Chebyshev's inequality:

It will give us a means of understanding precisely how the variance measures the variability about the expected value of a rv. If we know the probability distribution of a rv we may then compute E(X) and V(X), however the converse is NOT TRUE. Nevertheless, we can give a very useful upper or lower bound to such a probability.

Let X be a rv with $E(X) = \mu$ and let 'c' be any real number. Then, if E(X-c)<∞ and $\varepsilon > 0$, we have

$$P\{|X - c| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} E(X - c)^2$$

This is known as Chebyshev's inequality

Prove it!

Alternate forms:

(i). $P(|X - c| < \varepsilon) \geq 1 - \frac{1}{\varepsilon^2} E(X - c)^2$ (Complimentary event)

(ii). If $c = \mu$ ($=$ mean) then $P(|X - \mu| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$

(iii). If $c = \mu, \varepsilon = k\sigma;\ \sigma^2 = V(X)$ then $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

If variance V(X) is small, most of the probability distribution of X is concentrated near the mean, $E(X) = \mu$ and when V(X)=0 we have X=E(X), i.e. all the values assumed by the rv X coincide with its mean.

Correlation Coefficient:

Correlation coefficient between two rvs X and Y denoted by ρ_{XY} or ρ is a measure of the degree of linearity or linear relationship between two rvs and is given by,

$$\rho = \frac{Cov(XY)}{\sqrt{V(X)V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

If $\rho = 0$ then the rvs are said to be uncorrelated. But the converse need not be true i.e. if the rvs are uncorrelated then they need not be independent.

If $\rho = \pm 1$ then the rvs are said to be perfectly correlated.

$-$ve values of $\rho \Rightarrow X \uparrow Y \downarrow$

$+$ve values of $\rho \Rightarrow X \uparrow Y \uparrow$

Theorem 1:

The correlation coefficient between two rvs lies between -1 and +1.

i.e. $-1 \leq \rho \leq +1$,

Prove it!

Theorem 2:
The Correlation coefficient is independent of change of origin and scale.
i.e., $\rho_{UV} = \pm\rho_{XY}$; where U=a+bX; V=c+dY

Prove it!

Theorem 3:
If X and Y are linearly related, then $\rho = \pm 1 (\equiv \rho^2 = 1)$ and conversely.

Prove it!

Problems

1. To Show that Cov(XY)=0 does not necessarily imply that the rv are independent

Solution:
Let X be a rv with pdf
$X \sim f(x) = \begin{cases} \frac{1}{2}; & -1 \le x \le 1 \\ 0; & \text{elsewhere} \end{cases}$ (Can you identify the distribution?)

Let $Y=X^2$ (dependence is quadratic)

Now $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^{1} x \frac{1}{2} dx = 0$
$E(XY)=E(X^3) = \int_{-1}^{1} x^3 f(x) dx = \int_{-1}^{1} \frac{x^3}{2} dx = 0$
$\therefore \text{Cov}(XY) = E(XY) - E(X)E(Y) = 0$

2. Find the mean and the variance of probability distribution where $g(0) = \frac{16}{31}$, $g(1) = \frac{8}{31}$, $g(2) = \frac{4}{31}$, $g(3) = \frac{2}{31}$, $g(4) = \frac{1}{31}$

Solution:

Given function is a probability distribution (discrete)

Mean = E(X) = $\sum_{x=0}^{4} x g(x) = 0.\frac{16}{31} + 1.\frac{8}{31} + 2.\frac{4}{31} + 3.\frac{2}{31} + 4.\frac{1}{31} = \frac{26}{31}$

Variance = V(X) = $E(X^2) - (E(X))^2$

where, $E(X^2) = \sum_{x=0}^{4} x^2 g(x) = 0.\frac{16}{31} + 1.\frac{8}{31} + 4.\frac{4}{31} + 9.\frac{2}{31} + 16.\frac{1}{31} = \frac{58}{31}$

$\therefore V(X) = \frac{58}{31} - \left(\frac{26}{31}\right)^2 = \frac{1122}{961}$

3. A fair die is tossed 72 times. Given that X: No. of times 6 appears. Evaluate $E(X^2)$?

   Solve it!

4. Suppose a rv X has mean 10 and variance 25, for what +ve values of 'a' and 'b' does the rv Y= aX-b have expectation (mean) 0 and variance 1.

   Solve it!

5. The rv (X,Y) has a joint pdf given by $f(x,y) = \begin{cases} x+y; & 0 \le (x,y) \le 1 \\ 0; & \text{elsewhere} \end{cases}$.

   Find correlation coefficient ρ_{XY}.

Solution: We have

$$\rho_{XY} = \frac{\mathrm{Cov}(XY)}{\sqrt{V(X)V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$$

$$E(XY) = \int_0^1 \int_0^1 xy(x+y)dxdy = \int_0^1 y\left[\frac{x^3}{3} + \frac{x^2y}{2}\right]_0^1 dy = \int_0^1 y\left[\frac{1}{3} + \frac{y}{x}\right]$$

$$= \int_0^1 \left[\frac{4}{3} + \frac{y^2}{x}\right]dy = \frac{y^2}{6} + \frac{y^3}{6}\Big]_0^1 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$E(X) = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)dydx = \int_{-\infty}^{\infty} xg(x)dx$$

where, $g(x) = \int_0^1 (x+y)dy = x + \frac{1}{2}$

$$\therefore E(X) = \int_0^1 x\left(x + \frac{1}{2}\right)dx = \left[\frac{x^3}{3} + \frac{x^2}{4}\right]_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$E(Y) = \int_0^1 yh(y)dy$$

$$h(y) = \int_0^1 (x+y)dx = y + \frac{1}{2}$$

$$\therefore E(Y) = \int_0^1 y\left(y + \frac{1}{2}\right) = \frac{7}{12}$$

$$V(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \int_0^1 x^2\left(x + \frac{1}{2}\right)dx = \frac{x^4}{6} + \frac{x^2}{4} = \frac{10}{24} = \frac{5}{12}$$

$$V(X) = \frac{5}{12} - \frac{49}{144} = \frac{11}{144} = V(Y)$$

$$\rho_{XY} = \frac{\left(\frac{1}{3}\right) - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right)}{\sqrt{\left(\frac{11}{144}\right)\left(\frac{11}{144}\right)}} = \frac{\left(\frac{1}{3}\right) - \left(\frac{49}{144}\right)}{\frac{11}{144}} = \frac{\frac{-1}{144}}{\frac{11}{144}} = \frac{-1}{11}$$

6. Find ρ_{XY}, given the joint pdf of (X,Y) as $f(xy) = \begin{cases} 2 - x - y; & 0 \leq (xy) \leq 1 \\ 0 & ; \text{otherwise} \end{cases}$

   Solve it!

7. Two independent variates X_1, X_2 have means 5, 10 and variances 4, 9 respectively. Find covariance between $U=3X_1+4X_2$; $V=3X_1-X_2$

   Solve it!

8. Let X_1, X_2, X_3 be uncorrelated rvs having same standard deviation. Find the correlation coefficient between U and V where $U=X_1+X_2$ and $V=X_2+X_3$

   Solve it!

9. If X,Y,Z are uncorrelated rvs with V(X)=25, V(Y)=144 and V(Z)=81. Find ρ_{UV} where U=X+Y and V=Y+Z.

   Solve it!