ECE 302: Probability and Applications[1]

Week 9 Topics

- Conditional pmf and pdf
  - Total Probability & Conditional Expectation
  - Mean Square Estimation
- One Function of Two Random Variables
- Transformation of Two Random Variables

1 Conditional Probability and Conditional Expectation

Many random variables of practical interest are not independent: The output Y of a communication channel must depend on the input X in order to convey information; consecutive samples of a waveform that varies slowly are likely to be close in value and hence are not independent. In this section we are interested in computing the probability of events concerning the random variable Y given that we know $X = x$. We are also interested in the expected value of Y given $X = x$. We show that the notions of conditional probability and conditional expectation are extremely useful tools in solving problems, even in situations where we are only concerned with one of the random variables.

1.1 Conditional Probability

The definition of conditional probability allows us to compute the probability that Y is in A given that we know that $X = x$:

$$P[Y \text{ in } A \mid X = x] = \frac{P[Y \text{ in } A, X = x]}{P[X = x]} \text{ for } P[X = x] > 0. \tag{1}$$

Case 1: $\boldsymbol{X}$ Is a Discrete Random Variable

For X and Y discrete random variables, **the conditional pmf of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$** is defined by:

$$p_Y(y \mid x) = P[Y = y \mid X = x] = \frac{P[X = x, Y = y]}{P[X = x]} = \frac{p_{X,Y}(x, y)}{p_X(x)} \tag{2}$$

for x such that $P[X = x] > 0$. We define $p_Y(y \mid x) = 0$ for x such that $P[X = x] = 0$. Note that $p_Y(y \mid x)$ is a function of y over the real line, and that $p_Y(y \mid x) > 0$ only for y in a discrete set $\{y_1, y_2, \ldots\}$.

The conditional pmf satisfies all the properties of a pmf, that is, it assigns nonnegative values to every y and these values add to 1 . Note from Eq. (2) that $p_Y(y \mid x_k)$ is simply the cross section of $p_{X,Y}(x_k, y)$ along the column $X = x_k$ column, but normalized by the probability $p_X(x_k)$.

The probability of an event A given $X = x_k$ is found by adding the pmf values of the outcomes in A:

$$P\left[Y \text{ in } A \mid X = x_k\right] = \sum_{y_j \text{ in } A} p_Y\left(y_j \mid x_k\right) \tag{3}$$

If X and Y are independent, then the joint pmf factors into the product of the marginal pmfs, so

$$p_Y\left(y_j \mid x_k\right) = \frac{P\left[X = x_k, Y = y_j\right]}{P\left[X = x_k\right]} = P\left[Y = y_j\right] = p_Y\left(y_j\right) \tag{4}$$

In other words, if X and Y are independent then knowledge that $X = x_k$ does not affect the probability of events A involving Y.

Equation (2) implies that the joint pmf $p_{X,Y}(x, y)$ can be expressed as the product of a conditional pmf and a marginal pmf:

$$p_{X,Y}\left(x_k, y_j\right) = p_Y\left(y_j \mid x_k\right) p_X\left(x_k\right) \text{ and } p_{X,Y}\left(x_k, y_j\right) = p_X\left(x_k \mid y_j\right) p_Y\left(y_j\right)$$

This expression is very useful when we can view the pair (X, Y) as being generated sequentially, e.g., first X, and then Y given $X = x$. We find the probability that Y is in A as follows:

$$\begin{aligned} P[Y \text{ in } A] &= \sum_{\text{all } x_k} \sum_{y_j \text{ in } A} p_{X,Y}(x_k, y_j) \\ &= \sum_{\text{all } x_k} \sum_{y_j \text{ in } A} p_Y\left(y_j \mid x_k\right) p_X\left(x_k\right) \\ &= \sum_{\text{all } x_k} p_X\left(x_k\right) \sum_{y_j \text{ in } A} p_Y\left(y_j \mid x_k\right) \\ &= \sum_{\text{all } x_k} P\left[Y \text{ in } A \mid X = x_k\right] p_X\left(x_k\right) \end{aligned} \tag{5}$$

Equation (5) is simply a restatement of the theorem on total probability. In other words, to compute $P[Y \text{ in } A]$ we can first compute $P\left[Y \text{ in } A \mid X = x_k\right]$ and then "average" over X_k.

Example 1: Loaded Dice

Find $p_Y(y \mid 5)$ in the loaded dice experiment considered in Example 15 in Week 7.

In the aforementioned Example we found that $p_X(5) = 1/6$. Therefore:

$$p_Y(y \mid 5) = \frac{p_{X,Y}(5,y)}{p_X(5)} \text{ and so } p_Y(5 \mid 5) = 2/7 \text{ and}$$
$$p_Y(1 \mid 5) = p_Y(2 \mid 5) = p_Y(3 \mid 5) = p_Y(4 \mid 5) = p_Y(6 \mid 5) = 1/7$$

Clearly this die is not fair, that is, loaded.

Example 2: Number of Defects in a Region; Random Splitting of Poisson Counts

The total number of defects X on a chip is a Poisson random variable with mean α. Each defect has a probability p of falling in a specific region R and the location of each defect is independent of the locations of other defects. Find the pmf of the number of defects Y that fall in the region R.

We can imagine performing a Bernoulli trial each time a defect occurs with a "success" occurring when the defect falls in the region R. If the total number of defects is $X = k$, then Y is a binomial random variable with parameters k and p:

$$p_Y(j \mid k) = \begin{cases} 0 & j > k \\ \binom{k}{j} p^j (1-p)^{k-j} & 0 \le j \le k \end{cases}$$

From Eq. (5) and noting that $k \ge j$, we have

$$\begin{aligned} p_Y(j) &= \sum_{k=0}^{\infty} p_Y(j \mid k) p_X(k) = \sum_{k=j}^{\infty} \frac{k!}{j!(k-j)!} p^j (1-p)^{k-j} \frac{\alpha^k}{k!} e^{-\alpha} \\ &= \frac{(\alpha p)^j e^{-\alpha}}{j!} \sum_{k=j}^{\infty} \frac{\{(1-p)\alpha\}^{k-j}}{(k-j)!} \\ &= \frac{(\alpha p)^j e^{-\alpha}}{j!} e^{(1-p)\alpha} = \frac{(\alpha p)^j}{j!} e^{-\alpha p}. \end{aligned}$$

Thus Y is a Poisson random variable with mean αp.

Suppose Y is a continuous random variable. Eq. (1) can be used to define **the conditional cdf of Y given $X = x_k$** :

$$F_Y(y \mid x_k) = \frac{P[Y \le y, X = x_k]}{P[X = x_k]}, \quad \text{for } P[X = x_k] > 0 \tag{6}$$

It is easy to show that $F_Y(y \mid x_k)$ satisfies all the properties of a cdf. The **conditional pdf of $\boldsymbol{Y}$ given $\boldsymbol{X = x_k}$**, if the derivative exists, is given by

$$f_Y(y \mid x_k) = \frac{d}{dy} F_Y(y \mid x_k) \tag{7}$$

If X and Y are independent, $P[Y \leq y, X = X_k] = P[Y \leq y] P[X = X_k]$ so $F_Y(y \mid x) = F_Y(y)$ and $f_Y(y \mid x) = f_Y(y)$.

The probability of event A given $X = x_k$ is obtained by integrating the conditional pdf:

$$P[Y \text{ in } A \mid X = x_k] = \int_{y \text{ in } A} f_Y(y \mid x_k)\, dy \tag{8}$$

We obtain $P[Y \text{ in } A]$ using Eq. 5.

Example 3: Binary Communications System

The input X to a communication channel assumes the values +1 or -1 with probabilities 1/3 and 2/3. The output Y of the channel is given by $Y = X + N$, where N is a zero-mean, unit variance Gaussian random variable. Find the conditional pdf of Y given $X = +1$, and given $X = -1$. Find $P[X = +1 \mid Y > 0]$.

The conditional cdf of Y given $X = +1$ is:

$$\begin{aligned} F_Y(y \mid +1) &= P[Y \leq y \mid X = +1] = P[N + 1 \leq y] \\ &= P[N \leq y - 1] = \int_{-\infty}^{y-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \end{aligned}$$

where we noted that if $X = +1$, then $Y = N + 1$ and Y depends only on N. Thus, if $X = +1$, then Y is a Gaussian random variable with mean 1 and unit variance. Similarly, if $X = -1$, then Y is Gaussian with mean -1 and unit variance.

The probabilities that $Y > 0$ given $X = +1$ and $X = -1$ is:

$$P[Y > 0 \mid X = +1] = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2} dx = \int_{-1}^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 - Q(1) = 0.841.$$

$$P[Y > 0 \mid X = -1] = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x+1)^2/2} dx = \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = Q(1) = 0.159$$

Applying Eq. (5), we obtain:

$$P[Y > 0] = P[Y > 0 \mid X = +1]\frac{1}{3} + P[Y > 0 \mid X = -1]\frac{2}{3} = 0.386.$$

From Bayes' theorem we find:

$$P[X = +1 \mid Y > 0] = \frac{P[Y > 0 \mid X = +1]P[X = +1]}{P[Y > 0]} = \frac{(1 - Q(1))/3}{(1 + Q(1))/3} = 0.726$$

We conclude that if $Y > 0$, then $X = +1$ is more likely than $X = -1$. Therefore the receiver should decide that the input is $X = +1$ when it observes $Y > 0$.

In the previous example, we made an interesting step that is worth elaborating on because it comes up quite frequently: $P[Y \leq y \mid X = +1] = P[N + 1 \leq y]$, where $Y = X + N$. Let's take a closer look:

$$\begin{aligned} P[Y \leq z \mid X = x] &= \frac{P[\{X + N \leq z\} \cap \{X = x\}]}{P[X = x]} = \frac{P[\{x + N \leq z\} \cap \{X = x\}]}{P[X = x]} \\ &= P[x + N \leq z \mid X = x] = P[N \leq z - x \mid X = x] \end{aligned}$$

In the first line, the events $\{X + N \leq z\}$ and $\{x + N \leq z\}$ are quite different. The first involves the two random variables X and N, whereas the second only involves N and consequently is much simpler. We can then apply an expression such as Eq. (5) to obtain $P[Y \leq z]$. The step we made in the example, however, is even more interesting. Since X and N are independent random variables, we can take the expression one step further:

$$P[Y \leq z \mid X = x] = P[N \leq z - x \mid X = x] = P[N \leq z - x]$$

The independence of X and N allows us to dispense with the conditioning on x altogether!

Case 2: *X* Is a Continuous Random Variable

If X is a continuous random variable, then $P[X = x] = 0$ so Eq. (1) is undefined for all x. If X and Y have a joint pdf that is continuous and nonzero over some region of the plane, we define the **conditional cdf of $\boldsymbol{Y}$ given $\boldsymbol{X = x}$** by the following limiting procedure:

$$F_Y(y \mid x) = \lim_{h \to 0} F_Y(y \mid x < X \leq x + h) \tag{9}$$

The conditional cdf on the right side of Eq. 9 is:

$$\begin{aligned} F_Y(y \mid x < X \leq x + h) &= \frac{P[Y \leq y, x < X \leq x + h]}{P[x < X \leq x + h]} \\ &= \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}\left(x', y'\right) dx' dy'}{\int_{x}^{x+h} f_X\left(x'\right) dx'} = \frac{\int_{-\infty}^{y} f_{X,Y}\left(x, y'\right) dy' h}{f_X(x) h} \end{aligned} \tag{10}$$

As we let h approach zero, Eqs. (9) and (10) imply that

$$F_Y(y \mid x) = \frac{\int_{-\infty}^{y} f_{X,Y}\left(x, y'\right) dy'}{f_X(x)} \tag{11}$$

The conditional pdf of $\boldsymbol{Y}$ given $\boldsymbol{X = x}$ is then:

$$f_Y(y \mid x) = \frac{d}{dy} F_Y(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{12}$$

It is easy to show that $f_Y(y \mid x)$ satisfies the properties of a pdf. We can interpret $f_Y(y \mid x) dy$ as the probability that Y is in the infinitesimal strip defined by $(y, y + dy)$ given that X is in the infinitesimal strip defined by $(x, x + dx)$, as shown in Fig. 1.

The probability of event A given $X = x$ is obtained as follows:

$$P[Y \text{ in } A \mid X = x] = \int_{y \text{ in } A} f_Y(y \mid x) dy \tag{13}$$

There is a strong resemblance between Eq. (2) for the discrete case and Eq. (12) for the continuous case. Indeed many of the same properties hold. For example, we obtain the multiplication rule from Eq. (12):

$$f_{X,Y}(x, y) = f_Y(y \mid x) f_X(x) \text{ and } f_{X,Y}(x, y) = f_X(x \mid y) f_Y(y) \tag{14}$$

If X and Y are independent, then $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ and $f_Y(y \mid x) = f_Y(y)$, $f_X(x \mid y) = f_X(\text{x})$, $F_Y(y \mid x) = F_Y(y)$, and $F_X(x \mid y) = F_X(x)$.

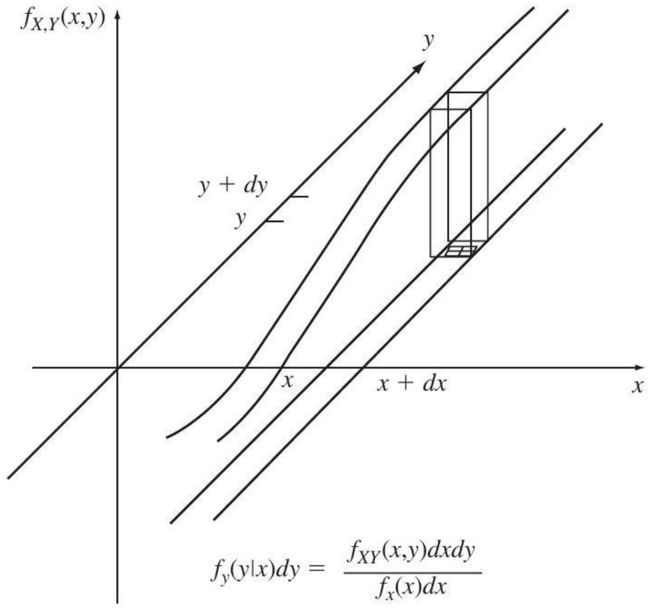

Figure 1: Interpretation of conditional pdf.

By combining Eqs. (13) and (14), we can show that:

$$P[Y \text{ in } A] = \int_{-\infty}^{\infty} P[Y \text{ in } A \mid X = x] f_X(x) dx \tag{15}$$

You can think of Eq.(15) as the "continuous" version of the theorem on total probability. The following examples show the usefulness of the above results in calculating the probabilities of complicated events.

Example 4

Let X and Y be the random variables in Example 2 of Week 8. Find $f_X(x \mid y)$ and $f_Y(y \mid x)$.

Using the marginal pdf's obtained in the Example, we have

$$\begin{aligned} f_X(y \mid x) &= \tfrac{2e^{-x}e^{-y}}{2e^{-2y}} = e^{-(x-y)} && \text{for } x \geq y \\ f_Y(y \mid x) &= \tfrac{2e^{-x}e^{-y}}{2e^{-x}(1-e^{-x})} = \tfrac{e^{-y}}{1-e^{-x}} && \text{for } 0 < y < x \end{aligned}$$

The conditional pdf of X is an exponential pdf shifted by y to the right. The conditional pdf of Y is an exponential pdf that has been truncated to the interval $[0, x]$.

Example 5: Number of Arrivals During a Customer's Service Time

The number N of customers that arrive at a service station during a time t is a Poisson random variable with arrival rate β. The time T required to service each customer is an exponential random variable with parameter α. Find the pmf for the number N that arrive during the service time T of a specific customer. Assume that the customer arrivals are independent of the customer service time.

Equation (15) holds even if Y is a discrete random variable, thus

$$\begin{aligned} P[N=k] &= \int_0^\infty P[N=k \mid T=t] f_T(t)dt \\ &= \int_0^\infty \frac{(\beta t)^k}{k!} e^{-\beta t} \alpha e^{-\alpha t} dt \\ &= \frac{\alpha \beta^k}{k!} \int_0^\infty t^k e^{-(\alpha+\beta)t} dt \end{aligned}$$

Let $r = (\alpha+\beta)t$, then

$$\begin{aligned} P[N=k] &= \frac{\alpha\beta^k}{k!(\alpha+\beta)^{k+1}} \int_0^\infty r^k e^{-r} dr \\ &= \frac{\alpha\beta^k}{(\alpha+\beta)^{k+1}} = \left(\frac{\alpha}{(\alpha+\beta)}\right)\left(\frac{\beta}{(\alpha+\beta)}\right)^k \end{aligned}$$

where we have used the fact that the last integral is a gamma function and is equal to $k!$. Thus N is a geometric random variable with probability of "success" $\alpha/(\alpha+\beta)$. Each time a customer arrives we can imagine that a new Bernoulli trial begins where "success" occurs if the customer's service time is completed before the next arrival.

Example 6

X is selected at random from the unit interval; Y is then selected at random from the interval$(0, X)$. Find the cdf of Y.

When $X = x, Y$ is uniformly distributed in $(0, x)$ so the conditional cdf given $X = x$ is

$$P[Y \le y \mid X = k] = \begin{cases} y/x & 0 \le y \le x \\ 1 & x < y \end{cases}$$

Equation (15) and the above conditional cdf yield:

$$F_Y(y) = P[Y \le y] = \int_0^1 P[Y \le y \mid X = x] f_X(x) dx =$$
$$= \int_0^y 1 dx' + \int_y^1 \frac{y}{x'} dx' = y - y \ln y$$

where we expressed the interval from 0 to 1, into two parts, from 0 to y and then from y to 1. The corresponding pdf is obtained by taking the derivative of the cdf:

$$f_Y(y) = -\ln y \quad 0 \le y \le 1$$

Example 7: Maximum A Posteriori Receiver

For the communications system in Example 3, find the probability that the input was $X = +1$ given that the output of the channel is $Y = y$.

This is a tricky version of Bayes' rule. We condition on the event $\{y < Y \le y + \Delta\}$ instead of $\{Y = y\}$:

$$P[X = +1 \mid y < Y < y + \Delta] = \frac{P[y < Y < y + \Delta \mid X = +1] P[X = +1]}{P[y < Y < y + \Delta]}$$
$$= \frac{f_Y(y \mid +1)\Delta(1/3)}{f_Y(y \mid +1)\Delta(1/3) + f_Y(y \mid -1)\Delta(2/3)}$$
$$= \frac{\frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}(1/3)}{\frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}(1/3) + \frac{1}{\sqrt{2\pi}} e^{-(y+1)^2/2}(2/3)}$$
$$= \frac{e^{-(y-1)^2/2}}{e^{-(y-1)^2/2} + 2e^{-(y+1)^2/2}} = \frac{1}{1 + 2e^{-2y}}$$

The above expression is equal to 1/2 when $y_T = 0.3466$. For $y > y_T, X = +1$ is more likely, and for $y < y_T, X = -1$ is more likely. A receiver that selects the input X that is the most likely given $Y = y$ is called a maximum a posteriori receiver.

1.2 Conditional Expectation

The **conditional expectation** of $\boldsymbol{Y}$ given $\boldsymbol{X = x}$ is defined by

$$E[Y \mid x] = \int_{-\infty}^{\infty} y f_Y(y \mid x) dy \tag{16}$$

In the special case where X and Y are both discrete random variables we have:

$$E\left[Y \mid x_k\right]=\sum_{y_j} y_j p_Y\left(y_j \mid x_k\right) \tag{17}$$

Clearly, $E[Y \mid x]$ is simply the center of mass associated with the conditional pdf or pmf.

The conditional expectation $E[Y \mid x]$ can be viewed as defining a function of $x: g(x)=E[Y \mid x]$. It therefore makes sense to talk about the random variable $g(X)=E[Y \mid X]$. We can imagine that a random experiment is performed and a value for X is obtained, say $X=x_0$, and then the value $g\left(x_0\right)=E\left[Y \mid x_0\right]$ is produced. We are interested in $E[g(X)]=E[E[Y \mid X]]$. In particular, we now show that

$$E[Y]=E[E[Y \mid X]] \tag{18}$$

where the right-hand side is

$$\begin{array}{ll} E[E[Y \mid X]]=\int_{-\infty}^{\infty} E[Y \mid x] f_X(x) d x & X \text{ continuous} \\ E[E[Y \mid X]]=\sum_{x_k} E\left[Y \mid x_k\right] p_X\left(x_k\right) & X \text{ discrete.} \end{array}$$

We prove Eq. (18) for the case where X and Y are jointly continuous random variables, then

$$\begin{aligned} E[E[Y \mid X]] & =\int_{-\infty}^{\infty} E[Y \mid x] f_X(x) d x \\ & =\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_Y(y \mid x) d y f_X(x) d x \\ & =\int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X, Y}(x, y) d x d y \\ & =\int_{-\infty}^{\infty} y f_Y(y) d y=E[Y] \end{aligned}$$

The above result also holds for the expected value of a function of Y :

$$E[h(Y)]=E[E[h(Y) \mid X]]$$

In particular, the kth moment of Y is given by

$$E\left[Y^k\right]=E\left[E\left[Y^k \mid X\right]\right]$$

Example 8: Average Number of Defects in a Region

Find the mean of Y in Example 2 using conditional expectation.

$$E[Y] = \sum_{k=0}^{\infty} E[Y \mid X = k] P[X = k] = \sum_{k=0}^{\infty} kpP[X = k] = pE[X] = p\alpha$$

The second equality uses the fact that $E[Y \mid X = k] = kp$ since Y is binomial with parameters k and p. Note that the second to the last equality holds for any pmf of X. The fact that X is Poisson with mean α is not used until the last equality.

Example 9: Binary Communications Channel

Find the mean of the output Y in the communications channel in Example 3.

Since Y is a Gaussian random variable with mean +1 when $X = +1$, and -1 when $X = -1$, the conditional expected values of Y given X are:

$$E[Y \mid +1] = 1 \quad \text{and} \quad E[Y \mid -1] = -1$$

Equation (5) implies

$$E[Y] = \sum_{k=0}^{\infty} E[Y \mid X = k] P[X = k] = +1(1/3) - 1(2/3) = -1/3$$

The mean is negative because the $X = -1$ inputs occur twice as often as $X = +1$.

Example 10: Average Number of Arrivals in a Service Time

Find the mean and variance of the number of customer arrivals N during the service time T of a specific customer in Example 5.

N is a Poisson random variable with parameter βt when $T = t$ is given, so the first two conditional moments are:

$$E[N \mid T = t] = \beta t \quad E\left[N^2 \mid T = t\right] = (\beta t) + (\beta t)^2$$

The first two moments of N are obtained from Eq. (18):

$$\begin{aligned} E[N] &= \int_0^{\infty} E[N \mid T = t] f_T(t) dt = \int_0^{\infty} \beta t f_T(t) dt = \beta E[T] \\ E\left[N^2\right] &= \int_0^{\infty} E\left[N^2 \mid T = t\right] f_T(t) dt = \int_0^{\infty} \left\{\beta t + \beta^2 t^2\right\} f_T(t) dt \\ &= \beta E[T] + \beta^2 E\left[T^2\right] \end{aligned}$$

The variance of N is then

$$\begin{aligned}\mathrm{VAR}[N] &= E\left[N^2\right] - (E[N])^2 \\ &= \beta^2 E\left[T^2\right] + \beta E[T] - \beta^2 (E[T])^2 \\ &= \beta^2 \,\mathrm{VAR}[T] + \beta E[T]\end{aligned}$$

Note that if T is not random (i.e., $E[T]$ = constant and $\mathrm{VAR}[T] = 0$) then the mean and variance of N are those of a Poisson random variable with parameter $\beta E[T]$. When T is random, the mean of N remains the same but the variance of N increases by the term $\beta^2 \,\mathrm{VAR}[T]$, that is, the variability of T causes greater variability in N. Up to this point, we have intentionally avoided using the fact that T has an exponential distribution to emphasize that the above results hold for any service time distribution $f_T(t)$. If T is exponential with parameter α, then $E[T] = 1/\alpha$ and $\mathrm{VAR}[T] = 1/\alpha^2$, so

$$E[N] = \frac{\beta}{\alpha} \quad \text{and} \quad \mathrm{VAR}[N] = \frac{\beta^2}{\alpha^2} + \frac{\beta}{\alpha}$$

2 Functions of Two Random Variables

Quite often we are interested in one or more functions of the random variables associated with some experiment. For example, if we make repeated measurements of the same random quantity, we might be interested in the maximum and minimum value in the set, as well as the sample mean and sample variance. In this section we present methods of determining the probabilities of events involving functions of two random variables.

2.1 One Function of Two Random Variables

Let the random variable Z be defined as a function of two random variables:

$$Z = g(X, Y) \tag{19}$$

The cdf of Z is found by first finding the equivalent event of $\{Z \leq z\}$, that is, the set $R_z = \{\mathbf{x} = (x, y) \text{ such that } g(\mathbf{x}) \leq z\}$, then

$$F_z(z) = P\left[\mathbf{X} \text{ in } R_z\right] = \iint_{(x,y) \in R_z} f_{X,Y}\left(x', y'\right) dx' dy' \tag{20}$$

The pdf of Z is then found by taking the derivative of $F_z(z)$.

Example 11: Sum of Two Random Variables

Let $Z = X + Y$. Find $F_Z(z)$ and $f_Z(z)$ in terms of the joint pdf of X and Y.

The cdf of Z is found by integrating the joint pdf of X and Y over the region of the plane corresponding to the event $\{Z \leq z\}$, as shown in Fig. 2.

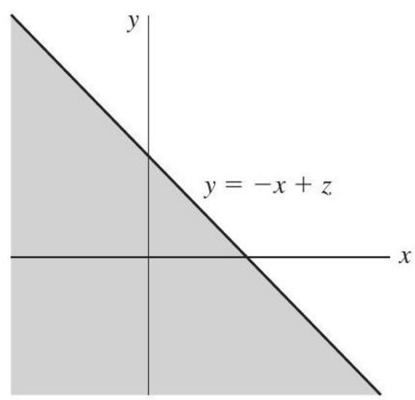

Figure 2: $P[Z \leq z] = P[X + Y \leq z]$.

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x'} f_{X,Y}(x', y')\, dy' dx'$$

The pdf of Z is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x', z - x')\, dx' \tag{21}$$

Thus the pdf for the sum of two random variables is given by a superposition integral.

If X and Y are independent random variables, then the joint pdf factors and the pdf of Z is given by the convolution integral of the marginal pdf's of X and Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x')\, f_Y(z - x')\, dx' \tag{22}$$

We will later show how transform methods are used to evaluate convolution integrals such as Eq. (22).

Example 12: Sum of Nonindependent Gaussian Random Variables

Find the pdf of the sum $Z = X + Y$ of two zero-mean, unit-variance Gaussian random variables with correlation coefficient $\rho = -1/2$.

The pdf of Z is obtained by substituting the pdf for the joint Gaussian random variables into the superposition integral found in Example 11:

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}\left(x', z - x'\right) dx' \\
&= \frac{1}{2\pi\left(1-\rho^2\right)^{1/2}} \int_{-\infty}^{\infty} e^{-\left[x'^2 - 2\rho x'(z-x') + (z-x')^2\right]/2\left(1-\rho^2\right)} dx' \\
&= \frac{1}{2\pi(3/4)^{1/2}} \int_{-\infty}^{\infty} e^{-\left(x'^2 - x'z + z^2\right)/2(3/4)} dx'
\end{aligned}$$

After completing the square of the argument in the exponent we obtain

$$f_Z(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

Thus the sum of these two nonindependent Gaussian random variables is also a zero-mean, unit-variance Gaussian random variable.

Example 13: A System with Standby Redundancy

A system with standby redundancy has a single key component in operation and a duplicate of that component in standby mode. When the first component fails, the second component is put into operation. Find the pdf of the lifetime of the standby system if the components have independent exponentially distributed lifetimes with the same mean.

Let T_1 and T_2 be the lifetimes of the two components, then the system lifetime is $T = T_1 + T_2$, and the pdf of T is given by Eq. (22). The terms in the integrand are

$$\begin{aligned}
f_{T_1}(x) &= \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \\
f_{T_2}(z - x) &= \begin{cases} \lambda e^{-\lambda(z-x)} & z - x \geq 0 \\ 0 & x > z \end{cases}
\end{aligned}$$

Note that the first equation sets the lower limit of integration to 0 and the second equation sets the upper limit to z. Equation (22) becomes

$$\begin{aligned} f_T(z) &= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z} \end{aligned}$$

Thus T is an Erlang random variable with parameter $m = 2$.

The conditional pdf can be used to find the pdf of a function of several random variables. Let $Z = g(X, Y)$, and suppose we are given that $Y = y$, then $Z = g(X, y)$ is a function of one random variable. Therefore we can use the methods developed in previously for single random variables to find the pdf of Z given $Y = y : f_Z(z \mid Y = y)$. The pdf of Z is then found from

$$f_Z(z) = \int_{-\infty}^{\infty} f_Z(z \mid y') f_Y(y') dy'$$

Example 14

Let $Z = X/Y$. Find the pdf of Z if X and Y are independent and both exponentially distributed with mean one.

Assume $Y = y$, then $Z = X/y$ is simply a scaled version of X. Therefore from Example 9 Week 6

$$f_Z(z \mid y) = |y| f_X(yz \mid y)$$

The pdf of Z is therefore

$$f_Z(z) = \int_{-\infty}^{\infty} |y'| f_X(y'z \mid y') f_Y(y') dy' = \int_{-\infty}^{\infty} |y'| f_{X,Y}(y'z, y') dy'$$

We now use the fact that X and Y are independent and exponentially distributed with mean one:

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} y' f_X(y'z) f_Y(y') dy' \quad z > 0 \\ &= \int_0^{\infty} y' e^{-y'z} e^{-y'} dy' \\ &= \frac{1}{(1+z)^2} \quad z > 0. \end{aligned}$$

3 Transformations of Two Random Variables

Let X and Y be random variables associated with some experiment, and let the random variables Z_1 and Z_2 be defined by two functions of $\mathbf{X} = (X, Y)$:

$$Z_1 = g_1(\mathbf{X}) \quad \text{and} \quad Z_2 = g_2(\mathbf{X})$$

We now consider the problem of finding the joint cdf and pdf of Z_1 and Z_2. The joint cdf of Z_1 and Z_2 at the point $\mathbf{z} = (z_1, z_2)$ is equal to the probability of the region of $\mathbf{x}$ where $g_k(\mathbf{x}) \leq z_k$ for $k = 1, 2$:

$$F_{z_1,z_2}(z_1, z_2) = P\left[g_1(\mathbf{X}) \leq z_1, g_2(\mathbf{X}) \leq z_2\right] \tag{23}$$

If X, Y have a joint pdf , then

$$F_{z_1,z_2}(z_1, z_2) = \iint_{\mathbf{x'}:\mathbf{g_k}(\mathbf{x'})\leq\mathbf{z_k}} f_{X,Y}(x', y')\, dx'dy' \tag{24}$$

Example 15

Let the random variables W and Z be defined by

$$W = \min(X, Y) \quad \text{and} \quad Z = \max(X, Y)$$

Find the joint cdf of W and Z in terms of the joint cdf of X and Y.

Equation (23) implies that

$$F_{W,Z}(wz) = P[\{\min(X, Y) \leq w\} \cap \{\max(X, Y) \leq z\}]$$

The region corresponding to this event is shown in Fig. 3. From the figure it is clear that if $z > w$, the above probability is the probability of the semi-infinite rectangle defined by the point (z, z) minus the square region denoted by A. Thus if $z > w$,

$$\begin{aligned}
F_{W,Z}(w, z) =& F_{X,Y}(z, z) - P[A] \\
=& F_{X,Y}(z, z) \\
& - \{F_{X,Y}(z, z) - F_{X,Y}(w, z) - F_{X,Y}(z, w) + F_{X,Y}(w, w)\} \\
=& F_{X,Y}(w, z) + F_{X,Y}(z, w) - F_{X,Y}(w, w)
\end{aligned}$$

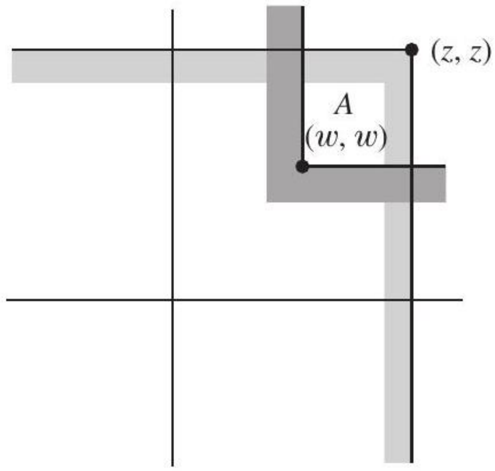

Figure 3: $\{\min(X,Y) \le w = \{X \le w\} \cup \{Y \le w\}$ and $\{\max(X,Y) \le z = \{X \le z\} \cap \{Y \le z\}$

If $z < w$ then

$$F_{W,Z}(w,z) = F_{X,Y}(z,z)$$

Example 16: Radius and Angle of Independent Gaussian Random Variables

Let X and Y be zero-mean, unit-variance independent Gaussian random variables. Find the joint cdf and pdf of R and Θ, the radius and angle of the point (X,Y):

$$R = \left(X^2+Y^2\right)^{1/2} \qquad \Theta = \tan^{-1}(Y/X)$$

The joint cdf of R and Θ is:

$$F_{R,\Theta}(r_0,\theta_0) = P\left[R \le r_0, \Theta \le \theta_0\right] = \iint_{(x,y)\in R_{(r_0,\theta_0)}} \frac{e^{-\left(x^2+y^2\right)/2}}{2\pi} dxdy$$

where

$$R_{(r_0,\theta_0)} = \left\{(x,y) : \sqrt{x^2+y^2} \le r_0, 0 < \tan^{-1}(Y/X) \le \theta_0\right\}$$

The region R_{r_0,θ_0} is the pie-shaped region in Fig. 4. We change variables from Cartesian to polar coordinates to obtain:

$$F_{R,\Theta}(r_0,\theta_0) = P[R \le r_0, \Theta \le \theta_0] = \int_0^{r_0}\int_0^{\theta_0} \frac{e^{-r^2/2}}{2\pi} r dr d\theta \tag{25}$$

$$= \frac{\theta_0}{2\pi}\left(1 - e^{-r_0^2/2}\right), \quad 0 < \theta_0 < 2\pi \quad 0 < r_0 < \infty \tag{26}$$

R and Θ are independent random variables, where R has a Rayleigh dis-

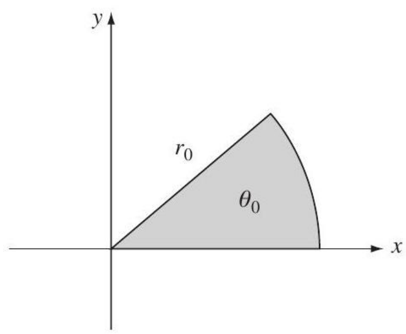

Figure 4: Region of integration R_{r_0,θ_0} in Example 16.

tribution and Θ is uniformly distributed in $(0, 2\pi)$. The joint pdf is obtained by taking partial derivatives with respect to r and θ :

$$f_{R,\Theta}(r,\theta) = \frac{\partial^2}{\partial r \partial \theta} \frac{\theta}{2\pi}\left(1 - e^{-r^2/2}\right)$$
$$= \frac{1}{2\pi}\left(re^{-r^2/2}\right), \quad 0 < \theta < 2\pi \quad 0 < r < \infty$$

This transformation maps every point in the plane from Cartesian coordinates to polar coordinates. We can also go backwards from polar to Cartesian coordinates. First we generate independent Rayleigh R and uniform Θ random variables. We then transform R and Θ into Cartesian coordinates to obtain an independent pair of zero-mean, unit-variance Gaussians. Neat!

3.1 pdf of Linear Transformations

The joint pdf of $\mathbf{Z}$ can be found directly in terms of the joint pdf of $\mathbf{X}$ by finding the equivalent events of infinitesimal rectangles. We consider the linear transformation of two random variables:

$$V = aX + bY \qquad \text{or} \qquad \begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$
$$W = cX + eY$$

Denote the above matrix by A. We will assume that A has an inverse, that is, it has determinant $|ae - bc| \neq 0$, so each point (v, w) has a unique corresponding point (x, y) obtained from

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix} \tag{27}$$

Consider the infinitesimal rectangle shown in Fig. 5. The points in this rectangle are mapped into the parallelogram shown in the figure. The infinitesimal rectangle and the parallelogram are equivalent events, so their probabilities must be equal. Thus

$$f_{X,Y}(x, y)dxdy \simeq f_{V,W}(v, w)dP$$

where dP is the area of the parallelogram. The joint pdf of V and W is thus given by

$$f_{V,W}(v, w) = \frac{f_{X,Y}(x, y)}{\left|\frac{dP}{dxdy}\right|} \tag{28}$$

where x and y are related to (v, w) by Eq. (27). Equation (28) states that the joint pdf of V and W at (v, w) is the pdf of X and Y at the corresponding point (x, y), but rescaled by the "stretch factor" $dP/dxdy$. It can be shown that $dP = (|ae - bc|)dxdy$, so the "stretch factor" is

$$\left|\frac{dP}{dxdy}\right| = \frac{|ae - bc|(dxdy)}{(dxdy)} = |ae - bc| = |A|$$

where $|A|$ is the determinant of A.

The above result can be written compactly using matrix notation. Let the vector $\mathbf{Z}$ be

$$\mathbf{Z} = A\mathbf{X}$$

where A is an $n \times n$ invertible matrix. The joint pdf of $\mathbf{Z}$ is then

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{f_{\mathbf{x}}\left(A^{-1}\mathbf{z}\right)}{|A|} \tag{29}$$

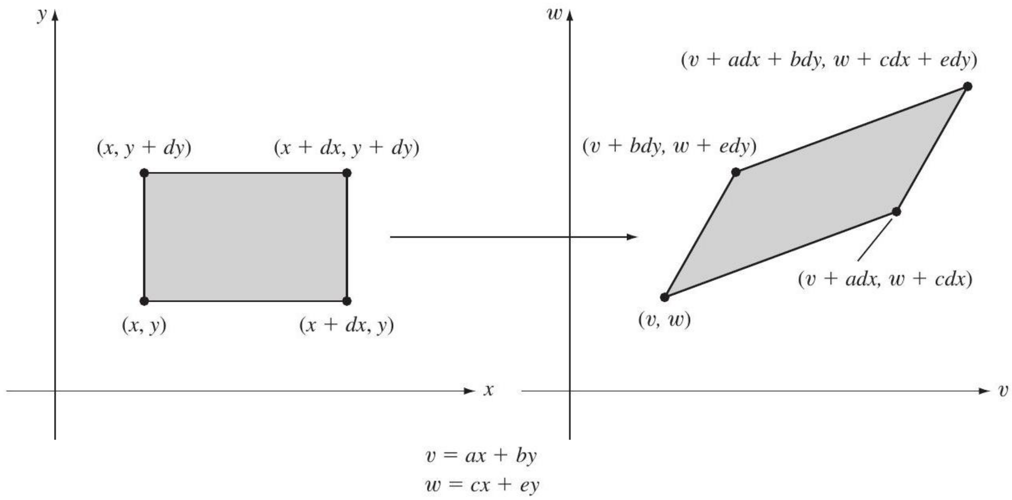

Figure 5: Image of an infinitesimal rectangle under a linear transformation.

Example 17: Linear Transformation of Jointly Gaussian Random Variables

Let X and Y be the jointly Gaussian random variables introduced in Example 4 Week 8. Let V and W be obtained from (X, Y) by

$$\begin{bmatrix} V \\ W \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}$$

Find the joint pdf of V and W.

The determinant of the matrix is $|A| = 1$, and the inverse mapping is given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} V \\ W \end{bmatrix}$$

so $X = (V - W)/\sqrt{2}$ and $Y = (V + W)/\sqrt{2}$. Therefore the pdf of V and W is

$$f_{V,W}(v, w) = f_{X,Y}\left(\frac{v - w}{\sqrt{2}}, \frac{v + w}{\sqrt{2}}\right)$$

where

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\left(x^2 - 2\rho xy + y^2\right)/2\left(1 - \rho^2\right)}$$

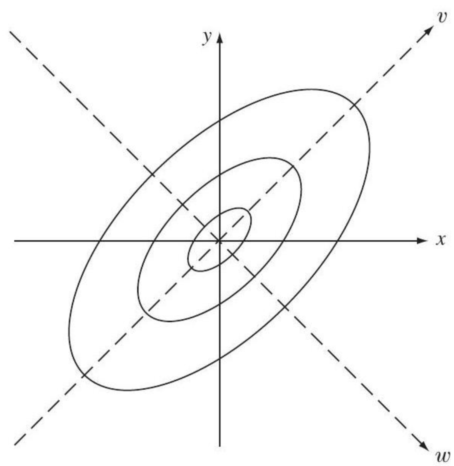

Figure 6: Contours of equal value of joint Gaussian pdf discussed in Example 5.45.

By substituting for x and y, the argument of the exponent becomes

$$\frac{(v-w)^2/2 - 2\rho(v-w)(v+w)/2 + (v+w)^2/2}{2\left(1-\rho^2\right)}$$
$$= \frac{v^2}{2(1+\rho)} + \frac{w^2}{2(1-\rho)}$$

Thus

$$f_{V,W}(v,w) = \frac{1}{2\pi\left(1-\rho^2\right)^{1/2}} e^{-\left\{\left[v^2/2(1+\rho)\right]+\left[w^2/2(1-\rho)\right]\right\}}$$

It can be seen that the transformed variables V and W are independent, zero-mean Gaussian random variables with variance $1+\rho$ and $1-\rho$, respectively. Figure 6 shows contours of equal value of the joint pdf of (X, Y). It can be seen that the pdf has elliptical symmetry about the origin with principal axes at 45° with respect to the axes of the plane. We will later show that the above linear transformation corresponds to a rotation of the coordinate system so that the axes of the plane are aligned with the axes of the ellipse.