_N; \theta) = f(\mathbf{x}; \theta)$ which depends on an unknown non-random parameter θ.
- Note that $f(\mathbf{x}; \theta)$ is called the likelihood function. If $X_1, X_2, \ldots, X_N$ are discrete, then the likelihood function will be a joint probability mass function.
- $L(\mathbf{x}; \theta) = \ln f(\mathbf{x}; \theta)$ is the log likelihood function.
- In discrete case, $f(\mathbf{x}; \theta)$ is replaced by $p(\mathbf{x}; \theta)$.
- As a function of the random variables, the likelihood and log-likelihood functions are random variables.

Maximum Likelihood Principle

- Select that value of θ which maximizes $f(x_1, x_2, ..., x_N; \theta)$. Thus the maximum likelihood estimator $\hat{\theta}_{MLE}$ is such an estimator that

$$f(x_1, x_2, ..., x_N; \hat{\theta}_{MLE}) \geq f(x_1, x_2, ..., x_N; \theta), \forall \theta.$$

- If the likelihood function is differentiable with respect to θ, then $\hat{\theta}_{MLE}$ is given by $\left.\frac{\partial f\ (\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}_{\text{MLE}}} = 0$

- Since $L(x\ ;\theta) = \ln(f(x\ ;\theta))$ is a monotonic function of the argument, it is convenient to express the MLE conditions in terms of the log-likelihood function

$$\left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}_{\text{MLE}}} = 0$$

Maximum Likelihood Principle …

- If we have k unknown parameters given by

$$\boldsymbol{\theta} = [\theta_1 \quad \theta_2 \ldots \theta_k]'$$

then the MLE is given by a set of equations.

$$\left.\frac{\partial f(\mathbf{x};\theta)}{\partial \theta_1}\right]_{\theta_1=\hat{\theta}_{1MLE}} = \left.\frac{\partial f(\mathbf{x};\theta)}{\partial \theta_2}\right]_{\theta_2=\hat{\theta}_{2MLE}} = \ldots. = \left.\frac{\partial f(\mathbf{x};\theta)}{\partial \theta_k}\right]_{\theta_k=\hat{\theta}_{kMLE}} = 0$$

- In terms of $\mathrm{L}(\mathbf{x};\theta)$, the set of equations are given by

$$\left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_1}\right]_{\theta_1=\hat{\theta}_{1MLE}} = \left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_2}\right]_{\theta_2=\hat{\theta}_{2MLE}} = \ldots. = \left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_k}\right]_{\theta_k=\hat{\theta}_{kMLE}} = 0$$

Example: MLE for the λ parameter of iid Poisson samples

Let $X_1, X_2, \ldots, X_N$ be iid random variables with $X_i \sim Poi(\lambda)$. Find MLE for λ .

Solution:

$$p(\mathbf{x};\lambda) = p(x_1, x_2, \ldots, x_N;\lambda)$$

$$= \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i\,!}$$

$$\therefore L(\mathbf{x};\lambda) = \ln p(\mathbf{x};\lambda)$$

$$= -N\lambda + \sum_{i=1}^{N} x_i \ln\lambda + \text{terms not involving } \lambda$$

$$\left.\frac{\partial L}{\partial o}\right]_{\hat{\lambda}_{MLE}} = 0$$

$$\Rightarrow -N + \sum_{i=1}^{N} \frac{x_i}{\hat{\lambda}_{MLE}} = 0$$

$$\hat{\lambda}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Example: MLE for multiple parameters

Let $X_1, X_2, \ldots, X_N$ be iid random variables with $X_i \sim N(\mu,\ \sigma^2)$. Find MLE for μ and σ^2.

Solution:

$$f(\mathbf{x};\mu,\sigma^2) = f(x_1, x_2, \ldots, x_N; \mu, \sigma^2)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\therefore L(\mathbf{x};\mu,\sigma^2) = \ln f(\mathbf{x};\mu,\sigma^2)$$

$$= -N(\ln\sqrt{2\pi}\sigma) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\left.\frac{\partial L}{\partial \mu}\right]_{\hat{\mu}_{MLE}} = 0$$

$$\Rightarrow \sum_{i=1}^{N}(x_i - \hat{\mu}_{MLE}) = 0$$

Example …

$$\left.\frac{\partial L}{\partial \sigma^2}\right]_{\hat{\sigma}^2_{MLE}} = 0$$

$$\Rightarrow -\frac{N}{\hat{\sigma}^2_{MLE}} + \frac{\sum_{i=1}^{N}(x_i - \hat{\mu}_{MLE})^2}{\hat{\sigma}^4_{MLE}} = 0$$

Solving we get

$$\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and}$$

$$\hat{\sigma}_{MLE}{}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{\mu}_{MLE}\right)^2$$

Example : Non-differentiable likelihood function

Let $X_1, X_2, \ldots, X_N$ be iid random samples with the PDF

$$f(x;\theta) = \frac{1}{2} e^{-|x-\theta|} \quad -\infty < x < \infty.$$

Show that $median\ (x_1, x_2, \ldots, x_N)$ is the MLE for θ.

$$f(x_1, x_2, \ldots, x_N; \theta) = \frac{1}{2^N} e^{-\sum_{i=1}^{N} |x_i - \theta|}$$

$$L(\mathbf{x};\theta) = \ln f(\mathbf{x};\theta)$$

$$= -N \ln 2 - \sum_{i=1}^{N} |x_i - \theta|$$

$\sum_{i=1}^{N} |x_i - \theta|$ is minimized by $median\ (x_1, x_2, \ldots, x_N)$

$$\therefore \hat{\theta}_{MLE} = median\ (x_1, x_2, \ldots, x_N)$$

Summary

- Most of the estimation methods are based on some optimality criteria. The optimality criterion tries to optimize some functions of the observed samples with respect to the unknown parameter to be estimated.
- MM estimation involves relating the moments with the parameters and then substituting the moments by the sample moments.
- MM estimators are obtained by solving the set of equations

$$\begin{aligned}
\hat{\mu}_1 &= h_1(\theta_1, \theta_2, \ldots, \theta_K) \\
\hat{\mu}_2 &= h_2(\theta_1, \theta_2, \ldots, \theta_K) \\
&\vdots \\
\hat{\mu}_k &= h_k(\theta_1, \theta_2, \ldots, \theta_K)
\end{aligned}$$

- MM estimators may or may not satisfy the desired properties of a good estimator.

Summary…

- The maximum likelihood estimator $\hat{\theta}_{MLE}$ is such an estimator that

$$f(x_1, x_2, \ldots, x_N; \hat{\theta}_{MLE}) \geq f(x_1, x_2, \ldots, x_N; \theta), \forall \theta.$$

- If the likelihood function is differentiable with respect to θ, then $\hat{\theta}_{MLE}$ is given

by $$\left.\frac{\partial f\ (\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}_{\mathrm{MLE}}} = 0$$

or equivalently $$\left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}_{\mathrm{MLE}}} = 0$$

- If we have k unknown parameters $\boldsymbol{\theta} = [\theta_1 \quad \theta_2 \ldots \theta_k]'$

then the MLEs are given by a set of equations:

$$\left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_1}\right]_{\theta_1=\hat{\theta}_{1MLE}} = \left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_2}\right]_{\theta_2=\hat{\theta}_{2MLE}} = \ldots. = \left.\frac{\partial \mathrm{L}(\mathbf{x};\theta)}{\partial \theta_k}\right]_{\theta_k=\hat{\theta}_{kMLE}} = 0$$

THANK YOU