STATISTICS-I FOR PGTRB

Collection of results and problems

Professor Academy

Chennai

This version of the notes is updated on March 8, 2021.

Contents

1 Random variables

Definition 1.1. *If a real variable X be associated with the outcome of a random experiment, then since the values which X takes depend on chance, it is called a* ***random variable or a stochastic variable or simply a variate****.*

If in a random experiment, the event corresponding to a number a occurs, then the corresponding random variable X is said to assume the value a and the probability of the event is denoted by $P(X = a) = p(a)$. That is, $P(X = a) = P\{w|X(w) = x\}$. This is known as ***probability mass function(pmf) (or) probability density function(pdf)*** *of X. The pdf must satisfy*
(i) $p(x_i) \geq 0 \quad for\, all\, i$ *(ii)* $\sum_{i=1}^{\infty} p(x_i) = 1$

For example, in a random experiment of tossing three coins, X denote the number of heads, then X is a random variable.

TTT	*TTH*	*THT*	*HTT*	*THH*	*HTH*	*HHT*	*HHH*
0	*1*	*1*	*1*	*2*	*2*	*2*	*3*

The pdf of X is given by

X	*0*	*1*	*2*	*3*
P(X=x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

If a random variable takes a finite set of values, it is called a ***discrete variate****. On the other hand, if it assumes an infinite number of uncountable values, it is called a* ***continuous variate****.*

The distribution function $F(x)$ of the discrete variate X is defined by $F(x) = P(X \leq x) = \sum_{i=1}^{x} p(x_i)$ where x is any integer. The graph of $F(x)$ will be stair step form. The distribution function is also sometimes called **cumulative distribution function(cdf)**. In continuous RV, $F(x) = \int_{-\infty}^{x} f(x)dx$ In the above example,

X	0	1	2	3
F(x)	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{7}{8}$	1

Note 1.2. *1. $F(x)$ is non-decreasing function*
2. $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$
3. $P(a \le x \le b) = \int_a^b f(x)dx = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = F(b) - F(a)$.

Problem 1.3. *A random variable X has the following probability distribution :*

x :	0	1	2	3	4	5	6	7
$p(x)$:	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(i) Find k,(Ans:$\frac{1}{10}$)
(ii) Evaluate $P(X < 6), P(X \ge 6)$, and $P(0 < X < 5)$,(Ans:$\frac{81}{100}, \frac{19}{100}, \frac{4}{5}$)
(iii) If $P(X \le c) > \frac{1}{2}$, find the minimum value of c,(Ans: 4) and
(iv) Determine the distribution function of X.
($F(x)$: $0 \quad \frac{1}{10} \quad \frac{3}{10} \quad \frac{5}{10} \quad \frac{4}{5} \quad \frac{81}{100} \quad \frac{83}{100} \quad 1$)

(1)**The mean value** (μ) of the probability distribution of a variate X is commonly known as its **expectation and is denoted by** $E(X)$. If $f(x)$ is the probability density function of the variate X, then

$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \Sigma_{-\infty}^{\infty} xf(x)$

Theorem 1.4.

- *$E(X+Y) = E(X) + E(Y)$*
- *Let $X_1, X_2, \ldots, X_n$ be any n random variables and if $a_1, a_2, ;!, a_n$ are any n constants, then*
$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$
provided all the expectations exist.
- *If $X \ge 0$ then $E(X) \ge 0$*
- *$|E(X)| \le E|X|$ provided the expectations exist.*

Note 1.5. *1. $E(X)$ exists iff $E|X|$ exists.*

2. Now
$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}; \quad -\infty < x < \infty$$
which is p.d.f. of Standard ***Cauchy distribution****.*

$$\int_{-\infty}^{\infty} |x| f(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{|x|}{1+x^2}dx = \frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}dx$$

($\because$ *Integrand is an even function of* x)

$$= \frac{1}{\pi} \quad \left|\log\left(1+x^2\right)\right|_0^\infty \to \infty$$

Since this integral does not converge to finite limit, $E(X)$ does not exist.

(2) **Variance** of a distribution is given by

$$\sigma^2 = \sum_i (x_i - \mu)^2 f(x_i)$$
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

where σ is the **standard deviation** of the distribution.

Theorem 1.6.

$$V(aX+b) = a^2 V(X)$$

where a and b are constants.

(3) **The rth moment about the mean** (denoted by μ_r) is defined by

$$\mu_r = \Sigma_i (x_i - \mu)^r f(x_i)$$
$$\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx$$

(4) **The rth moment about zero or simply rth moment** (denoted by μ_r') is defined by

$$\mu_r = \Sigma_i (x_i)^r f(x_i)$$
$$\mu_r = \int_{-\infty}^{\infty} (x)^r f(x)dx$$

(5) **Mean deviation** from the mean is given by

$$\Sigma |x_i - \mu| f(x_i)$$
$$\int_{-\infty}^{\infty} |x-\mu| f(x)dx$$

(6) **The moment generating function (mgf) of X** denoted by $M_X(t)$ is defined as $E(e^{tX})$.

The moments can be found from mgf as $\mu_r' = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}$.

(7) **The characteristic function of X** denoted by $\phi_X(t)$ is defined as $E(e^{itX})$.

(8) **The cumulunt generating function of X** denoted by $K_X(t)$ is defined as $\log M_X(t)$.

(9) **The probability generating function (p.g.f.) of X** denoted by $P_x(t)$, is defined as $P_x(t) = p_0 + p_1 t + p_2 t^2 + \ldots = \sum_{n=0}^{\infty} p_n t^n = E\left(t^x\right)$

The probability can be found from pgf as $\left[\frac{d^k}{dt^k} P_X(t)\right]_{t=0} = k! p_k, k = 1, 2, \ldots n.$

(10) **Median** is the point which divides the entire distribution in two equal parts. In case of continuous distribution, median is the point which divides the total area into two equal parts. Thus if M is the median, then

$$\int_a^M f(x) dx = \int_M^b f(x) dx = \frac{1}{2}$$

Example 1.7. *In a lottery, m tickets are drawn at a time out of n tickets numbered from 1 to n. Find the expected value of the sum of the numbers on the tickets drawn.*
Solution. *Let $x_1, x_2, \ldots, x_n$ be the variables representing the numbers on the first, second, ..., n th ticket. The probability of drawing a ticket out of n tickets being in each case $1/n$, we have*

$$E\left(x_{\mathrm{i}}\right) = 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + 3 \cdot \frac{1}{n} + \ldots + n \cdot \frac{1}{n} = \frac{1}{2}(n+1)$$

$\therefore$ *expected value of the sum of the numbers on the tickets drawn*

$$\begin{aligned} & = E\left(x_1 + x_2 + \ldots + x_m\right) = E\left(x_1\right) + E\left(x_2\right) + \ldots + E\left(x_m\right) \\ & = m E\left(x_i\right) = \tfrac{1}{2} m(n+1) \end{aligned}$$

Example 1.8. *Consider*

x	:	-3	6	9
$P(X=x)$	:	1/6	1/2	1/3

Find $E(X)$ and $E\left(X^2\right)$. Hence evaluate $E(2X+1)^2$. (ANS: $\frac{11}{2}$, $\frac{93}{2}$,209)

Problem 1.9. *If X be a random variable with probability generating function $P_X(t)$, find the probability generating function of (i) $X+2$ and (ii) $2X$.*
(ANS: $t^2 P_X(t)$ and $P_X(t^2)$)

Theorem 1.10. $M_{cX}(t) = M_X(ct), c$ *being a constant.*

Theorem 1.11. *The moment generating function of the sum of a number of independent random variables is equal to the product of their respective moment gener-*

ating functions.

That is, $M_{X_i+X_2+\ldots+X_0}(t) = M_{X_1}(t) M_1\left({}_2(t) \ldots M_{X_n}(t)\right.$

Theorem 1.12. ***Cauchy-Schwartz Inequality:*** *If X and Y are random variables taking. real values, then*

$$[E(XY)]^2 \leq E\left(X^2\right) \cdot E\left(Y^2\right)$$

Theorem 1.13. ***Jenson's Inequality:***. *If g is continuous and convex function on the interval, and X is a random variable whose values are in I with probability 1, then*

$$E[g(X)] \geq g[E(X)]$$

provided the expectations exist.

Theorem 1.14. *If g is a continuous and concave function on the interval I and X is a r.v. whose values are in I with probability 1, then*

$$E[g(X)], \leq g[E(X)]$$

provided the expectations exist.

Theorem 1.15. *If $X \geq 0$, $E[\log(X)] \leq \log[E(X)]$, if exist.*

Theorem 1.16. ***Chebychev's Inequality:*** *If X is a random variable with mean μ and variance σ^2, then for any positive number k, we have*

$$\begin{aligned} & P\{|X-\mu| \geq k\sigma\} \leq 1/k^2 \\ \text{or} \quad & P\{|X-\mu \mid < k\sigma\} \geq 1-(1/k^2) \end{aligned}$$

Problem 1.17. *If X is a r.v. such that $E(X)=3$ and $E\left(X^2\right)=13$, use Chebychev's inequality to determine a lower bound for $P(-2<X<8)$.(ANS: $P \geq \frac{21}{25}$ as $k=2.5$)*

2 Standard distributions

Definition 2.1. ***Bernoulli distribution:*** *A random variable X which takes two values 0 and 1, with probabilities q and p respectively, i.e., $P(X=1)=p$, $P(X=$*

8438760401

$0) = q, q = 1 - p$ *is called a Bernoulli variate and is said to have Bernoulli distribution.*

Definition 2.2. ***Binomial distribution:*** *Consider a sct of* n *independent Bernoullian trials (* n *being finite), in :vhich the probability '* p *' of success in any trial is constant for each trial. Then* $q = 1 - p$*, is the probability of failure in any trial. The probability distribution of the number of successes, so,obtained is alled! the Binomial probability distribution, for the obvious rason that the probabilities.*

A random variable X *is said to follow binomial distribution if it assumes only non-negative values and its'probability mass function is given by*

$$P(X = x) = p(x) = \begin{cases} \binom{n}{x} p^x \dot{q}^{n-x}; x = 0, 1, 2, \ldots, n; q = 1 - p \\ 0, \ \text{otherwise} \end{cases}$$

The two independent constants n *and* p *in the distribution, e known as the parameters of the distribution. '* n *' is also, sometimes, known as the degree of the binoznial distribution.*

In symbol, we write $X \sim B(n, p)$.

Remark 2.3. ***Conditions for Binomial Distribution.*** *We get the binomial distribution under the following experimental-conditions.*
(i) Each trial results in two mutually disjoint outcomes, termed as success and failure.
(ii) The number of trials ' n *' is finite.*
(iii) The'trials are independent of each other.
(iv) The probability of success ' p *' is constant for each trial.*

Applications of Binomial distribution. *This distribution is applied to problems concerning : (i) Number of defectives in a sample from production line, (ii) Estimation of reliability of systems, (iii) Number of rounds fired from a gun hitting a target, (iii) Radar detection.*

Theorem 2.4. *Let* $X \sim B(n, p)$.

- *Mean of* $X = np$.

- *Var of* $X = npq$.
- $Var \leq mean$.
- *mgf of X is*
  $M_X(t)=E\left(e^{tx}\right)=\Sigma^n C_x p^x q^{n-x} e^{tx} \quad =\Sigma^n C_x\left(pe^t\right)^x q^{n-x}=\left(q+pe^t\right)^n$
- *skewness* $=(1-2p)/\sqrt{(npq)}$, *kurtosis* $=3+\frac{1-6pq}{npq}$
- *Mode of* $X = \begin{cases} [(n+1)p] & \text{if (n+1)p is not an integer} \\ (n+1)p \,\&\, (n+1)p-1 & \text{if (n+1)p is an integer} \end{cases}$
- *If* $Y=n-X$, $Y \sim B(n,q)$.
- *pgf*, $P_X(t)=(pt+q)^n$
- *characteristic function:*
  $\varphi_X(t)=E\left(e^{it}\right)=\sum_{x=0}^{n} e^{it}p(x)=\left(q+pe^{it}\right)^n$

Theorem 2.5. ***(Sum of binomial with common p, is again binomial:)*** *If* $X_i,(i=1,2,\ldots,k)$ *are independent binomial variates with parameters* $(n_i,p),(i=1,2,\ldots,k)$ *then their* sum $\sum_{i=1}^{k} X_i \sim B\left(\sum_{i=1}^{k} n_i, p\right)$.

Theorem 2.6.

Problem 2.7. *The m.g.f. of a r.v.* X *is* $\left(\frac{2}{3}+\frac{1}{3}e^t\right)^9$. *Find mean and variance.*
Hint: $X \sim B\left(n=9, p=\frac{1}{3}\right)$.

Problem 2.8. *Determine the binomial distribution for which mean* $=2(variance)$ *and* $mean+variance=3$. *Also find* $P(X \leq 3)$.

Problem 2.9. *Random variable* X *follows binomial distribution 'with parameters* $n=40$ *and* $p=\frac{1}{4}\cdot$ *Use Chebyhev's inequality to find bounds for (i)* $P[|X-10|<8];(ii)P[|X-10|>10]$
(Ans. (i) $113/128$ *(lower bound), (ii) 3/40 (upper bound).)*