Chapter 13

Conditional Probability Density Functions

13.1 Introduction

A discussion of conditional probability mass functions (PMFs) was given in Chapter 8. The motivation was that many problems are stated in a conditional format so that the solution must naturally accommodate this conditional structure. In addition, the use of conditioning is useful for simplifying probability calculations when two random variables are statistically dependent. In this chapter we formulate the analogous approach for probability density functions (PDFs). A potential stumbling block is that the usual conditioning event $X = x$ has probability zero for a continuous random variable. As a result the conditional PMF cannot be extended in a straightforward manner. We will see, however, that using care, a conditional PDF can be defined and will prove to be useful.

13.2 Summary

The conditional PDF is defined in (13.3) and can be used to find conditional probabilities using (13.4). The conditional PDF for a standard bivariate Gaussian PDF is given by (13.5) and is seen to retain its Gaussian form. The joint, conditional, and marginal PDFs are related to each other as summarized by Properties 13.1–13.5. A conditional CDF is defined by (13.6) and is evaluated using (13.7). The use of conditioning can simplify probability calculations as described in Section 13.5. A version of the law of total probability is given by (13.12) and is evaluated using (13.13). An optimal predictor for the outcome of a random variable based on the outcome of a second random variable is given by the mean of the conditional PDF as defined by (13.14). An example is given for the bivariate Gaussian PDF in which the predictor becomes linear (actually affine). To generate realizations of two jointly distributed continuous random variables the procedure based on conditioning and described in Section 13.7 can be used. Lastly, an application to determining mortality rates for retirement planning is described in Section 13.8.

13.3 Conditional PDF

Recall that for two jointly *discrete* random variables X and Y, the conditional PMF is defined as

$$p_{Y|X}[y_j|x_i] = \frac{p_{X,Y}[x_i, y_j]}{p_X[x_i]} \qquad j = 1, 2, \ldots . \tag{13.1}$$

This formula gives the probability of the event $Y = y_j$ for $j = 1, 2, \ldots$ once we have observed that $X = x_i$. Since $X = x_i$ has occurred, the only joint events with a nonzero probability are $\{(x, y) : x = x_i, y = y_1, y_2, \ldots\}$. As a result we divide the joint probability $p_{X,Y}[x_i, y_j] = P[X = x_i, Y = y_j]$ by the probability of the reduced sample space, which is $p_X[x_i] = P[X = x_i, Y = y_1] + P[X = x_i, Y = y_2] + \cdots = \sum_{j=1}^{\infty} p_{X,Y}[x_i, y_j]$. This division assures us that

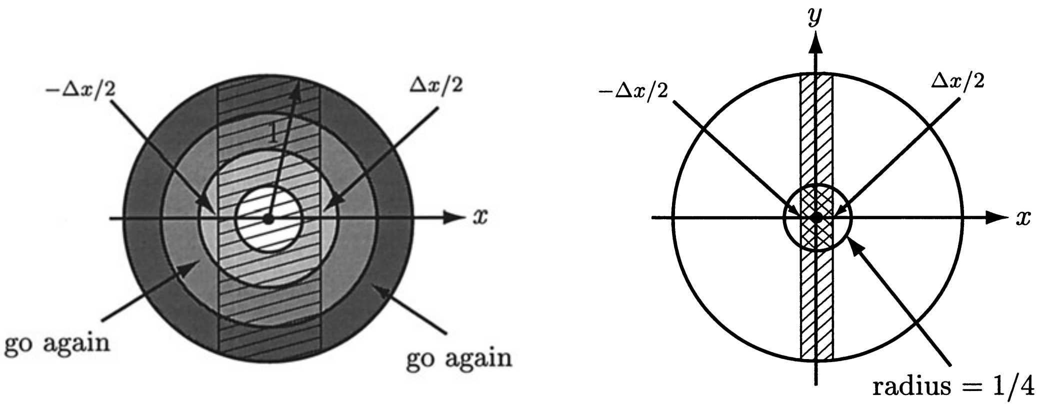
$$\begin{aligned}
\sum_{j=1}^{\infty} p_{Y|X}[y_j|x_i] &= \sum_{j=1}^{\infty} \frac{p_{X,Y}[x_i, y_j]}{p_X[x_i]} \\
&= \frac{\sum_{j=1}^{\infty} p_{X,Y}[x_i, y_j]}{p_X[x_i]} \\
&= \frac{\sum_{j=1}^{\infty} p_{X,Y}[x_i, y_j]}{\sum_{j=1}^{\infty} p_{X,Y}[x_i, y_j]} = 1.
\end{aligned}$$

In the case of *continuous* random variables X and Y a problem arises in defining a *conditional PDF*. If we observe $X = x$, then since $P[X = x] = 0$, the use of a formula like (13.1) is no longer valid due to the division by zero. Recall that our original definition of the conditional probability is

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

which is undefined if $P[B] = 0$. How should we then proceed to extend (13.1) for continuous random variables?

We will motivate a viable approach using the example of the circular dartboard described in Section 12.3. In particular, we consider a revised version of the dart throwing contest. Referring to Figure 12.2 the champion dart player realizes that the novice presents little challenge. To make the game more interesting the champion proposes the following modification. If the novice player's dart lands outside the region $|x| \leq \Delta x/2$, then the novice player gets to go again. He continues until his dart lands within the region $|x| \leq \Delta x/2$ as shown cross-hatched in Figure 13.1a. The novice dart player even gets to pick the value of Δx. Hence, he reasons that it



(a) Dartboard (b) Sample space

Figure 13.1: Revised dart throwing game. Only dart outcomes in the cross-hatched region are counted.

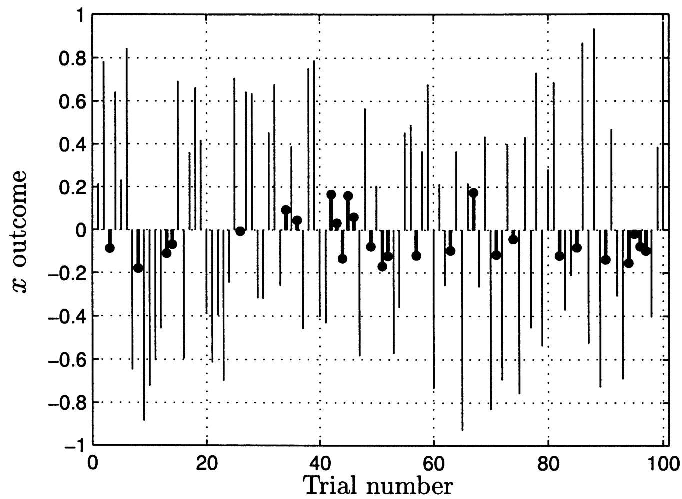
should be small to exclude regions of the dart board that are outside the bullseye circle. As a result, he chooses a Δx as shown in Figure 13.1b, which allows him to continue throwing darts until one lands within the cross-hatched region. The champion, however, has taken a course in probability and so is not worried. In fact, in Problem 12.30 the probability of the champion's dart landing in the bullseye area was shown to be 0.8646. To find the probability of the novice player obtaining a bullseye, we recall that his dart is equally likely to land anywhere on the dartboard. Hence, using conditional probability we have that

$$P[\text{bullseye}| -\Delta x/2 \leq X \leq \Delta x/2] = \frac{P[\text{bullseye}, -\Delta x/2 \leq X \leq \Delta x/2]}{P[-\Delta x/2 \leq X \leq \Delta x/2]}.$$

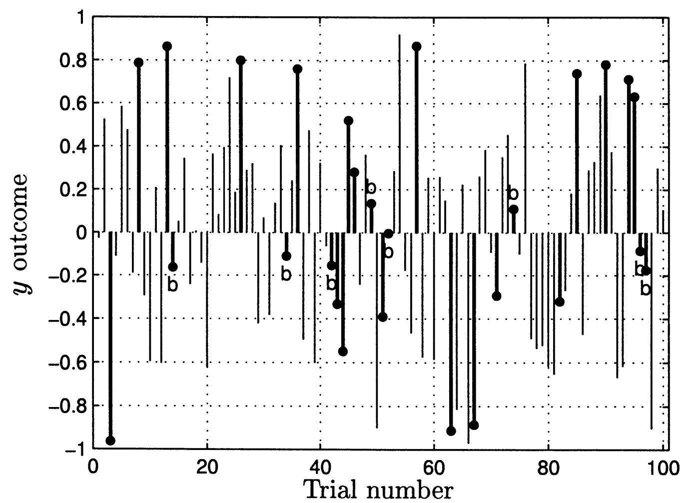
Since $\Delta x/2$ is small, we can assume that it is much less than $1/4$ as shown in Figure 13.1b. Therefore, we have that the cross-hatched regions can be approximated by rectangles and so

$$\begin{aligned} & P[\text{bullseye}| -\Delta x/2 \leq X \leq \Delta x/2] \\ = & \frac{P[\text{double cross-hatched region}]}{P[\text{double cross-hatched region}] + P[\text{single cross-hatched region}]} \\ = & \frac{\Delta x(1/2)/\pi}{\Delta x(2)/\pi} \quad \text{(probability = rectangle area/dartboard area)} \\ = & 0.25 < 0.865. \end{aligned} \tag{13.2}$$

Hence, the revised strategy will still allow the champion to have a higher probability of winning for any Δx, no matter how small it is chosen. Even though $P[X = 0] = 0$, the conditional probability is well defined even as $\Delta x \to 0$ (but not equal 0). Some typical outcomes of this game are shown in Figure 13.2, where it is assumed the novice player has chosen $\Delta x/2 = 0.2$. In Figure 13.2a are shown the outcomes of X



(a) All x outcomes – those with $|x| \leq 0.2$ are shown as dark lines



(b) Dark lines are y outcomes for which $|x| \leq 0.2$, b indicates a bullseye ($\sqrt{x^2+y^2} \leq 1/4$) for the outcomes with $|x| \leq 0.2$

Figure 13.2: Revised dart throwing game outcomes.

for the novice player. Only those for which $|x| \leq \Delta x/2 = 0.2$, which are shown as the darker lines, are kept. In Figure 13.2b the outcomes of Y are shown with the kept outcomes shown as the dark lines. Those outcomes that resulted in a bullseye are shown with a "b" over them. Note that there were 27 out of 100 outcomes that had $|x|$ values less than or equal to 0.2 (see Figure 13.2a), and of these, 8 outcomes resulted in a bullseye (see Figure 13.2b). Hence, the estimated probability of landing in either the single or double cross-hatched region of Figure 13.1b is $27/100 = 0.27$, while the theoretical probability is approximately $\Delta x(2)/\pi = 0.4(2)/\pi = 0.254$. Also, the estimated conditional probability of a bullseye is from Figure 13.2b, $8/27 = 0.30$ while from (13.2) the theoretical probability is approximately equal to 0.25. (The approximations are due to the use of rectangular approximations to the cross-hatched regions, which only become exact as $\Delta x \to 0$.) We will use the same strategy to define a conditional PDF. Let $A = \{(x,y) : \sqrt{x^2+y^2} \leq 1/4\}$, which is the bullseye region. Then

$$
\begin{aligned}
&P[A|\,|X| \leq \Delta x/2] \\
&= \frac{P[A, |X| \leq \Delta x/2]}{P[|X| \leq \Delta x/2]} && \text{(definition of cond. prob.)} \\
&= \frac{P[\{(x,y) : |x| \leq \Delta x/2, |y| \leq \sqrt{1/16 - x^2}\}]}{P[\{(x,y) : |x| \leq \Delta x/2, |y| \leq 1\}]} && \left(\frac{\text{double cross-hatched area}}{\text{cross-hatched area}}\right) \\
&= \frac{P[\{(x,y) : |x| \leq \Delta x/2, |y| \leq \sqrt{1/16 - x^2}\}]}{P[\{x : |x| \leq \Delta x/2\}]} \\
&= \frac{\int_{-\Delta x/2}^{\Delta x/2} \int_{-\sqrt{1/16-x^2}}^{\sqrt{1/16-x^2}} p_{X,Y}(x,y)\,dy\,dx}{\int_{-\Delta x/2}^{\Delta x/2} p_X(x)dx}.
\end{aligned}
$$

As $\Delta x \to 0$, we can write

$$
\begin{aligned}
&P[A|\,|X| \leq \Delta x/2] \\
&\approx \frac{\int_{-\Delta x/2}^{\Delta x/2} \int_{-1/4}^{1/4} p_{X,Y}(x,y)\,dy\,dx}{\int_{-\Delta x/2}^{\Delta x/2} p_X(x)dx} && \text{(since } \sqrt{1/16 - x^2} \approx 1/4 \text{ for } |x| \leq \Delta x/2) \\
&\approx \frac{\int_{-1/4}^{1/4} p_{X,Y}(0,y)\Delta x dy}{p_X(0)\Delta x} && \text{(since } p_{X,Y}(x,y) \approx p_{X,Y}(0,y) \text{ for } |x| \leq \Delta x/2) \\
&= \int_{-1/4}^{1/4} \frac{p_{X,Y}(0,y)}{p_X(0)} dy.
\end{aligned}
$$

We now define $p_{X,Y}/p_X$ as the conditional PDF

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}. \tag{13.3}$$

Note that it is well defined as long as $p_X(x) \neq 0$. Thus, as $\Delta x \to 0$

$$P[A|\,|X| \leq \Delta x/2] = \int_{-1/4}^{1/4} p_{Y|X}(y|0)dy.$$

More generally, the conditional PDF allows us to compute probabilities as (see Problem 13.6)

$$P[a \leq Y \leq b|x - \Delta x/2 \leq X \leq x + \Delta x/2] = \int_a^b p_{Y|X}(y|x)dy.$$

This probability is usually written as

$$P[a \leq Y \leq b|X = x]$$

but the conditioning event should be understood to be $\{x : x-\Delta x/2 \leq X \leq x+\Delta x/2\}$ for Δx small. Finally, with this understanding, we have that

$$P[a \leq Y \leq b|X = x] = \int_a^b p_{Y|X}(y|x)dy \tag{13.4}$$

where $p_{Y|X}$ is defined by (13.3) and is termed the *conditional PDF*. The conditional PDF $p_{Y|X}(y|x)$ is the probability per unit length of Y when $X = x$ (actually $x - \Delta x/2 \leq X \leq x + \Delta x/2$) is observed. Since it is found using (13.3), it is seen to be a function of y *and* x. It should be thought of as a *family* of PDFs with y as the independent variable, and with a different PDF for each value x. An example follows.

Example 13.1 – Standard bivariate Gaussian PDF

Assume that (X, Y) have the joint PDF

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < y < \infty \end{array}$$

and note that the marginal PDF for X is given by

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right].$$

The conditional PDF is found from (13.3) as

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]}{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right]} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2}Q\right) \end{aligned}$$

where

$$\begin{aligned} Q &= \frac{x^2 - 2\rho xy + y^2}{1-\rho^2} - \frac{(1-\rho^2)x^2}{1-\rho^2} \\ &= \frac{y^2 - 2\rho xy + \rho^2 x^2}{1-\rho^2} \\ &= \frac{(y-\rho x)^2}{1-\rho^2}. \end{aligned}$$

As a result we have that the conditional PDF is

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(y-\rho x)^2\right] \tag{13.5}$$

and is seen to be Gaussian. This result, although of great importance, is not true in general. The form of the PDF usually changes from $p_Y(y)$ to $p_{Y|X}(y|x)$. We will denote this conditional PDF in shorthand notation as $Y|(X=x) \sim \mathcal{N}(\rho x, 1-\rho^2)$. As expected, the conditional PDF depends on x, and in particular the mean of the conditional PDF is a function of x. It is a valid PDF in that *for each x value*, it is nonnegative and integrates to 1 over $-\infty < y < \infty$. These properties are true in general. In effect, the conditional PDF depends on the outcome of X so that we use a different PDF for each X outcome. For example, if $\rho = 0.9$ and we observe $X = -1$, then to compute $P[-1 \le Y \le -0.8|X=-1]$ and $P[-0.1 \le Y \le 0.1|X=-1]$, we first observe from (13.5) that $Y|(X=-1) \sim \mathcal{N}(-0.9, 0.19)$. Then

$$\begin{aligned} P[-1 \le Y \le -0.8|X=-1] &= Q\left(\frac{-1-(-0.9)}{\sqrt{0.19}}\right) - Q\left(\frac{-0.8-(-0.9)}{\sqrt{0.19}}\right) = 0.1815 \\ P[-0.1 \le Y \le 0.1|X=-1] &= Q\left(\frac{-0.1-(-0.9)}{\sqrt{0.19}}\right) - Q\left(\frac{0.1-(-0.9)}{\sqrt{0.19}}\right) = 0.0223. \end{aligned}$$

Can you explain the difference between these values? (See Figure 13.3b where the dark lines indicate $y = 0$ and $y = -0.9$.) In Figure 13.3b the cross-section of the joint PDF is shown. Once the cross-section is normalized so that it integrates to one, it becomes the conditional PDF $p_{Y|X}(y|-1)$. This is easily verified since

$$\begin{aligned} p_{Y|X}(y|-1) &= \frac{p_{X,Y}(-1,y)}{p_X(-1)} \\ &= \frac{p_{X,Y}(-1,y)}{\int_{-\infty}^{\infty} p_{X,Y}(-1,y)dy}. \end{aligned}$$

$\diamondsuit$

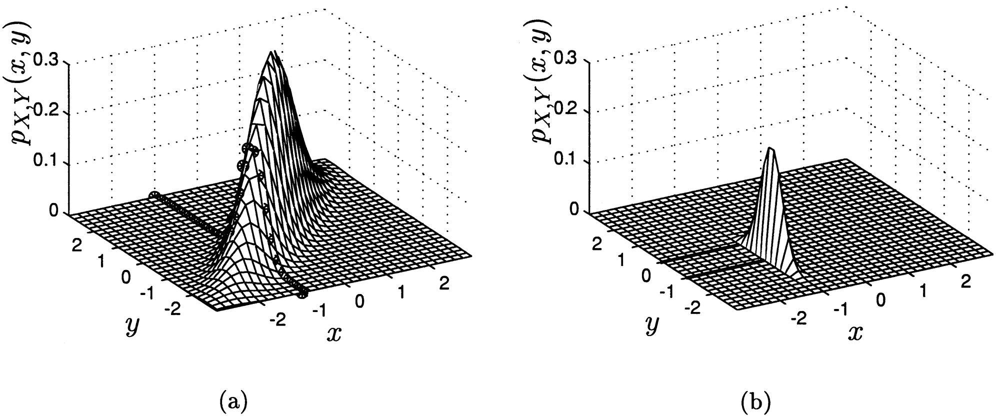


Figure 13.3: Standard bivariate Gaussian PDF and its cross-section at $x = -1$. The *normalized* cross-section is the conditional PDF.

13.4 Joint, Conditional, and Marginal PDFs

The relationships between the joint, conditional, and marginal PMFs as described in Section 8.4 also hold for the corresponding PDFs. Hence, we just summarize the properties and leave the proofs to the reader (see Problem 13.11).

Property 13.1 – Joint PDF yields conditional PDFs.

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{p_{X,Y}(x,y)}{\int_{-\infty}^{\infty} p_{X,Y}(x,y)dy} \\ p_{X|Y}(x|y) &= \frac{p_{X,Y}(x,y)}{\int_{-\infty}^{\infty} p_{X,Y}(x,y)dx} \end{aligned}$$

□

Property 13.2 – Conditional PDFs are related.

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$$

□

Property 13.3 – Conditional PDF is expressible using Bayes' rule.

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)p_Y(y)}{\int_{-\infty}^{\infty} p_{X|Y}(x|y)p_Y(y)dy}$$

□

Property 13.4 – Conditional PDF and its corresponding marginal PDF yields the joint PDF

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x) = p_{X|Y}(x|y)p_Y(y)$$

□

Property 13.5 – Conditional PDF and its corresponding marginal PDF yields the other marginal PDF

$$p_Y(y) = \int_{-\infty}^{\infty} p_{Y|X}(y|x)p_X(x)dx$$

□

A conditional CDF can also be defined. Based on (13.4) we have upon letting $a = -\infty$ and $b = y$

$$P[Y \leq y|X = x] = \int_{-\infty}^{y} p_{Y|X}(t|x)dt.$$

As a result the conditional CDF is defined as

$$F_{Y|X}(y|x) = P[Y \leq y|X = x] \tag{13.6}$$

and is evaluated using

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} p_{Y|X}(t|x)dt. \tag{13.7}$$

As an example, if $Y|(X = x) \sim \mathcal{N}(\rho x, 1 - \rho^2)$ as was shown in Example 13.1, we have that

$$F_{Y|X}(y|x) = 1 - Q\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right). \tag{13.8}$$

Finally, as previously mentioned in Chapter 12 two continuous random variables X and Y are independent if and only if the joint PDF factors as $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ or equivalently if the joint CDF factors as $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

This is consistent with our definition of the conditional PDF since if X and Y are independent

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{p_{X,Y}(x,y)}{p_X(x)} \\ &= \frac{p_X(x)p_Y(y)}{p_X(x)} = p_Y(y) \end{aligned} \tag{13.9}$$

(and similarly $p_{X|Y} = p_X$). Hence, the conditional PDF no longer depends on the observed value of X, i.e., x. This means that the knowledge that $X = x$ has occurred does not affect the PDF of Y (and thus does not affect the probability of events defined on $\mathcal{S}_Y$). Similarly, from (13.7), if X and Y are independent

$$\begin{aligned} F_{Y|X}(y|x) &= \int_{-\infty}^{y} p_{Y|X}(t|x)dt \\ &= \int_{-\infty}^{y} p_Y(t)dt \quad \text{(from (13.9))} \\ &= F_Y(y). \end{aligned}$$

An example would be if $\rho = 0$ for the standard bivariate Gaussian PDF. Then since $Y|(X = x) \sim \mathcal{N}(\rho x, 1 - \rho^2) = \mathcal{N}(0, 1)$, we have that $p_{Y|X}(y|x) = p_Y(y)$. Also, from (13.8)

$$\begin{aligned} F_{Y|X}(y|x) &= 1 - Q\left(\frac{y - \rho x}{\sqrt{1 - \rho^2}}\right) \\ &= 1 - Q(y) = F_Y(y). \end{aligned}$$

Another example follows.

Example 13.2 – Lifetime PDF of spare lightbulb

A professor uses the overhead projector for his class. The time to failure of a new bulb X has the exponential PDF $p_X(x) = \lambda \exp(-\lambda x)u(x)$, where x is in hours. A new spare bulb also has a time to failure Y that is modeled as an exponential PDF. However, the time to failure of the spare bulb depends upon how long the spare bulb sits unused. Assuming the spare bulb is activated as soon as the original bulb fails, the time to activation is given by X. As a result, the *expected* time to failure of the spare bulb is decreased as

$$\frac{1}{\lambda_Y} = \frac{1}{\lambda(1 + \alpha x)}$$

where $0 < \alpha < 1$ is some factor that indicates the degradation of the unused bulb with storage time. The expected time to failure of the spare bulb decreases as the

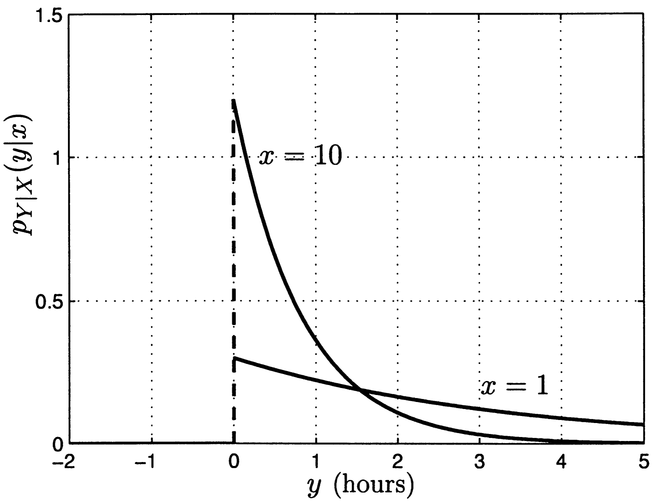


Figure 13.4: Conditional PDF for lifetime of spare bulb. Dependence is on time to failure x of original bulb.

original bulb is used longer (and hence the spare bulb must sit unused longer). Thus, we model the time to failure of the spare bulb as

$$\begin{aligned} p_{Y|X}(y|x) &= \lambda_Y \exp(-\lambda_Y y)u(y) \\ &= \lambda(1+\alpha x)\exp\left[-\lambda(1+\alpha x)y\right]u(y). \end{aligned}$$

This conditional PDF is shown in Figure 13.4 for $1/\lambda = 5$ hours and $\alpha = 0.5$. We now wish to determine the *unconditional PDF* of the time to failure of the spare bulb which is $p_Y(y)$. It is expected that the probability of failure of the spare bulb will increase than if the spare bulb were used rightaway or for $x = 0$. Note that if $x = 0$, then $p_{Y|X} = p_X$, which says that the spare bulb will fail with the same PDF as the original bulb. Using Property 13.5 we have

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} p_{Y|X}(y|x)p_X(x)dx \\ &= \int_0^{\infty} \lambda(1+\alpha x)\exp\left[-\lambda(1+\alpha x)y\right]\lambda\exp(-\lambda x)dx \\ &= \lambda^2\exp(-\lambda y)\int_0^{\infty}(1+\alpha x)\exp\left[-\lambda(\alpha y+1)x\right]dx \\ &= \lambda^2\exp(-\lambda y)\left[\int_0^{\infty}\exp(ax)dx + \alpha\int_0^{\infty}x\exp(ax)dx\right] \quad (\text{let } a = -\lambda(1+\alpha y)) \\ &= \lambda^2\exp(-\lambda y)\left[\left.\frac{\exp(ax)}{a}\right|_0^{\infty} + \alpha\left.\left(\frac{1}{a}x\exp(ax) - \frac{1}{a^2}\exp(ax)\right)\right|_0^{\infty}\right] \\ &= \lambda^2\exp(-\lambda y)\left[-\frac{1}{a} + \frac{\alpha}{a^2}\right] \\ &= \lambda^2\exp(-\lambda y)\left[\frac{1}{\lambda(\alpha y+1)} + \frac{\alpha}{[\lambda(\alpha y+1)]^2}\right]. \end{aligned}$$

or finally

$$p_Y(y) = \lambda^2 \exp(-\lambda y) \left[\frac{1}{\lambda(\alpha y + 1)} + \frac{\alpha}{[\lambda(\alpha y + 1)]^2} \right] u(y).$$

This is shown in Figure 13.5 for $1/\lambda = 5$ hours and $\alpha = 0.5$ along with the PDF $p_X(x)$ of the time to failure of the original bulb. As expected the probability of the

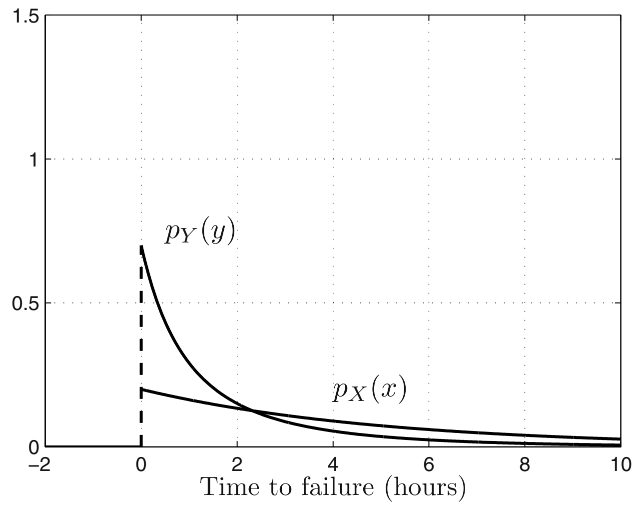


Figure 13.5: PDFs for time to failure of original bulb X and spare bulb Y.

spare bulb failing before 2 hours is greatly increased.

$\diamondsuit$

Finally, note that the conditional PDF is obtained by differentiating the conditional CDF. From (13.7) we have

$$p_{Y|X}(y|x) = \frac{\partial F_{Y|X}(y|x)}{\partial y}. \tag{13.10}$$

13.5 Simplifying Probability Calculations Using Conditioning

Following Section 8.5 we can easily find the PDF of $Z = g(X, Y)$ if X and Y are independent by using conditioning. We shall not repeat the argument other than to summarize the results and give an example. The procedure is

1. Fix $X = x$ and let $Z|(X = x) = g(x, Y)$.

2. Find the PDF of $Z|(X = x)$ using the standard approach for a transformation of a single random variable from Y to Z.

3. Uncondition the conditional PDF to yield the desired PDF $p_Z(z)$.

Example 13.3 – PDF for ratio of independent random variables

Consider the function $Z = Y/X$ where X and Y are independent random variables. In Problem 12.39 we asserted that

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(xz)|x|dx.$$

We now derive this using the aforementioned approach. First recall that if $Z = aY$ for a a constant, then $p_Z(z) = p_Y(z/a)/|a|$ (see Example 10.5). Now we have that

$$Z|(X = x) = \left.\frac{Y}{X}\right| (X = x) = \left.\frac{Y}{x}\right| (X = x)$$

so that with $a = 1/x$ and noting the independence of X and Y, we have

$$p_{Z|X}(z|x) = p_{Y|X}(zx)|x| = p_Y(zx)|x| \tag{13.11}$$

and thus

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} p_{Z,X}(z, x)dx && \text{(marginal PDF from joint PDF)} \\ &= \int_{-\infty}^{\infty} p_{Z|X}(z|x) p_X(x)dx && \text{(definition of conditional PDF)} \\ &= \int_{-\infty}^{\infty} p_Y(zx)|x| p_X(x)dx && \text{(from (13.11))} \\ &= \int_{-\infty}^{\infty} p_X(x) p_Y(xz)|x|dx. \end{aligned}$$

Note that without the independence assumption, we could *not* assert that $p_{Y|X} = p_Y$ in (13.11).

◇

In general to compute probabilities of events it is advantageous to use conditioning arguments whether or not X and Y are independent. The analogous result to (8.28) is (see Problem 13.15)

$$P[Y \in A] = \int_{-\infty}^{\infty} P[Y \in A|X = x] p_X(x)dx. \tag{13.12}$$

This is another form of the theorem of *total probability.* It can also be written as

$$P[Y \in A] = \int_{-\infty}^{\infty} \left[\int_A p_{Y|X}(y|x)dy \right] p_X(x)dx \tag{13.13}$$

where we have used (13.4) and replaced $\{y : a \leq y \leq b\}$ by the more general set A. The formula of (13.13) is analogous to (8.27) for discrete random variables. An example follows.

Example 13.4 – Probability of error for a digital communication system
Consider the PSK communication system shown in Figure 2.14. The probability of error was shown in Section 10.6 to be

$$P_e = P[W \leq -A/2] = Q(A/2)$$

since the noise sample $W \sim \mathcal{N}(0,1)$. In a wireless communication system such as is used in cellular telephone, the received amplitude A varies with time due to multipath propagation [Rappaport 2002]. As a result, it is usually modeled as a Rayleigh random variable whose PDF is

$$p_A(a) = \begin{cases} \frac{a}{\sigma_A^2} \exp\left(-\frac{1}{2\sigma_A^2}a^2\right) & a \geq 0 \\ 0 & a < 0. \end{cases}$$

We wish to determine the probability of error if A is a Rayleigh random variable. Thus, we need to evaluate $P[W + A/2 \leq 0]$ if $W \sim \mathcal{N}(0,1)$, A is a Rayleigh random variable, and we assume W and A are independent. A straightforward approach is to first find the PDF of $Z = W + A/2$, and then to integrate $p_Z(z)$ from $-\infty$ to 0. Alternatively, it is simpler to use (13.12) as follows.

$$\begin{aligned} P_e &= P[W \leq -A/2] \\ &= \int_{-\infty}^{\infty} P[W \leq -A/2|A = a]p_A(a)da \quad \text{(from (13.12))} \\ &= \int_{-\infty}^{\infty} P[W \leq -a/2|A = a]p_A(a)da \quad \text{(since } A = a \text{ has occurred).} \end{aligned}$$

But since W and A are independent, $P[W \leq -a/2|A = a] = P[W \leq -a/2]$ and thus

$$P_e = \int_{-\infty}^{\infty} P[W \leq -a/2]p_A(a)da.$$

Using $P[W \leq -a/2] = Q(a/2)$ we have

$$P_e = \int_0^{\infty} Q(a/2)\frac{a}{\sigma_A^2} \exp\left(-\frac{1}{2\sigma_A^2}a^2\right) da.$$

Unfortunately, this is not easily evaluated in closed form.

◊

13.6 Mean of Conditional PDF

For a conditional PDF the mean is given by the usual mean definition except that the PDF now depends on x. We therefore have the definition

$$E_{Y|X}[Y|x] = \int_{-\infty}^{\infty} y p_{Y|X}(y|x)dy \tag{13.14}$$

which is analogous to (8.29) for discrete random variables. We also expect and it follows that (see Problem 13.19 and also the discussion in Section 8.6)

$$E_X[E_{Y|X}[Y|X]] = E_Y[Y] \tag{13.15}$$

where $E_{Y|X}[Y|X]$ is given by (13.14) except that the value x is now replaced by the random variable X. Therefore, $E_{Y|X}[Y|X]$ is viewed as a function of the random variable X. As an example, we saw that for the bivariate Gaussian PDF $Y|(X = x) \sim \mathcal{N}(\rho x, 1 - \rho^2)$. Hence, $E_{Y|X}[Y|x] = \rho x$, but regarding the mean of the conditional PDF as a function of the random variable X we have that $E_{Y|X}[Y|X] = \rho X$. To see that (13.15) holds for this example

$$E_X[E_{Y|X}[Y|X]] = E_X[\rho X] = \rho E_X[X] = 0$$

since the marginal PDF of X for the standard bivariate Gaussian PDF was shown to be $\mathcal{N}(0,1)$. Also, since $Y \sim \mathcal{N}(0,1)$ for the standard bivariate Gaussian PDF, $E_Y[Y] = 0$, and we see that (13.15) is satisfied.

The mean of the conditional PDF arises in optimal prediction, where it is proven that the minimum mean square error (MMSE) prediction of Y given $X = x$ has been observed is $E_{Y|X}[Y|x]$ (see Problem 13.17). This is optimal over all predictors, linear and *nonlinear*. For the standard Gaussian PDF, however, the optimal prediction turns out to be linear since $E_{Y|X}[Y|x] = \rho x$. More generally, it can be shown that if X and Y are jointly Gaussian with PDF given by (12.35), then

$$\begin{aligned}
E_{Y|X}[Y|x] &= E_Y[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(x - E_X[X]) \\
&= \mu_Y + \frac{\rho\sigma_X\sigma_Y}{\sigma_X^2}(x - \mu_X) \\
&= \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X).
\end{aligned}$$

(See also Problem 13.20.)

13.7 Computer Simulation of Jointly Continuous Random Variables

In a manner similar to that described in Section 8.7 we can generate realizations of a continuous random vector (X, Y) using the relationship

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x).$$

(Of course, if X and Y are independent, we can generate X based on $p_X(x)$ and Y based on $p_Y(y)$). Consider as an example the standard bivariate Gaussian PDF. We know that $Y|(X = x) \sim \mathcal{N}(\rho x, 1 - \rho^2)$ and $X \sim \mathcal{N}(0,1)$. Hence, we can generate realizations of (X, Y) as follows.

Step 1. Generate $X = x$ according to $\mathcal{N}(0,1)$.

Step 2. Generate $Y|(X = x)$ according to $\mathcal{N}(\rho x, 1-\rho^2)$.

This procedure is conceptually simpler than what we implemented in Section 12.11 and much more general. There we used (12.53). Referring to (12.53), if we let $\mu_W = \mu_Z = 0$, $\sigma_W^2 = \sigma_Z^2 = 1$ and make the replacements of W, Z, X, Y with X, Y, U, V, we have

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \tag{13.16}$$

where $U \sim \mathcal{N}(0,1)$, $V \sim \mathcal{N}(0,1)$, and U and V are independent. The transformation of (13.16) can be used to generate realizations of a standard bivariate Gaussian random vector. It is interesting to note that in this special case the two procedures for generating bivariate Gaussian random vectors lead to the identical algorithm. Can you see why from (13.16)?

As an example of the conditional PDF approach, if we let $\rho = 0.9$, we have the plot shown in Figure 13.6. It should be compared with Figure 12.21 (note that in Figure 12.21 the means of X and Y are 1). The MATLAB code used to generate

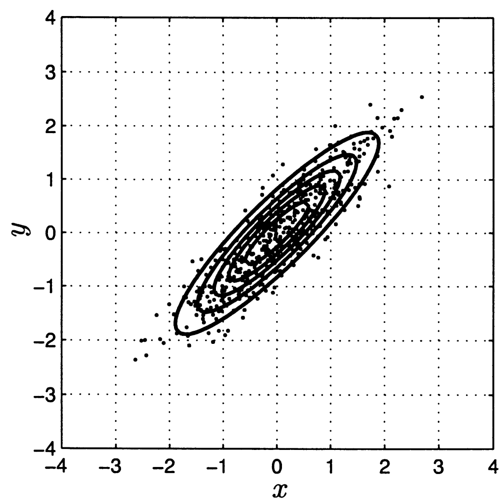


Figure 13.6: 500 outcomes of standard bivariate Gaussian random vector with $\rho = 0.9$ generated using conditional PDF approach.

realizations of a standard bivariate Gaussian random vector using conditioning is given below.

```
randn('state',0) % set random number generator to initial value
rho=0.9;
M=500; % set number of realizations to generate
for m=1:M
  x(m,1)=randn(1,1); % generate realization of N(0,1) random
                     % variable (Step 1)
```

```
  ygx(m,1)=rho*x(m)+sqrt(1-rho^2)*randn(1,1); % generate
                                               % Y|(X=x) (Step 2)
end
```

13.8 Real-World Example – Retirement Planning

Professor Staff, who teaches too many courses a semester, plans to retire at age 65. He will have accumulated a total of $500,000 in a retirement account and wishes to use the money to live on during his retirement years. He assumes that his money will earn enough to offset the decrease in value due to inflation. Hence, if he lives to age 75 he could spend $50,000 a year and if he lives to age 85, then he could spend only $25,000 a year. How much should he figure on spending per year?

Besides the many courses Professor Staff has taught in history, English, mathematics, and computer science, he has also taught a course in probability. He therefore reasons that if he spends s dollars a year and lives for Y years during his retirement, then the probability that $500,000 - sY < 0$ should be small. Here sY is the total money spent during his retirement. In other words, he desires

$$P[500,000 - sY < 0] = 0.5. \tag{13.17}$$

He chooses 0.5 for the probability of outliving his retirement fund. This acknowledges the fact that choosing a lower probability will lead to an overly conservative approach and a small amount of expendable funds per year as we will see shortly. Equivalently, he requires that

$$P\left[Y > \frac{500,000}{s}\right] = 0.5. \tag{13.18}$$

As an example, if he spends $s = 50,000$ per year, then the probability he lives more than $500,000/s = 10$ years should be 0.5.

It should now be obvious that (13.18) is actually the right-tail probability or *complementary CDF* of the years lived in retirement. This type of information is of great interest not only to retirees but also to insurance companies who pay annuities. An annuity is a payment that an insurance company pays annually to an investor for the remainder of his/her life. The amount of the payment depends upon how much the investor originally invests, the age of the investor, and the insurance company's belief that the investor will live for so many years. To quantify answers to questions concerning years of life remaining, the *mortality rate*, which is the distribution of years lived past a given age is required. If Y is a continuous random variable that denotes the years lived past age $X = x$, then the mortality rate can be described by first defining the *conditional CDF*

$$F_{Y|X}(y|x) = P[Y \leq y|X = x].$$

For example, the probability that a person will live at least 10 more years if he is currently 65 years old is given by

$$P[Y > 10|X = 65] = 1 - F_{Y|X}(10|65)$$

which is the complementary CDF or the *right-tail probability* of the conditional PDF $p_{Y|X}(y|x)$. It has been shown that for Canadian citizens the conditional CDF is well modeled by [Milevsky and Robinson 2000]

$$F_{Y|X}(y|x) = 1 - \exp\left[\exp\left(\frac{x-m}{l}\right)\left(1 - \exp\left(\frac{y}{l}\right)\right)\right] \qquad y \geq 0 \tag{13.19}$$

where $m = 81.95, l = 10.6$ for males and $m = 87.8, l = 9.5$ for females. As an example, if $F_{Y|X}(y|x) = 0.5$, then you have a 50% chance of living more than y years if you are currently x years old. In other words, 50% of the population who are x years old will live more than y years and 50% will live less than y years. The number of years y is the *median* number of years to live. (Recall that the median is the value at which the probability of being less than or equal to this value is 0.5.) From (13.19) this will be true when

$$0.5 = \exp\left[\exp\left(\frac{x-m}{l}\right)\left(1 - \exp\left(\frac{y}{l}\right)\right)\right]$$

which results in the remaining number of years lived by 50% of the population who are currently x years old as

$$y = l\ln\left[1 - \exp\left(-\left(\frac{x-m}{l}\right)\right)\ln 0.5\right]. \tag{13.20}$$

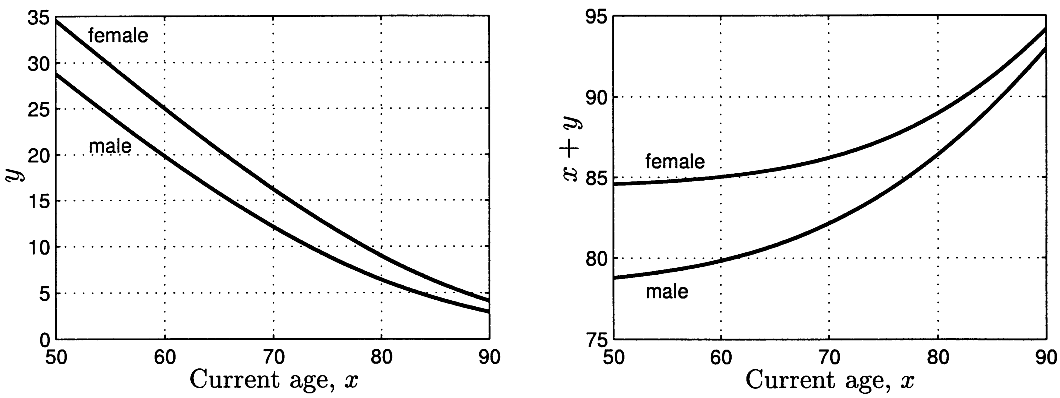
This is plotted in Figure 13.7a versus the current age x for males and females. In Figure 13.7a the median number of years left is shown while in Figure 13.7b the median life expectancy (which is $x + y$) is given.

Returning to Professor Staff, he can now determine how much money he can afford to spend each year. Since the probability of outliving one's retirement funds is a conditional probability based on current age, we rewrite (13.18) as

$$P\left[Y > \frac{500,000}{s}\middle| X = x\right] = P_L$$

where we allow the probability to be denoted in general by P_L. Since he will retire at age $x = 65$, we have from (13.19) that he will live more than y years with a probability of P_L given by

$$P_L = \exp\left[\exp\left(\frac{65-m}{l}\right)\left(1 - \exp\left(\frac{y}{l}\right)\right)\right]. \tag{13.21}$$



(a) Years to live, y

(b) Life expectancy, $x + y$

Figure 13.7: Mortality rates.

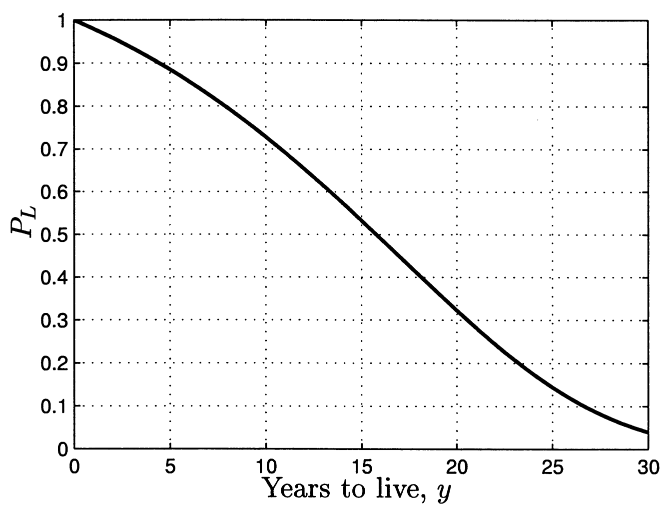


Figure 13.8: Probability P_L of exceeding y years in retirement for male who retires at age 65.

Assuming Professor Staff is a male, we use $m = 81.95, l = 10.6$ in (13.21) to produce a plot P_L versus y as shown in Figure 13.8. If the professor is overly conservative, he may want to assure himself that the probability of outliving his retirement fund is only about 0.1. Then, he should plan on living another 27 years, which means that his yearly expenses should not exceed $\$500,000/27 = \$18,500$. If he is less conservative and chooses a probability of 0.5, then he can plan on living about 15 years. Then his yearly expenses should not exceed $\$500,000/15 \approx \$33,000$.

References

Milevsky, M.A., C. Robinson, "Self-Annuitization and Ruin in Retirement," *North American Actuarial Journal*, Vol. 4, pp. 113–129, 2000.

Rappaport, T.S., *Wireless Communications, Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, 2002.

Problems

13.1 (w,c) In this problem we simulate on a computer the dartboard outcomes of the novice player for the game shown in Figure 13.1a. To do so, generate two independent $\mathcal{U}(-1,1)$ random variables to serve as the x and y outcomes. Keep only the outcomes (x,y) for which $\sqrt{x^2+y^2} \leq 1$ (see Problem 13.23 for why this produces a uniform joint PDF within the unit circle). Then, of the kept outcomes retain only the ones for which $\Delta x/2 \leq 0.2$ (see Figure 13.2a). Finally, estimate the probability that the novice player obtains a bullseye and compare it to the theoretical value. Note that the theoretical value of 0.25 as given by (13.2) is actually an approximation based on the areas in Figure 13.1b being rectangular.

13.2 (⌣̈) (w) Determine if the proposed conditional PDF

$$p_{Y|X}(y|x) = \begin{cases} c\exp(-y/x) & y \geq 0, x > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a valid conditional PDF for some c. If so, find the required value of c.

13.3 (w) Is the proposed conditional PDF

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y-x)^2\right] \qquad -\infty < y < \infty, -\infty < x < \infty$$

valid? If so, and if $X \sim \mathcal{N}(0,1)$, design an experiment that will produce the random variables X and Y.

13.4 (⌣̈) (f) If

$$p_{X,Y}(x,y) = \begin{cases} 2\exp[-(x+y)] & 0 \leq y \leq x, x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

find $p_{Y|X}(y|x)$.

13.5 (w) Plot the joint PDF

$$p_{X,Y}(x,y) = \begin{cases} 2x & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Next determine by inspection the conditional PDF $p_{Y|X}(y|x)$. Recall that the conditional PDF is just the normalized cross-section of the joint PDF.

13.6 (t) In this problem we show that

$$\lim_{\Delta x \to 0} P[a \le Y \le b | x - \Delta x/2 \le X \le x + \Delta x/2] = \int_a^b p_{Y|X}(y|x)dy.$$

To do so first show that

$$\lim_{\Delta x \to 0} P[a \le Y \le b | x - \Delta x/2 \le X \le x + \Delta x/2]$$

$$= \int_a^b \lim_{\Delta x \to 0} \frac{\int_{x-\Delta x/2}^{x+\Delta x/2} p_{X,Y}(x,y)dx}{\int_{x-\Delta x/2}^{x+\Delta x/2} p_X(x)dx} dy.$$

13.7 (f) Determine $P[Y > \frac{1}{2}|X = 0]$ if the joint PDF is given as

$$p_{X,Y}(x,y) = \begin{cases} 2x & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

13.8 (⌣) (f) If $X \sim \mathcal{U}(0,1)$ and $Y|(X = x) \sim \mathcal{U}(0,x)$, find the joint PDF for X and Y and also the marginal PDF for Y.

13.9 (f,t) For the standard bivariate Gaussian PDF find the conditional PDFs $p_{Y|X}$ and $p_{X|Y}$ and compare them. Explain your results. Are your results true in general?

13.10 (⌣) (f) If the joint PDF $p_{X,Y}$ is uniform over the region $0 < y < x$ and $0 < x < 1$ and zero otherwise, find the conditional PDFs $p_{Y|X}$ and $p_{X|Y}$.

13.11 (t) Prove Properties P13.1–13.5.

13.12 (f) Determine the PDF of Y/X if $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ and X and Y are independent. Do so by using the conditioning approach.

13.13 (t) Prove that the PDF of $X+Y$, where X and Y are independent is given as a convolution integral (see (12.14)). Do so by using the conditioning approach.

13.14 (⌣) (w) A game of darts is played using the linear dartboard shown in Figure 3.8. If two novice players throw darts at the board and each one's dart is equally likely to land anywhere in the interval $(-1/2, 1/2)$, prove that the probability of player 2 winning is $1/2$. Hint: Let X_1 and X_2 be the outcomes and use $Y = |X_2| - |X_1|$ and $X = X_1$ in (13.12).

13.15 (t) Prove (13.12) by starting with (13.4).

13.16 (⌣) **(w)** A resistor is chosen from a bin of 10 ohm resistors whose distribution satisfies $R \sim \mathcal{N}(10, 0.25)$. A $i = 1$ amp current source is applied to the resistor and the subsequent voltage V is measured with a voltmeter. The voltmeter has an error E that is modeled as $E \sim \mathcal{N}(0,1)$. Find the probability that $V > 10$ volts if an 11 ohm resistor is chosen. Note that $V = iR + E$. What assumption do you need to make about the dependence between R and E?

13.17 (t) In this problem we prove that the minimum mean square error estimate of Y based on $X = x$ is given by $E_{Y|X}[Y|x]$. First let the estimate be denoted by $\hat{Y}(x)$ since it will depend in general on the outcome of X. Then note that the mean square error is

$$\begin{aligned}
\text{mse} &= E_{X,Y}[(Y - \hat{Y}(X))^2] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \hat{Y}(x))^2 p_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \hat{Y}(x))^2 p_{Y|X}(y|x) p_X(x)\,dx\,dy \\
&= \int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} (y - \hat{Y}(x))^2 p_{Y|X}(y|x)\,dy\right]}_{J(\hat{Y}(x))} p_X(x)\,dx.
\end{aligned}$$

Now we can minimize $J(\hat{Y}(x))$ for each value of x since $p_X(x) \geq 0$. Complete the derivation by differentiating $J(\hat{Y}(x))$ and setting the result equal to zero. Consider $\hat{Y}(x)$ as a constant (since x is assumed fixed inside the inner integral) in doing so. Finally justify all the steps in the derivation.

13.18 (f) For the joint PDF given in Problem 13.10 find the minimum mean square error estimate of Y given $X = x$. Plot the region in the x-y plane for which the joint PDF is nonzero and also the estimated value of Y versus x.

13.19 (t) Prove (13.15).

13.20 (w,c) If a bivariate Gaussian PDF has a mean vector $[\mu_X\, \mu_Y]^T = [1\,2]^T$ and a covariance matrix

$$\mathbf{C} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

plot the contours of constant PDF. Next find the minimum mean square error prediction of Y given $X = x$ and plot it on top of the contour plot. Explain the significance of the plot.

13.21 (⌣) **(w)** A random variable X has a Laplacian PDF with variance σ^2. If the variance is chosen according to $\sigma^2 \sim \mathcal{U}(0,1)$, what is average variance of the random variable?

13.22 (c) In this problem we use a computer simulation to illustrate the known result that $E_{Y|X}[Y|x] = \rho x$ for (X, Y) distributed according to a standard bivariate Gaussian PDF. Using (13.16) generate $M = 10,000$ realizations of a standard bivariate Gaussian random vector with $\rho = 0.9$. Then let $A = \{x : x_0 - \Delta x/2 \leq x \leq x_0 + \Delta x/2\}$ and discard the realizations for which x is not in A. Finally, estimate the mean of the conditional PDF by taking the sample mean of the remaining realizations. Choose $\Delta x/2 = 0.1$ and $x_0 = 1$ and compare the theoretical value of $E_{Y|X}[Y|x]$ to the estimated value based on your computer simulation.

13.23 (t) We now prove that the procedure described in Problem 13.1 will produce a random vector (X, Y) that is uniformly distributed within the unit circle. First consider the polar equivalent of (X, Y), which is (R, Θ), so that the conditional CDF is given by

$$P[R \leq r, \Theta \leq \theta | R \leq 1] \qquad 0 \leq r \leq 1, 0 \leq \theta < 2\pi.$$

But this is equal to

$$\frac{P[R \leq r, R \leq 1, \Theta \leq \theta]}{P[R \leq 1]} = \frac{P[R \leq r, \Theta \leq \theta]}{P[R \leq 1]}.$$

(Why?) Next show that

$$P[R \leq r, \Theta \leq \theta | R \leq 1] = \frac{\theta r^2}{2\pi}$$

and differentiate with respect to r and then θ to find the joint PDF $p_{R,\Theta}(r, \theta)$ (which is actually a conditional joint PDF due to the conditioning on the value of R being $r \leq 1$). Finally, transform this PDF back to that of (X, Y) to verify that it is uniform within the unit circle. Hint: You will need the result

$$\det\left(\frac{\partial(r, \theta)}{\partial(x, y)}\right) = \frac{1}{\det\left(\frac{\partial(x,y)}{\partial(r,\theta)}\right)}.$$

13.24 (⌣) (f,c) For the conditional CDF of years left to live given current age, which is given by (13.19), find the conditional PDF. Plot the conditional PDF for a Canadian male who is currently 50 years old and also for one who is 75 years old. Next find the average life span for each of these individuals. Hint: You will need to use a computer evaluation of the integral for the last part.

13.25 (t) Verify that the conditional CDF given by (13.19) is a valid CDF.