Random Variables

Random Variables

Study Material for Week 6

Lecture Six

- **Normal Distribution**

Any quantity whose variation depends on random causes is distributed according to the normal law. A large number of distributions approximate to the normal distribution. Discovered by De Moivre in 1733.

Let X be a binomial variable with mean np and standard deviation $\sqrt{npq}$.

Define $$Z = \frac{X - np}{\sqrt{npq}} \qquad \text{---------- (i)}$$

Then Z is a binomial r.v. with mean 0 and standard deviation 1.

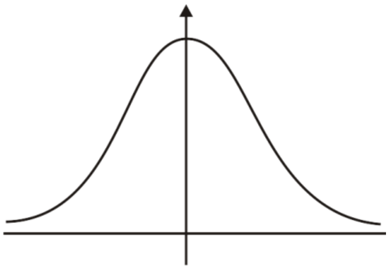
As $n \to \infty$, Z can take values between $-\infty$ to ∞. In limiting form as $n \to \infty$ when neither p or q is very small, Z follows a normal distribution. A normal random variable Z is a continuous random variable. The probability density function of a given random variable is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where, $\mu =$ mean and $\sigma =$ standard deviation.

Note that $Z = \frac{x-\mu}{\sigma}$ then $E(Z) = 0$ & $\text{Var}(Z) = 1$.

In this case, $Z(0, 1)$ is known as **standard normal variate** whose pdf is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

- **Properties of normal distribution**

1. The normal curve is bell-shaped and symmetric about its mean.

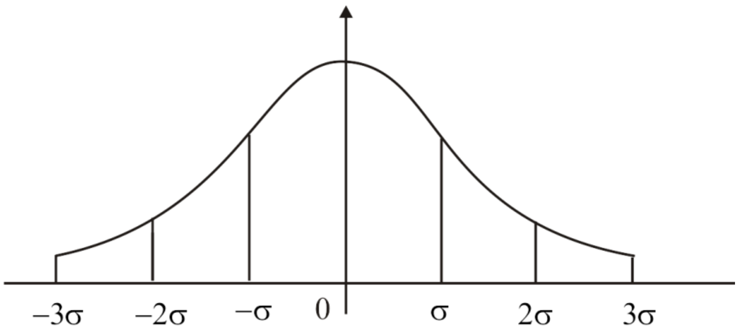


2. The probability of x lying between x_1 & x_2 is given by the area under the normal curve from x_1 to x_2.

   In terms of standard normal variate, $p(0 \le z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz$.

   This integral is called the probability integral or the error function.

Random Variables



Note that:

i) The area under the normal curve between the ordinates $x = \mu - \sigma$ & $x = \mu + \sigma$ is 0.6826, i.e. 68.5%. $\therefore$ Approximately $\frac{2}{3}$ of the values lie within these limits.

ii) The area under the normal curve between $x = \mu - 2\sigma$ & $x = \mu + 2\sigma$ is 0.9544 nearly 95.5% which implies that $4\frac{1}{2}\%$ of values lie inside these limits.

iii) 99.73% values lie inside $x = \mu - 3\sigma$ to $x = \mu + 3\sigma$.

iv) 95% values lie inside $x = \mu - 1.96\sigma$ to $x = \mu + 1.96\sigma$, i.e. only 5% lie outside these limits.

v) 99% values lie inside $x = \mu - 2.58\sigma$ to $x = \mu + 2.58\sigma$, i.e. only 1% values lie outside these limits.

\# Here onwards we will denote the p.d.f of standard normal variable z by symbol $f(z)$.

\# Values of p.d.f of standard normal variate can be obtained from statistical tables.

Examples

1. A certain number of articles manufactured one batch were classified into 3 categories according to a particular characteristic, being less than 50, between 50 – 60 and > 60. If this characteristic is known to be normally distributed, determine the mean and s.d. for this batch if 60%, 35% and 5% were found in these categories.

Let $Z = \frac{X - \mu}{\sigma}$

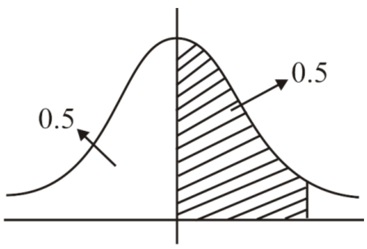
$$P(Z_1) = P\left(Z > \frac{60-\mu}{\sigma}\right) = 0.6$$

$$P(Z > Z_2) = P\left(Z > \frac{60-\mu}{\sigma}\right) = 0.05$$

$$A_1 \to Z_1 = \frac{50-\mu}{\sigma} = 0.1$$

$$A_2 \to Z_2 = \frac{60-\mu}{\sigma} = 0.05$$

$$6 = 7.543,\ \mu = 48.092$$



$$\int_{-\infty}^{Z_1} f(z)dz = 0.6 \quad \Rightarrow \quad 0.5 + \int_0^{Z_1} f(z)dz = 0.6 \quad \int_{-\infty}^{Z_1} f(z)dz = 0.1 \Rightarrow Z_1 = 0.26 .$$

2. In a normal distribution, 31% items are under 45 and 8% over 64. Find mean and standard deviation

   Standard normal variable $Z = \dfrac{X - \mu}{6}$

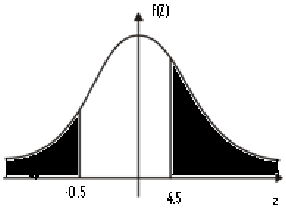
$$Z_1 = \frac{45-\mu}{\sigma}, P(Z < Z_1) = \int_{-\infty}^{Z_1} f(z)dz, \; 0.31 = \int_{-\infty}^{0} f(z)dz + \int_0^{Z_1} f(z)dz \quad \therefore \int_0^{Z_1} f(z)dz = -0.19 .$$

From table, $Z_1 = -0.5$

$$Z_2 = \frac{64-\mu}{\sigma}, \; P(Z > Z_2) = \int_{-Z_2}^{\infty} f(z)dz = \int_0^{\infty} f(z)dz - \int_0^{Z_2} f(z)dz$$

$$\int_0^{0.5} f(Z)dZ = 0.19 \quad \therefore \int_0^{Z_2} f(z)dz = 0.5 - 0.08 = 0.42 .$$ From table, $Z_2 = 1.4$

$$\frac{45-\mu}{\sigma} = -0.5 \;\&\; \frac{64-\mu}{\sigma} = 1.4 \qquad \therefore \mu = 50 \;\&\; \sigma = 10$$



3. A manufacturer of air-mail envelopes knows from experience that the weight of envelope is normally distributed with mean 1.95 gm and s.d. 0.05 gm. About how many envelopes weighing 2 gm or more 2.05 gm or more can be expected in a given packet of 100 envelopes.

$\mu = 1.95 \;\&\; \sigma = 0.05$

$$P\left(Z \geq \frac{2-1.95}{0.05}\right) = P(Z \geq 1) = \int_{-1}^{\infty} f(z)dz = \int_0^{\infty} f(z)dz - \int_0^{1} f(z)dz = 0.5 - 0.3413 = 0.1587$$

$\therefore$ No. of envelopes weighing $> 2\,\text{gm} = 100 \times 0.1587 = 15.87 \approx 16$

$$P\left(Z \geq \frac{2.05-1.95}{0.05}\right) = P(Z \geq 2) = \int_2^{\infty} f(z)dz = \int_0^{\infty} f(z)dz - \int_0^{2} f(z)dz = 0.5 - 0.4772 = 0.0228$$

$\therefore$ No. of envelopes weighing >2.05 $= 2.28 \approx 2$

4. The mean height of 500 students is 151 cm and the s.d. is 15 cm. Assuming that heights are normally distributed, find how many students' height lie between 120 and 155.

$$P\left(\frac{120-151}{15}\right)<Z<\left(\frac{155-151}{15}\right)=P(-2.0667<Z<0.2667).$$

$$=P(-2.07<Z<0.27)$$

$$=\int_{-2.07}^{0.27} f(z)dz=\int_{-2.07}^{0} f(z)dz+\int_{0}^{0.27} f(z)dz=0.4808+0.1064=0.5872$$

No. of students $=296.3\approx 296$

Problem Session

Q.1		Attempt the following
	1)	In a certain examination the percentage of candidates passing and getting distinction were 45 and 9 respectively. Estimate the average marks obtained by the candidates, the minimum pass and distinction marks being 40 and 75 respectively. (Assume the distribution of marks to be normal.)
	2)	5000 candidates appeared in a certain paper carrying maximum of 100 marks. It was found that marks were normally distributed with mean 39.5 and standard deviation 12.5. Determine approximately the number of candidates who secured a first class for which minimum of 60 marks is necessary
	3)	In a production of iron rods let the diameter X be normally distributed with mean 2 in. and standard deviation 0.008 in. (a) What percentage of defectives can we expect if we set the tolerance limits at 2 ± 0.02 in? (b) How should we set the tolerance limits to allow for 4% defectives?