Probability and Statistics

(September - 21)

Random vectors: $(\Omega, \mathcal{F}, P)$

$\underline{X} : \Omega \longrightarrow \mathbb{R}^r$

$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$ where $X_1, X_2, \ldots, X_r$ are random variables.

discrete random vector, joint p.m.f., marginal p.m.f.s.

Independence of r.v.s.

$X_1, \cdots, X_r$ are r.v.s. with marginal p.m.f.s $f_{X_1}(x_1), \ldots, f_{X_r}(x_r)$ resp. and joint p.m.f. is given by $f_{X_1,\ldots,X_r}(x_1,\ldots,x_r)$. Then $X_1, \cdots, X_r$ are called as mutually independent, if

$$f_{X_1,\ldots,X_r}(x_1,\ldots,x_r) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_r}(x_r)$$

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix} \leftarrow \text{r-dimensional random vector.}$$

$r=1$; pmf → one-dimensional array (vector)

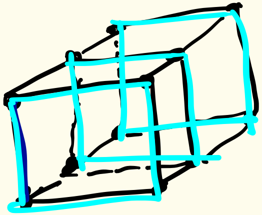
$r=2$; pmf → two-dimensional array (matrix)

$r=3$; pmf → three-dimensional array (tensor)

$=4$

5

$\vdots$



$\Omega = \{H, T\}$

$X_1 = 0 \quad \forall \omega \in \Omega$

$X_2 = \begin{cases} 0 & \omega = H \\ 1 & \omega = T \end{cases}$

$X_3 = \begin{cases} -100 & \omega = T \\ 100 & \omega = H \end{cases}$

Sum of <u>independent</u> random variable. (discrete)

Let X and Y be two independent r.v.s.

$$R_X = \{x_1, x_2, \dots\}$$

Interested in the event

$$\{X+Y = z\} \quad \forall z$$

$$\| $$

$$\bigcup_i \{X = x_i,\ Y = z - x_i\}$$ ← This union is disjoint.

$$P(X+Y=z) = P\left(\bigcup_i \{X=x_i,\ Y=z-x_i\}\right) = \sum_i P(X=x_i,\ Y=z-x_i)$$

independence.

$$= \sum_i P(X=x_i)\,P(Y=z-x_i)$$

$$\boxed{f_{X+Y}(z) = \sum_x f_X(x)\, f_Y(z-x)}$$

Expectation: (discrete case)

Let $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$ be a given discrete random vector

with joint p.m.f. $f_{X_1, X_2, \ldots, X_r}(x_1, \ldots, x_r)$

Let $h(x_1, \ldots, x_r)$ be a f^n of $x_1, \ldots, x_r$

$$\boxed{E(h(X_1, \ldots, X_r)) = \sum_{x_1, \ldots, x_r} h(x_1, \ldots, x_r) f_{X_1, \ldots, X_r}(x_1, \ldots, x_r)}$$

for $r=2$; $h(x_1, x_2) = x_1 + x_2$

$$E(X_1 + X_2) = \sum_{x_1} \sum_{x_2} (x_1 + x_2) f_{X_1, X_2}(x_1, x_2)$$

$r=2, \quad h(x_1,x_2)=x_1$

$$E(X_1) = \sum_{x_1}\sum_{x_2} x_1 f_{X_1,X_2}(x_1,x_2)$$

$$= \sum_{x_1} x_1 \boxed{\sum_{x_2} f_{X_1,X_2}(x_1,x_2)} \leftarrow \text{marginal of } X_1 \text{ which } f_{X_1}(x_1)$$

$$\boxed{E(X_1) = \sum_{x_1} x_1 f_{X_1}(x_1)}$$

In general, take $h(x_1,\dots,x_r)=x_i$ for $1\le i\le r$

$$E(X_i) = \sum x_i f_{X_i}(x_i)$$

$f_{X_i}(x_i)$: marginal pmf of X_i

Let X and Y be independent discrete random variables with joint pmf $f_{XY}(x,y)$.

Let $h(x,y) = xy$

$$E(XY) = \sum_x \sum_y xy\, f_{XY}(x,y)$$

independence of X & Y.

$$= \sum_x \sum_y xy\, f_X(x)\, f_Y(y)$$

$$= \sum_x x\, f_X(x) \sum_y y\, f_Y(y)$$

$$\boxed{E(XY) = E(X)\,E(Y)}$$

independence of X and Y $\Rightarrow$ $E(XY) = E(X)E(Y)$

Suppose $h(x,y) = \varphi_1(x)\,\varphi_2(y)$

Assume that x & y are independent

$$E(h(x,y)) = E(\varphi_1(x)\,\varphi_2(y))$$

$$= \sum_x \sum_y \varphi_1(x)\,\varphi_2(y)\, f_{XY}(x,y)$$

indep. of x, y

$$= \sum_x \sum_y \varphi_1(x)\,\varphi_2(y)\, f_X(x)\, f_Y(y)$$

$$= \sum_x \varphi_1(x)\, f_X(x) \sum_y \varphi_2(y)\, f_Y(y)$$

$$\boxed{E(\varphi_1(x)\,\varphi_2(y)) = E(\varphi_1(x))\, E(\varphi_2(y))}$$

$\blacksquare$

Ex: Let X and Y be indep. discrete r.v.s. with joint pmf $f_{XY}(x,y)$.

MGF of $X+Y$.

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right)$$

$$= E\left(e^{tX} e^{tY}\right)$$

previous page indep. result.

$$= E\left(e^{tX}\right) E\left(e^{tY}\right)$$

$$\boxed{M_{X+Y}(t) = M_X(t)\, M_Y(t)}$$

Generalising this result, $X_1, \dots, X_n$ are mutually indep. discrete r.v.s.

$$\boxed{M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)}$$

Ex: Let $X_i \sim \text{Bernoulli}(p)$ for $i = 1, 2, \dots, n$

be iid (independent and identically distributed)

Let $Y = \sum_{i=1}^{n} X_i$

$$M_Y(t) = M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= \prod_{i=1}^{n} \left((1-p) + pe^t\right)$$

$M_Y(t) = \left((1-p) + pe^t\right)^n$ is the MGF of Binomial(n,p)

Binomial (n, p) = sum of 'n' independent Bernoulli (p).

Important observation:

X and Y are r.v.s. with joint pmf $f_{XY}(x,y)$ and marginal pmfs $f_X(x)$ & $f_Y(y)$ resp.

i) (X and Y are indepent) $\Longleftrightarrow$ $\boxed{f_{XY}(x,y) = f_X(x) \cdot f_Y(y)}$

ii) $\begin{pmatrix} X \text{ and } Y \text{ are} \\ \text{independent} \end{pmatrix} \Rightarrow E(XY) = E(X)E(Y)$
$\not\Leftarrow$
(already proved!!)

Q: Is the converse true??

A: NO. see the following example.

Ex: $\begin{pmatrix} X \\ Y \end{pmatrix}$ is a discrete random vector with range $R_{\begin{pmatrix} X \\ Y \end{pmatrix}} = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$ with each outcome equally likely.

$$p(x,y) = \frac{1}{4} \quad \text{for } \begin{pmatrix} x \\ y \end{pmatrix} \in R_{\begin{pmatrix} X \\ Y \end{pmatrix}}$$

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

$E(X) = 0$; $E(Y) = 0$; observe $XY = 0 \Rightarrow E(XY) = 0$

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

Observe: $P(X=0, Y=0) = 0 \neq P(X=0) \cdot P(Y=0) = \frac{1}{2} \times \frac{1}{2}$

$= \frac{1}{4}$

$\Rightarrow$ X and Y are $\boxed{\text{NOT}}$ independent.

9

Recall: $E(X+Y) = E(X) + E(Y)$

$E(aX) = a E(X)$

$E(X+c) = E(X) + c$

Similarly results for variances as well.

Sum of variances

Let X and Y be discrete r.v.s.

$$\text{var}(X+Y) = E(X+Y - E(X+Y))^2$$

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

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

$$\text{var}(X+Y) = \underbrace{\boxed{E(X-E(X))^2}}_{\text{var}(X)} + \underbrace{\boxed{E(Y-E(Y))^2}}_{\text{var}(Y)}$$

$$+ 2 \underbrace{E[(X-E(X))(Y-E(Y))]}_{\text{cov}(X,Y)}$$

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y)$$

Note: $\mathrm{Cov}(X,Y) = E\,(X - E(X))(Y - E(Y))$

$$= E\,(XY - X\,E(Y) - Y\,E(X) + E(X)E(Y))$$

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

$$\boxed{\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)}$$

Corrollary: If X and Y are independent, then $\mathrm{cov}(X,Y) = 0$.

pf: Since X & Y are indep $\Rightarrow E(XY) = E(X)E(Y)$

$$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0 \quad \blacksquare$$

1) Converse is <u>NOT</u> true.

$$\mathrm{cov}(x, y) = 0 \not\Rightarrow x \text{ and } Y \text{ are indepent.}$$

2) If x and Y are independent,

$$\mathrm{var}(x+y) = \mathrm{var}(x) + \mathrm{var}(Y)$$

$$\mathrm{var}(ax) = E(ax - E(ax))^2$$

$$= E(ax - aE(x))^2$$

$$= E\left[a^2(x - E(x))^2\right]$$

$$= a^2 E(x - E(x))^2$$

$$\boxed{\mathrm{Var}(ax) = a^2 \mathrm{var}(x)}$$

$$\boxed{\mathrm{var}(c) = 0 \text{ where } c \text{ is a const.}}$$

$$\mathrm{var}(x+c) = \mathrm{var}(x)$$

Let $x_1, x_2, \cdots, x_n$ be independent random variables with variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$ resp.

$$var(x_1 + x_2 + \cdots + x_n) = var(x_1) + var(x_2) + \cdots + var(x_n)$$
$$= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

Let $x_1, \cdots, x_n$ be iid with mean μ and variance σ^2.

Define: $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

$$E(\bar{x}) = E\left(\frac{x_1 + \cdots + x_n}{n}\right)$$

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

$$= \frac{1}{n}E(x_1) + \frac{1}{n}E(x_2) + \cdots + \frac{1}{n}E(x_n)$$

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

$$\boxed{E(\bar{x}) = \mu}$$

$$\mathrm{var}(\bar{x}) = \mathrm{var}\left(\frac{x_1 + \cdots + x_n}{n}\right) = \mathrm{var}\left(\frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n}\right)$$

$$= \frac{1}{n^2}\mathrm{var}(x_1 + \cdots + x_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(x_i) = \frac{1}{n^2}\cdot n\sigma^2$$

$$\boxed{\mathrm{var}(\bar{x}) = \sigma^2/n}$$

Correlation coefficient:

Let X and Y be two discrete r.v.s. Then correlation coefficient $\rho(X,Y)$ (rho) is defined as

$$\rho_{XY} = \rho = \rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{SD(X)\, SD(Y)}$$

$$= \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\,\sqrt{\operatorname{Var}(Y)}}$$

Observe: X and Y indep. $\Rightarrow \rho_{XY} = 0$

Converse is NOT true.

Theorem: (Cauchy - Schwartz inequality)

Let X and Y be random variables with finite second order moments.

$$[E(XY)]^2 \leq E(X^2)\,E(Y^2)$$

Furthermore, equality holds when

$$P(Y=0)=1$$

or $P(Y=aX)=1$ for some $a \in \mathbb{R}$

Pf: Easy to verify (write down if you are not sure) equality in case $P(Y=0)=1$ & $P(Y=aX)=1$ for $a \in \mathbb{R}$.

In order to prove inequality,

for any $\lambda \in \mathbb{R}$

$$0 \le \underbrace{E(X-\lambda Y)^2}_{} = \lambda^2 E(Y^2) - 2\lambda E(XY) + E(X^2)$$

Since the above expression is quadratic in λ and $E(Y^2) > 0$, the minimum value of this quadratic expression is achieved at λ^* is

$$2\lambda^* E(Y^2) - 2E(XY) = 0$$

$$\Rightarrow \lambda^* = \frac{E(XY)}{E(Y^2)}$$

The minimum value of this quadratic expression at $\lambda^* = \dfrac{E(xy)}{E(y^2)}$ is given by

$$\left[\frac{E(xy)}{E(y^2)}\right]^2 E(y^2) - 2\,\frac{E(xY)}{E(y^2)}\,E(xy) + E(x^2) \geq 0$$

$$\Rightarrow \quad -\frac{[E(xy)]^2}{E(y^2)} + E(x^2) \geq 0$$

$$\Rightarrow \quad \boxed{[E(xy)]^2 \leq E(x^2)\,E(y^2)}$$

∎

Importance of Schwartz inequality:

Apply this inequality on $X - E(X)$ and $Y - E(Y)$, we get,

$$\left[E\left(X - E(X)\right)\left(Y - E(Y)\right)\right]^2 \leq E\left(X - E(X)\right)^2 E\left(Y - E(Y)\right)^2$$

$$\Rightarrow \quad [\mathrm{cov}(X,Y)]^2 \leq \mathrm{var}(X)\,\mathrm{var}(Y)$$

$$\Rightarrow \quad \left[\frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)\,\mathrm{var}(Y)}\right]^2 \leq 1 \qquad \text{or} \quad \boxed{|\rho_{XY}| \leq 1}$$

$$\Rightarrow \quad \boxed{\rho_{XY}^2 \leq 1} \quad \Rightarrow \quad \boxed{-1 \leq \rho_{XY} \leq 1}$$

Further,

$$\rho_{xy} = \pm 1 \iff P(X = aY) = 1$$

WLLN (weak law of large numbers)

Recall: If X is a non-negative random variable with finite expectation, then for $t > 0$

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Chebyshev inequality: $P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$

Let $X_1, X_2, \cdots, X_n$ be i.i.d. r.v.s.

Let $\mu = \text{mean} = E(X_i)$ and $\sigma^2 = \text{var}(X_i)$

Define: $S_n = X_1 + \cdots + X_n$

$$E\left(\frac{S_n}{n}\right) = \mu \qquad \left(\text{Note: } \frac{S_n}{n} = \bar{X}\right)$$

$$\text{var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$$

Chebyshev inequality $\Rightarrow$

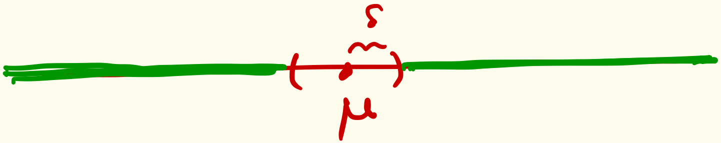
$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) \leq \frac{\text{var}(S_n/n)}{\delta^2} = \frac{\sigma^2}{n\delta^2}$$

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) \leq \frac{\sigma^2}{n\delta^2}$$

In particular,

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) = 0}$$

Weak Law of large number.



Hypergeometric pmf. (discrete)

Population of r objects

Type I objects r_1

Type II objects $r - r_1 = r_2$

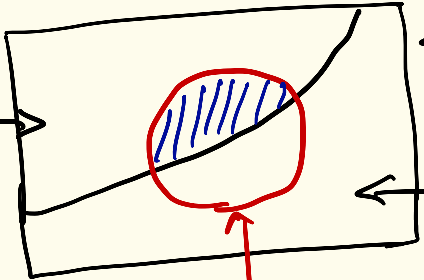
Let a sample of size n is chosen from this population. $(n \leq r)$

X : number of objects of Type I in the random sample.

entrie population 'r'

Type I

r_1



Type II

$r - r_1$

sample of size 'n'

$$P(X = x) = \frac{\binom{r_1}{x}\binom{r - r_1}{n - x}}{\binom{r}{n}} \qquad x = 0, 1, 2, \cdots, n$$

$$= 0 \qquad \text{o.w.}$$

Population (r)

Type I: r_1

Type II: $r - r_1$

A sample of size n is drawn from 'r'.

<u>Define</u>: X_i : i^{th} indicator random variable indicating whether the i^{th} object in the sample is of Type I.

0	0	1		1	
X_1	X_2	X_3	$\ldots$	X_n	$\leftarrow$ random sample of size n.
1	2	3		n	

clearly, $X = \sum_{i=1}^{n} X_i$

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

Note: $E(X_i) = \frac{r_1}{r}$

$$\Rightarrow \boxed{E(X) = n\frac{r_1}{r}}$$