innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

BITS Pilani
Pilani Campus

Chapter 7 (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.

Unbiased estimator:

An estimator $\hat{\theta}$ is an unbiased estimator for a population parameter θ if and only if

$$E\left(\hat{\theta}\right)=\theta.$$

Sample Mean

Ex.7.1. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a distribution with mean μ. The sample mean, $\bar{X}$, is an unbiased estimator for μ.

Sol. Here $X_1, X_2, \cdots, X_n$ is a random sample of size n from a distribution with mean μ. Then, the sample mean, $\bar{X}$, is

$$\bar{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

$$\begin{aligned} E\left(\bar{X}\right) &= \frac{1}{n} E\left(X_1 + X_2 + \cdots + X_n\right) \\ &= \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + E\left(X_3\right) + \cdots + E\left(X_n\right)\right] \\ &= \mu \end{aligned}$$

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

Variance of Sample Mean

Ex.7.2. Show that $\sigma_{\bar{X}}^2 = \mathrm{Var}\left(\bar{X}\right) = \dfrac{\sigma^2}{n}$

- From this theorem, it follows that larger the sample size, sample mean can be expected to lie closer to the population mean.
- Thus choosing large sample makes estimation more reliable.

Variance of Sample Mean

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

Ex.7.3. Show that sample variance S^2 of a random sample of size n from a population X is an unbiased estimator for population variance.

$$E\left(S^2\right) = \sigma_X^2$$

More Efficient Unbiased Estimator

A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimator of the parameter θ than the statistic $\hat{\theta}_2$ if

(a) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ

(b) variance of the sampling distribution of the first estimator is no larger than that of the second and is smaller for at least one value of θ .

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.7.4. Use method of moments to estimate the parameter of exponential distribution.

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

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.7.1. 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.7.2. 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.7.3. 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.7.4. 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.7.5. 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.7.6. 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.

Examples: MLE

Ex.7.5. 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.7.6. Use MLE to estimate parameters of exponential distribution

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

Sol.

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

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

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.7.7. Use MLE to estimate the parameter of Poisson distribution.

HW.7.8. Use MLE to estimate the parameter of Gaussian distribution.

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

Note that the ML estimator for σ^2 is $\boxed{\hat{\sigma}^2=\dfrac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n}=\dfrac{n-1}{n}S^2.}$

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

Example: MLE

innovate achieve lead

HW7.9. 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. $\hat{\alpha}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}\,?$

HW7.10. 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

HW7.11. Use MLE to estimate the parameters of inverse Gaussian distribution

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

HW7.12. Use MLE to estimate the parameters of lognormal distribution

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

Examples: MLE

Ex.7.7. Use MLE to estimate the parameters of Weibull distribution

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

Sol.

Step 1: The log – likelihood function is

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

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

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

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

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

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

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

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

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Self Reading Material:
Estimation of Tiger Population in India

https://esajournals.onlinelibrary.wiley.com/doi/epdf/10.1890/11-2110.1

Section 7.3: Functions of RV

Theorem 7.3.1 : Let X and Y be random variables with moment generating functions $m_X(t)$ and $m_Y(t)$, respectively. If $m_X(t) = m_Y(t)$ for all t in some open interval about 0, then X and Y have the same distribution.

Theorem 7.3.2 : Let X_1 and X_2 be independent random variables with moment generating functions $m_{X_1}(t)$ and $m_{X_2}(t)$, respectively. Let $Y = X_1 + X_2$. The moment generating function for Y is given by :

$$m_Y(t) = m_{X_1}(t).\, m_{X_2}(t)$$

Ex 7.3.2 : (Distribution of the sum of independent normally distributed random variables)

Let $X_1, X_2, X_3,\ldots, X_n$ be independent normal random variables with means $\mu_1, \mu_2, \mu_3,\ldots, \mu_n$ and variances $\sigma^2_1, \sigma^2_2, \sigma^2_3,\ldots,\sigma^2_n$ respectively.

Let $Y = X_1+X_2+X_3+\ldots+X_n$. Note that the moment generating function for X_i is given by:

$$m_{X_i}(t) = e^{(\mu_i t + (\sigma^2_i t^2/2))} \quad i = 1,2,3,\ldots,n$$

and the moment generating function for *Y* is (why?)

$$m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t) = \exp\left[\left(\sum_{i=1}^{n}\mu_i\right)t + \left(\sum_{i=1}^{n}\sigma_i^2\right)\frac{t^2}{2}\right]$$

The function on the right is nothing but the moment generating function for a random variable Y with mean $\mu = \sum_{i=1}^{n}\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n}\sigma_i^2$

Section 7.3: Functions of RV

Theorem 7.3.3:

Let *X* be a random variable with moment generating function $m_x(t)$. Let $Y = \alpha + \beta X$. The moment generating function for Y is

$$m_Y(t) = e^{\alpha t} m_x(\beta t)$$

Section 7.3: Functions of RV

Theorem 7.3.4: (Distribution of $\overline{X}$ -normal population)

Let X_1, X_2, X_3,…, X_n be a random sample of size ‘n’ from a normal distribution with mean μ and variance σ^2.

Then $\overline{X}$ is normally distributed with mean μ and variance σ^2/n.

HW 7.13 (Q 39: Distribution of a sum of independent random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be a collection of independent random variables with moment generating functions $m_{X_i}(t)$ (i=1,2,3,.....,n, respectively). Let a_0, $a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n.$$

Show that the moment generating function for Y is given by

$$m_Y(t) = e^{a_0 t} \prod_{i=1}^{n} m_{X_i}(a_i t)$$

HW 7.14 (Q 41: Distribution of a linear combination of independent normally distributed random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means μ_i and σ_i^2 $(i=1,2,3,\ldots,n,$ respectively). Let a_0, $a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

Show that Y is normal with mean

$$\mu = a_0 + \sum_{i=1}^{n} a_i \mu_i \text{, and variance}$$

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

HW 7.15 (Q 44: Distribution of a sum of independent chi-squared random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be independent chi-squared random variables with $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$ degrees of freedom ,respectively.

Let $Y = X_1 + X_2 \ldots + X_n$.

Show that Y is a chi-squared random variable with degrees of freedom where $\gamma = \sum \gamma_i$

Homework

HW 7.16 (Q 52) Consider the random variable X with density given by

$$f(x) = 1/\theta; \quad 0 < x < \theta$$

(a) Find E[X].

(b) Find the method of moments estimator for θ.

(c) Find the method of moments estimate for θ based on these data

1, 0.5, 1.4, 2.0, 0.25

Homework

innovate achieve lead

HW 7.17 (Q 58). Let $X_1,X_2,X_3,\ldots.,X_{100}$ be a random sample of size 100 from gamma distribution with $\alpha=5$ and $\beta=3$.

(a) Find the mgf of $Y=\sum_{i=1}^{100} X_i$

(b) What is the distribution of *Y*?

(c) Find the mgf of $\overline{X}=Y/n$

(d) What is the distribution of $\overline{X}$?

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Interval Estimation

- A point estimate cannot be expected to provide the exact value (close value) of the population parameter.
- **Usually**, an interval estimate can be obtained by adding and subtracting a margin of error to the point estimate. Then,

$$\boxed{\text{Interval Estimate} = \text{Point Estimate} +/- \text{Margin of Error}}$$

- Interval estimation provides us information about **how close** the point estimate is to the value of the parameter.
- Why we use the term ***confidence interval?***

Interval (CI) Estimation

– Instead of considering a statistic as a point estimator, we may use *random intervals* to trap the parameter.

– In this case, the end points of the interval are RVs and we can talk about the probability that it traps the parameter value.

Confidence Interval : A 100(1- α)% confidence interval for a parameter is a random interval $[L_1,L_2]$ such that

$$P[L_1 \le \theta \le L_2] = 1-\alpha$$, regardless the value of θ.

Theorem 7.4.1: Interval estimation for μ: σ known

Let $X_1,\ X_2,\ \cdots, X_n$ be a random sample from a normal population with mean μ $\left(\text{unknown}\right)$ and the variance σ^2 $\left(\text{known}\right)$. Then, using Thm 7.3.4,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N\left(0,1\right)$$

Taking two points $\pm z_{\alpha/2}$ symmetrically about the origin, we get

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

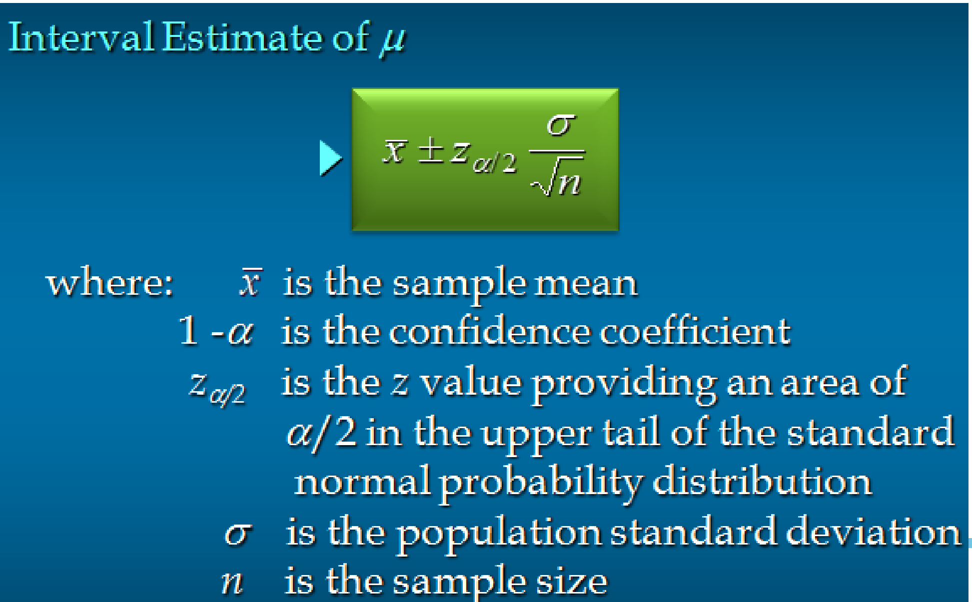
Here $\left(1-\alpha\right)$ is known as confidence level, and α is the level of significance.

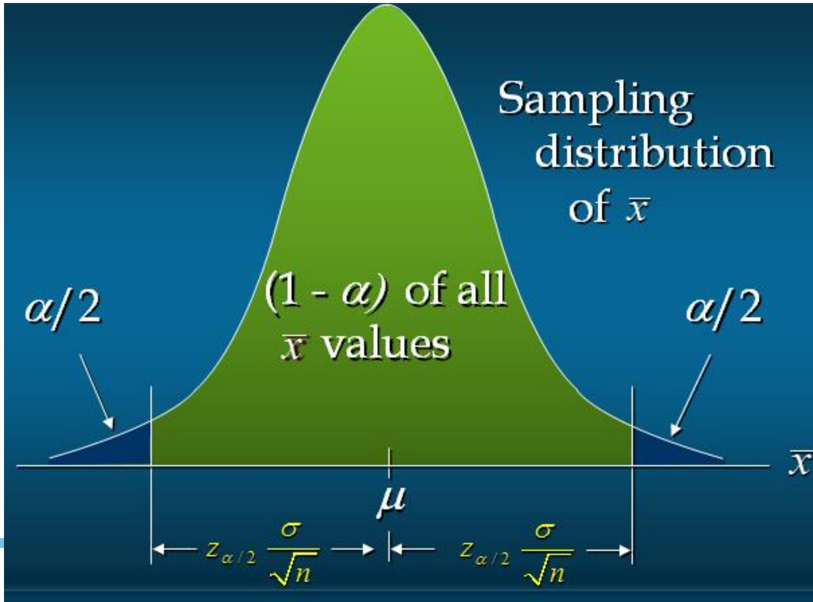
Interval estimation for μ: σ known

$$P\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)=1-\alpha$$

Hence, the confidence interval for population mean μ having confidence level $100\times(1-\alpha)\%$ is given as $\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$.

The endpoints of the confidence interval is called confidence limits.





Interval estimation for μ: σ known

Most commonly used confidence levels:

Confidence Level	α	$\alpha/2$	Table Look-up Area	$z_{\alpha/2}$
90%	.10	.05	.9500	1.645
95%	.05	.025	.9750	1.960
99%	.01	.005	.9950	2.576

Hence, 95% CI for μ is given as $\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \right)$.

That is, $P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \right) = 0.95$

Practice Problems

HW 7.18 (Q 53) Studies have shown that the random variable X, the processing time required to do a multiplication on a new 3-D computer, is normally distributed with mean μ and standard deviation 2 microseconds. A random sample of 16 observations is to be taken

(a) These data are obtained

42.65	45.15	39.32	44.44
41.63	41.54	41.59	45.68
46.50	41.35	44.37	40.27
43.87	43.79	43.28	40.70

Based on these data, find an unbiased estimate for μ.

(b) Find a 95% confidence interval for μ.

Practice Problems

HW7.19. The mean of a sample size 50 from a normal population is observed to be 15.68. If the s.d. of the population is 3.27, find (a) 80% (b) 95%, (c) 99% confidence interval for the population mean. Can you find out the respective margin of errors? What is the length of CI for each case?

Sol. (b)

Step 1: Here $n = 50$, $\bar{x} = 15.68$, σ=3.27, and $\alpha = 0.05$. We need CI for μ.

Step 2: As $\alpha = 0.05,$ we need to find $z_{\alpha/2}$ such that $P\left(Z < z_{\alpha/2}\right)$=0.975.

From cumulative normal distribution table, we see $z_{\alpha/2} = 1.96.$

Step 3: The CI for $\mu\left(\sigma \text{ known}\right)$ is $\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{0.025}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{0.025}\right) = (14.77, 16.59)$

Impracticality of Assumptions in CI

In practice, we usually face mainly two problems in application of previous C.I. formula.

- What if the population is not normal?
- What if the population variance is unknown?

Need to introduce Central Limit Theorem.

What is the beauty and importance of CLT?

Central Limit Theorem (CLT)

- Regardless of the population distribution model, as the sample size increases, the sample mean tends to be normally distributed around the population mean, and its standard deviation shrinks as *n* increases.
- Two conditions must be satisfied to apply CLT – (a) samples must be i.i.d. (b) sample size must be large enough (usually, n ≥ 30, but depends on problem!)

Let $X_1, X_2, \cdots, X_n$ be a random sample (that is, X_i are i.i.d) of size n from a distribution with mean μ and variance σ^2. Then for large *n*, sample mean $\bar{X}$ is approximately normal with mean μ and variance σ^2/n; $\bar{X} \to N\left(\mu, \sigma/\sqrt{n}\right)$

Furthermore, for large *n*, the random variable $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$.

CLT: Examples

- A certain brand of tires has a mean life of 25,000 km with a s.d. of 1600 km. What is the probability that the mean life of 64 tires is less than 24,600 km?

- A hawker sells dolls earns at various prices with a mean of Rs.700 and s.d. of Rs. 250. During Christmas week, assume that he sells 60 dolls. What is the probability that he earns Rs. 45000 or more in that week?

Central Limit Theorem (CLT)

- **Randomization** – we assume that **samples constitute a random sample (i.i.d.)** from the population.

- **Large enough sample size – how large is large?**

  - If the population is **normal**, then the sampling distribution $\bar{X}$ will also be normal, no matter what is the sample size.
  - If the population is approximately **symmetric**, the distribution becomes approximately normal for relatively small values of n.
  - When the population is **skewed**, the sample size must be **at least 30** before the sampling distribution of $\bar{X}$ becomes approximately normal.

CLT: Step by Step

Step 1: Identify parts of the problem. Your question should state:

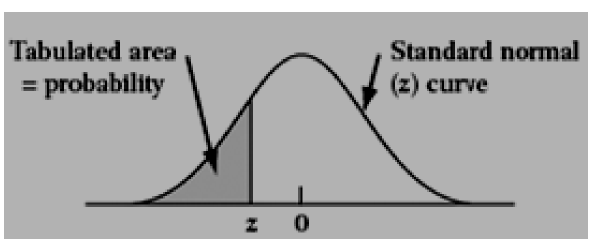
- The mean (average or μ)
- The standard deviation (σ)
- The sample size (n)

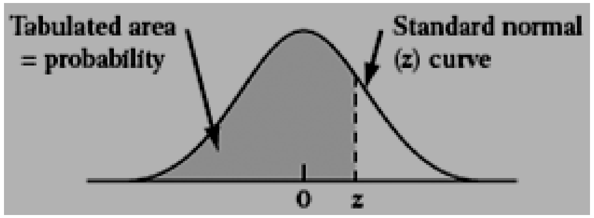
Step 2: Find $\bar{X}$ and express the problem in terms of "greater than" or "less than" the sample mean $\bar{X}$.

Step 3: Use CLT to find the distribution of $\bar{X}$ and $\bar{X} \rightarrow N\left(\mu, \sigma/\sqrt{n}\right)$

Step 4: Convert the normal variate $\bar{X}$ to a standard normal variate $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Now you may draw a graph, centre with the 0 (mean of ***Z***) and shade the appropriate area to find the required probability.





Problem Solving

HW7.20. A certain brand of tyres has a mean life of 25,000 km with a s.d. of 1600 km. What is the probability that the mean life of 64 tyres is less than 24,600 km?

Sol.

Step 1: Here $X_1, X_2, \cdots, X_{64}$ constitute a random sample, and it is given that $E(X_i) = 25,000$ and $\sigma_{X_i} = 1600.$

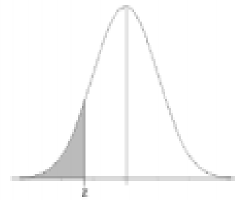
Step 2: $\bar{X} = \frac{1}{64}\sum_{i=1}^{64} X_i$ and our interest is to find out $P(\bar{X} < 24600)$.

Step 3: As we have a random sample of size 64 (sufficiently large n), we can use CLT to find the distribution of $\bar{X}$, that is, $\bar{X} \sim N\left(25000, \frac{1600}{\sqrt{64}}\right) \Rightarrow \bar{X} \sim N(25000, 200)$

Step 4: $P(\bar{X} < 24600) = P\left(\frac{\bar{X} - 25000}{200} < \frac{24600 - 25000}{200}\right)$

$$= P(Z < -2) = 0.0228 \text{ (using standard normal cdf table)}$$

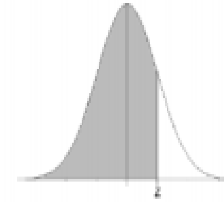
Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Problem Solving

innovate achieve lead

HW7.21. A hawker sells dolls at varying prices with a mean of Rs.700 and s.d.of Rs. 250. During Christmas week, assume that he sells 60 dolls. What is the probability that he earns Rs. 45000 or more in that week?

Step 1: Here $X_1, X_2, \cdots, X_{60}$ constitute a random sample, and it is given that $E(X_i) = 700$ and $\sigma_{X_i} = 250.$

Step 2: $\bar{X} = \frac{1}{60}\sum_{i=1}^{60} X_i$ and our aim is to find out $P\left(\sum_{i=1}^{60} X_i > 45000\right)$ i.e., $P(\bar{X} > 750).$

Step 3: As we have a random sample of size 60 (sufficiently large *n*), we can use CLT to find the distribution of $\bar{X}$, that is, $\bar{X} \sim N\left(700, \frac{250}{\sqrt{60}}\right) \Rightarrow \bar{X} \sim N(700, 32.27)$

Step 4: $P(\bar{X} > 750) = P\left(\frac{\bar{X} - 700}{32.27} > \frac{750 - 700}{32.27}\right) = P(Z > 1.55)$

$= P(Z < -1.55) = 0.0606$ (using standard normal cdf table)

Problem Solving

HW7.22. A population of 30 year – old males has a mean salary of Rs. 75000 with a standard deviation of Rs. 10000. If a sample of 100 men is taken, what is the probability that their mean salary will be less than Rs. 77500 ?

HW7.23. A certain group of welfare recipients receives pension benefits of Rs. 45000 per month with a standard deviation of Rs. 7500. If a random sample of 25 people is taken, what is the probability that their mean pension benefit will be greater than Rs. 47000 or less than Rs. 43000 per month ?

HW7.24. A certain population of dogs weigh an average of 5 kg, with a standard deviation of 2 kg. If 40 dogs are chosen at random, what is the probability they have an average weight of greater than 6.0 kg or less than 4.5 kg ?

Problem Solving

HW 7.25 (Q 49) When fission occurs, many of the nuclear fragments formed have too many neutrons for stability. Some of these neutrons are expelled almost instantaneously. These observations are obtained on X, the number of neutrons released during fission of plutonium-239:

3	2	2	2	2	3	3	3
3	3	3	3	4	3	2	3
3	2	3	3	3	3	3	1
3	3	3	3	3	3	3	3
3	3	2	3	3	3	3	3

(a) Is X normally distributed? Explain. (b), (c), (d)