Conditional probability If X and Y are jointly distributed random variables then the conditional probability density for X given Y is

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Similarly, $$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Bayes' theorem. From the above equations we obtain

$$f(y|x) = \frac{f(x|y)\,f_Y(y)}{f_X(x)}$$

Law of total probability: Recall $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, then

$$f_X(x) = \int_{-\infty}^{\infty} f(x|y)\,f_Y(y)\,dy$$

Similarly, $$f_Y(y) = \int_{-\infty}^{\infty} f(y|x)\,f_X(x)\,dx \Rightarrow f(y|x) = \frac{f(x|y)\,f(y)}{\int_{-\infty}^{\infty} f(x|y)\,f_Y(y)\,dy}$$

Observe that the joint density completely specifies the marginal densities and the conditional densities

Application. Suppose a signal X is perturbed by additive noise N, so we observe $Y = X + N$. We want to find the conditional density of X given the observed Y, because the most probable values of $X|Y$ may be a good guess of the true value of X

From Bayes' theorem $$f(x|y) = \frac{f(y|x)\,f_X(x)}{f_Y(y)}$$

If we assume a fixed value for X then the only randomness on Y is N which has density $f_N(n)$. Thus, since $N = Y - X$ $f(y|x) = f_N(n) = f_N(y-x)$

Then $f(x|y) = \dfrac{f_N(y-x) f_X(x)}{f_Y(y)}$

As an specific example consider $X \sim Exp(b) \Rightarrow f_X(x) = b\exp(-bx)$
and $N \sim N(0, \sigma_N^2) \Rightarrow f_N(n) = \dfrac{1}{\sqrt{2\pi}\,\sigma_N} \exp\left(-\dfrac{n^2}{2\sigma_N^2}\right)$

Such process might arise as a signal from a space probe in wich the time intervals between counts of high-energy particles are converted to voltage amplitudes for porpouses of transmission back to Earth Then

$$f(x|y) = \frac{1}{f_Y(y)} \left\{ \frac{1}{\sqrt{2\pi}\,\sigma_N} \exp\left[-\frac{(y-x)^2}{2\sigma_N^2} \right] b\exp(-bx) \right\}$$

$$= \frac{b}{f_Y(y)\sqrt{2\pi}\,\sigma_N} \exp\left\{ -\frac{1}{2\sigma_N^2} \left[x^2 - 2(y - b\sigma_N^2)x + y^2 \right] \right\}$$

As we are interested in finding $\max_x \{f(x|y)\}$ we don't need to evaluate $f_Y(y)$ since it's only a constant for an observed y. The function $f(x|y)$ reaches its maximum where the argument for the exponential reaches its minimum, thus

$$\max_x \{f(x|y)\} = \min_x \{ x^2 - 2(y - b\sigma_N^2)x + y^2 \}$$

Differentiating and equating to zero we get: $2x - 2(y - b\sigma_N^2) = 0$
$\Rightarrow x = y - b\sigma_N^2$ which is a maximum provided $y - b\sigma_N^2 > 0$. In this case, if y_1 is observed, a good estimate $\hat{X}$ of X is $\hat{X} = y_1 - b\sigma_N^2$. If $y_1 - b\sigma_N^2 < 0$ the appropiate estimate for X is $\hat{X} = 0$. Note that if $\sigma_N^2 \to 0$ then $X \to y_1$

Independence. Random variables X an Y are said to be independent if $F(x,y) = F_X(x) F_Y(y)$

Equivalently, their joint density also factors because

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 [F_X(x) F_Y(y)]}{\partial x \partial y} = \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y} F_X(x) F_Y(y)\right] = \frac{\partial}{\partial x} F_X(x) \frac{\partial}{\partial y} F_Y(y)$$

$$= f_X(x) f_Y(y)$$

Later on we'll see that a particularly important function for which we will want to compute the expected value is $g(X,Y) = XY$. If X and Y are independent then

$$E[g(X,Y)] = E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y)\, dx dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\, dx dy$$

$$= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy = E[X] E[Y]$$

The conditional distributions also simplify

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

and similarly $f(y|x) = f_Y(y)$

The intuitive idea behind this result is that knowledge about one of the random variables gives no information about the value of the other. The result is used more often the other way around: when we know that $X_1, X_2, \dots, X_n$ are independent random variables we can construct their joint density by multiplying their marginal densities.

Covariance and Correlation

Definition. Covariance If X and Y are jointly distributed random variables with expectation μ_X and μ_Y respectively, then the covariance of X and Y is

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

which can be simplified as follows:

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E[XY] - E[X]\mu_Y - E[Y]\mu_X + \mu_X\mu_Y \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

In particular $Cov(X,Y) = 0$ if X and Y are independent (the converse is not always true, normality is needed).

Properties

$Cov(aW+bX, cY+dZ) = ac\,Cov(W,Y) + bc\,Cov(X,Y) + ad\,Cov(W,Z) + bd\,Cov(X,Z)$

$Var(X) = Cov(X,X) \implies Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$Var(X+Y) = Var(X) + Var(Y)$ if X and Y are independent