Chap 3 : Two Random Variables

Chap 3.1: Distribution Functions of Two RVs

In many experiments, the observations are expressible not as a single quantity, but as a family of quantities. For example to record the height and weight of each person in a community or the number of people and the total income in a family, we need two numbers.

Let X and Y denote two random variables based on a probability model . Then

$$P(x_1 < X(\xi) \le x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx$$

and

$$P(y_1 < Y(\xi) \le y_2) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)\,dy$$

What about the probability that the pair of RVs (X, Y) belongs to an arbitrary region D? In other words, how does one estimate, for example

$$P[(x_1 < X(\xi) \le x_2) \cap (y_1 < Y(\xi) \le y_2)] = ?$$

Towards this, we define the joint probability distribution function of X and Y to be

$$F_{XY}(x, y) \quad = \quad P(X \le x, Y \le y) \ge 0 \tag{1}$$

where x and y are arbitrary real numbers.

Properties

1.

$$F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \quad F_{XY}(+\infty, +\infty) = 1 \tag{2}$$

since $(X(\xi) \leq -\infty, Y(\xi) \leq y) \subset (X(\xi) \leq -\infty)$, we get

$$F_{XY}(-\infty, y) \leq P(X(\xi) \leq -\infty) = 0$$

Similarly, $(X(\xi) \leq +\infty, Y(\xi) \leq +\infty) = \Omega$, we get $F_{XY}(+\infty, +\infty) = P(\Omega) = 1$.

2.

$$P(x_1 < X(\xi) \leq x_2, Y(\xi) \leq y) = F_{XY}(x_2, y) - F_{XY}(x_1, y) \tag{3}$$

$$P(X(\xi) \leq x, y_1 < Y(\xi) \leq y_2) = F_{XY}(x, y_2) - F_{XY}(x, y_1) \tag{4}$$

To prove (3), we note that for $x_2 > x_1$

$$(X(\xi) \leq x_2, Y(\xi) \leq y) = (X(\xi) \leq x_1, Y(\xi) \leq y) \cup (x_1 < X(\xi) \leq x_2, Y(\xi) \leq y)$$

and the mutually exclusive property of the events on the right side gives

$$P(X(\xi) \leq x_2, Y(\xi) \leq y) = P(X(\xi) \leq x_1, Y(\xi) \leq y) + P(x_1 < X(\xi) \leq x_2, Y(\xi) \leq y)$$

which proves (3). Similarly (4) follows.

3.

$$\begin{aligned} P(x_1 < X(\xi) \leq x_2, y_1 < Y(\xi) \leq y_2) &= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) \\ &\quad - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1) \end{aligned} \tag{5}$$

This is the probability that (X, Y) belongs to the rectangle in Fig. 3. To prove (5), we can make use of the following identity involving mutually exclusive events on the right side.

$$\begin{aligned} (x_1 < X(\xi) \leq x_2, Y(\xi) \leq y_2) &= (x_1 < X(\xi) \leq x_2, Y(\xi) \leq y_1) \\ &\quad \cup (x_1 < X(\xi) \leq x_2, y_1 < Y(\xi) \leq y_2) \end{aligned}$$

This gives

$$\begin{aligned} P(x_1 < X(\xi) \leq x_2, Y(\xi) \leq y_2) &= P(x_1 < X(\xi) \leq x_2, Y(\xi) \leq y_1) \\ &\quad + P(x_1 < X(\xi) \leq x_2, y_1 < Y(\xi) \leq y_2) \end{aligned}$$

and the desired result in (5) follows by making use of (3) with $y = y_2$ and y_1 respectively.

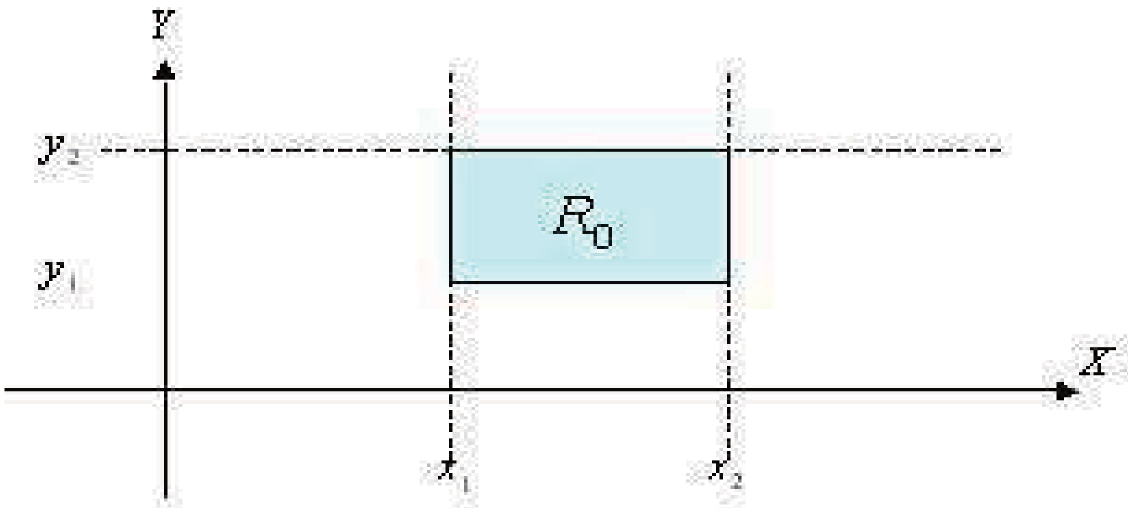

Figure 1: Two dimensional RV.

Joint Probability Density Function (Joint pdf)

By definition, the joint pdf of X and Y is given by

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} \tag{6}$$

and hence we obtain the useful formula

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\, du\, dv \tag{7}$$

Using (2), we also get

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) \text{ dxdy} = 1 \tag{8}$$

$$P((X,Y) \in D) = \int \int_{(x,y)\in D} f_{XY}(x,y)\, dx\, dy \tag{9}$$

Marginal Statistics

In the context of several RVs, the statistics of each individual ones are called marginal statistics. Thus $F_X(x)$ is the marginal probability distribution function of X, and $f_X(x)$ is the marginal pdf of X. It is interesting to note that all marginal can be obtained from the joint pdf. In fact

$$F_X(x) = F_{XY}(x, +\infty) \qquad F_Y(y) = F_{XY}(+\infty, y) \tag{10}$$

Also

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx. \tag{11}$$

To prove (10), we can make use of the identity

$$(X \leq x) = (X \leq x) \cap (Y \leq +\infty)$$

so that

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \leq \infty) = F_{XY}(x, +\infty) \tag{12}$$

To prove (11), we can make use of (7) and (10), which gives

$$F_X(x) = F_{XY}(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{XY}(u, y) dy\, du$$

and taking derivative with respect to x , we get

$$\boxed{f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy} \tag{13}$$

If X and Y are discrete RVs, then $p_{ij} = P(X = x_i, Y = y_j)$ represents their joint pmf, and their respective marginal pmfs are given by

$$\boxed{P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}} \tag{14}$$

and

$$\boxed{P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij}} \tag{15}$$

Assuming that $P(X = x_i, Y = y_j)$ is written out in the form of a rectangular array, to obtain $P(X = x_i)$ from (14), one needs to add up all the entries in the i-th row.

$$\begin{array}{c|cccccc}
 & & & & \sum_{j} p_{ij} & & \\
\hline
 & p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\
 & p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i} p_{ij} & p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mn}
\end{array}$$

Figure 2: Illustration of marginal pmf.

Examples

From (11) and (12), the joint CDF and/or the joint pdf represent complete information about the RVs, and their marginal pdfs can be evaluated from the joint pdf. However, given marginals, (most often) it will not be possible to compute the joint pdf.

Example 1: Given

$$f_{XY}(x,y) = \begin{cases} \text{c} & 0 < x < y < 1 \\ 0 & \text{o.w.} \end{cases} \tag{16}$$

Obtain the marginal pdfs $f_X(x)$ and $f_Y(y)$.

Solution: It is given that the joint pdf $f_{XY}(x,y)$ is a constant in the shaded region in Fig. 3. We can use (8) to determine that constant c. From (8)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dx\,dy = \int_{y=0}^{1}\left(\int_{x=0}^{y} c\cdot dx\right)dy = \int_{y=0}^{1} cy\,dy = \frac{cy^2}{2}\Big|_0^1 = \frac{c}{2} = 1$$

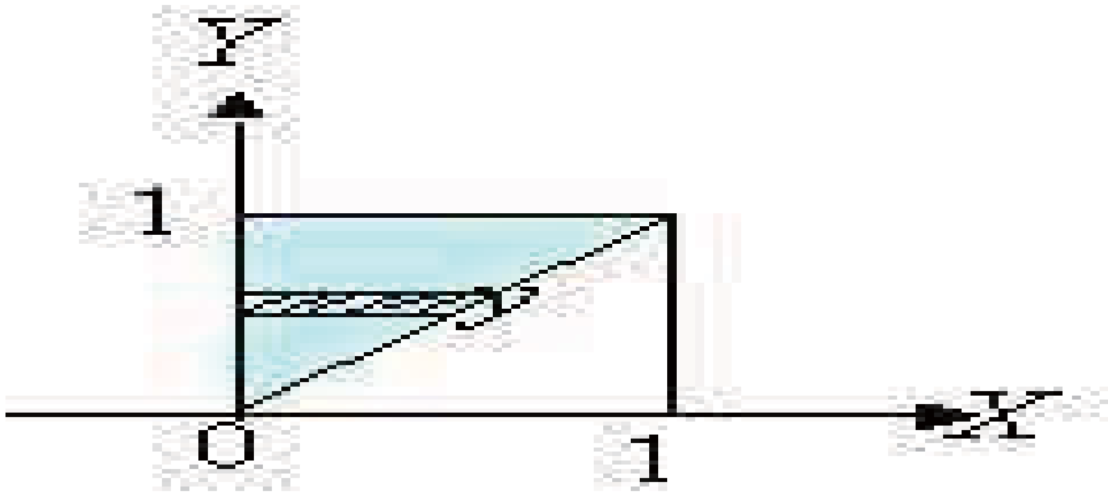

Figure 3: Diagram for the example.

Thus $c = 2$. Moreover

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_{y=x}^{1} 2dy = 2(1-x) \quad 0 < x < 1$$

and similarly,

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_{x=0}^{y} 2\,dx = 2y \quad 0 < y < 1$$

Clearly, in this case given $f_X(x)$ and $f_Y(y)$ as above, it will not be possible to obtain the original joint pdf in (16).

Example 2: X and Y are said to be jointly normal (Gaussian) distributed, if their joint pdf has the following form:

$$\begin{aligned} f_{XY}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\left[\frac{(x-\mu_X)^2}{\sigma_X^2}-\frac{2\rho(x-\mu_x)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \\ &-\infty<x<\infty, -\infty<y<\infty, |\rho|<1 \end{aligned} \tag{17}$$

By direct integration, it can be shown that

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] \sim N(\mu_X, \sigma_X^2)$$

and similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] \sim N(\mu_Y, \sigma_Y^2)$$

Following the above notation, we will denote (17) as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Once again, knowing the marginals in above alone doesn't tell us everything about the joint pdf in (17). As we show below, the only situation where the marginal pdfs can be used to recover the joint pdf is when the random variables are statistically independent.

Independence of RVs

Definition: The random variables X and Y are said to be statistically independent if

$$P[(X(\xi) \leq x) \cap (Y(\xi) \leq y)] = P(X(\xi) \leq x) \cdot P(Y(\xi) \leq y)$$

- For continuous RVs,

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y) \tag{18}$$

  or equivalently, if X and Y are independent, then we must have

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \tag{19}$$

- If X and Y are discrete-type RVs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \quad \text{for all } i, j \tag{20}$$

Equations (18)-(20) give us the procedure to test for independence. Given $f_{XY}(x, y)$, obtain the marginal pdfs $f_X(x)$ and $f_Y(y)$ and examine whether one of the equations in (18) or (20) is valid. If so, the RVs are independent, otherwise they are dependent.

- Returning back to Example 1, we observe by direct verification that

$f_{XY}(x, y) \neq f_X(x) \cdot f_Y(y)$. Hence X and Y are dependent RVs in that case.

- It is easy to see that such is the case in the case of Example 2 also, unless in other words, two jointly Gaussian RVs as in (17) are independent if and only if the fifth parameter $\rho = 0$.

Expectation of Functions of RVs

If X and Y are random variables and $g(\cdot)$ is a function of two variables, then

$$\begin{aligned} E[g(X,Y)] &= \sum_y \sum_x g(x,y)\, p(x,y) \quad \text{discrete case} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy \quad \text{continuous case} \end{aligned}$$

If $g(X,Y) = aX + bY$, then we can obtain

$$E[aX + bY] = aE[X] + bE[Y]$$

Example 3 At a party N men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who select their own hats.

Solution: let X denote the number of men that select their own hats, we can compute $E[X]$ by noting that

$$X = X_1 + X_2 + \cdots + X_N$$

where X_i is the indicator RV, given as

$$P[X_i = 1] = P\{i\text{th man selects his own hat}\} = 1/N$$

So

$$E[X_i] = 1P[X_i = 1] + 0P[X_i = 0] = 1/N$$

Therefore, $E[X] = E[X_1] + E[X_2] + \cdots + E[X_N] = 1$. No matter how many people are at the party, on the average, exactly one of the men will select his own hat.

If X and Y are independent, then for any functions $h(\cdot)$ and $g(\cdot)$

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)] \tag{21}$$

And

$$Var(X+Y) = Var(X) + Var(Y) \tag{22}$$

Example 4: Random variables X_1 and X_2 are independent and identically distributed with probability density function

$$f_X(x) = \begin{cases} 1 - x/2 & 0 \le x \le 2 \\ 0 & \text{o.w.} \end{cases}$$

Find

- The joint pdf $f_{X_1,X_2}(x_1, x_2)$
- The cdf of $Z = \max(X_1, X_2)$.

Solution: (a) since X_1 and X_2 are independent,

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\cdot f_{X_2}(x_2) = \begin{cases} \left(1-\frac{x_1}{2}\right)\left(1-\frac{x_2}{2}\right) & 0 \le x_1 \le 2, 0 \le x_2 \le 2 \\ 0 & o.w. \end{cases}$$

(b) Let $F_X(x)$ denote the CDF of both X_1 and X_2. The CDF of $Z = \max(X_1, X_2)$ is found by observing that $Z \le z$ iff $X_1 \le z$ and $X_2 \le z$. That is

$$P(Z \le z) = P(X_1 \le z, X_2 \le z) = P(X_1 \le z)P(X_2 \le z) = [F_X(z)]^2$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x \left(1 - \frac{t}{2}\right)\,dt = x - x^2/4 \quad 0 \leqslant x \leqslant 2$$

Thus, for $0 \le z \le 2$,

$$F_Z(z) = \left(z - \frac{z^2}{4}\right)^2$$

The complete CDF of Z is

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \left(z - \frac{z^2}{4}\right)^2 & 0 \leq z \leq 2 \\ 1 & o.w. \end{cases}$$

Example 5: Given

$$f_{XY}(x,y) = \begin{cases} xy^2e^{-y} & 0 < y < \infty, 0 < x < 1 \\ 0 & \text{o.w.} \end{cases}$$

Determine whether X and Y are independent.

Solution:

$$\begin{aligned} f_X(x) &= \int_0^{+\infty} f_{XY}(x,y)dy = x\int_0^{+\infty} y^2e^{-y}dy = x\int_0^{+\infty} -y^2\, de^{-y} \\ &= x\left(-y^2e^{-y}\Big|_0^{\infty} + 2\int_0^{+\infty} ye^{-y}dy\right) = 2x, \quad 0 < x < 1 \end{aligned}$$

Similarly

$$f_Y(y) = \int_0^1 f_{XY}(x,y)dx = \frac{y^2}{2}e^{-y}, \quad 0 < y < \infty$$

In this case

$$f_{XY}(x,y) = f_X(x)\cdot f_Y(y)$$

and hence X and Y are independent random variables.

Chap 3.2 : Correlation, Covariance, Moments and CF

$\boxed{\text{Correlation}}$: Given any two RVs X and Y, define

$$\begin{aligned} E[X^m Y^n] & \qquad m,n\text{th joint moment} \\ E[XY] = Corr(X,Y) = R_{XY} & \qquad \text{correlation of } X \text{ and } Y \\ E[(X-\mu_X)^m (Y-\mu_Y)^n] & \qquad m,n\text{th central joint moment} \\ E[(X-\mu_X)(Y-\mu_Y)] = Cov(X,Y) = K_{XY} & \qquad \text{covariance of } X \text{ and } Y \end{aligned}$$

$\boxed{\text{Covariance}}$: Given any two RVs X and Y, define

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{23}$$

By expanding and simplifying the right side of (23), we also get

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y) \tag{24}$$

Correlation coefficient between X and Y.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \qquad -1 \leq \rho_{XY} \leq 1 \tag{25}$$

$$Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y \tag{26}$$

Uncorrelated RVs: If $\rho_{XY} = 0$, then X and Y are said to be uncorrelated RVs. If X and Y are uncorrelated, then

$$E(XY) = E(X)\,E(Y) \tag{27}$$

Orthogonality: X and Y are said to be orthogonal if

$$E(XY) = 0 \tag{28}$$

From above, if either X or Y has zero mean, then orthogonality implies uncorrelatedness and vice-versa.

Suppose X and Y are independent RVs,

$$E(XY) = E(X)\,E(Y) \tag{29}$$

therefore from (27), we conclude that the random variables are uncorrelated. **Thus independence implies uncorrelatedness** ($\rho_{XY} = 0$)**. But the inverse is generally not true.**

Example 6: Let $Z = aX + bY$. Determine the variance of Z in terms of σ_X, σ_Y and ρ_{XY}.

Solution:

$$\mu_Z = E(Z) = E(aX + bY) = a\mu_X + b\mu_Y$$

and

$$\begin{aligned}
\sigma_Z^2 &= Var(Z) = E[(Z - \mu_Z)^2] = E\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\} \\
&= a^2 E[(X - \mu_X)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)] + b^2 E[(Y - \mu_Y)^2] \\
&= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2
\end{aligned}$$

In particular if X and Y are independent, then $\rho_{XY} = 0$, and the above equation reduces to

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Thus the variance of the sum of independent RVs is the sum of their variances ($a = b = 1$).

$\boxed{\text{Moments}}$:

$$E[X^kY^m] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^m f_{XY}(x,y)dx\,dy \tag{30}$$

represents the joint moment of order (k,m) for X and Y.

$\boxed{\text{Joint Characteristic Function}}$: following the one random variable case, we can define the joint characteristic function between two random variables which will turn out to be useful for moment calculations. The joint characteristic function between X and Y is defined as

$$\Phi_{XY}(\omega_1,\omega_2) = E\left(e^{j(X\omega_1+Y\omega_2)}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j(\omega_1 x+\omega_2 y)} f_{XY}(x,y)dx\,dy \tag{31}$$

From this and the two-dimensional inversion formula for Fourier transforms, it follows that

$$f_{XY}(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_{XY}(\omega_1,\omega_2)e^{-j(\omega_1 x+\omega_2 y)}d\omega_1\,d\omega_2 \tag{32}$$

Note that

$$|\Phi_{XY}(\omega_1,\omega_2)| \leq \Phi_{XY}(0,0) = 1$$

If X and Y are independent RVs, then from (31), we obtain

$$\Phi_{XY}(\omega_1, \omega_2) = E\left(e^{j(\omega_1 X)}\right) E\left(e^{j(\omega_2 Y)}\right) = \Phi_X(\omega_1)\, \Phi_Y(\omega_2) \tag{33}$$

Also

$$\Phi_X(\omega) = \Phi_{XY}(\omega, 0) \qquad \Phi_Y(\omega) = \Phi_{XY}(0, \omega) \tag{34}$$

Independence If the RV X and Y are independent, then

$$E\left[e^{j(\omega_1 X + \omega_2 Y)}\right] = E[e^{j\omega_1 X}] \cdot E[e^{j\omega_2 Y}]$$

From this it follows that

$$\Phi(\omega_1, \omega_2) = \Phi_X(\omega_1) \cdot \Phi_Y(\omega_2)$$

conversely, if above equation is true, then the random variables X and Y are independent.

Product Characteristic functions are useful in determining the pdf of linear combinations of RVs. If the RV X and Y are independent and $Z = X + Y$, then

$$E\left[e^{j\omega Z}\right] = E\left[e^{j\omega(X+Y)}\right] = E\left[e^{j\omega X}\right] \cdot E\left[e^{j\omega Y}\right]$$

Hence,

$$\Phi_Z(\omega) = \Phi_X(\omega) \cdot \Phi_Y(\omega)$$

From above, the characteristic function of RV Z is equal to the product between the characteristic function of X and the characteristic function of Y.

Example 7: X and Y are independent Poisson RVs with parameters λ_1 and λ_2 respectively, let

$$Z = X + Y$$

Then

$$\Phi_Z(\omega) = \Phi_X(\omega)\,\Phi_Y(\omega)$$

From earlier results

$$\Phi_X(\omega) = e^{\lambda_1(e^{j\omega}-1)} \quad \Phi_Y(\omega) = e^{\lambda_2(e^{j\omega}-1)}$$

so that

$$\Phi_Z(\omega) = e^{(\lambda_1+\lambda_2)(e^{j\omega}-1)} \quad \sim \quad P(\lambda_1 + \lambda_2)$$

i.e., sum of independent Poisson RVs is also a Poisson random variable.

Chap 3.3 : Gaussian RVs and Central Limit Theorem

From (17), X and Y are said to be jointly Gaussian if their joint pdf has the form in (17):

$$\begin{aligned} f_{XY}(x,y) \quad &= \quad \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\left[\frac{(x-\mu_X)^2}{\sigma_X^2}-\frac{2\rho(x-\mu_x)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \\ &-\infty < x < \infty, -\infty < y < \infty, |\rho| < 1 \end{aligned}$$

By direct substitution and simplification, we obtain the joint characteristic function of two jointly Gaussian RVs to be

$$\Phi_{XY}(\omega_1,\omega_2) = E\left(e^{j(\omega_1 X+\omega_2 Y)}\right) = e^{j(\mu_X\omega_1+\mu_Y\omega_2)-\frac{1}{2}(\sigma_X^2\omega_1^2+2\rho\sigma_X\sigma_Y\omega_1\omega_2+\sigma_Y^2\omega_2^2)} \tag{35}$$

Letting $\omega_2 = 0$ in (35), we get

$$\Phi_X(\omega_1) = \Phi_{XY}(\omega_1,0) = e^{j\mu_X\omega_1-\frac{1}{2}\sigma_X^2\omega_1^2} \tag{36}$$

From (17) by direct computation, it is easy to show that for two jointly Gaussian random variables

$$Cov(X,Y) = \rho\sigma_X\sigma_Y$$

Hence from definition of ρ, ρ in $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ represents the actual correlation coefficient of the two jointly Gaussian RVs in (17). Notice that $\rho = 0$ implies

$$f_{XY}(X, Y) = f_X(x)\, f_Y(y)$$

Thus if X and Y are jointly Gaussian, uncorrelatedness does imply independence between the two random variables. Gaussian case is an exception where the two concepts imply each other.

Example 8: Let X and Y be jointly Gaussian RVs with parameters $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Define $Z = aX + bY$, determine $f_Z(z)$.

Solution: In this case we can make use of characteristic function to solve this problem

$$\Phi_Z(\omega) = E\left(e^{jZ\omega}\right) = E\left(e^{j(aX+bY)\omega}\right) = E\left(e^{jXa\omega + jYb\omega}\right) = \Phi_{XY}(a\omega, b\omega) \tag{37}$$

From (35) with ω_1 and ω_2 replaced by $a\omega$ and $b\omega$ respectively we get

$$\Phi_Z(\omega) = e^{j(a\mu_X + b\mu_Y)\omega - \frac{1}{2}(a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2)^2\omega^2} = e^{j\mu_Z\omega - \frac{1}{2}\sigma_Z^2\omega^2} \tag{38}$$

where

$$\mu_Z = a\mu_X + b\mu_Y \qquad \sigma_Z^2 = a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2$$

Notice that (38) has the same form as (36), and hence we conclude that $Z = aX + bY$ is also Gaussian with mean and variance as above, which also agrees with previous example.

From the previous example, we conclude that any linear combination of jointly Gaussian RVs generates a new Gaussian RV. In other words, linearity preserves Gaussianity.

Gaussian random variables are also interesting because of the following result.

Central Limit Theorem

Suppose $X_1, X_2, \cdots, X_n$ are a sequence of independent, identically distributed (i.i.d) random variables, each with mean μ and variance σ^2. Then the distribution of

$$Y = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \tag{39}$$

Tends to be standard normal as $n \to \infty$

$$Y \to N(0, 1) \tag{40}$$

The central limit theorem states that a large sum of independent random variables each with finite variance tends to behave like a normal random variable. Thus the individual pdfs become unimportant to analyze the collective sum behavior. If we model the noise phenomenon as the sum of a large number of independent random variables (eg: electron motion in resistor components), then this theorem allows us to conclude that noise behaves like a Gaussian RV.

This theorem holds for any distribution of the X_i's; herein lies its power.

Review Gaussian approximation to sum of binomial RVs

The Normal Approximation: Suppose $n \to \infty$ with p held fixed. Then for k in the neighborhood of np, we can approximate

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}}\, e^{-(k-np)^2/2npq}$$

Now, thinking that $Y = X_1 + X_2 + \cdots + X_n$, each X_i is a Bernoulli RV with parameter p. Then Y follows binomial distribution. From central limit theorem,

$$\frac{Y - E[Y]}{\sqrt{Var(Y)}} = \frac{Y - np}{\sqrt{np(1-p)}}$$

approaches the standard normal distribution as n approaches ∞. The normal approximation will be generally quite good for values of n satisfying $np(1-p) \geq 10$.

$\boxed{\text{Example 9}}$: The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, at which point it is replaced by a new one. Assuming a stockpile of 25 such batteries, the lifetimes of which are independent, approximate the probability that over 1100 hours of use can be obtained.

Solution: if we let X_i denote the lifetime of the ith battery to be put in use, and $Y = X_1 + X_2 + \cdots + X_{25}$. Then we want to find $P(Y > 1100)$.

Chap 3.4 : Conditional Probability Density Functions

For any two events A and B, we have defined the conditional probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0 \tag{41}$$

Noting that the probability distribution function $F_X(x)$ is given by $F_X(x) = P\{X(\xi) \leq x\}$, we may define the conditional distribution of the RV X given the event B as

$$F_X(x|B) = P\{X(\xi) \leq x|B\} = \frac{P\{(X(\xi) \leq x) \cap B\}}{P(B)} \tag{42}$$

In general, event B describes some property of X. Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$F_X(+\infty|B) = \frac{P\{(X(\xi) \leq +\infty) \cap B\}}{P(B)} = \frac{P(B)}{P(B)} = 1 \tag{43}$$

$$F_X(-\infty|B) = \frac{P\{(X(\xi) \leq -\infty) \cap B\}}{P(B)} = \frac{P(\phi)}{P(B)} = 0$$

Furthermore, since for $x_2 \geq x_1$

$$(X(\xi) \leq x_2) = (X(\xi) \leq x_1) \cup (x_1 < X(\xi) \leq x_2)$$

$$P(x_1 < X(\xi) \leq x_2|B) = \frac{P\{(x_1 < X(\xi) \leq x_2) \cap B\}}{P(B)} = F_X(x_2|B) - F_X(x_1|B). \quad (44)$$

The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x|B) = \frac{d}{dx} F_X(x|B)$$

we obtain

$$\boxed{F_X(x|B) = \int_{-\infty}^{x} f_X(u|B)\, du} \quad (45)$$

Using above equation, we can also have

$$\boxed{P(x_1 < X(\xi) \leq x_2|B) = \int_{x_1}^{x_2} f_X(x|B)\, dx} \quad (46)$$

$\boxed{\text{Example 10}}$: Toss a coin and $X(T) = 0, X(H) = 1$. Suppose $B = \{H\}$. Determine $F_X(x|B)$. (Suppose q is the probability of landing a tail)

Solution: From earlier example, $F_X(x)$ has the following form shown in Fig. 4(a). We need $F_X(x|B)$ for all x.

- For $x < 0$, $\{X(\xi) \leq x\} = \phi$, so that $\{(X(\xi) \leq x) \cap B\} = \phi$ and $F_X(x|B) = 0$.
- For $0 \leq x < 1$, $\{X(\xi) \leq x\} = \{T\}$, so that

  $$\{(X(\xi) \leq x) \cap B\} = \{T\} \cap \{H\} = \phi$$

  and $F_X(x|B) = 0$.
- For $x \geq 1$, $\{X(\xi) \leq x\} = \Omega$, and

  $$\{(X(\xi) \leq x) \cap B\} = \Omega \cap \{B\} = \{B\}$$

  and

  $$F_X(x|B) = \frac{P(B)}{P(B)} = 1$$

The conditional CDF is shown in Fig. 4(b).

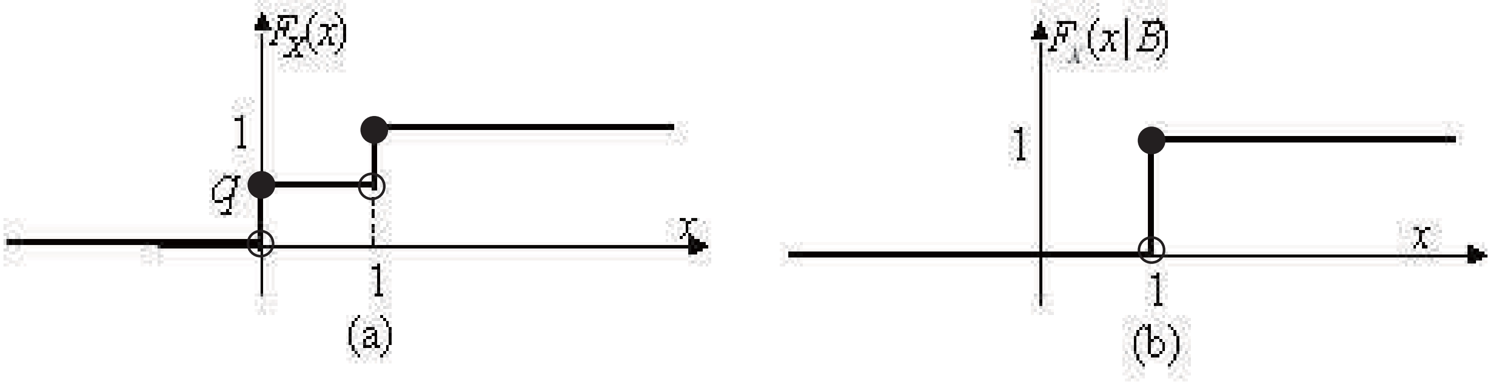

Figure 4: Condition CDF for Example 10.

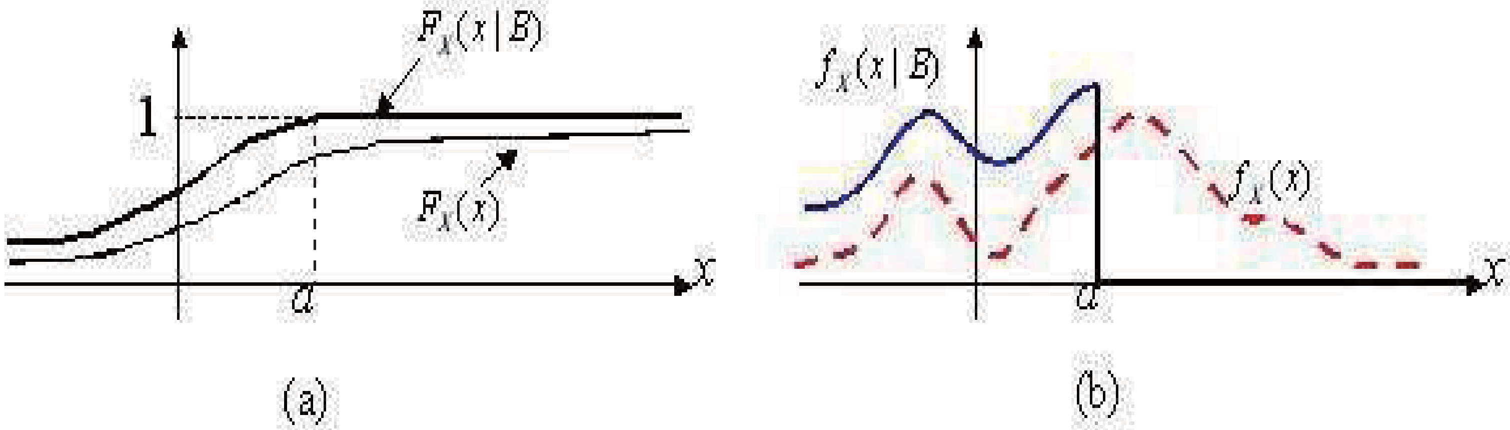

Figure 5: Condition CDF and pdf for Example 11.

Example 11 : Given $F_X(x)$, suppose $B = \{X(\xi) \leq a\}$. Find $f_X(x|B)$.

Solution: We will first determine $F_X(x|B)$ as,

$$F_X(x|B) = \frac{P\{(X \leq x) \cap (X \leq a)\}}{P(X \leq a)}$$

- For $x < a$,

$$F_X(x|B) = \frac{P(X \leq x)}{P(X \leq a)} = \frac{F_X(x)}{F_X(a)}$$

- For $x \geq a$, $(X \leq x) \cap (X \leq a) = (X \leq a)$, so that $F_X(x|B) = 1$.

Thus, the conditional CDF and pdf are given as below (shown in Fig. 5)

$$F_X(x|B) = \begin{cases} \frac{F_X(x)}{F_X(a)} & x < a \\ 1 & x \geq a \end{cases}$$

and hence

$$f_X(x|B) = \frac{d}{dx} F_X(x|B) = \begin{cases} \frac{f_X(x)}{F_X(a)} & x < a \\ 0 & \text{o.w.} \end{cases}$$

Example 12: Let B represent the event $\{a < X(\xi) \leq b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x|B)$ and $f_X(x|B)$.

Solution:

$$
\begin{aligned}
F_X(x|B) &= P\{X(\xi) \leq x|B\} = \frac{P\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\}}{P(a < X(\xi) \leq b)} \\
&= \frac{P\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\}}{F_X(b) - F_X(a)}
\end{aligned}
$$

- For $x < a$, we have $\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\} = \phi$ and hence $F_X(x|B) = 0$.
- For $a \leq x < b$, we have $\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\} = \{a < X(\xi) \leq x\}$ and hence
$$F_X(x|B) = \frac{P(a < X(\xi) \leq x)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}$$
- For $x \geq b$, we have $\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\} = \{a < X(\xi) \leq b\}$ so that $F_X(x|B) = 1$

Therefore, the density function shown below and given as

$$f_X(x|B) = \begin{cases} \frac{f_X(x)}{F_X(b)-F_X(a)} & a < x \leq b \\ 0 & \text{o.w.} \end{cases}$$

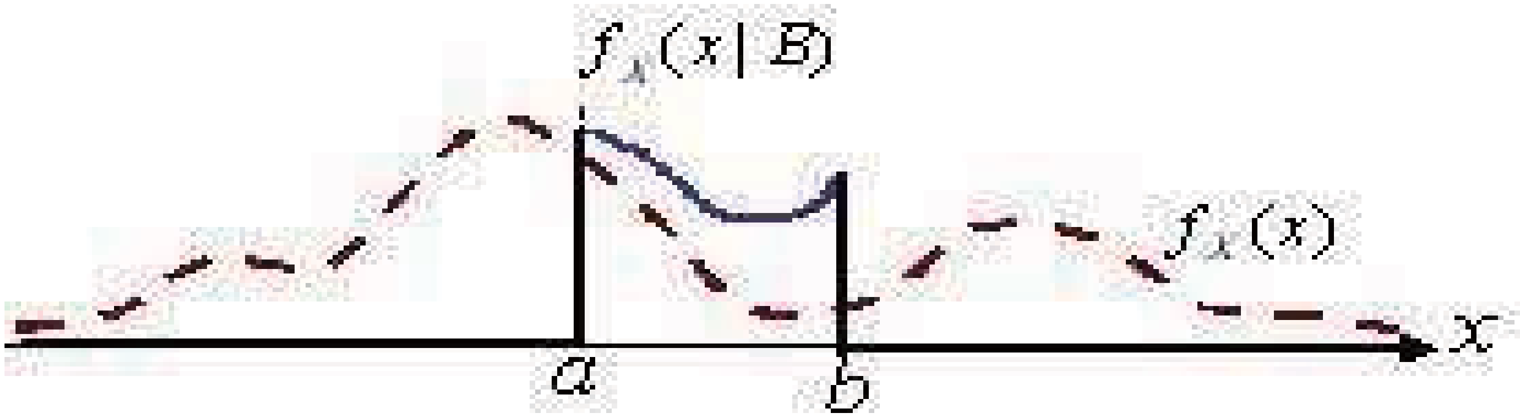

Figure 6: Condition pdf for Example 12.

B is related to another RV

In Summary, conditional (on events) CDF is defined as

$$F_X(x|B) = P(X(\xi) \leq x|B) = \frac{P[(X(\xi) \leq x) \cap B]}{P(B)}$$

Suppose, we let $B = \{y_1 < Y(\xi) \leq y_2\}$. We can get

$$\begin{aligned} F_X(x|y_1 < Y \leq y_2) &= \frac{P(X(\xi) \leq x, y_1 < Y(\xi) \leq y_2)}{P(y_1 < Y(\xi) \leq y_2)} \\ &= \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)} \end{aligned}$$

The above equation can be rewrite as

$$F_X(x|y_1 < Y \leq y_2) = \frac{\int_{-\infty}^{x} \int_{y_1}^{y_2} f_{XY}(u, v) \,\text{dvdu}}{\int_{y_1}^{y_2} f_Y(v) \, dv}$$

Compare ($P(A|B) = P(AB)/P(B)$).

We have examined how to condition a mass/density function by the occurrence of an event B, where event B describes some property of X. Now we focus on the special case in which the event B has the form of $X = x$ or $Y = y$. Learning $Y = y$ changes the likelihood that $X = x$. For example, conditional PMF is defined as:

$$P_{X|Y}(x|y) = P[X = x|Y = y]$$

To determine, the limiting case $F_X(x|Y = y)$, we can let $y_1 = y$ and $y_2 = y + \Delta y$, then this gives

$$F_X(x|y < Y \leq y + \Delta y) = \frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta y} f_{XY}(u,v)\mathrm{dvdu}}{\int_{y}^{y+\Delta y} f_Y(v)\,dv} = \frac{\int_{-\infty}^{x} f_{XY}(u,v)\,du\Delta y}{f_Y(y)\Delta y}$$

and hence in the limit

$$F_X(x|Y = y) = \lim_{\Delta y \to 0} F_X(x|y < Y \leq y + \Delta y) = \frac{\int_{-\infty}^{x} f_{XY}(u,v)\,du}{f_Y(y)}$$

To remind about the conditional nature on the left hand side, we shall use the subscript $X|Y$

(instead of X) there. Thus

$$\boxed{F_{X|Y}(x|Y=y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}} \tag{47}$$

Differentiating above with respect to x, we get

$$\boxed{f_{X|Y}(x|Y=y) = \frac{f_{XY}(x,y)}{f_Y(y)}} \tag{48}$$

It is easy to see that the conditional density represents a valid probability density function. In fact

$$f_{X|Y}(x|Y=y) = \frac{f_{XY}(x,y)}{f_Y(y)} \geq 0$$

and

$$\int_{-\infty}^{\infty} f_{X|Y}(x|Y=y)dx = \frac{\int_{-\infty}^{\infty} f_{XY}(x,y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

Therefore, the conditional density indeed represents a valid pdf, and we shall refer to it as the conditional pdf of the RV X given $Y=y$. We may also write

$$\boxed{f_{X|Y}(x|Y=y) = f_{X|Y}(x|y)} \tag{49}$$

and

$$\boxed{f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}} \tag{50}$$

and similarly

$$\boxed{f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}} \tag{51}$$

If the RVs X and Y are independent, then $f_{XY}(x,y) = f_X(x)f_Y(y)$ and the conditional density reduces to

$$f_{X|Y}(x|y) = f_X(x) \qquad f_{Y|X}(y|x) = f_Y(y) \tag{52}$$

implying that the conditional pdfs coincide with their unconditional pdfs. This makes sense, since if X and Y are independent RVs, information about Y shouldn't be of any help in updating our knowledge about X.

In the case of discrete-type RVs, conditional density reduces to

$$\boxed{P(X = x_i|Y = y_j) = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}} \tag{53}$$

$\boxed{\text{Example 13}}$: Given

$$f_{XY}(x,y) = \begin{cases} k & 0 \leq x \leq y \leq 1 \\ 0 & \text{o.w.} \end{cases}$$

determine $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$

Solution: The joint pdf is given to be a constant in the shadowed region. This gives

$$\int\int f_{XY}(x,y)\, dx\, dy = \int_0^1 \int_0^y k\ dx\, dy = \int_0^1 k\, y\ dy = \frac{k}{2} = 1 \Rightarrow k = 2$$

Similarly

$$f_X(x) = \int f_{XY}(x,y)\, dy = \int_x^1 k\, dy = k(1-x), \quad 0 \leq x \leq 1$$

and

$$f_Y(y) = \int f_{XY}(x,y)\, dx = \int_0^y k\, dx = k\, y, \quad 0 \leq y \leq 1$$

Therefore

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{y}, \quad 0 \leq x \leq y$$

and

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{1-x}, \qquad x \leqslant y \leqslant 1$$

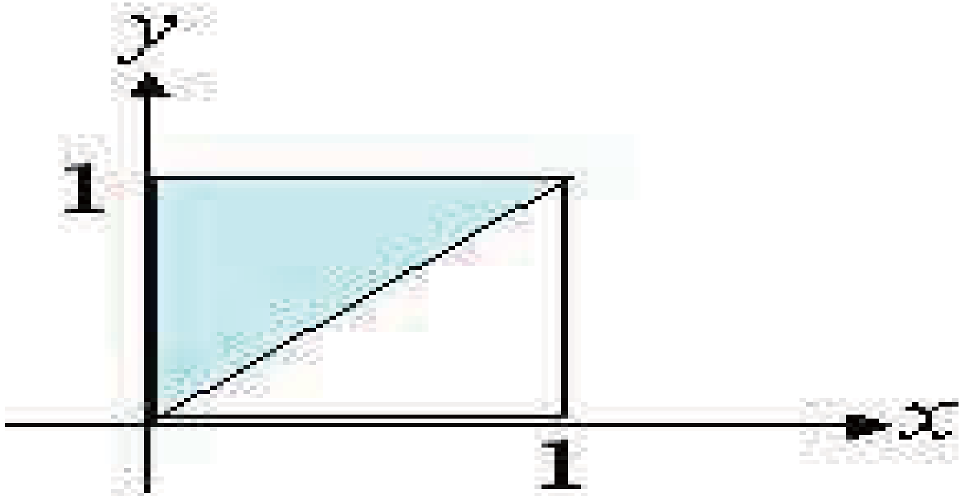

Figure 7: Joint pdf of X and Y (Example 13).

Example 14: Let R be a uniform random variable with parameters 0 and 1. Given $R = r$, X is a uniform random variable with parameters 0 and r. Find the conditional pdf of R given X, $f_{R|X}(r|x)$).

Solution: conditional density of X given R is

$$f_{X|R}(x|r) = \begin{cases} \frac{1}{r} & 0 \le x \le r \\ 0 & \text{o.w.} \end{cases}$$

since

$$f_R(r) = \begin{cases} 1 & 0 \le r \le 1 \\ 0 & \text{o.w.} \end{cases}$$

it follows that the joint pdf of R and X is

$$f_{R,X}(r,x) = f_{X|R}(x|r)\, f_R(r) = \begin{cases} \frac{1}{r} & 0 \le x \leqslant r \leqslant 1 \\ 0 & \text{o.w.} \end{cases}$$

Now, we can find the marginal pdf of X as

$$f_X(x) = \int_{-\infty}^{\infty} f_{R,X}(r,x)dr = \int_{x}^{1} \frac{1}{r}\,dr = -\ln x \qquad 0 < x \lessgtr 1$$

From the definition of the condition pdf

$$f_{R|X}(r|x) = \frac{f_{R,X}(r,x)}{f_X(x)} = \begin{cases} -\frac{1}{r\ln x} & x \le r \lessgtr 1 \\ 0 & \text{o.w.} \end{cases}$$

Chap 3.5 : Conditional Mean

We can use the conditional pdfs to define the conditional mean. More generally, applying definition of expectation to conditional pdfs we get

$$E[g(X)|B] = \int_{-\infty}^{\infty} g(x) f_X(x|B)\, dx$$

Using a limiting argument, we obtain

$$\boxed{\mu_{X|Y} = E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx} \tag{54}$$

to be the conditional mean of X given $Y = y$. Notice that $E(X|Y = y)$ will be a function of y. Also

$$\boxed{\mu_{Y|X} = E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy} \tag{55}$$

In a similar manner, the conditional variance of X given $Y = y$ is given by

$$\boxed{Var(X|Y) = \sigma^2_{X|Y} = E(X^2|Y=y) - [E(X|Y=y)]^2 = E[(x-\mu_{X|Y})^2|Y=y]} \tag{56}$$

$\boxed{\text{Example 15}}$: Let

$$f_{XY}(x,y) = \begin{cases} 1 & 0 < |y| < x < 1 \\ 0 & \text{o.w.} \end{cases}$$

Determine $E(X|Y)$ and $E(Y|X)$.

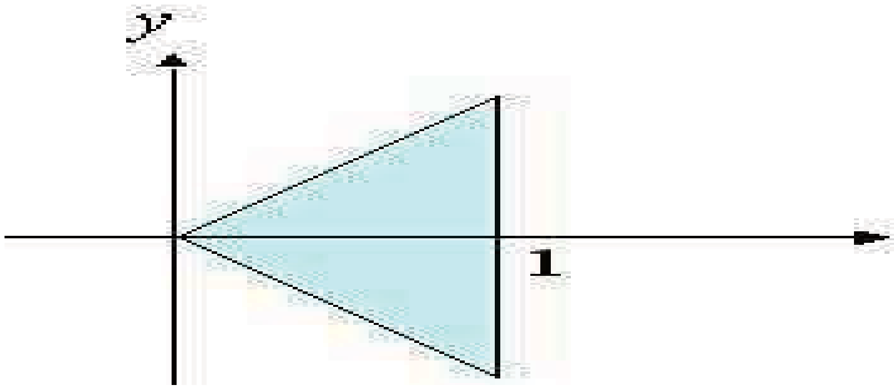

Figure 8: Example for conditional expectation.

Solution: As Fig. 8 shows, $f_{XY}(x,y) = 1$ in the shadowed area, and zero elsewhere. From there

$$f_X(x) = \int_{-x}^{x} f_{XY}(x,y)\,dy = 2x \quad 0 < x < 1$$

and

$$f_Y(y) = \int_{|y|}^{1} 1\, dx = 1 - |y| \quad |y| < 1$$

This gives

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{1-|y|} \qquad |y| < x < 1$$

and

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{2x} \quad 0 < |y| < x$$

Hence

$$\begin{aligned} E(X|Y) &= \int x f_{X|Y}(x|y)\, dx = \int_{|y|}^{1} \frac{x}{1-|y|}\, dx = \frac{1}{1-|y|} \frac{x^2}{2}\bigg|_{|y|}^{1} \\ &= \frac{1-|y|^2}{2(1-|y|)} = \frac{1+|y|}{2} \quad |y| < 1 \\ E(Y|X) &= \int y\, f_{Y|X}(y|x)\, dy = \int_{-x}^{x} \frac{y}{2x}\, dy = \frac{1}{2x} \frac{y^2}{2}\bigg|_{-x}^{x} = 0 \quad 0 < x < 1 \end{aligned}$$

It is possible to obtain an interesting generalization of the conditional mean formulas as

$$E[g(X)|Y = y)] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx \tag{57}$$

Example 16 : **Poisson sum of Bernoulli random variables:** Let $X_i, i = 1, 2, 3, \cdots,$ represent independent, identically distributed Bernoulli random variables with

$$P(X_i = 1) = p \qquad P(X_i = 0) = 1 - p = q$$

and N a Poisson random variable with parameter λ that is independent of all X_i. Consider the random variables

$$Y_1 = \sum_{i=1}^{N} X_i, \qquad Y_2 = N - Y_1$$

Show that Y_1 and Y_2 are independent Poisson random variables.

Solution : the joint probability mass function of Y_1 and Y_2 can be solved as

$$\begin{aligned} P(Y_1 = m, Y_2 = n) &= P(Y_1 = m, N - Y_1 = n) = P(Y_1 = m, N = m + n) \\ &= P(Y_1 = m | N = m + n) P(N = m + n) \\ &= P\left(\sum_{i=1}^{N} X_i = m | N = m + n\right) P(N = m + n) \\ &= P\left(\sum_{i=1}^{m+n} X_i = m\right) P(N = m + n) \end{aligned}$$

Note that $\sum_{i=1}^{m+n} X_i \sim B(m+n, p)$ and X_i's are independent of N

$$\begin{aligned} P(Y_1 = m, Y_2 = n) &= \left(\frac{(m+n)!}{m!\,n!} p^m\, q^n\right)\left(e^{-\lambda}\,\frac{\lambda^{m+n}}{(m+n)!}\right) \\ &= \left(e^{-p\lambda}\,\frac{(p\lambda)^m}{m!}\right)\left(e^{-q\lambda}\,\frac{(q\lambda)^n}{n!}\right) \\ &= P(Y_1 = m)\cdot P(Y_2 = n) \end{aligned}$$

Thus,

$$Y_1 \sim P(p\lambda) \qquad Y_2 \sim P(q\lambda)$$

and Y_1 and Y_2 are independent random variables. Thus if a bird lays eggs that follow a Poisson random variable with parameter λ, and if each egg survives with probability p, then the number of baby birds that survive also forms a Poisson random variable with parameter $p\lambda$.

$\boxed{\text{Example 17}}$: Suppose that the number of people who visit a yoga academy each day is a Poisson RV. with mean λ. Suppose further that each person who visits is, independently, female with probability p or male with probability $1-p$. Find the joint probability that exactly n women and m men visit the academy today.

Solution: Let N_1 denote the number of women, and N_2 the number of men, who visit the academy today. Also, let $N = N_1 + N_2$ be the total number of people who visit. Conditioning on N gives

$$P(N_1 = n, N_2 = m) = \sum_{i=0}^{\infty} P[N_1 = n, N_2 = m | N = i] P(N = i)$$

Because $P[N_1 = n, N_2 = m | N = i] = 0$ when $n + m \neq i$, therefore

$$\begin{aligned} P(N_1 = n, N_2 = m) &= P[N_1 = n, N_2 = m | N = n + m] e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \\ &= \binom{n+m}{n} p^n (1-p)^m e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \\ &= e^{-\lambda p} \frac{(\lambda p)^n}{n!} \times e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^m}{m!} \\ &= P(N_1 = n) \times P(N_2 = m) \end{aligned} \tag{58}$$

We can conclude that N_1 and N_2 are independent Poisson RVs with respectively means λp and $\lambda(1-p)$. Therefore, example 16 and 17 showed an important result: when each of a

Poisson number of events is independently classified either as being type 1 with probability p or type 2 with probability $1 - p$, then the number of type 1 and type 2 events are independent Poisson random variables.

Computing Expectation by Conditioning

- $E[X] = E[E[X|Y]]$

$$E[X] = \sum_y E[X|Y=y]P(Y=y) \qquad \text{Y is discrete} \qquad (59)$$

$$E[X] = \int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy \qquad \text{Y is continuous} \qquad (60)$$

Proof: (see textbook for X and Y discrete case)

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy = E[E[X|Y]]
\end{aligned}$$

Computing Expectation by Conditioning – Examples

Example 18 (The expectation of the sum of a random number of random variables) Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the number of workers injured in each accident are independent RVs with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?

Solution: Letting N denote the number of accidents and X_i the number of injured in the ith accident, $i = 1, 2, ...$, then the total number of injuries can be expressed as $\sum_{i=1}^{N} X_i$. Now

$$E\left[\sum_{1}^{N} X_i\right] = E\left[E\left[\sum_{1}^{N} X_i | N\right]\right]$$

But

$$
\begin{aligned}
E\left[\sum_{1}^{N} X_i | N = n\right] &= E\left[\sum_{1}^{n} X_i | N = n\right] \\
&= E\left[\sum_{1}^{n} X_i\right] \qquad \text{by independence of } X_i \text{ and } N \\
&= nE[X]
\end{aligned}
$$

which is

$$
E\left[\sum_{1}^{N} X_i | N\right] = NE[X]
$$

and thus

$$
E\left[\sum_{1}^{N} X_i\right] = E[NE[X]] = E[N]E[X]
$$

Therefore, the expected number of injuries during a week equals $4 \times 2 = 8$.

Example 19 (The mean of a geometric distribution) A coin, having probability p of coming up head, is to be successively flipped until the first head appears. What is the expected number of flips required?

Solution: Letting N be the number of flips required, and let

$$Y = \begin{cases} 1, & \text{if the first flip results in a head} \\ 0. & \text{if the first flip results in a tail} \end{cases}$$

Now

$$\begin{aligned} E[N] &= E[N|Y=1]P[Y=1] + E[N|Y=0]P[Y=0] \\ &= pE[N|Y=1] + (1-p)E[N|Y=0] = p + (1-p)(1+E[N]) \end{aligned}$$

and thus

$$E[N] = 1/p$$

compare with $E[N] = \sum_1^\infty np(n) = \sum_1^\infty n \times p(1-p)^{n-1}$

Example 20 A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to his mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Letting X denote the time until the miner reaches safety, and let Y denote the door he initially chooses. Now

$$\begin{aligned} E[X] &= E[X|Y=1]P[Y=1] + E[X|Y=2]P[Y=2] + E[X|Y=3]P[Y=3] \\ &= \frac{1}{3}\left(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]\right) \end{aligned}$$

However

$$\begin{aligned} E[X|Y=1] &= 2 \qquad\qquad E[X|Y=2] = 3 + E[X] \\ E[X|Y=3] &= 5 + E[X] \end{aligned}$$

and thus

$$E[X] = \frac{1}{3}(2 + 3 + E[X] + 5 + E[X]) \qquad \text{leads to} \qquad E[X] = 10 \text{ hours.}$$

Computing Probability by Conditioning

Let E denote an arbitary event ,

for any RV Y. Therefore,

$$
\begin{aligned}
P[E] &= \sum_y P[E|Y=y]P(Y=y) \quad \text{if } Y \text{ is discrete} && (61) \\
&= \int_{-\infty}^{\infty} P[E|Y=y]f_Y(y)dy \quad \text{if } Y \text{ is continuous} && (62)
\end{aligned}
$$

Example 21 Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $P(X < Y)$

Solution: Conditioning on the value of Y yields

$$\begin{aligned} P(X < Y) &= \int_{-\infty}^{\infty} P[X < Y|Y = y] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P[X < y|Y = y] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P(X < y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy \end{aligned}$$

where

$$F_X(y) = \int_{-\infty}^{y} f_X(x) dx$$