BITS Pilani
Pilani Campus

MATH F113
Probability and Statistics

Dr. Shivi Agarwal
Department of Mathematics

7.3. Functions of Random Variables- Distribution of sample mean

The basic principle which allows us to determine the distribution many statistics:

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 :

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

innovate achieve lead

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

Let X_1, X_2, X_3,…, X_n be independent normal random variables with means μ_1, μ_2, μ_3,…, μ_n and variances $\sigma^2{}_1$, $\sigma^2{}_2$,$\sigma^2{}_3$,…,$\sigma^2{}_n$ respectively.

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

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

and the moment generating function for Y is

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

The function on the right is the moment generating function for a random variable with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma^2_i$

Theorem 7.3.3 : Let X random variable with moment generating function $m_x(t)$.

Let $Y = \alpha + \beta X$. The moment generating function for Y is

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

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

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size 'n' from a normal distribution with mean μ and variance σ^2.

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

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

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

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,…,n, respectively). Let $a_0, a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n.$$

Show that Y is normal with mean

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

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

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 $\gamma = \sum \gamma_i$

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

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 $\overline{X}$?

7.4

- Interval Estimation and Central Limit Theorem:
  - 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 r.v.s but 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 :

$$P[L_1 \leq \theta \leq L_2] = 1 - \alpha$$

regardless of the value of θ.

Notes :

- Here θ is not a random variable. Its value is fixed, though unknown. The interval is random. The meaning of the probability is that for several choices of this interval, proportion of those which will contain θ is $(1- \alpha)$.

- From a particular sample we can read off a particular confidence interval. Here we can say with confidence $(1- \alpha)100\%$ that this interval contains θ. Since here nothing remains random, we talk of confidence rather than probability.

Theorem 7.4.1 : [100(1- α)% confidence interval on μ when σ^2 is known]

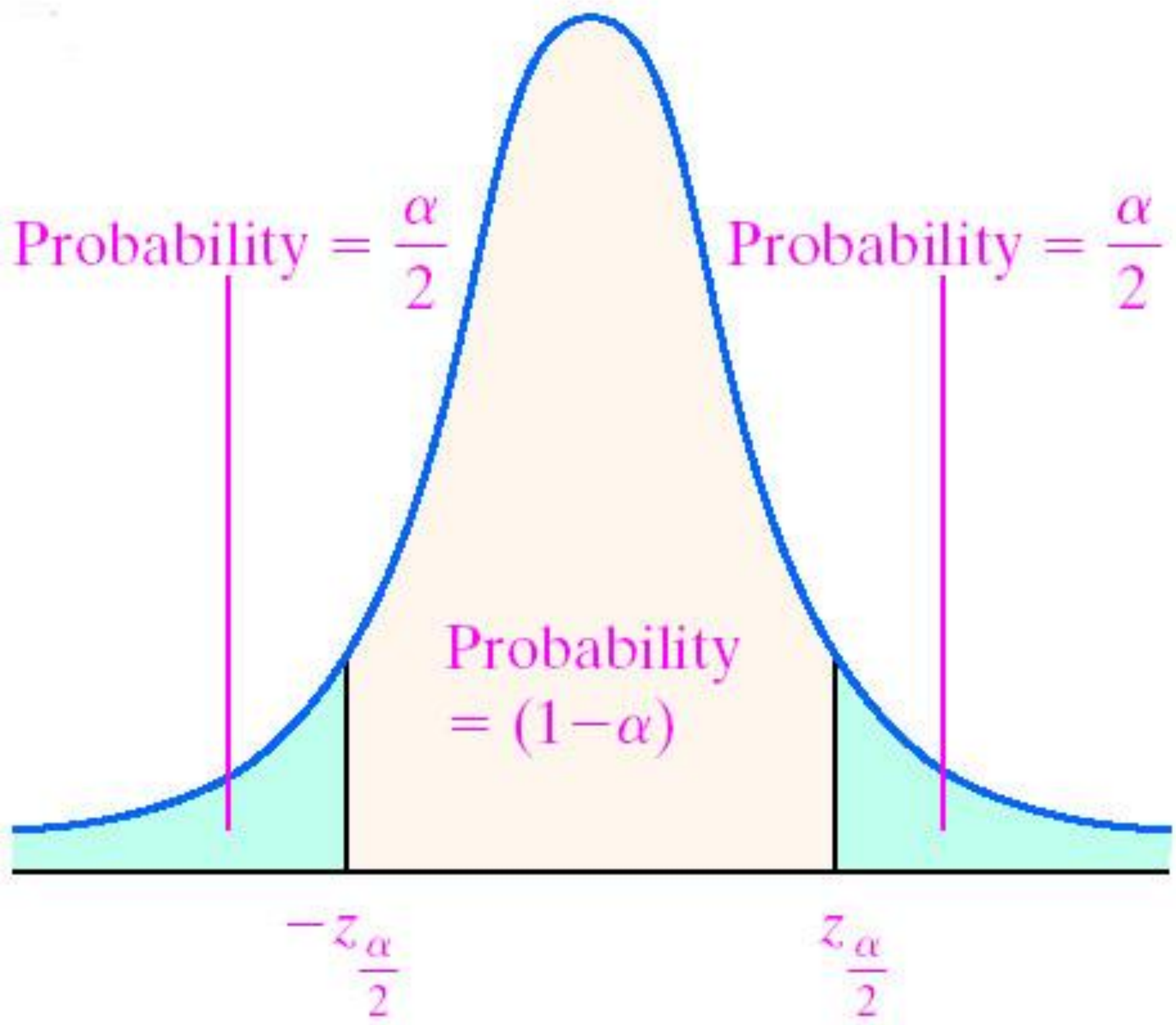
Let X_1, X_2, X_3,…, X_n be a random sample of size ‘n’ from a normal distribution with mean μ (unknown) and variance σ^2 (known).

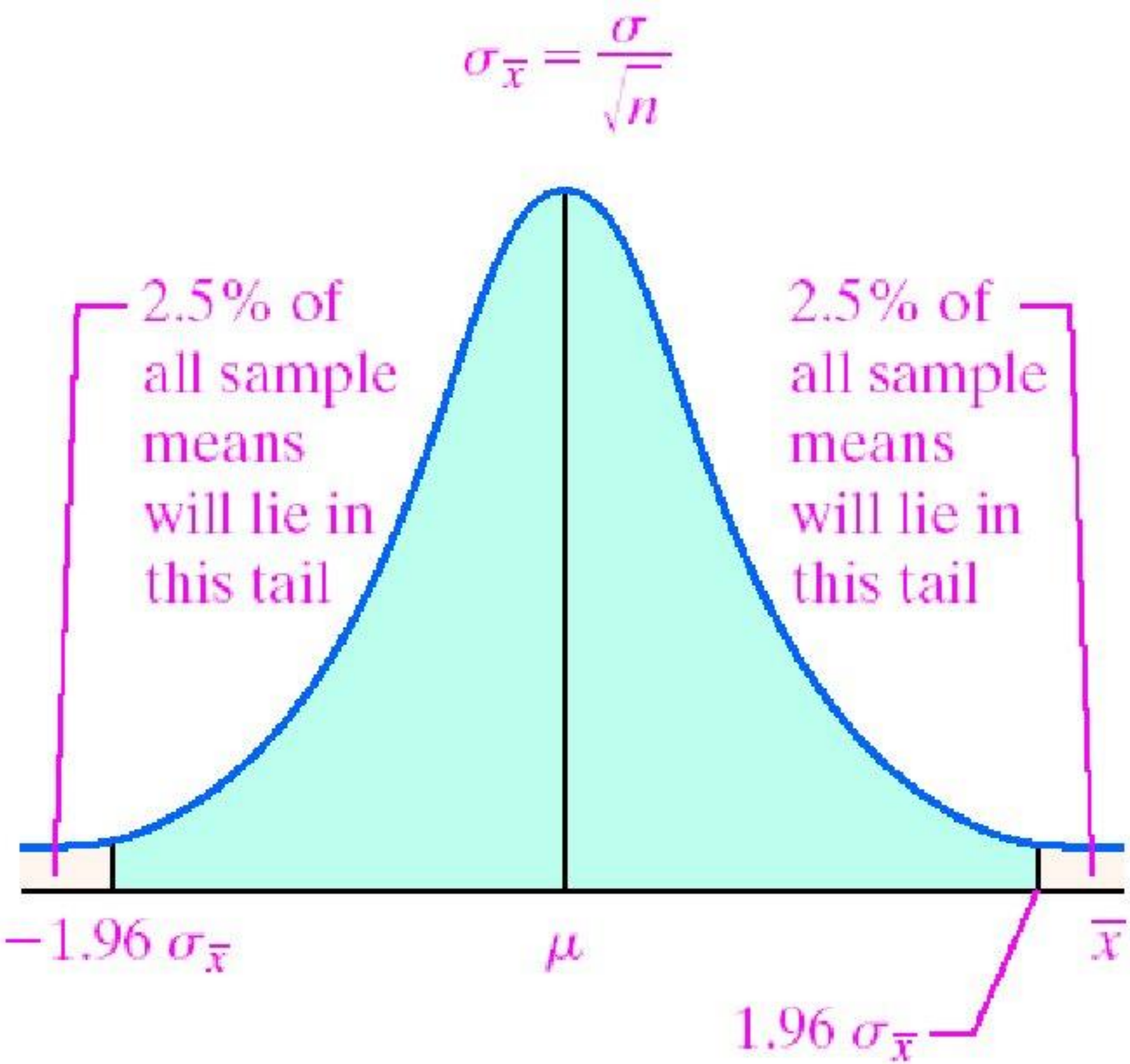
A 100(1- α)% confidence interval on μ is given by:

$$\bar{X} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$$

$$\left[\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \qquad \bar{X} + z_{\alpha/2}\sigma/\sqrt{n} \right]$$

$$i.e.\ \bar{X} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}$$


Probability $= \frac{\alpha}{2}$
Probability $= \frac{\alpha}{2}$
Probability
$= (1-\alpha)$
$-z_{\frac{\alpha}{2}}$
$z_{\frac{\alpha}{2}}$


$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
2.5% of all sample means will lie in this tail
2.5% of all sample means will lie in this tail
$-1.96\,\sigma_{\bar{x}}$
μ
$\bar{x}$
$1.96\,\sigma_{\bar{x}}$

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 (unknown) and the variance σ^2 (known). Then, using Thm 7.3.4,

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

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

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

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.

Would you be surprised to read that the average time required to process a multiplication on this system is 42.2 microseconds? Explain, based on the C.I.

Impracticality of assumptions

In practice, we face 2 problems in application of this C.I. formula and need some remedies.

- The population is not normal
- Population variance is unknown.

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 big enough ($n \geq 25$)

Theorem 7.4.2 : (Central Limit Theorem) Let X_1, X_2, X_3,…, X_n be a random sample of size 'n' from a distribution with mean μ and variance σ^2.

Then for large 'n', $\bar{X}$ is approximately normal with mean μ and variance σ^2/n.

Furthermore for large 'n', the random variable $(\bar{X} - \mu)/ (\sigma/\sqrt{n})$ is approximately standard normal.

The approx works satisfactorily for $n \geq 25$.

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)

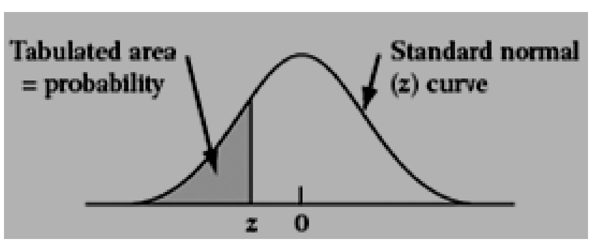
Step 2: Find $\bar{X}$ and express the problem in terms of "greater than" or "less than" the sample mean $\bar{X}$.

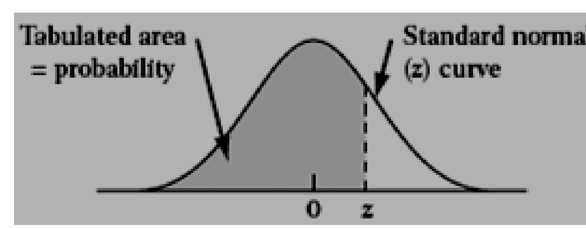
Step 3: Use CLT to find the distribution of $\bar{X}$ and $\bar{X} \to N\left(\mu, \sigma/\sqrt{n}\right)$

Step 4: Convert the normal variate $\bar{X}$ to a standard normal variate

Now you may draw a graph, center with the 0 (mean of ***Z***) $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$

and shade the appropriate area to find the required probability.





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) Estimate the mean no. of neutrons expelled during fission of plutonium-239.

(c) Assume that $\sigma = 0.5$. Find a 99% confidence interval on μ. What theorem justifies the procedure you used to construct this interval

Problem Solving

Ex. 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?

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

$= P(Z < -2) = 0.0228$ (using standard normal cdf table)

Q 55. (Central Limit Theorem) In an attempt to approximate the proportion p of improperly sealed packages produced on an assembly line, a random sample of 100 packages is selected and inspected . Let

$X_i=1$ if the i^{th} package selected is improperly sealed

$X_i=0$ otherwise

(a) What is the distribution of $\bar{X}$?

(b) Based on Central Limit Theorem , what is the approximate distribution of $\bar{X}$?

(c) When the experiment is conducted, we observe five improperly sealed packages. Find a point estimate for the proportion of improperly sealed packages being produced on this assembly line.