11. Conditional Density Functions and Conditional Expected Values

As we have seen in section 4 conditional probability density functions are useful to update the information about an event based on the knowledge about some other related event (refer to example 4.7). In this section, we shall analyze the situation where the related event happens to be a random variable that is dependent on the one of interest.

From (4-11), recall that the distribution function of X given an event B is

$$F_X(x \mid B) = P\big(X(\xi) \le x \mid B\big) = \frac{P\big((X(\xi) \le x) \cap B\big)}{P(B)}. \qquad (11\text{-}1)$$

Suppose, we let

$$B = \{ y_1 < Y(\xi) \le y_2 \}. \tag{11-2}$$

Substituting (11-2) into (11-1), we get

$$F_X(x \mid y_1 < Y \le y_2) = \frac{P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big)}{P(y_1 < Y(\xi) \le y_2)}$$

$$= \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)}, \tag{11-3}$$

where we have made use of (7-4). But using (3-28) and (7-7) we can rewrite (11-3) as

$$F_X(x \mid y_1 < Y \le y_2) = \frac{\int_{-\infty}^{x} \int_{y_1}^{y_2} f_{XY}(u, v)\,du\,dv}{\int_{y_1}^{y_2} f_Y(v)\,dv}. \tag{11-4}$$

To determine, the limiting case $F_X(x \mid Y = y)$, we can let $y_1 = y$ and $y_2 = y + \Delta y$ in (11-4).

This gives

$$F_X(x \mid y < Y \le y+\Delta y) = \frac{\int_{-\infty}^{x}\int_{y}^{y+\Delta y} f_{XY}(u,v)\,du\,dv}{\int_{y}^{y+\Delta y} f_Y(v)\,dv} \approx \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du\;\Delta y}{f_Y(y)\,\Delta y} \tag{11-5}$$

and hence in the limit

$$F_X(x \mid Y = y) = \lim_{\Delta y \to 0} F_X(x \mid y < Y \le y+\Delta y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \tag{11-6}$$

(To remind about the conditional nature on the left hand side, we shall use the subscript $X \mid Y$ (instead of X) there). Thus

$$F_{X|Y}(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \tag{11-7}$$

Differentiating (11-7) with respect to x using (8-7), we get

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)}. \tag{11-8}$$

It is easy to see that the left side of (11-8) represents a valid probability density function. In fact

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)} \geq 0 \tag{11-9}$$

and

$$\int_{-\infty}^{+\infty} f_{X|Y}(x \mid Y = y) dx = \frac{\int_{-\infty}^{+\infty} f_{XY}(x, y) dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1, \tag{11-10}$$

where we have made use of (7-14). From (11-9) - (11-10), (11-8) indeed represents a valid p.d.f, and we shall refer to it as the conditional p.d.f of the r.v X given $Y = y$. We may also write

$$f_{X|Y}(x \mid Y = y) = f_{X|Y}(x \mid y). \tag{11-11}$$

From (11-8) and (11-11), we have

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}, \tag{11-12}$$

and similarly

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)}. \qquad (11\text{-}13)$$

If the r.vs X and Y are independent, then $f_{XY}(x, y) = f_X(x) f_Y(y)$ and (11-12) - (11-13) reduces to

$$f_{X|Y}(x \mid y) = f_X(x), \quad f_{Y|X}(y \mid x) = f_Y(y), \qquad (11\text{-}14)$$

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

In the case of discrete-type r.vs, (11-12) reduces to

$$P\left(X = x_i \mid Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}. \qquad (11\text{-}15)$$

Next we shall illustrate the method of obtaining conditional p.d.fs through an example.

Example 11.1: Given

$$f_{XY}(x,y)=\begin{cases} k, & 0<x<y<1, \\ 0, & \text{otherwise} \end{cases}, \qquad (11\text{-}16)$$

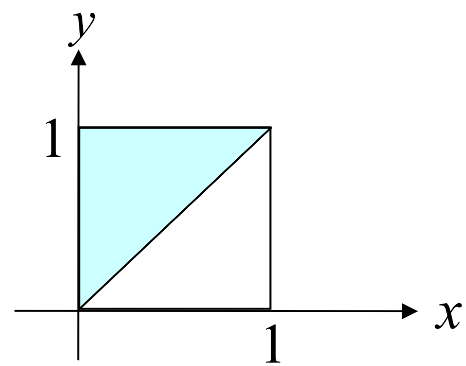


Fig. 11.1

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

Solution: The joint p.d.f is given to be a constant in the shaded region. This gives

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

Similarly

$$f_X(x)=\int f_{XY}(x,y)dy = \int_x^1 k\; dy = k\,(1-x), \quad 0<x<1, \qquad (11\text{-}17)$$

and

$$f_Y(y)=\int f_{XY}(x,y)dx = \int_0^y k\; dx = k\; y, \quad 0<y<1. \qquad (11\text{-}18)$$

From (11-16) - (11-18), we get

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{y}, \quad 0 < x < y < 1, \tag{11-19}$$

and

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{1-x}, \quad 0 < x < y < 1. \tag{11-20}$$

We can use (11-12) - (11-13) to derive an important result. From there, we also have

$$f_{XY}(x,y) = f_{X|Y}(x \mid y) f_Y(y) = f_{Y|X}(y \mid x) f_X(x) \tag{11-21}$$

or

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)}. \tag{11-22}$$

But

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y)dy \tag{11-23}$$

and using (11-23) in (11-22), we get

$$f_{YX}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y) dy}. \tag{24}$$

Equation (11-24) represents the p.d.f version of Bayes' theorem. To appreciate the full significance of (11-24), one need to look at communication problems where observations can be used to update our knowledge about unknown parameters. We shall illustrate this using a simple example.

Example 11.2: An unknown random phase θ is uniformly distributed in the interval $(0, 2\pi)$, and $r = \theta + n$, where $n \sim N(0, \sigma^2)$. Determine $f(\theta \mid r)$.

Solution: Initially almost nothing about the r.v θ is known, so that we assume its a-priori p.d.f to be uniform in the interval $(0, 2\pi)$.

In the equation $r = \theta + n,$ we can think of n as the noise contribution and r as the observation. It is reasonable to assume that θ and n are independent. In that case

$$f(r \mid \boldsymbol{\theta} = \theta) \sim N(\theta, \sigma^2) \tag{11-25}$$

since it is given that $\boldsymbol{\theta} = \theta$ is a constant, $r = \theta + n$ behaves like n. Using (11-24), this gives the a-posteriori p.d.f of θ given r to be (see Fig. 11.2 (b))

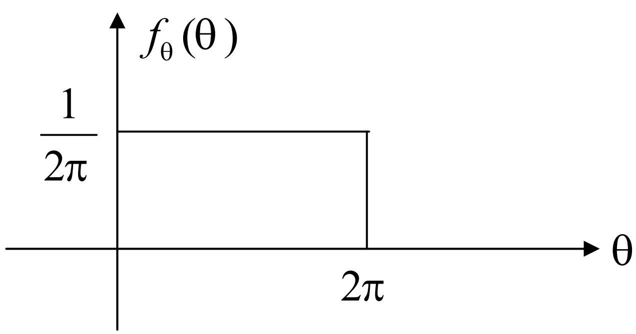
$$f(\theta \mid r) = \frac{f(r \mid \theta) f_\theta(\theta)}{\int_0^{2\pi} f(r \mid \theta) f_\theta(\theta) d\theta} = \frac{e^{-(r-\theta)^2/2\sigma^2}}{\frac{1}{2\pi}\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}$$

$$= \varphi(r) e^{-(\theta - r)^2/2\sigma^2}, \qquad 0 < \theta < 2\pi, \tag{11-26}$$

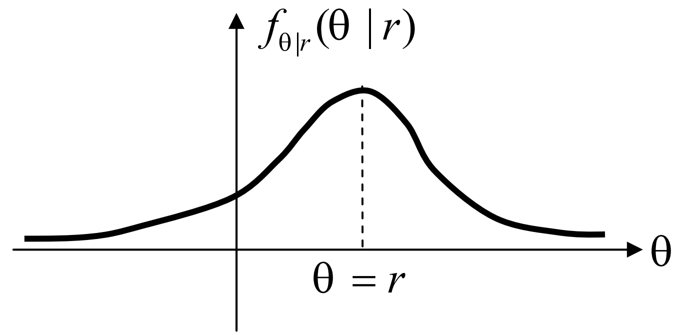
where

$$\varphi(r) = \frac{2\pi}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}.$$

Notice that the knowledge about the observation r is reflected in the a-posteriori p.d.f of θ in Fig. 11.2 (b). It is no longer flat as the a-priori p.d.f in Fig. 11.2 (a), and it shows higher probabilities in the neighborhood of $\theta = r$.



(a) a-priori p.d.f of θ



(b) a-posteriori p.d.f of θ

Fig. 11.2

Conditional Mean:

We can use the conditional p.d.fs to define the conditional mean. More generally, applying (6-13) to conditional p.d.fs we get

$$E\left(g(X)\,|\,B\right)=\int_{-\infty}^{+\infty} g(x) f_X(x\,|\,B)dx. \qquad (11\text{-}27)$$

and using a limiting argument as in (11-2) - (11-8), we get

$$\mu_{X|Y} = E\left(X\,|\,Y=y\right)=\int_{-\infty}^{+\infty} x\, f_{X|Y}(x\,|\,y)\,dx \qquad (11\text{-}28)$$

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

$$\mu_{Y|X} = E\left(Y\,|\,X=x\right)=\int_{-\infty}^{+\infty} y\, f_{Y|X}(y\,|\,x)\;dy. \qquad (11\text{-}29)$$

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

$$Var(X\,|\,Y)=\sigma^2_{X|Y} = E\left(X^2\,|\,Y=y\right)-\left(E(X\,|\,Y=y)\right)^2$$

$$= E\left((X-\mu_{X|Y})^2\,|\,Y=y\right). \qquad (11\text{-}30)$$

we shall illustrate these calculations through an example.

Example 11.3: Let

$$f_{XY}(x,y)=\begin{cases}1, & 0<|y|<x<1,\\ 0, & \text{otherwise}\ .\end{cases} \qquad (11\text{-}31)$$

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

Solution: As Fig. 11.3 shows, $f_{XY}(x,y)=1$ in the shaded area, and zero elsewhere.

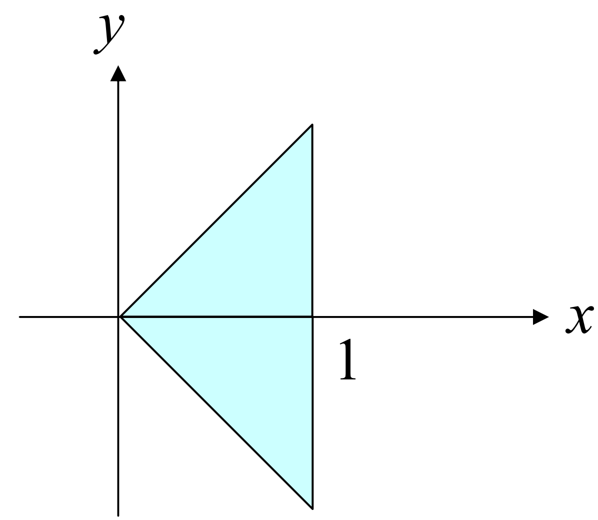


Fig. 11.3

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

$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{1}{1-|y|}, \quad 0<|y|<x<1, \qquad (11\text{-}32)$$

and

$$f_{Y|X}(y\,|\,x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{1}{2x}, \quad 0<|y|<x<1. \qquad (11\text{-}33)$$

Hence $E(X \mid Y) = \int x \, f_{X|Y}(x \mid y) dx = \int_{|y|}^{1} \frac{x}{(1-|y|)} dx$

$$= \frac{1}{(1-|y|)} \left. \frac{x^2}{2} \right|_{|y|}^{1} = \frac{1-|y|^2}{2(1-|y|)} = \frac{1+|y|}{2}, \quad |y|<1. \qquad (11\text{-}34)$$

$$E(Y \mid X) = \int y f_{Y|X}(y \mid x) dy = \int_{-x}^{x} \frac{y}{2x} dy = \frac{1}{2x} \left. \frac{y^2}{2} \right|_{-x}^{x} = 0, \quad 0<x<1. \qquad (11\text{-}35)$$

It is possible to obtain an interesting generalization of the conditional mean formulas in (11-28) - (11-29). More generally, (11-28) gives

$$E\left(g(X) \mid Y=y\right) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y) dx \ . \qquad (11\text{-}36)$$

But

$$E\left(g(X)\right) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx = \int_{-\infty}^{+\infty} g(x) \int_{-\infty}^{+\infty} f_{XY}(x,y) dy dx$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x) f_{XY}(x,y) dx dy = \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y) dx}_{E(g(X)|Y=y)} \, f_Y(y) dy$$

$$= \int_{-\infty}^{+\infty} E\left(g(X) \mid Y=y\right) f_Y(y) dy = E\left\{E\left(g(X) \mid Y=y\right)\right\} . \qquad (11\text{-}37)$$

Obviously, in the right side of (11-37), the inner expectation is with respect to X and the outer expectation is with respect to Y. Letting $g(X) = X$ in (11-37) we get the interesting identity

$$E(X) = E\{E(X \mid Y = y)\}, \tag{11-38}$$

where the inner expectation on the right side is with respect to X and the outer one is with respect to Y. Similarly, we have

$$E(Y) = E\{E(Y \mid X = x)\}. \tag{11-39}$$

Using (11-37) and (11-30), we also obtain

$$Var(X) = E\big(Var(X \mid Y = y)\big). \tag{11-40}$$

Conditional mean turns out to be an important concept in estimation and prediction theory. For example given an observation about a r.v X, what can we say about a related r.v Y? In other words what is the best predicted value of Y given that $X = x$? It turns out that if "best" is meant in the sense of minimizing the mean square error between Y and its estimate $\hat{y}$, then the conditional mean of Y given $X = x$, i.e., $E(Y \mid X = x)$ is the best estimate for Y (see Lecture 16 for more on Mean Square Estimation).

We conclude this lecture with yet another application of the conditional density formulation.

Example 11.4 : **Poisson sum of Bernoulli random variables**

Let $X_i, \quad i = 1, 2, 3, \cdots$ 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

$$Y = \sum_{i=1}^{N} X_i, \qquad Z = N - Y. \tag{11-41}$$

Show that Y and Z are independent Poisson random variables.

Solution : To determine the joint probability mass function of Y and Z, consider

$$\begin{aligned} P(Y = m,\ Z = n) &= P(Y = m,\ N - Y = n) = P(Y = m,\ N = m + n) \\ &= P(Y = m \mid N = m + n) P(N = m + n) \\ &= P(\sum_{i=1}^{N} X_i = m \mid N = m + n)\, P(N = m + n) \\ &= P(\sum_{i=1}^{m+n} X_i = m) P(N = m + n) \qquad (11\text{-}42) \end{aligned}$$

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

$$= \left(\frac{(m+n)!}{m!n!} p^m q^n \right) \left(e^{-\lambda} \frac{\lambda^{m+n}}{(m+n)!} \right)$$

$$= \left(e^{-p\lambda} \frac{(p\lambda)^m}{m!} \right) \left(e^{-q\lambda} \frac{(q\lambda)^n}{n!} \right)$$

$$= P(Y = m)P(Z = n). \tag{11-43}$$

Thus

$$Y \sim P(p\lambda) \quad \text{and} \quad Z \sim P(q\lambda) \tag{11-44}$$

and Y and Z 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 chicks that survive also forms a Poisson random variable with parameter $p\lambda$.