Ch. 3. Discrete Random variables & Their Probability Distributions

Why to study the probability distribution

The probability of an observed sample is needed to make inferences about a population.

The sample observations are frequently expressed as numerical events that corresponds to the values of the r.v's. Hence we need know the probabilities of these numerical events.

Certain types of r.v's. occur frequently in practise, so it is useful to know the probability for each value of a r.v.. This collection of probabilities is called probability Distribution.

1. Discrete Random Variable

A random variable (r.v.) is a real-valued function defined over a sample space.

A r.v. *Y* is said to be discrete if it can assume only a finite or countably infinite number of distinct values.

2. Probability Distribution (p.d.) for a Discrete R.V.

Y can be represented by a formula, a table, or a graph which provides the probabilities $p(y) = P(Y = y)$ corresponding to each and every value of y.

For any discrete p.d. the following holds

- $0 \leq p(y) \leq 1$ for all y.
- $\sum_y p(y) = 1$.

3. Expected Value of a R.V.

Let Y be a discrete r.v. with the p.d. $p(y)$. Then the expected value of Y:

$$E(Y) = \sum_y y\, p(y).$$

Note : $E(Y) = \mu$, population mean of the r.v. Y.

Expected Value of a function of a R.V.

Let $g(Y)$ be a function of a discrete r.v. Y with p.f. $p(y)$. Then

$$E[g(Y)] = \sum_y g(y)\, p(y).$$

Variance of a R.V.

$$V(Y) = \sigma^2 = E[(Y-\mu)^2] = E(Y^2) - \mu^2.$$

and standard deviation $= +\sqrt{V(Y).}$

More Results on Expected Value

If c is a constant, then

$$E(c) = c.$$

$$E[c\, g(Y)] = c\, E[g(Y)]$$

$$E[g_1(Y) + g_2(Y) + ..] = E[g_1(Y)] + E[g_2(Y)] + ..$$

4. Binomial Probability Distribution

A Binomial Experiment

- n identical and independent trials.
- Each trial results in one of two outcomes, either a success S or a failure F.
- The probability of success on a single trial is p and remains the same from trial to trial. The probability of a failure $q = 1 - p$.
- The random variable of interest is Y, the number of successes observed in n trials.

Binomial Distribution

A r.v. Y is said to have a binomial distribution [denoted by $B(n,p)$] based on n trials with success probability p $(0 \leq p \leq 1)$ iff

$$p(y) = \binom{n}{y} p^y q^{n-y}, \ y = 0, 1, 2, \ldots, n.$$

Note:

$$a. \sum p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y q^{n-y} = (q+p)^n = 1.$$

b. The binomial p.d. has many applications in sampling for defective in industrial quality control, sampling of consumer preference or voting populations, and in many other physical situations.

c. Let Y be a $B(n,p)$. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

5. Geometric probability Distribution

Consider an experiment which involves identical and independent trials, each of which can result in one of two outcomes : success or failure. The probability of success in a trial is p and is constant from trial to trial. Then, the geometric random variable Y is the number of trial on which the *first* success occurs.

A r.v. Y is said to have a geometric p.d. [denoted by $G(p)$] iff

$$p(y) = q^{y-1}p, \quad y = 1, 2, 3, \ldots, \quad 0 \le p \le 1.$$

Note:

a. The experiment could end with the first trial if a success is observed on the first trial or the experiment can go on indefinitely.

b. The sample space S for the experiment contains the countably infinite set of sample points:

$$E_1 : S,\ E_2 : FS,\ E3 : FFS, \ldots, E_k : (FF\ldots F)S.$$

c. Let Y be a $G(p)$. Then

$$\mu = E(Y) = \frac{1}{p} \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

6. Negative Binomial Probability Distribution

Consider an experiment with identical and independent trials, each of which can result in one of two outcomes : success or failure. The probability of success in a trial is p and is constant from trial to trial. We are interested in knowing the number of the trial on which the second, third, ... success occurs.

The random variable Y, the number of trial on which the r^{th} success occurs $(r = 2, 3, 4, \ldots)$, has the *negative binomial* distribution.

A r.v. Y is said to have a negative binomial p.d. [denoted by $NB(r, p)$] iff

$$p(y) = \binom{y-1}{r-1} q^{y-r} p^r, \quad y = r, r+1, .., 0 \leq \text{p} \leq 1.$$

Note:

Let Y be a $NB(r, p)$. Then

$$\mu = E(Y) = \frac{r}{p} \text{ and } \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

7. Hypergeometric Probability Distribution

In the binomial experiment if trials are not independent and probability of success p does not remain the same from trial to trial, we use the hypergeometric probability distribution for the r.v. Y.

Suppose that a population contains a finite number N of elements that possess one of two characteristics. Thus r of the elements may be red and $b = N - r$ black. A sample of n elements is randomly selected from the population and the random variable of interest is Y, the number of red elements in the sample. This r.v. Y has the hypergeometric probability distribution iff

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, \ y = 0, 1, \ldots, n$$

and $y \leq r$ and $n - y \leq N - r$.

If Y is a r.v. with a hypergeometric

distribution, then

$$\mu = E(Y) = \frac{nr}{N},$$

$$\sigma^2 = V(Y) = n(\frac{r}{N})(\frac{N-r}{N})(\frac{N-n}{N-1}).$$

Note: a. Defining $p = \frac{r}{N}$ and $q = 1 - p = \frac{N-r}{N}$, we can express

$$\mu = E(Y) = np,$$

$$\sigma^2 = V(Y) = npq(\frac{N-n}{N-1}).$$

b. For fixed n as $N \to \infty$, $\frac{N-n}{N-1} \to 1$ and $V(Y) = npq$.

8. Poisson Probability Distribution

The Poisson probability distribution provides a good model for the p.d. of the number Y of rare events that occur in space, time , volume or any other dimension (for example: automobile or industrial or other types of accidents in a given unit of time, number of phone calls handled by a switchboard in a time interval, number of radioactive particles that decay in a particular time period, number of errors a typist makes in typing a page).

A r.v. Y has a Poisson probability

distribution [denoted by $Y \sim P(\lambda)$] iff

$$p(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \; y = 0, 1, \ldots, \;\; \lambda > 0.$$

If Y is a r.v. with a Poisson p.d. with parameter λ, then

$$\mu = E(Y) = \lambda,$$

$$\sigma^2 = V(Y) = \lambda.$$

9. Moments and Moment-generating Functions

Moments

Like the two numerical descriptive measures μ and σ that locate the center and describe the spread of the values of a r.v., we define a set of numerical descriptive measures, called *moments*, that uniquely determine the p.d. of a random variable.

The $i-th\ moment$ of a r.v. Y taken about the $origin$ is defined to be

$$\mu_i' = E(Y^i)$$

The $i-th\ moment$ of a r.v. Y taken about its $mean$, or the $i-th\ central\ moment$ of Y, is defined to be

$$\mu_i = E[(Y-\mu)^i]$$

Moment-Generating Function

The moment-generating function for a r.v. packages all the moments for a r.v. into one simple expression.

The moment-generating function $m(t)$ for a r.v. Y is defined to be $E(e^{ty})$.

If $m(t)$ exists, then for any positive integer k,

$$\frac{d^k m(t)}{dt^k}\Big|_{t=0} = m^k(0) = \mu_k^{'}.$$

Thus, to obtain $\mu_k^{'}$, find the $k-th$ derivative of $m(t)$ and set $t = 0$.

11. Tchebysheff's Theorem

The empirical rule provides good approximation to probabilities over certain intervals if the probability or population histogram is roughly bell- shaped and mean and variance are known.

When the shape of distribution is not bell-shape, the empirical may not provide the useful approximations to the probabilities of interest.

Then, the following result known as Tchebysheff's theorem can be used to determine the lower bound for the probability that the r.v. Y with mean μ and standard deviation σ,falls in an interval $\mu \pm k\sigma,\ k > 1$:

$$P(|Y-\mu)| < k\sigma) \geq 1 - \frac{1}{k^2}, \text{ or}$$

$$P(|Y-\mu)| \geq k\sigma) \geq \frac{1}{k^2}, \text{ or}$$

$$P(\mu - k\sigma < Y < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Note: The result applies for any probability distribution, whether bell-shaped or not.