X & Y are Independent

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

Then $\rightarrow$ $E(XY) = E(X)E(Y)$

CDF

$$F(x,y) = \sum^{y}\sum^{x} P(X=x, Y=y) = P(X \le x, Y \le y)$$

Bivariate Random Variables X & Y

Discrete

Joint PMF

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

(i) $P(X=x, Y=y) \geq 0$

(ii) $\sum_y \sum_x P(X=x, Y=y) = 1$

Marginal PMFs

$$P(X=x) = \sum_y P(X=x, Y=y)$$

$$P(Y=y) = \sum_x P(X=x, Y=y)$$

Expectation,

$$E(g(x,y)) = \sum_y \sum_x g(x,y) P(X=x, Y=y)$$

$$E(g(x)) = \sum_x g(x) P(X=x)$$

$$E(g(y)) = \sum_y g(y) P(Y=y)$$

Continuous

Joint PDF

$$f(x,y)$$

(i) $f(x,y) \geq 0$

(ii) $\int_y \int_x f(x,y)\,dx\,dy = 1$

$$f(x) = \int_y f(x,y)\,dy$$

$$f(y) = \int_x f(x,y)\,dx$$

Marginal PDFs

$$f(x,y) = f(x)f(y)$$

X & Y are Independent

$$E(g(x,y)) = \int_y \int_x g(x,y) f(x,y)\,dx\,dy$$

$$E(g(x)) = \int_x g(x) f(x)\,dx$$

$$E(g(y)) = \int_y g(y) f(y)\,dy$$

$$E(XY) = E(X)E(Y)$$

CDF

$$F(x,y) = \int^{y}\int^{x} f(x,y)\,dx\,dy$$

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

Conditional

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

X & Y are Independent

Discrete | **Continuous**

$$P(X=x, Y=y) = P(X=x)P(Y=y) \qquad f(x,y) = f(x)\,f(y)$$

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} \qquad f(x \mid y) = \frac{f(x,y)}{f(y)}$$

$$= P(X=x) \qquad = f(x)$$

Likewise

$$P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)} \qquad f(y \mid x) = \frac{f(x,y)}{f(x)}$$

$$= P(Y=y) \qquad = f(y)$$

$$E(XY) = E(X)E(Y) \qquad E(XY) = E(X)E(Y)$$

Expectations

Discrete | **Continuous**

$$E(g(Y) \mid X=x) = \sum_{y} g(y)\, P(Y=y \mid X=x) \qquad \int_{y} g(y)\, f(y \mid x)\, dy$$

$$E(g(X) \mid Y=y) = \sum_{x} g(x)\, P(X=x \mid Y=y) \qquad \int_{x} g(x)\, f(x \mid y)\, dx$$

V. Imp

Thm. Let X & Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by

$$M_Z(t) = M_X(t)\, M_Y(t)$$

Proof.

$$M_Z(t) = E(e^{tz}) = E(e^{t(x+y)})$$

$$= E(e^{tx}\, e^{ty})$$

$$= E(e^{tx})\, E(e^{ty})$$

$$= M_X(t)\, M_Y(t)$$

Example Given X & Y are Independent random variables,

Let $X \sim N(\mu_1, \sigma_1^2)$

$Y \sim N(\mu_2, \sigma_2^2)$

That $Z = X + Y \sim$?

$$M_Z(t) = M_X(t)\, M_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}\, e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$$

$$= e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$ which is MGF of normal distn.

$$\therefore \quad Z \sim N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$$

Eg. Let $X_1, X_2, \ldots X_n$ are independent variables.

Further $X_i \sim$ Bernoulli(p) distribution

$$Y = \sum X_i \sim ?$$

Sol^n.

$$M_Y(t) = M_{X_1}(t)\, M_{X_2}(t) \ldots M_{X_n}(t)$$

$$= (q + pe^t)(q + pe^t) \ldots (q + pe^t)$$

$$= (q + pe^t)^n$$

which is MGF of Binomial (n, p) distribution

$\therefore$ $Y \sim$ Binomial (n, p)

Eg Let $X_1, X_2, \ldots X_n$ are independent random variables.

Further $X_i \sim$ Poisson(λ_i) distribution.

Then $Y = \sum X_i \sim ?$

Sol^n.

$$M_Y(t) = M_{X_1}(t)\, M_{X_2}(t) \ldots M_{X_n}(t)$$

$$= e^{\lambda_1(e^t - 1)}\, e^{\lambda_2(e^t - 1)} \ldots e^{\lambda_n(e^t - 1)}$$

$$= e^{\sum_{i=1}^{n} \lambda_i (e^t - 1)}$$

which is again MGF of Poisson$\left(\sum_{i=1}^{n} \lambda_i\right)$

$\therefore$ $Y \sim$ Poisson$\left(\sum_{i=1}^{n} \lambda_i\right)$

Note: if $X_i \sim \text{Poisson}(\lambda)$

Then $Y = \sum X_i \sim \text{Poisson}(n\lambda)$

Q: Given X & Y are independent random variables
if $X \sim \text{Binomial}(n_1, p)$

$Y \sim \text{Binomial}(n_2, p)$

Then $Z = X + Y \sim ?$

$$M_Z(t) = M_X(t)\, M_Y(t)$$

$$= (q + pe^t)^{n_1} (q + pe^t)^{n_2}$$

$$= (q + pe^t)^{n_1 + n_2}$$ which is again PMF of Binomial$(n_1 + n_2, p)$

$\therefore Z \sim \text{Binomial}(n_1 + n_2, p)$

Covariance

$$\text{Cov}(X, Y) = E\big((X - E(X))(Y - E(Y))\big)$$

OR

$$E(XY) - E(X)E(Y)$$

Correlation

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Correlation Coefficient

* if X & Y are independent Then

$\text{Cov}(X, Y) = 0$

So

$\rho_{XY} = 0$

1ub

$$Var(X+Y) = E\left((X+Y)^2\right) - [E(X+Y)]^2$$

$$= E\left(X^2+Y^2+2XY\right) - [E(X)+E(Y)]^2$$

$$= E(X^2) + E(Y^2) + 2E(XY) - [E(X)]^2 - [E(Y)]^2 - 2E(X)E(Y)$$

$$= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2$$

$$+ 2\left[E(XY) - E(X)E(Y)\right]$$

$$= Var(X) + Var(Y) + 2\,Cov(X,Y)$$

<u>Likewise</u>

$$Var(X-Y) = Var(X) + Var(Y) - 2\,Cov(X,Y)$$

$X' = X+a, \quad Y' = Y+b$

$$Cov(X',Y') = \cancel{E(X'Y')}\; E\left((X'-E(X'))(Y'-E(Y'))\right)$$

$$= E\left((X+a-E(X+a))(Y+b-E(Y+b))\right)$$

$$= E\left((X-E(X))(Y-E(Y))\right)$$

$$= Cov(X,Y)$$

$\therefore$ Covariance is independent of change of origin

So Correlation is independent of change of origin

$\{ \because$ sd is also indepen

Let $x' = ax$, $y' = by$

Then $Cov(x', y') = E(x'y') - E(x')E(y')$

$$= E(ax\,by) - E(ax)\,E(by)$$

$$= ab\,E(xy) - ab\,E(x)\,E(y)$$

$$= ab\left[E(xy) - E(x)E(y)\right]$$

$$= ab\,Cov(x,y)$$

$\therefore$ Covariance is not independent of change of scale.

$$\rho_{x'y'} = \frac{Cov(x', y')}{\sigma_{x'}\,\sigma_{y'}} = \frac{ab\,Cov(x,y)}{a\sigma_x\,b\sigma_y} = \frac{ab\,Cov(x,y)}{ab\,\sigma_x\sigma_y}$$

$$= \frac{Cov(x,y)}{\sigma_x\sigma_y}$$

$$= \rho_{xy}$$

$\therefore$ Correlation is independent of change of scale

$Var(ax+by) = a^2 V(x) + b^2 V(y) + 2ab\, Cov(x,y)$

if x & y are independent Then

$V(ax+by) = a^2 V(x) + b^2 V(y)$

* v.imp

Prove That $-1 \le \rho_{xy} \le 1$

Now

$$V\left(\frac{x}{\sigma_x} \pm \frac{y}{\sigma_y}\right) = V\left(\frac{x}{\sigma_x}\right) + V\left(\frac{y}{\sigma_y}\right) \pm 2\,Cov\left(\frac{x}{\sigma_x}, \frac{y}{\sigma_y}\right)$$

$$= \frac{1}{\sigma_x^2} V(x) + \frac{1}{\sigma_y^2} V(y) \pm 2\frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$= 1 + 1 \pm 2\rho_{xy}$$

$$= 2 \pm 2\rho_{xy}$$

Since $V(\cdot) \geq 0$

$\therefore V\left(\frac{x}{\sigma_x} + \frac{y}{\sigma_y}\right) \geq 0$

$2 + 2\rho_{xy} \geq 0$

$2 \geq -2\rho_{xy}$

$1 \geq -\rho_{xy}$

$\cancel{\rho_{xy}}$

$-1 \le \rho_{xy}$

Also

$V\left(\frac{x}{\sigma_x} - \frac{y}{\sigma_y}\right) \geq 0$

$2 - 2\rho_{xy} \geq 0$

$2 \geq 2\rho_{xy}$

$1 \geq \rho_{xy}$

$\rho_{xy} \le 1$

$\therefore \boxed{-1 \le \rho_{xy} \le 1}$

For given data on X & Y

$$\mathrm{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} xy - \bar{x}\,\bar{y}$$

or

$$\frac{1}{n}\sum (x-\bar{x})(y-\bar{y})$$

$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{\mathrm{sd}(x)\ \mathrm{sd}(y)} = \frac{\frac{1}{n}\sum xy - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}\ \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}}$$

Regression line

Y on X $\quad (Y-\bar{y}) = b_{yx}(x-\bar{x})$

X on Y $\quad (x-\bar{x}) = b_{xy}(y-\bar{y})$

regression coefficient b_{xy} & b_{yx} $\rightarrow$ $b_{yx} = \dfrac{\mathrm{Cov}(x,y)}{V(x)}$, $\quad b_{xy} = \dfrac{\mathrm{Cov}(x,y)}{V(y)}$

$$b_{yx}\,b_{xy} = \frac{\mathrm{Cov}(x,y)}{V(x)}\ \frac{\mathrm{Cov}(x,y)}{V(y)}$$

$$= \frac{[\mathrm{Cov}(x,y)]^2}{[\mathrm{sd}(x)\ \mathrm{sd}(y)]^2} = \left[\frac{\mathrm{Cov}(x,y)}{\mathrm{sd}(x)\ \mathrm{sd}(y)}\right]^2$$

$$= [\mathrm{Corr}(x,y)]^2$$

$$\therefore\ \mathrm{Corr}(x,y) = \sqrt{b_{yx}\,b_{xy}}$$

Bivariate Normal Distribution

The Bivariate normal PDF with means μ_x and μ_y, variances σ_x^2 and σ_y^2, and Correlation ρ, is given by

$$f(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma_x\sigma_y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

$-\infty < x < \infty$

$-\infty < y < \infty$

$-\infty < \mu_x < \infty$

$-\infty < \mu_y < \infty$

$\sigma_x > 0$

$\sigma_y > 0$

Here $\rho = \rho_{xy}$

Note if x & y are independent The $\rho = 0$

$$\therefore f(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma_x\sigma_y}\, e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi}\,\sigma_y}\, e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

$$= f(x)\, f(y)$$

when $X \sim N(\mu_x, \sigma_x^2)$

$Y \sim N(\mu_y, \sigma_y^2)$

Random Sample $X_1, X_2, \ldots X_n$

i.i.d

Independent & Identically distributed

→ marginal pdf or pmf of each X_i is the same function.

Sample Mean — $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sample variance — $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

v.v.Imp

Thm Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

a) $E(\bar{X}) = \mu$

b) $V(\bar{X}) = \sigma^2/n$

c) $E(S^2) = \sigma^2$

Proof.

a) $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$

$\left\{ E(X_1) = E(X_2) = \ldots = \mu \right.$

$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$

$= \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right]$

$= \frac{1}{n} n\mu = \mu$

$$V(\bar{x}) = \frac{1}{n} V\left(\frac{1}{n} \sum x_i\right)$$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2} V(x_1 + x_2 + \text{—} + x_n)$$

$$= \frac{1}{n^2} \left[V(x_1) + V(x_2) + \text{—} + V(x_n) \right]$$

$\left\{ \because \text{ of Independence} \right.$ $Cov(x_i, x_j) = 0$ $i \neq j$

$$= \frac{1}{n^2} \left[\sigma^2 + \sigma^2 + \text{——} + \sigma^2 \right]$$

$\left\{ \because \text{ of identically distributed} \right.$ $V(x_1) = V(x_2) = \text{---} = \sigma^2$

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

c)

$$E(S^2) = E\left(\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]\right)$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

$$= \frac{1}{n-1} \left[E(x_1^2) + E(x_2^2) + \text{—} E(x_n^2) - n E(\bar{x}^2) \right]$$

$$= \frac{1}{n-1} \left[(\sigma^2 + \mu^2) + (\sigma^2 + \mu^2) + \text{—} + (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \right]$$

$$\because V(x) = E(x^2) - [E(x)]^2$$

$$\Rightarrow E(x^2) = V(x) + [E(x)]^2 = \sigma^2 + \mu^2$$

$$E(S^2) = \frac{1}{n-1}\left[n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$$

$$= \frac{1}{n-1}(n-1)\sigma^2$$

$$= \sigma^2$$

Note

$\bar{X}$ is an unbiased estimator of μ ; $E(\bar{x}) = \mu$

S^2 is an unbiased estimator of σ^2 ; $E(S^2) = \sigma^2$

Thm. if x is a random sample from a population with MGF $M_x(t)$. Then the MGF of

if $Y = \sum_{i=1}^{n} x_i$ Then $M_Y(t) = [M_x(t)]^n$

if $Y = \bar{X}$ Then $M_{\bar{x}}(t) = [M_x(t/n)]^n$

Proof

$Y = \sum x_i$

$$M_Y(t) = E(e^{tY}) = E(e^{t(x_1+x_2 \text{ --- } + x_n)})$$

$$= E(e^{tx_1} e^{tx_2} \text{ --- } e^{tx_n})$$

$$= E(e^{tx_1}) E(e^{tx_2}) \text{ --- } E(e^{tx_n}) \quad \{\because \text{ of Independent}$$

$$= M_{x_1}(t) M_{x_2}(t) \text{ --- } M_{x_n}(t)$$

$$= M_x(t) M_x(t) \text{ --- } M_x(t) \quad \{\because \text{ of Identically distributed}$$

$$= [M_x(t)]^n$$

Likewise

$$M_{\bar{x}}(t) = E(e^{t\bar{x}})$$

$$= E(e^{\frac{t}{n}\sum x_i})$$

$$= E\left(e^{\frac{x_1}{n}t + \frac{x_2}{n}t + \cdots \frac{x_n t}{n}}\right)$$

$$= E\left(e^{\frac{x_1}{n}t}\, e^{\frac{x_2}{n}t} \cdots e^{\frac{x_n}{n}t}\right)$$

$$= E(e^{\frac{x_1}{n}t})\, E(e^{\frac{x_2}{n}t}) \cdots E(e^{\frac{x_n}{n}t})$$

$$= M_{x_1}(\tfrac{t}{n})\, M_{x_2}(\tfrac{t}{n}) \cdots M_{x_n}(\tfrac{t}{n})$$

$$= M_x(\tfrac{t}{n})\, M_x(\tfrac{t}{n}) \cdots M_x(\tfrac{t}{n})$$

$$= \left[M_x(\tfrac{t}{n})\right]^n$$

V.V.Imp

Q. Suppose X is a random sample from a normal population $N(\mu, \sigma^2)$. Then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Sol^n.

$$M_{\bar{x}}(t) = \left[M_x(\tfrac{t}{n})\right]^n \qquad \left\{ \begin{array}{l} \because X \sim N(\mu, \sigma^2) \\ M_x(t) = e^{\mu t + \frac{t^2\sigma^2}{2}} \end{array} \right.$$

$$= \left[e^{\mu\frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2}}\right]^n$$

$$= e^{\mu + \frac{\sigma^2}{n}\frac{t^2}{2}}$$ which is mgf of normal distribution

$$M_{\bar{x}}(t) = e^{\mu t + \left(\frac{\sigma^2}{n}\right)\frac{t^2}{2}}$$

$$\therefore \quad \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Q. If x is a random sample from Bernoulli (p) distribution

Then $Y = \sum x_i \sim ?$

Soln

$$M_Y(t) = \left(M_x(t)\right)^n$$

$$= \left[q + pe^t\right]^n$$

which is MGF of Binomial (n, p) distribution

$$\therefore \quad Y = \sum x_i \sim \text{Binomial}(n, p)$$