Probability

5

Mean, Median and Mode

- $$\text{Mean } (\bar{X}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}$$

- $$\text{Median} = \frac{x_{n/2} + (x_{n/2} + 1)}{2}; \ n \text{ is even}$$
$$= x_{n+1/2}; \ n \text{ is odd}$$
$$= L_{\text{med}} + \left(\frac{N/2 - F}{f}\right) \times w$$

where L_{med} = lower limit of median class
$N = \Sigma f_i$
F = Cumulative frequency upto the median class (cumulative frequency of the preceding class)
f = Frequency of the median class

Mode : Value of 'x' corresponding to maximum frequency.

$$L_{\text{mode}} + \frac{f_m - f_1}{(2f_m - f_1 - f_2)} \times h$$

where L_{mode} = lower limit of modal class
f_m = frequency of modal class
f_1 = preceding frequency of modal class
f_2 = Following frequency of modal class

Note:

- Mode = 3 Median – 2 Mean [for Asymmetric distribution]
- Mean = Mode = Median [for Symmetric distribution]

Axioms of Probability

Let A and B be two events. Then

1. $P(\bar{A}) = 1 - P(A)$
2. $P(\phi) = 0;$
3. $P(A - B) = P(A) - P(A \cap B)$
4. $P(A - B) = P(A \cap \bar{B})$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $P(A \cup B) = P(A) + P(B)$; mutually exclusive events.
7. $P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$
8. $P(A \cap B) = P(A) \cdot P(B)$; independent events.
9. $P(A \cap B) = \phi$; mutually exclusive events.
10. $P(S) = 1$; S is sample space.
11. $\max(0, P(A) + P(B) - 1) \le P(A \cap B) \le \min(P(A), P(B))$
12. $\max(P(A), P(B)) \le P(A \cup B) \le \min(1, P(A) + P(B))$
13. $\max(0, P(A_1) + P(A_2) \ldots + P(A_n) - (n-1)) \le P(A_1 \cap A_2 \ldots \cap A_n)$
    $\le \min(P(A_1), P(A_2) \ldots, P(A_n))$
14. $\max(P(A_1), P(A_2) \ldots, P(A_n)) \le P(A_1 \cup A_2 \ldots \cup A_n)$
    $\le \min(1, P(A_1) + P(A_2) \ldots + P(A_n))$
15. $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
16. $P(A|B) = P(A)$; independent events.
17. $P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \ldots P(E_n)$; independent events.
18. Rule of total probability: $P(X) = \sum_{i=1}^{n} P(E_i)\, P(X|E_i)$
19. Baye's Theorem:

$$P(E_1|X) = \frac{P(E_1) \cdot P(X|E_1)}{P(E_1) \cdot P(X|E_1) + P(E_2) \cdot P(X|E_2) + P(E_3) \cdot P(X|E_3)}$$

In general,

$$P(E_i|X) = \frac{P(E_i) \cdot P(X|E_i)}{\sum_{j=1}^{n} P(E_j) \cdot P(X|E_j)}$$

Random Variable (Stochastic Variable)

Random variable assigns a real number to each possible outcome.

Let X be a discrete random variable, then

1. $F(x) = P(X \leq x)$ is called distribution function $\sum_{i=0}^{n} P(i)$ of X.
2. Mean or Expectation of $X = \mu = E(X) = \sum_{i=1}^{n} x_i \, P(x_i)$
3. Variance of $X = \sigma^2 = E(X^2) - [E(X)]^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$
4. Standard deviation of $X = \sigma = \sqrt{\text{Variance}}$
5. $\sum_{i=1}^{n} P(x_i) = 1$

Types of Random Variables

1. **Discrete Random Variable:** "Finite set of values" or "Countably infinite".
2. **Continous Random Variable (Non-discrete):** "Infinite number of uncountable values".

Discrete Distributions

1. **Binomial Distribution:** The probability that the event will happen exactly r times in n trials i.e. r successes and $n - r$ failures will occur.

   $P(X = r) = P(r) = \Sigma\, {}^nC_r p^r q^{n-r}$

   Mean $= E(x) = np$

   Variance $(\sigma^2) = V(x) = npq = np(1 - p)$

   S.D $(\sigma) = \sqrt{npq} = \sqrt{np(1-p)}$

   Where
   r = 0, 1, . . . , n
   $q = 1 - p$
   n = fixed number of trials
   p = probability of success

2. **Poisson Distribution:**

$$\boxed{P(X = x) = \sum \frac{e^{-\lambda} \cdot \lambda^x}{x!};\ x = 0, 1, 2, \ldots, \infty}$$

Where X = Discrete random variable

λ = Parameter of distribution (positive constant)

- Mean (μ) = Variance (σ^2) = λ
- $S.D = \sqrt{\lambda}$

Poisson distribution is a limiting case of binomial distribution as $n \to \infty$ and $p \to 0$.

X	X Counts	p(x)	Value of X	E(x)	V(x)
Discrete uniform	Outcomes that are equally likes (finite)	$\frac{1}{b-a+1}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of successes in n fixed trials	$\binom{n}{x}p^x(1-p)^{n-x}$	$x = 0, 1, \ldots, n$	np	np(1 − p)
Poisson	Number of arrivals in a fixed time period	$\frac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, \ldots$	λ	λ
Geometric	Number of trials up through 1st success	$(1-p)^{x-1}p$	$x = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number of trials up through kth success	$\binom{x-1}{k-1}(1-p)^{x-k}p^k$	$x = k, k+1, \ldots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hyper-geometric	Number of marked individuals in sample taken without replacement	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max(0, M+n-N) \le x \le \min(M, n)$	$n \cdot \frac{M}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$

Continuous Distribution

Let X be continous random variable. Then

(*i*) **Density functions:**

$$\text{(a) } P(X \le a) = \int_{-\infty}^{a} f(x)\cdot dx \text{, (b) } P(a \le X \le b) = \int_{a}^{b} f(x)\cdot dx$$

(*ii*) **Mean** = $E(x) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$

(*iii*) **Variance of X** = $V(X) = \int_{-\infty}^{\infty} [x - E(x)]^2 \cdot f(x)dx = E(x^2) - (E(x))^2$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2$$

(*iv*) $\int_{-\infty}^{\infty} f(x)dx = 1$

1. Uniform Distribution (Rectangular Distribution)

(*i*) Density function:

$$f(x) = \frac{1}{b-a}; \quad a \le x \le b$$

$$= 0 \text{ ; otherwise}$$

(*ii*) Cumulative function:

$$P(X \le x) = \int_{-\infty}^{x} f(x) \cdot dx = \begin{cases} 0 & ; \quad \text{if } x \le a \\ \dfrac{x-a}{b-a} & ; \quad \text{if } a \le x \le b \\ 1 & ; \quad \text{if } x > b \end{cases}$$

(*iii*) **Mean (μ)** = $(a + b)/2 = E(X)$

(*iv*) **Variance (σ^2)** = $(b - a)^2/12$

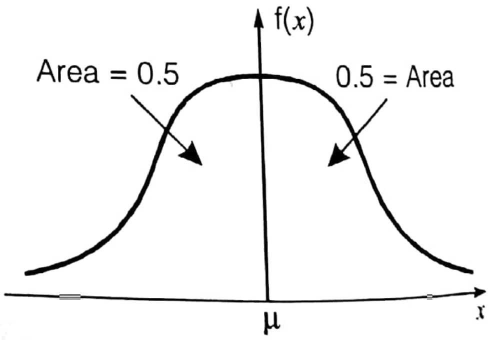
2. Normal Distribution

(*i*) Density function $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; -\infty \le x \le \infty, \sigma > 0, -\infty < \mu < \infty$

(*ii*) Normal distribution is symmetrical

(*iii*) Mean = μ; Variance = σ^2

(*iv*) $f(x) \ge 0$ for all x



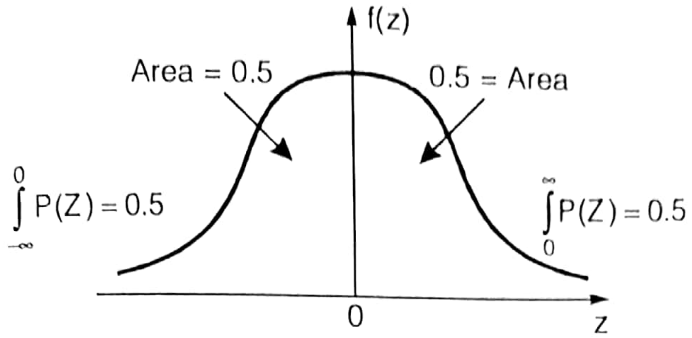
(*v*) $\displaystyle\int_{-\infty}^{\infty} f(x) \cdot dx = 1$

(*vi*) $P(Z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}; -\infty \le Z \le \infty$

and $Z = \dfrac{x-\mu}{\sigma}$

$Z = \dfrac{x-np}{\sqrt{npq}}$ (when approximating binomial by normal)

Z = Standard normal variate



3. Exponential Distribution

(*i*) Density function:

$$f(x) = \lambda \cdot e^{-\lambda x} \quad ; x > 0$$
$$= 0 \quad ; \text{ Otherwise}$$

(*ii*) Mean (μ) = $\frac{1}{\lambda} = S.D(\sigma)$

(*iii*) Variance (σ^2) = $\frac{1}{\lambda^2}$

X	X Measures	f(x)	Value of X	E(x)	V(x)
Continuous uniform	Outcomes with equal density (continuous)	$\frac{1}{b-a}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential	Time between events time until an event	$\lambda e^{-\lambda x}$	$x \ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal	Values with a bell-shaped distribution(continuous)	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$	μ	σ
Standard normal (Z)	Standard scores	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$	$Z = \frac{x-\mu}{\sigma}$	0	1
Binomial approximation	Number of successes in a fixed time period (large average)	Approx, normal if np ≥ 5 and n(1 – p) ≥ 5 by CLT	$Z = \frac{x-np}{\sqrt{np(1-p)}}$	np	np(1 – p)
Poisson approximation	Number of occurrences in a fixed time period (large average)	Approx normal if $\lambda > 30$	$Z = \frac{x-\lambda}{\sqrt{\lambda}}$	λ	λ

■■■