STATISTICS-I,II FOR PGTRB

Collection of results and problems

Professor Academy

Chennai

This version of the notes is updated on March 16, 2021.

Contents

1 Normal Distribution

Definition 1.1. *A random variable X is said to have* ***a normal distribution with parameters μ (called "mean") and σ^2 (called "variance")*** *if its density function is given by the probability law:*

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2\right]$$

$$\text{or } f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

In symbol, we write $X \sim N(\mu, \sigma^2)$.

Definition 1.2. *If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a* ***standard normal variate*** *with $E(Z) = 0$ and $\operatorname{Var}(Z) = 1$ and we write $Z \sim N(0,1)$.*

The p.d.f. of standard normal variate Z is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} - e^{-z^2/2}, -\infty < \dot{z} < \infty$$

and the corresponding distribution function, denoted by $\Phi(z)$ is given by

$$\begin{aligned} \Phi(z) = P(Z \leq z) &= \int_{-\infty}^{z} \varphi(u) du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du \end{aligned}$$

The definite integral

$$\phi(z) = \int_{0}^{z_1} \varphi(u) du$$

is known as ***normal probability integral*** *and gives the area under standard normal curve between the ordinates at $Z = 0$ and $Z = z_1$.*

Remark 1.3. *$\Phi(z)$ has the following important properties:*

- $\Phi(-z) = 1 - \Phi(z)$
- *whenever* $X \sim N(\mu, \sigma^2)$, $P(a \leq X, \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$
- *The graph of $f(x)$ is a famous 'bell - shaped' curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.*

Theorem 1.4. ***Normal Distribution as a Limiting form of Binomial Distribtion.*** *Normal distribution is another limiting form of the binomial distribution under the following conditions :*
(i) n, the number of trials is indefinitely large, i.e:, $n \to \infty$ and
(ii) neither p nor q is very small.
NOTE: Stirlings approximation is used in the proof.

the standard binomial variate:

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}}; X = 0, 1, 2, \ldots, n$$

follows $N(0,1)$ as $n \to \infty$.

Remark 1.5. *Characteristics of the Normal Distribution and Normal Probability Curve:*

- *The crve is bell shaped and symmetrical abut the line $x = \mu$.*
- *Mean, median and mode of the distribution coincide*
- *As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occurring at the point $x = \mu$, and given by $[p(x)]_{\max} = \frac{1}{\sigma\sqrt{2\pi}}$.*
- $\beta_1 = 0$ *and* $\beta_2 = 3$
- *x -axis is an asymptote to the curve.*
- *The points of inflexion ($f''(x) = 0$, and $f'''(x) \neq 0$) of the curve are given by*

$$\left[x = \mu \pm \sigma, f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}\right]$$

- *Q.D. : M.D. : S.D.::* $\frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma :: \frac{2}{3} : \frac{4}{5} : 1$
  $\Rightarrow$ *Q. D. : M.D.: S.D. :: 10: 12:* 15

- *If X and Y are independent standard normal variates, then $U = X + Y$ $V = X - Y$ are independently distributed, $U \sim N(0,2)$ and $V \sim N(0,2)$*

- ***Bernstein's Theorem. (Converse of above result)*** *If X and Y are independent and identically distributed random variables with finite varaince and if $U = X + Y$ and $V = X - Y$ are independent, then all r.v.'s $X, Y; U$ and V are normally distributed.*

- *MGF of $X \sim N(\mu, \sigma^2)$ is given by $M_X(t) = e^{\mu t + t^2\sigma^2/2}$, and $X \sim N(0,1)$ is* $\exp(t^2/2)$.

- *CGF is* $K_X(t) = \log_e M_X(t) = \mu t + \frac{t^2 \cdot \sigma^2}{2}$.

- *For the normal distribution all* ***odd order moments about mean vanish*** *and the* ***even order moments about mean*** *are given by the* ***recurrence relation***: $\mu_{2n} = \sigma^2(2n-1)\mu_{2n-2}$.
  Thus $\mu_{2n} = 1.3.5\ldots(2n-1)\sigma^{2n}$.

- ***A linear combination of independent normal variates is also a normal variate:*** *Let $X_i, (i = 1, 2, \ldots, n)$ be n independent normal variates with mean μ_i and variance σ_i^2 respectively. Then*

$$\sum_{i=1}^{n} a_i X_i \sim N\left[\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right]$$

  So, $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ *and* $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.

- *The definite integral $\phi(z) = \int_0^{z_1} \varphi(z)dz$ is known as* ***normal probability integral*** *and gives the area under standard normal curve between the ordinates at $Z = 0$ and $Z = z_1$.*

- *(i)* $P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1) = 2 \times 0 \cdot 3413 = 0.6826$, *(ii)* $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$ *and (iii)* $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

Problem 1.6. *If $X \sim N(50, 10)$ and $Y \sim N(60, 6)$ are independent then $Y - X$ is*
A) $N(10, 16)$ *B)* $N(-10, 16)$ *C)* $N(10, 4)$ *D)* $N(10, -4)$

Problem 1.7. *If X is normally distributed with mean 2 and variance 1 then* $P(|X - 2| < 1)$ *is*
A) 0.3413 B) 0.6826 C) 0.0228 D) 0.4772

Example 1.8. *For a certain normal distribution, the first moment about 10 and the fourth moment about 50 is* 48. *What is the arithmetic mean and standard deviation of the distribution ?*

Solution. We know that if μ_1' *is the first. moment about the point* $X = A$ *then arithmetic mean is given by:*

$$Mean = A + \mu_1'$$

We are given μ_1' *(about the point* $X = 10) = 40 \Rightarrow$ *Mean* $= 10 + 40 = 50$ *Also we are given* μ_4' *(about the point* $X_. = 50) = 48$, *i.e.,* $\mu_4 = 48$

$$(\because \; Mean = 50)$$

But for a normal distribution withstandard deviation σ_2, $\mu_4 = 3\sigma^4 \Rightarrow 3\sigma^4 = 48$ *i.e.,* $\sigma = 2$

Example 1.9. X *is normally distributed and the mean of* X *is 12 and S.D. is* 4. *Find out the probability of the following :*

(i) $X \geq 20$ *(ii)* $X \leq 20$, *and (iii)* $0 \leq X \leq 12$

Soln:

(i) $P(X \geq 20) =$ *?*

When $X = 20, \quad Z = \dfrac{20 - 12}{4} = 2$

$\therefore \quad P(X \geq 20) = \dot{P}(Z \geq 2) = 0.5 - P(0 \leq Z \leq 2) = 0.5 - 0.4772 = 0.0228$

(ii) $P(X \leq 20) = 1 - P(X \geq 20)$

$$= 1 - 0.0228 = 0.9772$$

(iii) $P(0 \leq X \leq 12) = P(-3 \leq Z \leq 0)$

$$= P(0 \leq Z \leq 3) = 0.49865$$

Example 1.10. X *is a normal rariate with mean 30 and S.D. 5. Find the probabilities that*

(i) $26 \leq X \leq 40$ *(ii)* $X \geq 45$, *and (iii)* $\mid X - 30 \mid > 5$. *Given* $\phi(0.8) = 0.2881$, $\phi(2) = 0.4772$, $\phi(3) = 0.49865$, $\phi(1) = 0.3413$.

(ANS: (i)0.7653, (ii) 0.00135, (iii) 0.3174)

Problem 1.11. *Two independent random variates* X *and* Y *are both normally distributed with means 1 and 2 and standard deviation 3 and 4 respectively. If* $Z = X - Y$, *then find the median, s.d. and mean of the distribution of* Z. *Find* $P(Z + 1 \leq 0)$.

Problem 1.12. *Which of the following is not true about normal distribution?*
A) median is equidistant from Q_1 and Q_3.
B) the sum of two independent standard normal Variate is a standard normal Variate
C) The points of inflexion are at $\mu \pm \sigma$
D) Curve is symmetrical about mode

Problem 1.13. *If the quartile deviation of a normal distribution is 15, then its mean deviation about mean is*
A) 12 B) 10 C) 18 D) 15

Example 1.14. *Let $X \sim N(\mu, \sigma^2)$. If $\sigma^2 = \mu^2, (\mu > 0)$, express $P(X < -\mu \mid X < \mu)$ in terms of cumulative distribution function of $N(0,1)$.*
Soln: $P(X < -\mu \mid X < \mu) = \frac{P(X<-\mu \cap X<\mu)}{P(X<\mu)} = \frac{P(X<-\mu)}{P(X<\mu)} = \frac{P(Z<-2)}{P(Z<0)} \quad \left(Z = \frac{X-\mu}{\sigma} = \frac{X-\mu}{\mu}\right)$
$= \frac{P(Z>2)}{(1/2)} = 2[1 - P(Z \le 2).] = 2[1 - \Phi(2)]$

Problem 1.15. *If a continuous RV X follows $N(0,4)$, find $P\{1 \le X \le 2\}$ and $P\{1 \le X \le 2/X \ge 1\}$.*
Given$\phi(1) = 0.3413$, $\phi(0.5) = 0.1915$.
(ANS: 0.1498, 0.4856)

Definition 1.16. *If the discrete RVX can take the values $0, 1, 2, \ldots,$ such that $P(X = i) = (n + i - 1)C_i p^n q^i, i = 0, 1, 2, \ldots,$ where $p + q = 1$, then X is said to follow a* ***Pascal (or negative binomial) distribution*** *with parameter n.*

If the pdf of a continuous RVX is $f(x) = \frac{1}{\Gamma(n)} e^{-x} x^{n-1}, 0 < x < \infty$ and $n > 0$, then X follows a ***gamma distribution*** *with parameter n.* ***Gamma distribution is a particular case of Erlang distribution****, the pdf of which is $f(x) = \frac{c^n}{\Gamma(n)} x^{n-1} e^{-cx}, 0 < x < \infty, n > 0, c > 0$*

An Erlang distribution with $n = 1$ [i.e., $f(x) = ce^{-cx}, 0 < x < \infty, c > 0$] is called an ***exponential (or negative exponential) distribution*** *with parameter c.*

If the pdf of a continuous RV X is $f(x) = \frac{\alpha}{\pi} \times \frac{1}{x^2 + \alpha^2}, \alpha > 0, -\infty < x < \infty$, then X follows a ***Cauchy distribution*** *with parameter a .*

The continuous random variable which is distributed according to the probability *law*

$$f(x) = \begin{cases} \frac{1}{B(\mu,v)} \cdot x^{\mu-1}(1-x)^{v-1}; (\mu,v) > 0, 0 < x < 1 \\ 0, \ otherwise \end{cases}$$

*(where $B(\mu,v)$ is the Beta function), is known as a **Beta variate of the first kind** with parameters μ and v and is referred to as $\beta_1(\mu,v)$ variate and its distribution is called **Beta distribution of the first kind**.*

The continuous random variable X which is distributed according to the probability law

$$f(x) = \begin{cases} \frac{1}{B(\mu,v)} \cdot \frac{x^{\mu-1}}{(1+x)^{\mu+v}}; (\mu,v) > 0, 0 < x < \infty \\ 0, \ otherwise \end{cases}$$

*is known as a **Beta variate of second kind** with parameters' μ and v and i denoted as $\beta_2(\mu,v)$ variate and its distribution is called **Beta distributen of second kind**.*

The probability that E_1 occurs x_1 times, E_2 occurs x_2 times ... and E_k occurs x_k times in n independent observations, is given by

$$p\left(x_1, x_2, \ldots, x_k\right) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

*where $\sum x_i = n$ is known as **Multinomial distribution**.*