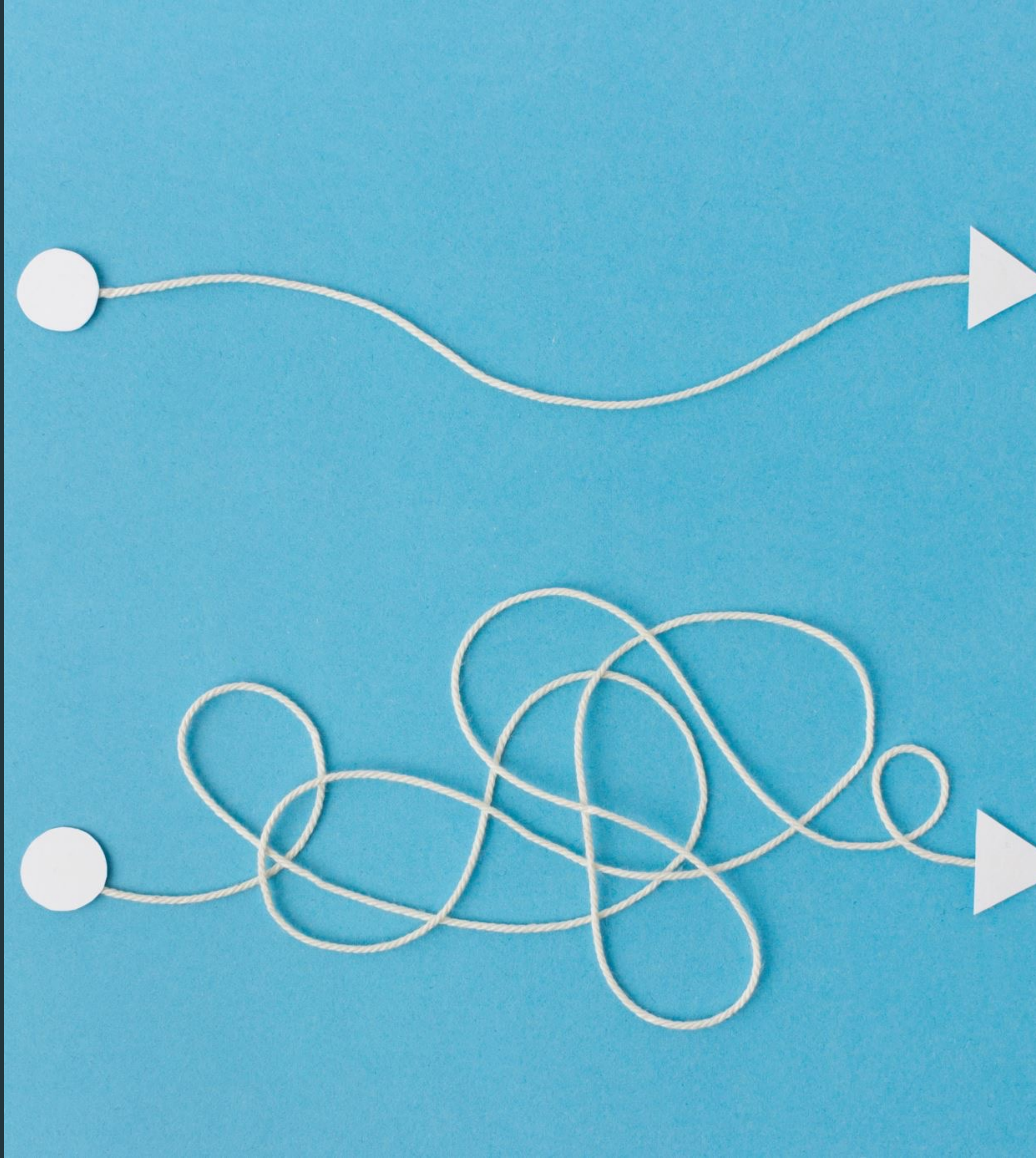
Week 2 – Lecture 2

Estimation of Parameters

and

Fitting of Probability Distributions

(cont.)



Objectives

- Methods of Moments
- Method of Maximum Likelihood
- Properties of MLEs
- Confidence Intervals from MLEs
- Properties of MLEs

Method of Moments

Definition:

- Let X be a random sample from a pmf or pdf $f(x)$. For k = 1, 2, 3,…, the k^th^ population moment, or k^th^ moment of the distribution $f(x)$, or k^th^ moment of the probability, is $E(X^k)$.

- Let $X_1, \ldots, X_n$ be a random sample from a pmf or pdf $f(x)$. The k^th^ sample moment is $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.

Method of Moments (cont.)

Definition: Let $X_1, \ldots, X_n$ be a random sample from a pmf or pdf $f(x; \theta_1, \theta_2, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown.

- The method of moments estimators (MME) $\boldsymbol{\hat{\theta}_1, \ldots, \hat{\theta}_m}$ are obtained by equating the first *m* sample moments to the corresponding first *m* population moments and solving for $\theta_1, \ldots, \theta_m$.

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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

$$E(X^3) = \frac{1}{n}\sum_{i=1}^{n} X_i^3$$

$$\cdots$$

$$E(X^m) = \frac{1}{n}\sum_{i=1}^{n} X_i^m$$

Method of Moments (cont.)

Example 1: Find MME for the Poisson distribution with parameter λ.

Example 2: Find MME for the binomial distribution with parameter p.

Example 3: Find MME for the normal distribution with parameters μ and σ.

Practice Problems:

1) Find MME for the exponential distribution with parameter λ.

2) Find MME for the Gamma distribution with parameters α and β.

Method of Maximum Likelihood

- The method of maximum likelihood was first introduced by R. A. Fisher, a geneticist and statistician, in the 1920s. Most statisticians recommend this method, at least when the sample size is large, since the resulting estimators have certain desirable efficiency properties.

Definition: Let $X_1, X_2, \ldots, X_n$ have joint pmf or pdf $f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m)$, where the parameters $\theta_1, \ldots, \theta_m$ have unknown values. When $x_1, \ldots, x_n$ are the observed sample values and f is regarded as a function of $\theta_1, \ldots, \theta_m$, it is called the likelihood function.

➢ The maximum likelihood estimates (mle) $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are those values of the θ_i that maximize the likelihood function, i.e.

$$f\left(x_1, \ldots, x_n; \hat{\theta}_1, \ldots, \hat{\theta}_m\right) \geq f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m)$$

for all $\theta_1, \ldots, \theta_m$.

➢ When the X_i are substituted in place of the x_i, the maximum likelihood estimators (MLEs) result.

Method of Maximum Likelihood (cont.)

Steps of Method of Maximum Likelihood:

1) Write out the likelihood function.
2) Maximize the likelihood function (using Calculus or other techniques).
3) The solution of the above maximization problem is MLE.

Likelihood Function of a Random Sample:

Let $X_1, X_2, \dots, X_n$ be a random sample from a pmf or pdf $f(x; \theta_1, \theta_2, \dots, \theta_m)$, where $\theta_1, \dots, \theta_m$ are parameters whose values are unknown. Then the likelihood function

$$L(\theta_1, \dots, \theta_m) = f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$$

$$= f(x_1; \theta_1, \theta_2, \dots, \theta_m) \cdot f(x_2; \theta_1, \theta_2, \dots, \theta_m) \dots f(x_n; \theta_1, \theta_2, \dots, \theta_m)$$

$$= \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \dots, \theta_m)$$

Method of Maximum Likelihood (cont.)

Note:

- We can maximize the log-likelihood function to avoid complicated calculations in the maximization problem of the likelihood function.

Log-Likelihood Function of a Random Sample:

Let $X_1, X_2, \ldots, X_n$ be a random sample from a pmf or pdf $f(x; \theta_1, \theta_2, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. Then the likelihood function

$$l(\theta_1, \ldots, \theta_m) = \log L(\theta_1, \ldots, \theta_m)$$
$$= \log[f(x_1; \theta_1, \theta_2, \ldots, \theta_m) \cdot f(x_2; \theta_1, \theta_2, \ldots, \theta_m) \ldots f(x_n; \theta_1, \theta_2, \ldots, \theta_m)]$$
$$= \sum_{i=1}^{n} \log f(x_i; \theta_1, \theta_2, \ldots, \theta_m)$$

Method of Maximum Likelihood (cont.) - Examples

1) Find MLE for the Poisson distribution with parameter λ.
2) Find MLE for the exponential distribution with parameter λ.
3) Find MLE for the normal distribution with parameters μ and σ.

Properties of MLEs

Properties:

1) Let $\hat{\theta}_1, \dots, \hat{\theta}_m$ be the MLEs of the parameters $\theta_1, \dots, \theta_m$. Then the MLE of any function $h(\theta_1, \dots, \theta_m)$ of these parameters is the function $h(\hat{\theta}_1, \dots, \hat{\theta}_m)$ of the MLEs.

2) Under very general conditions on the joint distribution of the sample, when the sample size n is large, the maximum likelihood estimator of any parameter θ is at least approximately unbiased ($E(\hat{\theta}) \approx \theta$) and has variance that is either as small as or nearly as small as can be achieved by any estimator. Stated another way, the MLE $\hat{\theta}$ is either exactly or at least approximately the MVUE of θ.

Confidence Intervals from MLEs

Definition: Fisher information for θ contained in a random sample $X_1, \dots, X_n$ is defined as

$$I_n(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2} l(\theta)\right)$$

- Fisher information provides a way to measure the amount of information that random variables contain about some parameter θ of the assumed probability distribution. If it is small, the random variables provide much information about θ.

Confidence Intervals from MLEs

- MLEs can be used to find a confidence interval for the parameter θ. The $(1 - \alpha)100\%$ confidence interval is given by

$$\hat{\theta}_{MLE} \pm \frac{z_{\alpha/2}}{\sqrt{I_n(\hat{\theta}_{MLE})}}$$

Example:

1) Find a confidence interval for the parameter λ of the Poisson distribution.
2) Given a random sample $X_1, \ldots, X_{36}$ with sample mean of 10. Assume they follow the exponential distribution with parameter λ. Construct a 95% confidence interval for the true value of the parameter λ.

Mean Square Error for MLEs

- The mean square error for MLEs are given by

$$MSE(\hat{\theta}) = E\left(\left(\hat{\theta} - \theta\right)^2\right) = Var(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2$$

$$= Variance + (Bias)^2$$

- Better estimators have lower MSE.

Compare The Efficiency of Two Estimators

- The efficiency of two estimators $\hat{\theta}_A$ and $\hat{\theta}_B$ is given by

$$eff(\hat{\theta}_A, \hat{\theta}_B) = \frac{MSE(\hat{\theta}_A)}{MSE(\hat{\theta}_B)}$$

- Using the efficiency of two estimators, we can compare them.
  - If $eff(\hat{\theta}_A, \hat{\theta}_B) < 1$, estimator $\hat{\theta}_A$ is better than estimator $\hat{\theta}_B$.
  - If $eff(\hat{\theta}_A, \hat{\theta}_B) > 1$, estimator $\hat{\theta}_B$ is better than estimator $\hat{\theta}_A$.
  - If $eff(\hat{\theta}_A, \hat{\theta}_B) = 1$, estimator $\hat{\theta}_A$ is as efficient as estimator $\hat{\theta}_B$.

Cramér-Rao Lower Bound

- If $\hat{\theta}$ is an unbiased estimator, then

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

- If the equality is achieved, $\hat{\theta}$ is said to be efficient.

Practice Problem:

Show that $\hat{\mu}_{MLE}$ for the normal distribution with mean μ and standard deviation σ is efficient.

Consistency of MLEs

Definition: Let $\{X_1, X_2, \dots, X_n\}$ be a sequence of observations. Let $\hat{\theta}_{\boldsymbol{n}}$ be the estimator using $\{X_1, X_2, \dots, X_n\}$. We say that $\hat{\theta}_n$ is consistent if $\hat{\theta}_n \xrightarrow{p} \theta$, i.e.,

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \to 0 \; as \; n \to \infty$$

- Let $X_1, X_2, \dots, X_n$ be a random sample from a distribution with a parameter θ. Then MLE of θ is consistent.

Asymptotic Normality of MLEs

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with a parameter θ. If $\hat{\theta}$ is the MLE of θ, then

$$(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{1}{\sqrt{I_n(\theta)}}\right)$$