innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

BITS Pilani
Pilani Campus

Chapter 6: Point Estimation

Sumanta Pasari

sumanta.pasari@pilani.bits-pilani.ac.in

Parameter Estimation

- Parameter estimation is one of the important steps in **statistical inference**.
- It belongs to the subject of estimation theory.
- Why do we require parameter estimation?
- What are the different estimation methods?
- What are the desirable properties of an estimator?
- How to judge “how good is my estimator”?
- **Two broad types – point estimation and interval estimation**

Estimator and estimate

- A statistic (which is a function of a random sample, and hence a random variable) used to estimate the population parameter θ is called a ***point estimator*** for θ and is denoted by $\hat{\theta}$

- The value of the point estimator on a particular sample of given size is called a point estimate for θ.

Desirable Properties

1. $\hat{\theta}$ to be unbiased for θ.

2. $\hat{\theta}$ to have a small variance for large sample size.

(MVUE: Minimum Variance Unbiased Estimator)

Unbiased estimator:

A point estimator $\hat{\theta}$ is an unbiased estimator for a population parameter θ if $E\left(\hat{\theta}\right)=\theta$.

Point Estimator

Comments.

1. The sample mean, $\bar{X}$, is an unbiased estimator for μ.

2. The sample variance, S^2, is an unbiased estimator for σ^2.

3. When X is binomial RV with parameters n and p, the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p.

Standard error of sample mean $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Minimum Variance Unbiased Estimator

Among all estimators of the parameter θ that are unbiased, choose the one that has the minimum variance. The resulting $\hat{\theta}$ is called the MVUE of θ.

Theorem :

Let $X_1,\ X_2,\ \ldots,\ X_n$ be a random sample from a normal population with mean μ and standard deviation σ. Then the estimator $\hat{\mu} = \bar{X}$ is the MVUE for μ.

Method of Moments

- In method of moments (MoM), we compare the observed sample moments (about origin) with the corresponding population moments (about origin).
- If there are *k*-parameters in the distribution, then first *k* sample moments will be compared with the first *k* population moments to yield *k* equations. The solution of this *k*-equations will provide the required estimated parameter values.

Example: Method of Moments

Ex.1. Use method of moments to estimate the parameter of exponential distribution.

$$f(x;\beta)=\frac{1}{\beta}e^{-\frac{x}{\beta}};\ \ x>0,\beta>0$$

Sol.

Step 1 : Find $E(X)=\beta.$

Step 2 : Find the first sample moment as $M_1=\frac{1}{n}\sum_{i=1}^{n}X_i$

Step 3 : Equate the first sample moment with the first population moment.

$$\beta=\frac{1}{n}\sum_{i=1}^{n}X_i\Rightarrow\boxed{\hat{\beta}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i}$$

Is the estimator $\hat{\beta}$ unbiased?

Example: Method of Moments

HW.1. Use MoM to estimate the parameter of Poisson distribution.

$$f\left(x;k\right)=\frac{e^{-k}k^{x}}{x!};\ x=0,1,2,\cdots \text{ and } k>0$$

Sol. $\hat{k}=\bar{X}$? Is there an alternative estimator of k ?

(hint: compare sample and population variance)

HW.2. Use MoM to estimate the parameters of Binomial distribution.

$$f\left(x;n,p\right)=\binom{n}{x}p^{x}\left(1-p\right)^{n-x};\ x=0,1,2,\cdots,n \text{ and } 0<p<1$$

Example: Method of Moments

HW.3. Use MoM to estimate the parameter of Rayleigh distribution.

$$f(x;\alpha)=\frac{x}{\alpha^2}\exp\left(-\frac{x^2}{2\alpha^2}\right);\alpha>0,x>0$$

Sol. $\hat{\alpha}=\bar{X}\sqrt{\frac{2}{\pi}}$? Is it an unbiased estimator?

HW.4. Use MoM to estimate the parameter of Maxwell distribution.

$$f(x;\alpha)=\sqrt{\frac{2}{\pi}}\frac{x^2}{\alpha^3}\exp\left[-\frac{1}{2}\left(\frac{x}{\alpha}\right)^2\right];\text{for }\alpha>0,x>0$$

Sol. $\hat{\alpha}=\frac{\bar{X}}{2}\sqrt{\frac{2}{\pi}}$? Is it an unbiased estimator?

Example: Method of Moments

HW.5. Use MoM to estimate the parameters of Gaussian distribution.

$$f(x;\mu,\sigma)=\begin{cases}\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; -\infty<x<\infty, -\infty<\mu<\infty, \sigma>0 \\ 0 \qquad ;\ \text{otherwise}\end{cases}$$

HW.6. Use MoM to estimate the parameter of gamma distribution.

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}};\ x>0,\ \alpha>0, \beta>0 \\ 0 \qquad ;\ \text{otherwise}\end{cases}$$

Maximum Likelihood Estimation

1. MLE is the most widely used parameter estimation method as on today.

2. The basic principle is to maximize the likelihood of the parameters, denoted by $L(\theta \mid x)$, as a function of the model parameters θ.

3. Note that the θ can be a single parameter or a vector of parameters; $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$.

4. The likelihood function $L(\theta \mid x)$ is defined as $L(\theta \mid x) = \prod_{i=1}^{n} f(x_i;\theta)$

5. As log is a one – to – one function, maximization of log – likelihood $(\ln L)$ is often preferred for computational ease.

The MLE method was recently proposed by Fisher in 1920s

Examples: MLE

Ex.2. Let $X_1,\ X_2,\ X_m$ be a random sample of size m from a binomial distribution of parameters n (known) and p. Find the maximum likelihood estimator for p. Is it an unbiased estimator?

Sol.

Step 1 : The log-likelihood function for binomial distribution is

$$L(p|x) = \prod_{i=1}^{m} f(x_i, p), \qquad 0 < p < 1$$

$$= \prod_{i=1}^{m}\left(\binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}\right) = \left(\prod_{i=1}^{m}\binom{n}{x_i}\right) p^{\sum_{i=1}^{m} x_i}(1-p)^{nm-\sum_{i=1}^{m} x_i}$$

$$\ln L(p|x) = \ln\left(\prod_{i=1}^{m}\binom{n}{x_i}\right) + \left(\sum_{i=1}^{m} x_i\right)\ln p + \left(nm - \sum_{i=1}^{m} x_i\right)\ln(1-p)$$

Examples: MLE

Step 2 : The corresponding log – likelihood equation is

$$\frac{\partial}{\partial p}\ln L(p|x)=0$$

$$\Rightarrow \frac{L'(p)}{L(p)}=\frac{\left(\sum_{i=1}^{m}x_i\right)}{p}-\frac{\left(nm-\sum_{i=1}^{m}x_i\right)}{1-p}=0$$

$$\Rightarrow \left(nm-\sum_{i=1}^{m}x_i\right)p=\left(\sum_{i=1}^{m}x_i\right)(1-p)$$

Step 3 : The estimator of p is then obtained as $\boxed{\hat{p}=\frac{\left(\sum_{i=1}^{m}X_i\right)}{nm}=\frac{\bar{X}}{n}}$

Why does this estimator maximize likelihood function?

Examples: MLE

Ex.3. Use MLE to estimate parameters of exponential distribution

$$f\left(x;\alpha\right)=\frac{1}{\alpha}e^{-\frac{x}{\alpha}};\ \ x>0,\alpha>0$$

Sol.

Step 1: The log – likelihood function for exponential distribution is

$$\ln L\left(\theta\,|\,x\right)=\ln L\left(\alpha;x_1,x_2,\ldots,x_n\right)=-n\ln\alpha-\sum_{i=1}^{n}\frac{x_i}{\alpha}$$

Step 2: The corresponding log – likelihood equation is

$$\frac{\partial}{\partial\alpha}\ln L=0$$

Step 3: The estimator of α is then obtained as $\boxed{\hat{\alpha}=\frac{1}{n}\sum_{i=1}^{n}X_i}$

Examples: MLE

HW.7. Use MLE to estimate the parameter of Poisson distribution.

HW.8. Use MLE to estimate the parameter of Gaussian distribution.

$$f(x;\mu,\sigma)=\begin{cases}\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & ;-\infty<x<\infty,-\infty<\mu<\infty,\sigma>0\\ 0 & ;\ \text{otherwise}\end{cases}$$

$$\boxed{\hat{\mu}=\bar{X}\quad and\quad \hat{\sigma}^2=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n}=\frac{n-1}{n}S^2.}$$

Thus M-L estimator for σ^2 is not unbiased.

Example: MLE

HW.9. Use MLE to estimate the parameter of Rayleigh distribution.

$$f(x;\alpha) = \frac{x}{\alpha^2}\exp\left(-\frac{x^2}{2\alpha^2}\right); \alpha > 0, x > 0$$

Sol. $\hat{\alpha} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$?

HW.10. Use MLE to estimate the parameter of Maxwell distribution.

$$f(x;\alpha) = \sqrt{\frac{2}{\pi}}\frac{x^2}{\alpha^3}\exp\left[-\frac{1}{2}\left(\frac{x}{\alpha}\right)^2\right]; \text{for } \alpha > 0, x > 0$$

Sol. $\hat{\alpha} = \sqrt{\frac{1}{3n}\sum_{i=1}^{n} X_i^2}$?

Example: MLE

HW.11. Use MLE to estimate the parameters of inverse Gaussian distribution

$$f(t;\lambda,\mu)=\sqrt{\frac{\lambda}{2\pi t^3}}\exp\left[-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right];\ t>0,\lambda>0,\mu>0$$

HW.12. Use MLE to estimate the parameters of lognormal distribution

$$f(t;\alpha,\beta)=\frac{1}{t\beta\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln t-\alpha}{\beta}\right)^2\right];t>0,\beta>0$$

Example: MLE

HW.12. Use MoM and MLE to estimate the parameter of exponential distribution.

$$f\left(x;\lambda\right)=e^{-\lambda x};\ \ x>0,\lambda>0$$

Discuss whether the estimator $\hat{\lambda}$ unbiased, in case of MoM and MLE.

HW.13. Let a random sample of size n is taken from a uniform distribution on $[0,\theta]$. Find $\hat{\theta}_{MLE}$. What is the distribution of $\hat{\theta}_{MLE}$? Is it an unbiased estimator of θ? If not, find one unbiased estimator.

Examples: MLE

Ex.4. Use MLE to estimate the parameters of Weibull distribution

$$f(t;\alpha) = \frac{\beta}{\alpha^{\beta}} t^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}\,;\ t > 0, \alpha > 0, \beta > 0$$

Sol.

Step 1 : The log – likelihood function is

$$\ln L(\theta \mid t) = \ln L(\alpha, \beta; t_1, t_2, \ldots, t_n)$$

$$= n \ln \beta - n\beta \ln \alpha + (\beta - 1)\sum_{i=1}^{n} \ln(t_i) - \sum_{i=1}^{n} \left({t_i}/{\alpha} \right)^{\beta}$$

Step 2 : The corresponding log – likelihood equations are

$$\frac{\partial}{\partial \alpha} \ln L = 0 \ \text{ and } \ \frac{\partial}{\partial \beta} \ln L = 0$$

Examples: MLE

This gives

$$\frac{\partial}{\partial \alpha}\ln L\left(\alpha,\beta;t_1,t_2,\ldots,t_n\right)=0 \Rightarrow \alpha^{\beta}-\frac{1}{n}\sum_{i=1}^{n} t_i^{\beta}=0$$

$$\frac{\partial}{\partial \beta}\ln L\left(\alpha,\beta;t_1,t_2,\ldots,t_n\right)=0 \Rightarrow \frac{n}{\beta}+\sum_{i=1}^{n}\left[1-\left(\frac{t_i}{\alpha}\right)^{\beta}\right]\ln\left(\frac{t_i}{\alpha}\right)=0$$

Step 3 : The estimates of α and β are then obtained from

$$\frac{1}{\beta}+\frac{1}{n}\sum_{i=1}^{n}\ln\left(t_i\right)-\frac{\sum_{i=1}^{n} t_i^{\beta}\ln\left(t_i\right)}{\sum_{i=1}^{n} t_i^{\beta}}=0 \text{ and } \alpha=\left(\frac{1}{n}\sum_{i=1}^{n} t_i^{\beta}\right)^{\frac{1}{\beta}}$$

How to solve now? (Need to learn more! Numerical techniques?)

Example: MLE

HW.14. Let a random sample of size n is taken from a uniform distribution on $[\alpha, \beta]$. Find $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$. What are the distributions of $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$? Are they unbiased estimators?

Estimating Functions of Parameters

innovate achieve lead

Let $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ be the MLEs of the parameters $\theta_1, \theta_2, \cdots, \theta_{\mathrm{m}}$. Then the MLE of any function $h(\theta_1, \theta_2, \cdots, \theta_{\mathrm{m}})$ of these parameters is the function $h(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m)$ of the MLEs.

HW.15. What is $\hat{\sigma}_{MLE}$ in a normal distribution? Is it an unbiased estimator?

HW.16. What is the MLE estimator for mean μ of a gamma (α, β) distribution? Is it an unbiased estimator of μ?

Recall: Sample Proportion

The statistic that estimates the parameter p, a proportion of a population that has some property, is the sample proportion

$$\hat{p} = \frac{\text{number in sample with the trait (success)}}{\text{sample size}} = \frac{X}{n}$$

Properties:

(i) As the sample size increases (n large), the sampling distribution of $\hat{p}$ becomes approximately normal (WHY?)

(ii) The mean of $\hat{p}$ is p, and variance of $\hat{p}$ is $\frac{p(1-p)}{n}$ (WHY?)

(iii) Can we get a point estimators of p? (See Ex. 6.15, page no. 258)

Example 6.15 (page 258)

A sample of ten new bike helmets manufactured by a certain company is obtained. Upon testing, it is found that the first, third, and tenth helmets are flawed, whereas the others are not.

Let p = P(flawed helmet), i.e., p is the proportion of all such helmets that are flawed.

Define (Bernoulli) random variables $X_1, X_2, \ldots, X_{10}$ by

$$X_1 = \begin{cases} 1 \text{ if 1st helmet is flawed} \\ 0 \text{ if 1st helmet isn't flawed} \end{cases} \quad \ldots \quad X_{10} = \begin{cases} 1 \text{ if 10th helmet is flawed} \\ 0 \text{ if 10th helmet isn't flawed} \end{cases}$$

Example 6.15 (page 258)

Then for the obtained sample, $X_1 = X_3 = X_{10} = 1$ and the other seven X_i's are all zero.

The probability mass function of any particular X_i is $p^{x_i}(1-p)^{1-x_i}$, which becomes p if $x_i = 1$ and $1 - p$ when $x_i = 0$.

Now suppose that the conditions of various helmets are independent of one another.

This implies that the X_i's are independent, so their joint probability mass function is the product of the individual pmf's.

Example 6.15 (page 258)

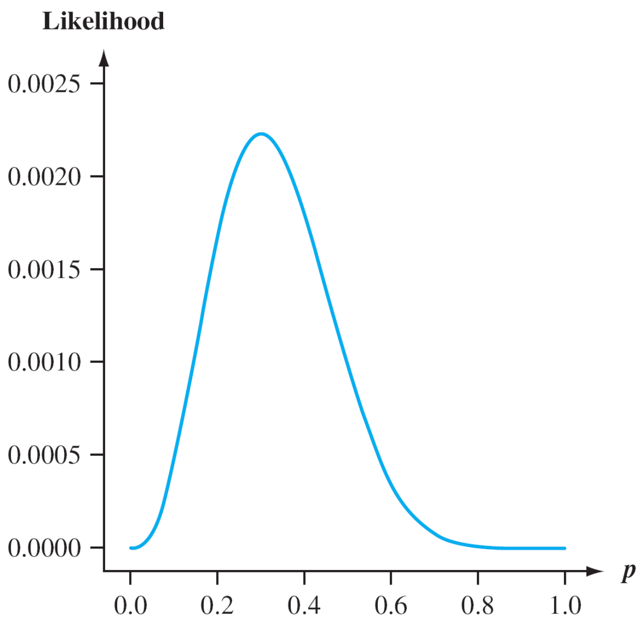
Thus the joint pmf evaluated at the observed X_i's is

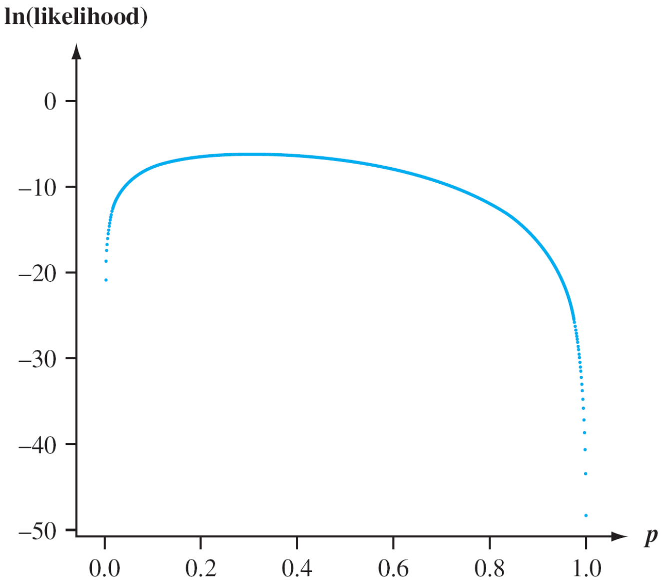
$$f(x_1, \ldots, x_{10}; p) = p(1-p)p \cdots p = p^3(1-p)^7 \qquad \textbf{(6.4)}$$

Suppose that $p = .25$. Then the probability of observing the sample that we actually obtained is $(.25)^3(.75)^7 = .002086$.

If instead $p = .50$, then this probability is $(.50)^3(.50)^7 = .000977$. For what value of p is the obtained sample most likely to have occurred? That is, for what value of p is the joint pmf (6.4) as large as it can be? What value of p maximizes (6.4)?

Example 6.15 (page 258)





That is, our point estimate is $\hat{p} = .30$. It is called the *maximum likelihood estimate* because it is the parameter value that maximizes the likelihood (joint pmf) of the observed sample.

Large Sample Behaviour of MLE

innovate achieve lead

Although the principle of maximum likelihood estimation has considerable intuitive appeal, the following proposition provides additional rationale for the use of mle's.

Proposition

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 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 approximately the MVUE of θ.