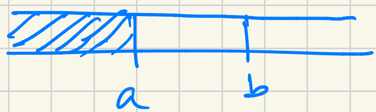
Continuous Random variable :

Cumulative distribution function (cdf) of a random variable X is given by,

$$F : \mathbb{R} \to [0,1] \qquad F(x) = P(X \leq x)$$

$$P\{a < X \leq b\} = P\{X \leq b\} - P(X \leq a) \quad \text{for } a<b$$
$$= F(b) - F(a).$$

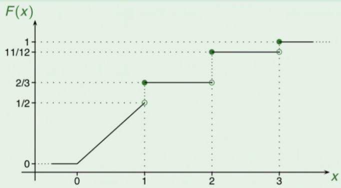


Properties of cdf (F)

- → non-decreasing
- → has $\lim_{x \to -\infty} F(x) = 0$ on the left
- → has $\lim_{x \to \infty} F(x) = 1$ on the right
- → is continuous from the right.

Distribution function

Example (…cont'd)



$$\mathbf{P}\{X \le 3\} = F(3) = 1,$$

$$\mathbf{P}\{X < 3\} = \lim_{x \nearrow 3} F(x) = \frac{11}{12},$$

$$\mathbf{P}\{X = 3\} = F(3) - \lim_{x \nearrow 3} F(x) = \frac{1}{12}.$$

$$P\{X \le 3\} = F(3) = 1$$

$$P\{X < 3\} = \lim_{x \to 3} F(x) = \frac{11}{12}$$

$$P\{X = 3\} = P\{X \le 3\} - P\{X < 3\}$$

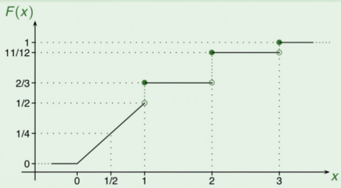
$$= F(3) - \lim_{x \to 3} F(x)$$

$$= 1 - \frac{11}{12}$$

$$= \frac{1}{12}$$

Distribution function

Example (…cont'd)

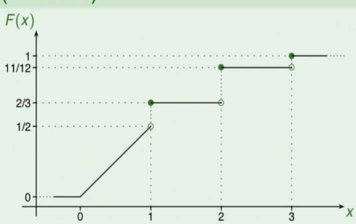


$$\mathbf{P}\left\{X \ge \frac{1}{2}\right\} = \mathbf{P}\left\{X > \frac{1}{2}\right\} = \frac{3}{4}.$$

$$P\{X \ge 1/2\} = P\{X > 1/2\} = 1 - \frac{1}{4} = 3/4$$

Distribution function

Example (…cont'd)



$$\mathbf{P}\{2 < X \le 4\} = F(4) - F(2) = 1 - \frac{11}{12} = \frac{1}{12},$$

$$\mathbf{P}\{2 \le X < 4\} = \mathbf{P}\{2 < X \le 4\} - \mathbf{P}\{X = 4\} + \mathbf{P}\{X = 2\}$$

$$= \frac{1}{12} - 0 + \left(\frac{11}{12} - \frac{2}{3}\right) = \frac{1}{3}.$$

$$P\{2 < X \le 4\} = P(4) - P(2)$$

$$= 1 - 11/12 = 1/12$$

$$P\{2 \le X < 4\} = P\{2 < X \le 4\} - P\{X = 4\} + P\{X = 2\}$$

$$= 1/12 - 0 + \left(P\{X \le 2\} - P\{X < 2\}\right)$$

$$= 1/12 + \left(\frac{11}{12} - 2/3\right)$$

$$= 1/12 + \left(\frac{11-8}{12}\right) = 1/12 + 1/4$$

$$= \frac{1+3}{12} = 1/3.$$

$P\{1 \leq x \leq 2\}$

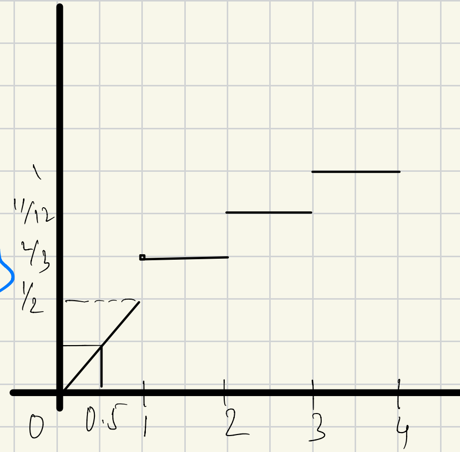
$\Rightarrow P\{x \geq 1\} + P\{1 < x \leq 2\}$

$\Rightarrow P\{x \leq 1\} - P\{x < 1\} + P\{1 < x \leq 2\}$

$\Rightarrow \frac{2}{3} - \frac{1}{2} + F(2) - F(1)$

$\Rightarrow \cancel{\frac{2}{3}} - \frac{1}{2} + \frac{11}{12} - \cancel{\frac{2}{3}}$

$\Rightarrow \frac{11-6}{12} = \frac{5}{12}$



for $P\{0 < x < 1\}$ there is a slope but, for $P\{1 < x < 2\}$ it's a flat line so probability will be 0.

☐ Density Function:

A probability density function f

→ is non-negative

→ has total integral $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$P(X \in B) = P(5 \leq X \leq 10) = \int_5^{10} \underline{f(x)}\,dx$$

↳ density function.

$B = [5, 10]$

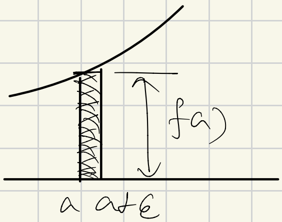
$$P(X = a) = \int_a^a f(x)\,dx = 0$$ when f is continuous.

Corollary:

for small ϵ,

$$P\{X \in (a, a+\epsilon]\} = \int_a^{a+\epsilon} f(x)\,dx \simeq f(a)\cdot\epsilon$$

$f(x) \rightarrow$ density function.



as ϵ is too small we can consider it as a rectangle.

** to get density function (pdf) from a distribution function (cdf),

$$f(a) = \frac{d}{da} F(a) \quad (a \in \mathbb{R})$$

Expectation:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx \quad \text{if integral exists.}$$

variance:

$$\operatorname{Var}(X) = E(X^2) - \{E(X)\}^2$$

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

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

$$E|X|^n = \int_{-\infty}^{\infty} |x|^n \cdot f(x)\,dx$$

** Let X be a continuous random variable, and g an $\mathbb{R} \rightarrow \mathbb{R}$ function then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx$$

Uniform distribution function:

for $\alpha < \beta$ where $\alpha, \beta \in \mathbb{R}$
we say that x has the uniform distribution over the interval (α, β)

$X \sim U(\alpha, \beta)$ if its density is given by

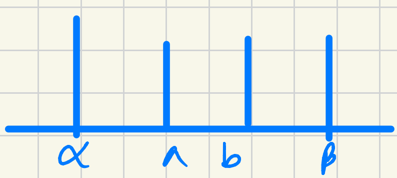
$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } x \in (\alpha, \beta) \\ 0 & \text{otherwise.} \end{cases}$$

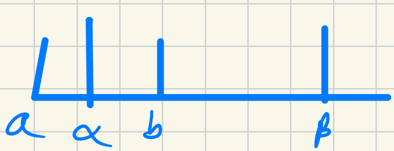
CDF

$$F(x) = \begin{cases} 0 & \text{if } x \leq \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \text{if } \alpha \leq x \leq \beta \\ 1 & \text{if } x > \beta \end{cases}$$

if $X \sim U(\alpha, \beta)$ and $\alpha < a < b < \beta$ then

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx = \frac{b-a}{\beta-\alpha} \text{ (area)}$$





$$\frac{b-\alpha}{\beta-\alpha} \text{ (area)}$$

Expectation

$$E(X) = \frac{\alpha+\beta}{2} \qquad var(X) = \frac{(\beta-\alpha)^2}{12}$$

$$E(x) = \int_{-\alpha}^{\alpha} x f(x)\,dx$$

$$= \int_{\alpha}^{\beta} x f(x)\,dx$$

$$= \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha}\,dx$$

$$= \frac{1}{2(\beta-\alpha)}\left[x^2\right]_{\alpha}^{\beta}$$

$$= \frac{1}{2(\beta-\alpha)}(\beta^2-\alpha^2)$$

$$= \frac{\beta+\alpha}{2}$$

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

$$E(x^2) = \int_{\alpha}^{\beta} x^2 f(x)\,dx$$

$$= \frac{1}{3(\beta-\alpha)}\left[x^3\right]_{\alpha}^{\beta}$$

$$= \frac{\beta^3-\alpha^3}{3(\beta-\alpha)}$$

$$= \frac{(\beta-\alpha)(\beta^2+\alpha\beta+\alpha^2)}{3(\beta-\alpha)}$$

$$= \frac{\alpha^2+\alpha\beta+\beta^2}{3}$$

$$\frac{\alpha^2+\alpha\beta+\beta^2}{3} - \frac{(\alpha+\beta)^2}{4}$$

$$= \frac{4\alpha^2+4\alpha\beta+4\beta^2 - 3\alpha^2-3\beta^2-6\alpha\beta}{12}$$

$$= \frac{\alpha^2+\beta^2-2\alpha\beta}{12}$$

$$= \frac{(\alpha-\beta)^2}{12}$$

Made with Goodnotes

Exponential Distribution:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

$\lambda > 0$

Figure 4.5 shows the PDF of exponential distribution for several values of λ.

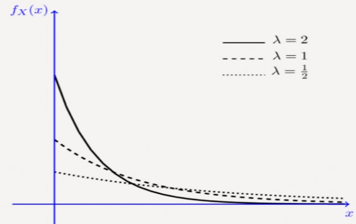


Fig.4.5 - PDF of the exponential random variable.

cdf

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

$$\int_0^x f(x)\,dx = \int_0^x \lambda e^{-\lambda x}\,dx = \lambda\left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^x = -\left[e^{-\lambda x} - 1\right] = 1 - e^{-\lambda x}$$

Proposition

The exponential is the only continuous non-negative memoryless distribution. That is, the only distribution with $X \geq 0$ *and*

$$\mathbf{P}\{X > t + s \,|\, X > t\} = \mathbf{P}\{X > s\} \qquad (\forall t,\, s \geq 0).$$

Suppose we have waited for time t. The chance of waiting an additional time s is the same as if we would start waiting anew. The distribution does not remember its past.

The memoryless property

Proof.

To prove that the Exponential distribution is memoryless,

$$\begin{aligned}\mathbf{P}\{X > t + s \,|\, X > t\} &= \frac{\mathbf{P}\{\{X > t + s\} \cap \{X > t\}\}}{\mathbf{P}\{X > t\}} \\ &= \frac{\mathbf{P}\{X > t + s\}}{\mathbf{P}\{X > t\}} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbf{P}\{X > s\}.\end{aligned}$$

To prove that the Exponential is the only one, ⇝ . □

Integration By parts

$$\int uv\,dx = u\int v\,dx - \int\left(\frac{du}{dx}\int v\,dx\right)dx$$

Expectation :

$$E(X) = \frac{1}{\lambda}$$

$$E(X) = \int_0^{\infty} x f(x)\,dx$$

$$= \int_0^{\infty} x \lambda e^{-\lambda x}\,dx$$

Consider $\lambda x = y$

$x = \frac{y}{\lambda}$

$dx = \frac{dy}{\lambda}$

$$= \int_0^{\infty} y e^{-y} \frac{dy}{\lambda} = \frac{1}{\lambda}\int_0^{\infty} y e^{-y}\,dy$$

$$= \frac{1}{\lambda}\left[y\int e^{-y}dy - \int -e^{-y}\,dy \right]_0^{\infty}$$

$$= \frac{1}{\lambda}\left[-ye^{-y} - e^{-y}\right]_0^{\infty}$$

$$= \frac{1}{\lambda}[1] = \frac{1}{\lambda}$$

$$var(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

variance

$$Var(X) = \frac{1}{\lambda^2}$$

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

$$E(X^2) = \int_0^{\infty} x^2 f(x)\,dx$$

$$= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx$$

$$= \int_0^{\infty} \frac{y}{\lambda} \times y \times e^{-y} \frac{dy}{\lambda}$$

$$= \frac{1}{\lambda^2}\int_0^{\infty} y^2 e^{-y}\,dy$$

$\lambda x = y$

$x = \frac{y}{\lambda}$

$$= \frac{1}{\lambda^2}\left[-y^2e^{-y} - \int 2y(-e^{-y})\,dy\right]_0^{\infty}$$

$$= \frac{1}{\lambda^2}\left[-y^2e^{-y} + 2\int y e^{-y}\,dy\right]_0^{\infty}$$

$$= \frac{1}{\lambda^2}\left[-ye^{-y}\right]_0^{\infty} + \frac{2}{\lambda^2}\boxed{\left[\int ye^{-y}dy\right]_0^{\infty}}$$

$\downarrow$

1

$$= \frac{2}{\lambda^2}$$

3. The memoryless property

Example

Suppose that the length of a phone call in a phone booth is exponentially distributed with mean 10 minutes. If we arrive to an already occupied booth (but there is no one else queuing), what is the probability that we'll wait between 10 and 20 minutes for the booth to free up?

Notice that by the memoryless property we don't care about how long the phone booth has been occupied for. (For all other distributions this would matter!). The remaining time of that person's phone call is $X \sim \text{Exp}(\lambda)$ with $\lambda = 1/\mathbf{E}X = 1/10$, and we calculate

$$\mathbf{P}\{10 < X \leq 20\} = F(20) - F(10)$$
$$= 1 - e^{-20/10} - (1 - e^{-10/10}) = e^{-1} - e^{-2} \simeq 0.233.$$

$E(X) = 10$ mins given

we know $E(X) = \frac{1}{\lambda}$ so, $\boxed{\lambda = \frac{1}{10}}$

$$P\{10 < X \leq 20\} = F(20) - F(10)$$
$$= \left(1 - e^{-\frac{1}{10} \times 20}\right) - \left(1 - e^{-\frac{1}{10} \times 10}\right)$$
$$= 1 - e^{-2} - 1 + e^{-1}$$
$$= e^{-1} - e^{-2} = \frac{1}{e} - \frac{1}{e^2}$$
$$= \frac{e - 1}{e^2}$$
$$= \frac{1.78}{(2.78)^2} \approx .233$$

$e = 2.78$

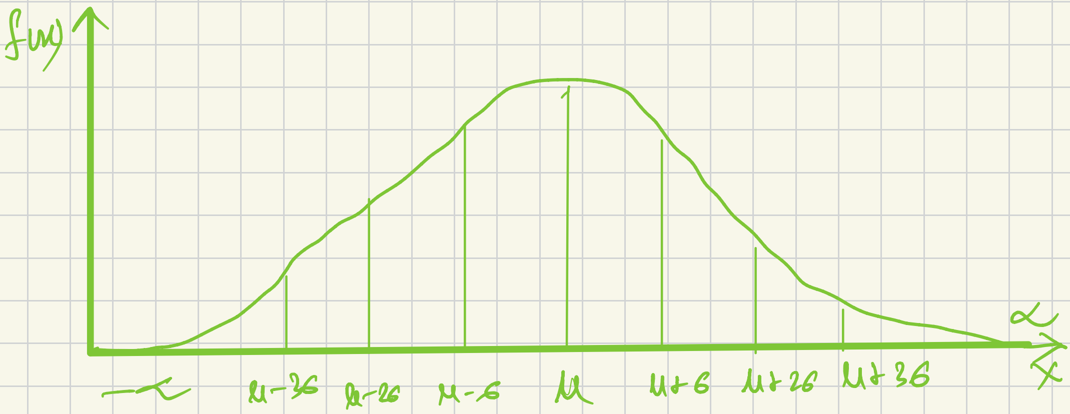
Normal Distribution:

Let $\mu \in \mathbb{R}$ & $\sigma > 0$ be real parameters.

$X \sim N(\mu, \sigma^2)$

density function is given by

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}}$$



when $\mu = 0$ & $\sigma^2 = 1$ then it is called standard normal distribution. It's density f^n is denoted by φ and its distribution function by Φ:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \quad (x \in \mathbb{R})$$

**

For any $z \in \mathbb{R}$,

$$\boxed{\phi(-z) = 1 - \phi(z)}$$

** The standard normal distribution is symmetric

if $X \sim N(0,1)$ & $-X \sim N(0,1)$

** If $X \sim N(\mu, \sigma^2)$, then its standardised version

$$\frac{X-\mu}{\sigma} \sim N(0,1).$$

Expectation

$$E(X) = np$$

$$\text{st. dev} = np(1-p)$$

Ex: Let X be normally distributed with mean 3 & variance 4. What is the chance that X is positive?

$$P(X > 0) = 1 - P(X \leq 0)$$

$$= 1 - P\left(\frac{X-3}{2} \leq \frac{0-3}{2}\right)$$

$$= 1 - P(Z \leq -1.5)$$

$$= 1 - \left(1 - P(Z \leq 1.5)\right)$$

$$= 1 - (1 - .9332)$$

$$= .9332$$

Made with Goodnotes

Theorem:

Fix p & let $X_n \sim \text{Binom}(n, p)$. Then for every fixed $a < b$ reals,

$$\lim_{n \to \infty} P\left\{ a < \frac{X_n - np}{\sqrt{np(1-p)}} \leq b \right\} = \phi(b) - \phi(a)$$

if n is very large & p is fixed,

Then $\dfrac{X_n - np}{\sqrt{np(1-p)}}$ is approximately $N(0,1)$ distributed.

Ex: The ideal size of a course is 150 students. On avg, 30% of those accepted will enroll, thus the organizations accept 450 students. What is the chance that more than 150 student will enroll?

$$np = 450 \times 30\% = 135$$

$$\text{St. dev} = \sqrt{np(1-p)} \approx 9.72$$

$$P(X > 150) = 1 - P(X \leq 150) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{150 - 135}{9.72}\right)$$

$$= 1 - P\left(Z \leq \frac{15}{9.72}\right) = 1 - P(Z \leq 1.54)$$

$$= 1 - .9382$$

$$= 0.0618$$

joint Distribution :

Y/X	0	1	2	3	$P_Y(\cdot)$
0	$\frac{^5C_3}{^{12}C_3}$	$\frac{^3C_1 \cdot {^5C_2}}{^{12}C_3}$	$\frac{^3C_2 \cdot {^5C_1}}{^{12}C_3}$	$\frac{1}{^{12}C_3}$	$\frac{^8C_3}{^{12}C_3}$
1	$\frac{^4C_1 \cdot {^5C_2}}{^{12}C_3}$	$\frac{^3C_1 \times {^4C_1} \times {^5C_1}}{^{12}C_3}$	$\frac{^3C_2 \times {^4C_1}}{^{12}C_3}$	0	
2	$\frac{^4C_2 \; {^5C_1}}{^{12}C_3}$	$\frac{^4C_2 \; {^3C_1}}{^{12}C_3}$	0	0	
3	$\frac{^4C_3}{^{12}C_3}$	0	0	0	
$P_X(\cdot)$					

red	white	Black
3	4	5

X = red

Y = white

5.1.1 Joint Probability Mass Function (PMF)

Remember that for a discrete random variable X, we define the PMF as $P_X(x) = P(X = x)$. Now, if we have two random variables X and Y, and we would like to study them jointly, we define the **joint probability mass function** as follows:

> The **joint probability mass function** of two discrete random variables X and Y is defined as
>
> $$P_{XY}(x, y) = P(X = x, Y = y).$$

Note that as usual, the comma means "and," so we can write

$$\begin{aligned} P_{XY}(x, y) &= P(X = x, Y = y) \\ &= P\big((X = x) \text{ and } (Y = y)\big). \end{aligned}$$

We can define the joint range for X and Y as

$$R_{XY} = \{(x, y) | P_{XY}(x, y) > 0\}.$$

In particular, if $R_X = \{x_1, x_2, \dots\}$ and $R_Y = \{y_1, y_2, \dots\}$, then we can always write

$$\begin{aligned} R_{XY} &\subset R_X \times R_Y \\ &= \{(x_i, y_j) | x_i \in R_X, y_j \in R_Y\}. \end{aligned}$$

In fact, sometimes we define $R_{XY} = R_X \times R_Y$ to simplify the analysis. In this case, for some pairs (x_i, y_j) in $R_X \times R_Y$, $P_{XY}(x_i, y_j)$ might be zero. For two discrete random variables X and Y, we have

> $$\sum_{(x_i, y_j) \in R_{XY}} P_{XY}(x_i, y_j) = 1$$

We can use the joint PMF to find $P\big((X, Y) \in A\big)$ for any set $A \subset \mathbb{R}^2$. Specifically, we have

> $$P\big((X, Y) \in A\big) = \sum_{(x_i, y_j) \in (A \cap R_{XY})} P_{XY}(x_i, y_j)$$

Note that the event $X = x$ can be written as $\{(x_i, y_j) : x_i = x, y_j \in R_Y\}$. Also, the event $Y = y$ can be written as $\{(x_i, y_j) : x_i \in R_X, y_j = y\}$. Thus, we can write

$$\begin{aligned} P_{XY}(x, y) &= P(X = x, Y = y) \\ &= P\big((X = x) \cap (Y = y)\big). \end{aligned}$$

Marginal PMFs

The joint PMF contains all the information regarding the distributions of X and Y. This means that, for example, we can obtain PMF of X from its joint PMF with Y. Indeed, we can write

$$\begin{aligned} P_X(x) &= P(X = x) \\ &= \sum_{y_j \in R_Y} P(X = x, Y = y_j) \qquad \text{law of total probablity} \\ &= \sum_{y_j \in R_Y} P_{XY}(x, y_j). \end{aligned}$$

Here, we call $P_X(x)$ the **marginal PMF** of X. Similarly, we can find the marginal PMF of Y as

$$P_Y(Y) = \sum_{x_i \in R_X} P_{XY}(x_i, y).$$

Marginal PMFs of X and Y:

$$P_X(x) = \sum_{y_j \in R_Y} P_{XY}(x, y_j), \qquad \text{for any } x \in R_X$$

$$P_Y(y) = \sum_{x_i \in R_X} P_{XY}(x_i, y), \qquad \text{for any } y \in R_Y \qquad (5.1)$$

Example 5.1
Consider two random variables X and Y with joint PMF given in Table 5.1.

Table 5.1 Joint PMF of X and Y in Example 5.1

	$Y=0$	$Y=1$	$Y=2$	
$X=0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{8+12+6}{48} = \frac{26}{48} = \frac{13}{24}$
$X=1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{6+8+8}{48} = \frac{22}{48} = \frac{11}{24}$
	$\frac{14}{48}$	$\frac{10}{24}$	$\frac{14}{48}$	1

Figure 5.1 shows $P_{XY}(x, y)$.

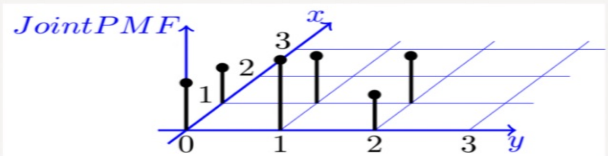


Figure 5.1: Joint PMF of X and Y (Example 5.1).

a. Find $P(X=0, Y \leq 1)$.
b. Find the marginal PMFs of X and Y.
c. Find $P(Y=1|X=0)$.
d. Are X and Y independent?

a) Find $P(X=0, Y\leq 1)$

$\Rightarrow P(X=0) * P(Y \leq 1)$

$\Rightarrow P(X=0) * \left[P(Y=0) + P(Y=1)\right]$

$\Rightarrow P(X=0) * P(Y=0) + P(X=0) * P(Y=1)$

$\Rightarrow \frac{1}{6} + \frac{1}{4} = \frac{4+6}{24} = \frac{10}{24} = \frac{5}{12}$ ✓

b) $R_X = \{0,1\}$ $R_Y = \{0,1,2\}$

$$P_X(x) = \begin{cases} \frac{13}{24} & x=0 \\ \frac{11}{24} & x=1 \\ 0 & \text{otherwise} \end{cases}$$

$$P_Y(y) = \begin{cases} \frac{14}{48} & y=0 \\ \frac{10}{24} & y=1 \\ \frac{14}{48} & y=2 \\ 0 & \text{otherwise} \end{cases}$$

Made with Goodnotes

c) $P(Y=1 \mid X=0)$

$$= \frac{P(Y=1) \cap P(X=0)}{P(X=0)} \rightarrow P(X=0,Y=0)+P(X=0,Y=1)+P(X=0,Y=2)$$

$$= \frac{6/24}{13/24} = \frac{6}{13}$$

d)

$P(X=0, Y=0) = \frac{1}{6}$

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

$P(X=0, Y=2)$

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

$P(X=1, Y=1)$

$P(X=1, Y=2)$

$P(X=0) \cdot P(Y=0)$

$= \frac{13}{24} \times \frac{14}{48}$

$P(X=0, Y=0) \neq P(X=0) \cdot P(Y=0)$

Independent R.V.

Random variables X, Y are independent, if

$P(X=x_i, Y=y_j) = P(X=x_i)\,P(Y=y_j)$

for all $x_i \in R_x$ & $y_j \in R_y$

So, X, Y are not independent.

#A

Example: The joint distribution for X + Y is given. Are X + Y independent?

$h(x,y)$

X \ Y	0	1	$f(x_i)$
0	0.1	0.3	0.4
1	0.5	0.1	0.6
$g(y_j)$	0.6	0.4	

need $h(0,0) = f(0)g(0)$

$h(0,1) = f(0)g(1)$

$h(1,0) = f(1)g(0)$

$h(1,1) = f(1)g(1)$

check $h(0,0) = 0.1$

$f(0)g(0) = 0.4 \cdot 0.6 = 0.24$

$h(0,0) \neq f(0)g(0)$ ∴ not independent.

Properties of Expectations

① Suppose that $a \le X \le b$ then

$$a \le E(X) \le b$$

Proof:

$$a \cdot 1 = \sum_i a \cdot p(x_i) \le \sum_i x_i p(x_i) \le \sum_i b\, p(x_i) \le b \cdot 1$$

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

Proof:
$$E(X \pm Y) = \sum_{i,j} (x_i \pm y_j) P(x_i, y_j)$$

$$= \sum_i \sum_j x_i P(x_i, y_j) \pm \sum_i \sum_j y_j P(x_i, y_j)$$ | we can go for marginal probability for X & Y.

$$= \sum_i x_i P_X(x_i) \pm \sum_j y_j P_Y(y_j) = E(X) \pm E(Y)$$ ✓

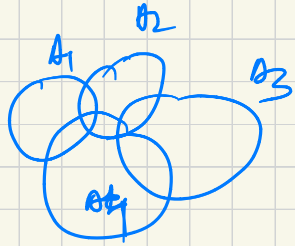
③ Let X & Y be r.v such that $X \le Y$

Then $E(X) \le E(Y)$

④ Let $X_1, X_2 - - X_n$ be identically distributed r.v with mean μ.
their sample mean is

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

□ Boole's Inequality:



$y = 1$ if atleast one A_i occurs.

$y = 0$ otherwise

another way to write is

$y = 1$ if $x \geq 1$ → no. of events occurs.

$y = 0$ if $x = 0$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

This is called Union Bound, or Boole's inequality.

Independence:

Let x & y be independent random variables, & g, h functions.

$$\boxed{E\big(g(x)\cdot h(y)\big) = Eg(x)\cdot Eh(y)}$$

** only true for independent random variable.

Covariance:

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

If x & y are independent then $\mathrm{cov}(x,y) = 0$ ✓

**

$\mathrm{cov}(x,y)=0$ doesn't mean x & y are independent.

Variance:

Let $x_1, x_2 \dots x_n$ be random variables. Then

$$\mathrm{var}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mathrm{Var}(x_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(x_i, x_j)$$

when x, y are independent then

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

☐ Cauchy-Schwarz inequality:

For every X & Y,

$$|E(XY)| \leq \sqrt{E(X^2)} \cdot \sqrt{E(Y^2)}$$

☐ Correlation coefficient:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \qquad -1 \leq \rho \leq 1$$

5. Correlation

Example

Rolling two dice, let X be the number shown on the first die, Y the one shown on the second die, $Z = X + Y$ the sum of the two numbers. Clearly, X and Y are independent, $\mathbf{Cov}(X, Y) = 0$, $\varrho(X, Y) = 0$. For X and Z,

$$\begin{aligned}
\mathbf{Cov}(X, Z) &= \mathbf{Cov}(X, X + Y) = \mathbf{Cov}(X, X) + \mathbf{Cov}(X, Y) \\
&= \mathbf{Var}X + 0 = \mathbf{Var}X; \\
\mathbf{Var}(Z) &= \mathbf{Var}(X + Y) = \mathbf{Var}X + \mathbf{Var}Y \qquad \text{indep.!} \\
&= \mathbf{Var}X + \mathbf{Var}X = 2\mathbf{Var}X; \\
\varrho(X, Z) &= \frac{\mathbf{Cov}(X, Z)}{\mathrm{SD}\,X \cdot \mathrm{SD}\,Z} = \frac{\mathbf{Var}X}{\sqrt{\mathbf{Var}X} \cdot \sqrt{2\mathbf{Var}X}} = \frac{1}{\sqrt{2}}.
\end{aligned}$$