Joint Probability Distribution

Author:CHEA MAKARA Date: July 2022

Summarise

1. $D = \{(x, y) : x \in D_x, y \in D_y\}$ the set of all possible values of rv.

2. Let X and Y be two $drv's$ on sample space S. The probability that $X = x$ and $Y = y$ is denoted by

$$p(x, y) = P(X = x, Y = y)$$

3. $p(x, y_)$ is called joint pmf if

   - $0 \leq p(x, y) \leq 1$.
   - $\sum \sum_{(x,y)\in D} p(x, y) = 1$.
   - $P[(X, Y) \in A] = \sum \sum_{(x,y)\in A} p(x, y)$, where $A \subset D$

4. marginal pmf denoted by $p_X(x)$ is given by
$P_X(x) = \sum_{y\in D_y} p(x, y)$ for each possible values x and
$P_Y(y) = \sum_{x\in D_x} p(x, y)$ for each possible value y

5. Let X and Y be two crv's. The joint pdf of function $f(x, y)$ satisfying $f(x, y) \geq 0$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1.$$

Then for any (measurable) set $A \subset \mathbb{R}^2, P((X, Y) \in A) = \int_A f(x, y)dxdy$.

> In particular, if $A = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$, then
> $P((X, Y) \in A) = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)dxdy$

6. The marginal pdf of X and Y is given by
$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ for $-\infty < x < \infty$ and
$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$ for $-\infty < y < \infty$

> Two rv X and Y are said to be independent if for every pair of x and y values
> $p(x, y) = p_X(x)p_Y(x)$ X, Y are drv and
> $f(x, y) = f_X(x)f_Y(y)$ X, Y are crv
> Otherwise $X$and Y are said to be dependent

7.
- If $X_1, X_2, \cdots, X_n$ are all drv the joint pmf of the variables is the function $p(x_1, x_2, \cdots, x_n) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$
- If $X_1, X_2, \cdots, X_n$ are all crv the joint pdf of the variables is the function $f(x_1, x_2, \cdots, x_n)$ such that for any n intervals $[a_1, b_1], \cdots, [a_n, b_n]$, $P(a_1 \le X_1 \le b_1, \cdots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \cdots, x_n) dx_n \cdots dx_1$

8. The random variables $X_1, X_2, \cdots, X_n$ are said to be independent if for every subset $X_{i_1}, X_{i_2}, \cdots, X_{i_n}$ of the variable (each pair,each triple, and so on), the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's .

9. X, Y are crv then the conditional probability density function of Y given that $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, -\infty < y < \infty$$

For X, Y are drv just replace pdf's by pmf's

10. The expected value is defined by

$$\mu_X = E[h(x,y)] = \begin{cases} \sum_x \sum_y h(x,y)p(x,y), & X, Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy, & X, Y \text{ are continuous} \end{cases}$$

11. The Covariance between two rv's X and Y is

$$\begin{aligned} Cov(X,Y) &= E[(X - \mu_X(Y - \mu_Y] \\ &= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x,y), & X, Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x,y)dxdy, & X, Y \text{ are continuous} \end{cases} \end{aligned}$$

Theorem 1 $Cov(X,Y) = E(XY) - \mu_X\mu_Y$

12. The correlation coefficient of X and Y , denoted by $Corr(X,Y)$, $\rho_{X,Y}$ or ρ is defined by

$$\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X . \sigma_Y}$$

Theorem2

- if a and c are either both positive or both negative , $Corr(aX+b, cY+d) = Corr(X,Y)$
- For any two rv's X and Y,$-1 \le Corr(X,Y) \le 1$.
- If X and Y are independent , then $\rho = 0$
- $\rho = 1 or - 1$ iff $Y = aX + b$ for some numbers a and b with $a \neq 0$

13. The rv's $X_1, X_2, \cdots, X_n$ are said to form a (simple) random sample of size n if

- The $X_i's$ are independent random variable.
- Every X_i has the same probability distribution .

14. Let $X_1, X_2, \cdots, X_n$ be a random sample from distribution with mean values μ and standard deviation σ and let

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \text{ (sample mean)},$$

Then

- $E(\bar{X}) = \mu_{\bar{x}} = \mu$
- $V(\bar{X}) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

> In addition , with $T_0 = X_1 + \cdots + X_n$ (The sample total).
> $E(T_0) = n\mu, V(T_0) = n\sigma^2$ and $\sigma_{T_0} = \sqrt{n}\sigma$

15. The Central Limit Theorem

> Theorem 4 if $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$

16. Theorem 5 (The Cental Limit Theorem(CLT))

> Let $X_1, \ldots, X_n$ be a random sample from a distribution mean μ and variance σ^2.Then if n is sufficiently large , $\bar{X}$ has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$ and T_0 also has approximately a normal distribution with $\mu_{T_0} = n\mu, \sigma^2_{T_0} = n\sigma^2$, The larger the value of n , the better the approximation

Rule of Thumb: If $n > 30$, the CLT can be applied.

17. Given a collection of n random variable $X_1, \cdots, X_n$ and n numerical constant $a_1, \cdots, a_n$ the rv

$Y = a_1X_1 + \cdots + a_nX_n = \displaystyle\sum_{i=1}^{n} a_iX_i$ is called the linear combination of the $X_i's$

Theorem 6 Let $X_1, \cdots, X_n$ have mean values $\mu_1 \cdots, \mu_n$ respectively and variance $\sigma_1^2, \cdots, \sigma_n^2$ respectively

- Whether or not the $X_i's$ are independent.
  $E(a_1X_1 + \cdots + a_nX_n) = a_1E(X_1) + \cdots + a_nE(X_n)$
- if $X_1, \cdots, X_n$ are independent,
  $V(a_1X_1) + \cdots + a_nX_n) = a_1^2V(X_1 + \cdots + a_n^2V(X_n)$
  $\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}$
- For any $X_1, \cdots, X_n$
  $V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)$