Chapter 11

Multiple Continuous Random Variables

Being versed at dealing with multiple random variables is an essential part of statistics, engineering and science. This is equally true for models based on discrete and continuous random variables. In this chapter, we focus on the latter and expand our exposition of continuous random variables to random vectors. Again, our initial survey of this topic revolves around conditional distributions and pairs of random variables. More complex scenarios will be considered in the later parts of the chapter.

11.1 Joint Cumulative Distributions

Let X and Y be two random variables associated with a same experiment. The *joint cumulative distribution function* of X and Y is defined by

$$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y) \quad x, y \in \mathbb{R}.$$

Keeping in mind that X and Y are real-valued functions acting on a same sample space, we can also write

$$F_{X,Y}(x, y) = \Pr\left(\{\omega \in \Omega | X(\omega) \leq x, Y(\omega) \leq y\}\right).$$

From this characterization, we can identify a few properties of the joint CDF;

$$\lim_{y\uparrow\infty} F_{X,Y}(x,y) = \lim_{y\uparrow\infty} \Pr\left(\{\omega \in \Omega | X(\omega) \leq x, Y(\omega) \leq y\}\right)$$
$$= \Pr\left(\{\omega \in \Omega | X(\omega) \leq x, Y(\omega) \in \mathbb{R}\}\right)$$
$$= \Pr\left(\{\omega \in \Omega | X(\omega) \leq x\}\right) = F_X(x).$$

Similarly, we have $\lim_{x\uparrow\infty} F_{X,Y}(x,y) = F_Y(y)$. Taking limits in the other direction, we get

$$\lim_{x\downarrow-\infty} F_{X,Y}(x,y) = \lim_{y\downarrow-\infty} F_{X,Y}(x,y) = 0.$$

When the function $F_{X,Y}(\cdot,\cdot)$ is totally differentiable, it is possible to define the *joint probability density function* of X and Y,

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial y \partial x}(x,y) \quad x,y \in \mathbb{R}. \qquad (11.1)$$

Hereafter, we refer to a pair of random variables as continuous if the corresponding joint PDF exists and is defined unambiguously through (11.1). When this is the case, standard calculus asserts that the following equation holds,

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(\xi,\zeta) d\zeta d\xi.$$

From its definition, we note that $f_{X,Y}(\cdot,\cdot)$ is a nonnegative function which integrates to one,

$$\iint_{\mathbb{R}^2} f_{X,Y}(\xi,\zeta) d\zeta d\xi = 1.$$

Furthermore, for any admissible set $S \subset \mathbb{R}^2$, the probability that $(X,Y) \in S$ can be evaluated through the integral formula

$$\Pr((X,Y) \in S) = \iint_{\mathbb{R}^2} \mathbf{1}_S(\xi,\zeta) f_{X,Y}(\xi,\zeta) d\zeta d\xi$$
$$= \iint_{S} f_{X,Y}(\xi,\zeta) d\zeta d\xi. \qquad (11.2)$$

In particular, if S is the cartesian product of two intervals,

$$S = \left\{(x,y) \in \mathbb{R}^2 \big| a \leq x \leq b, c \leq y \leq d\right\},$$

then the probability that $(X, Y) \in S$ reduces to the typical integral form

$$\Pr((X,Y) \in S) = \Pr(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(\xi, \zeta)d\zeta d\xi.$$

Example 94. *Suppose that the random pair (X, Y) is uniformly distributed over the unit circle. We can express the joint PDF $f_{X,Y}(\cdot, \cdot)$ as*

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \le 1 \\ 0 & otherwise. \end{cases}$$

We wish to find the probability that the point (X, Y) lies inside a circle of radius $1/2$.

Let $S = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \le 0.5\}$. The probability that (X, Y) belongs to S is given by

$$\Pr((X,Y) \in S) = \iint_{\mathbb{R}^2} \frac{\mathbf{1}_S(\xi, \zeta)}{\pi} d\xi d\zeta = \frac{1}{4}.$$

Thus, the probability that (X, Y) is contained within a circle of radius half is one fourth.

Example 95. *Let X and Y be two independent zero-mean Gaussian random variables, each with variance σ^2. For $(x, y) \in \mathbb{R}^2$, their joint PDF is given by*

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

We wish to find the probability that (X, Y) falls within a circle of radius r centered at the origin.

We can compute this probability using integral formula (11.2) applied to this particular problem. Let $R = \sqrt{X^2 + Y^2}$ and assume $r > 0$, then

$$\Pr(R \le r) = \iint_{R \le r} f_{X,Y}(x, y)dxdy = \iint_{R \le r} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} (x, y)dxdy$$

$$= \int_0^r \int_0^{2\pi} \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} r d\theta dr = 1 - e^{-\frac{r^2}{2\sigma^2}}.$$

The probability that (X, Y) is contained within a circle of radius r is $1 - e^{-\frac{r^2}{2\sigma^2}}$. Recognizing that R is a continuous random variable, we can write its PDF as

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad r \ge 0.$$

From this equation, we gather That R possesses a Rayleigh distribution with parameter σ^2.

11.2   Conditional Probability Distributions

Given non-vanishing event A, we can write the conditional CDF of random variable X as

$$F_{X|A}(x) = \Pr(X \le x | A) = \frac{\Pr(\{X \le x\} \cap A)}{\Pr(A)} \quad x \in \mathbb{R}.$$

Note that event A can be defined in terms of variables X and Y. For instance, we may use $A = \{Y \ge X\}$ as our condition. Under suitable conditions, it is equally straightforward to specify the conditional PDF of X given A,

$$f_{X|A}(x) = \frac{dF_{X|A}}{dx}(x) \quad x \in \mathbb{R}.$$

Example 96. *Let X and Y be continuous random variables with joint PDF*

$$f_{X,Y}(x,y) = \lambda^2 e^{-\lambda(x+y)} \quad x, y \ge 0.$$

We wish to compute the conditional PDF of X given $A = \{X \le Y\}$. To solve this problem, we first compute the probability of the event $\{X \le x\} \cap A$,

$$\Pr(\{X \le x\} \cap A) = \int_0^x \int_0^\xi f_{X,Y}(\xi,\zeta) d\zeta d\xi = \int_0^x \int_0^\xi \lambda^2 e^{-\lambda(\xi+\zeta)} d\zeta d\xi$$

$$= \int_0^x \lambda e^{-\lambda\xi} \left(1 - e^{-\lambda\xi}\right) d\xi = \frac{\left(1 - e^{-\lambda x}\right)^2}{2}.$$

By symmetry, we gather that $\Pr(A) = 1/2$ and, as such,

$$f_{X|A}(x) = 2\lambda e^{-\lambda x} \left(1 - e^{-\lambda x}\right) \quad x \ge 0.$$

One case of special interest is the situation where event A is defined in terms of the random variable X itself. In particular, consider the PDF of X conditioned on the fact that X belongs to an interval I. Then, $A = \{X \in I\}$ and the conditional CDF of X becomes

$$\begin{aligned}
F_{X|A}(x) &= \Pr(X \le x | X \in I) \\
&= \frac{\Pr(\{X \le x\} \cap \{X \in I\})}{\Pr(X \in I)} \\
&= \frac{\Pr(X \in (-\infty, x] \cap I)}{\Pr(X \in I)}.
\end{aligned}$$

Differentiating with respect to x, we obtain the conditional PDF of X,

$$f_{X|A}(x) = \frac{f_X(x)}{\Pr(X \in I)}$$

for any $x \in I$. In words, the conditional PDF of X becomes a scaled version of $f_X(\cdot)$ whenever $x \in I$, and it is equal to zero otherwise. Essentially, this is equivalent to re-normalizing the PDF of X over interval I, accounting for the partial information given by $X \in I$.

11.2.1 Conditioning on Values

Suppose that X and Y form a pair of random variables with joint PDF $f_{X,Y}(\cdot, \cdot)$. With great care, it is possible and desirable to define the conditional PDF of X, conditioned on $Y = y$. Special attention must be given to this situation because the event $\{Y = y\}$ has probability zero whenever Y is a continuous random variable. Still, when X and Y are jointly continuous and for any $y \in \mathbb{R}$ such that $f_Y(y) > 0$, we can defined the conditional PDF of X given $Y = y$ as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{11.3}$$

Intuitively, this definition is motivated by the following property. For small Δ_x and Δ_y, we can write

$$\Pr(x \leq X \leq x + \Delta_x | y \leq Y \leq y + \Delta_y)$$
$$= \frac{\Pr(x \leq X \leq x + \Delta_x, y \leq Y \leq y + \Delta_y)}{\Pr(y \leq Y \leq y + \Delta_y)}$$
$$\approx \frac{f_{X,Y}(x,y)\Delta_x\Delta_y}{f_Y(y)\Delta_y} = \frac{f_{X,Y}(x,y)}{f_Y(y)}\Delta_x.$$

Thus, loosely speaking, $f_{X|Y}(x|y)\Delta_x$ represents the probability that X lies close to x, given that Y is near y.

Using this definition, it is possible to compute the probabilities of events associated with X conditioned on a specific value of random variable Y,

$$\Pr(X \in S | Y = y) = \int_S f_{X|Y}(x|y)dx.$$

A word of caution is in order. The technical difficulty that surfaces when conditioning on $\{Y = y\}$ stems from the fact that $\Pr(Y = y) = 0$. Remember

that, in general, the notion of conditional probability is only defined for non-vanishing conditions. Although we were able to circumvent this issue, care must be taken when dealing with the conditional PDF of the form (11.3), as it only provides valuable insight when the random variables (X, Y) are jointly continuous.

Example 97. *Consider the experiment where an outcome (ω_1, ω_2) is selected at random from the unit circle. Let $X = \omega_1$ and $Y = \omega_2$. We wish to compute the conditional PDF of X given that $Y = 0.5$.*

First, we compute the marginal PDF of Y evaluated at $Y = 0.5$,

$$f_Y(0.5) = \int_{\mathbb{R}} f_{X,Y}(x, 0.5)dx = \int_{\frac{\sqrt{3}}{2}}^{\frac{\sqrt{3}}{2}} \frac{1}{\pi}dx = \frac{\sqrt{3}}{\pi}.$$

We then apply definition (11.3) to obtain the desired conditional PDF of X,

$$f_{X|Y}(x|0.5) = \frac{f_{X,Y}(x, 0.5)}{f_Y(0.5)} = \frac{\pi}{\sqrt{3}} f_{X,Y}(x, 0.5)$$

$$= \begin{cases} \frac{1}{\sqrt{3}} & |x| \leq \frac{\sqrt{3}}{2} \\ 0 & otherwise. \end{cases}$$

11.2.2   Conditional Expectation

The conditional expectation of a function $g(Y)$ is simply the integral of $g(Y)$ weighted by the proper conditional PDF,

$$E[g(Y)|X = x] = \int_{\mathbb{R}} g(y)f_{Y|X}(y|x)dy$$

$$E[g(Y)|S] = \int_{\mathbb{R}} g(y)f_{Y|S}(y)dy.$$

Note again that the function

$$h(x) = \mathrm{E}[Y|X = x]$$

defines a random variable since the conditional expectation of Y may vary as a function of X. After all, a conditional expectation is itself a random variable.

Example 98. *An analog communication system transmits a random signal over a noisy channel. The transmit signal X and the additive noise N are both*

standard Gaussian random variables, and they are independent. The signal received at the destination is equal to

$$Y = X + N.$$

We wish to estimate the value of X conditioned on $Y = y$.

For this problem, the joint PDF of X and Y is

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

and the conditional distribution of X given Y becomes

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{\frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2}\right)}{\frac{1}{2\sqrt{\pi}} \exp\left(-\frac{y^2}{4}\right)}$$

$$= \frac{1}{\sqrt{\pi}} \exp\left(-\frac{4x^2 - 4xy + y^2}{4}\right).$$

By inspection, we recognize that this conditional PDF is a Gaussian distribution with parameters $m = y/2$ and $\sigma^2 = 1/2$. A widespread algorithm employed to perform the desired task is called the minimum mean square error (MMSE) estimator. In the present case, the MMSE estimator reduces to the conditional expectation of X given $Y = y$, which is

$$\mathrm{E}[X|Y = y] = \int_{\mathbb{R}} x f_{X,Y}(x|y) dx = \frac{y}{2}.$$

11.2.3 Derived Distributions

Suppose X_1 and X_2 are jointly continuous random variables. Furthermore, let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$, where $g_1(\cdot, \cdot)$ and $g_2(\cdot, \cdot)$ are real-valued functions. Under certain conditions, the pair of random variables (Y_1, Y_2) will also be continuous. Deriving the joint PDF of (Y_1, Y_2) can get convoluted, a task we forgo. It requires the skillful application of vector calculus. Nevertheless, we examine the case where a simple expression for $f_{Y_1,Y_2}(\cdot, \cdot)$ exists.

Consider the scenario where the functions $g_1(\cdot, \cdot)$ and $g_2(\cdot, \cdot)$ are totally differentiable, with Jacobian determinant

$$J(x_1, x_2) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(x_1, x_2) & \frac{\partial g_1}{\partial x_2}(x_1, x_2) \\ \frac{\partial g_2}{\partial x_1}(x_1, x_2) & \frac{\partial g_2}{\partial x_2}(x_1, x_2) \end{bmatrix}$$

$$= \frac{\partial g_1}{\partial x_1}(x_1, x_2)\frac{\partial g_2}{\partial x_2}(x_1, x_2) - \frac{\partial g_1}{\partial x_2}(x_1, x_2)\frac{\partial g_2}{\partial x_1}(x_1, x_2) \neq 0.$$

Also, assume that the system of equations

$$g_1(x_1, x_2) = y_1$$
$$g_2(x_1, x_2) = y_2$$

has a unique solution. We express this solution using $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$. Then, the random variables (Y_1, Y_2) are jointly continuous with joint PDF

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{|J(x_1, x_2)|} \tag{11.4}$$

where $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$. Note the close resemblance between this equation and the derived distribution of (9.4). Looking back at Chapter 9 offers an idea of what proving this result entails. It also hints at how this equation can be modified to accommodate non-unique mappings.

Example 99. *An important application of (11.4) pertains to the properties of Gaussian vectors. Suppose that X_1 and X_2 are jointly continuous random variables, and let*

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

Define the mean of $\mathbf{X}$ by

$$\mathbf{m} = \mathrm{E}[\mathbf{X}] = \begin{bmatrix} \mathrm{E}[X_1] \\ \mathrm{E}[X_2] \end{bmatrix}.$$

and its covariance by

$$\Sigma = \mathrm{E}\left[(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T \right]$$
$$= \begin{bmatrix} \mathrm{E}\left[(X_1 - m_1)^2 \right] & \mathrm{E}[(X_1 - m_1)(X_2 - m_2)] \\ \mathrm{E}[(X_2 - m_2)(X_1 - m_1)] & \mathrm{E}\left[(X_2 - m_2)^2 \right] \end{bmatrix}.$$

Random variables X_1 and X_2 are said to be jointly Gaussian provided that their joint PDF is of the form

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m}) \right).$$

Assume that the random variables Y_1 and Y_2 are generated through the matrix equation

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = A\mathbf{X} + \mathbf{b},$$

where A is a 2×2 invertible matrix and $\mathbf{b}$ is a constant vector. In this case,
$\mathbf{X} = A^{-1}(\mathbf{Y} - \mathbf{b})$ *and the corresponding Jacobian determinant is*

$$J(x_1, x_2) = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = |A|.$$

Applying (11.4), *we gather that the joint PDF of* (Y_1, Y_2) *is expressed as*

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{2\pi |\Sigma|^{\frac{1}{2}} |A|} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m})\right) \\ &= \frac{1}{2\pi |A\Sigma A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(A^{-1}(\mathbf{y} - \mathbf{b}) - \mathbf{m}\right)^T \Sigma^{-1} \left(A^{-1}(\mathbf{y} - \mathbf{b}) - \mathbf{m}\right)\right) \\ &= \frac{1}{2\pi |A\Sigma A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{b} - A\mathbf{m})^T (A\Sigma A)^{-1} (\mathbf{y} - \mathbf{b} - A\mathbf{m})\right). \end{aligned}$$

Looking at this equation, we conclude that random variables Y_1 and Y_2 are also jointly Gaussian, as their joint PDF possesses the proper form. It should come as no surprise that the mean of $\mathbf{Y}$ is $\mathrm{E}[\mathbf{Y}] = A\mathbf{m} + \mathbf{b}$ and its covariance matrix is equal to

$$\mathrm{E}\left[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^T\right] = A\Sigma A.$$

In other words, a non-trivial affine transformation of a two-dimensional Gaussian vector yields another Gaussian vector. This admirable property generalizes to higher dimensions. Indeed, if $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ where A is an $n \times n$ invertible matrix and $\mathbf{X}$ is a Gaussian random vector, then $\mathbf{Y}$ remains a Gaussian random vector. Furthermore, to obtain the derived distribution of the latter vector, it suffices to compute its mean and covariance, and substitute the resulting parameters in the general form of the Gaussian PDF. Collectively, the features of joint Gaussian random vectors underly many contemporary successes of engineering.

11.3   Independence

Two random variables X and Y are mutually *independent* if their joint CDF is the product of their respective CDFs,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

for $x, y \in \mathbb{R}$. For jointly continuous random variables, this definition neces-
sarily implies that the joint PDF $f_{X,Y}(\cdot, \cdot)$ is the product of their marginal
PDFs,

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{dF_X}{dx}(x)\frac{dF_Y}{dy}(y) = f_X(x)f_Y(y)$$

for $x, y \in \mathbb{R}$. Furthermore, we gather from (11.3) that the conditional PDF
of Y given $X = x$ is equal to the marginal PDF of Y whenever X and Y are
independent,

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

provided of course that $f_X(x) \neq 0$. Additionally, if X and Y are indepen-
dent random variables, then the two events $\{X \in S\}$ and $\{Y \in T\}$ are also
independent,

$$\begin{aligned}
\Pr(X \in S, Y \in T) &= \int_S \int_T f_{X,Y}(x,y)dydx \\
&= \int_S f_X(x)dx \int_T f_Y(y)dy \\
&= \Pr(X \in S)\Pr(Y \in T).
\end{aligned}$$

Example 100. *Consider a random experiment where an outcome (ω_1, ω_2) is
selected at random from the unit square. Let $X = \omega_1$, $Y = \omega_2$ and $U = \omega_1 + \omega_2$.
We wish to show that X and Y are independent, but that X and U are not
independent.*

*We begin by computing the joint CDF of X and Y. For $x, y \in [0, 1]$, we
have*

$$F_{X,Y}(x,y) = \int_0^x \int_0^y d\zeta d\xi = xy = F_X(x)F_Y(y).$$

More generally, if $x, y \in \mathbb{R}^2$, we get

$$\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^x \int_{-\infty}^y \mathbf{1}_{[0,1]^2}(\xi, \zeta)d\zeta d\xi \\
&= \int_{-\infty}^x \mathbf{1}_{[0,1]}(\xi)d\xi \int_{-\infty}^y \mathbf{1}_{[0,1]}(\zeta)d\zeta = F_X(x)F_Y(y).
\end{aligned}$$

Thus, we gather that X and Y are independent.

Next, we show that X and U are not independent. Note that $F_U(1) = 0.5$ and $F_X(0.5) = 0.5$. Consider the joint CDF of X and U evaluated at $(0.5, 1)$,

$$F_{X,U}(0.5, 1) = \int_0^{\frac{1}{2}} \int_0^{1-\xi} d\zeta d\xi = \int_0^{\frac{1}{2}} (1 - \xi) d\xi$$
$$= \frac{3}{8} \neq F_X(0.5) F_U(1).$$

Clearly, random variables X and U are not independent.

11.3.1 Sums of Continuous Random Variables

As mentioned before, sums of independent random variables are frequently encountered in engineering. We therefore turn to the question of determining the distribution of a sum of independent continuous random variables in terms of the PDFs of its constituents. If X and Y are independent random variables, the distribution of their sum $U = X + Y$ can be obtained by using the *convolution* operator. Let $f_X(\cdot)$ and $f_Y(\cdot)$ be the PDFs of X and Y, respectively. The convolution of $f_X(\cdot)$ and $f_Y(\cdot)$ is the function defined by

$$(f_X * f_Y)(u) = \int_{-\infty}^{\infty} f_X(\xi) f_Y(u - \xi) d\xi$$
$$= \int_{-\infty}^{\infty} f_X(u - \zeta) f_Y(\zeta) d\zeta.$$

The PDF of the sum $U = X + Y$ is the convolution of the individual densities $f_X(\cdot)$ and $f_Y(\cdot)$,

$$f_U(u) = (f_X * f_Y)(u).$$

To show that this is indeed the case, we first consider the CDF of U,

$$F_U(u) = \Pr(U \leq u) = \Pr(X + Y \leq u)$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{u-\xi} f_{X,Y}(\xi, \zeta) d\zeta d\xi$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{u-\xi} f_Y(\zeta) d\zeta f_X(\xi) d\xi$$
$$= \int_{-\infty}^{\infty} F_Y(u - \xi) f_X(\xi) d\xi.$$

Taking the derivative of $F_U(u)$ with respect to u, we obtain

$$\frac{d}{du}F_U(u) = \frac{d}{du}\int_{-\infty}^{\infty} F_Y(u - \xi)f_X(\xi)d\xi$$
$$= \int_{-\infty}^{\infty} \frac{d}{du}F_Y(u - \xi)f_X(\xi)d\xi$$
$$= \int_{-\infty}^{\infty} f_Y(u - \xi)f_X(\xi)d\xi.$$

Notice the judicious use of the fundamental theorem of calculus. This shows that $f_U(u) = (f_X * f_Y)(u)$.

Example 101 (Sum of Uniform Random Variables). *Suppose that two numbers are independently selected from the interval $[0, 1]$, each with a uniform distribution. We wish to compute the PDF of their sum. Let X and Y be random variables describing the two choices, and let $U = X + Y$ represent their sum. The PDFs of X and Y are*

$$f_X(\xi) = f_Y(\xi) = \begin{cases} 1 & 0 \leq \xi \leq 1 \\ 0 & otherwise. \end{cases}$$

The PDF of their sum is therefore equal to

$$f_U(u) = \int_{-\infty}^{\infty} f_X(u - \xi)f_Y(\xi)d\xi.$$

Since $f_Y(y) = 1$ when $0 \leq y \leq 1$ and zero otherwise, this integral becomes

$$f_U(u) = \int_0^1 f_X(u - \xi)d\xi = \int_0^1 \mathbf{1}_{[0,1]}(u - \xi)d\xi.$$

The integrand above is zero unless $0 \leq u - \xi \leq 1$ (i.e., unless $u - 1 \leq \xi \leq u$). Thus, if $0 \leq u \leq 1$, we get

$$f_U(u) = \int_0^u d\xi = u;$$

while, if $1 < u \leq 2$, we obtain

$$f_U(u) = \int_{u-1}^1 d\xi = 2 - u.$$

If $u < 0$ or $u > 2$, the value of the PDF becomes zero. Collecting these results, we can write the PDF of U as

$$
f_U(u) = \begin{cases} u & 0 \leq u \leq 1, \\ 2 - u & 1 < u \leq 2, \\ 0 & \text{otherwise.} \end{cases}
$$

Example 102 (Sum of Exponential Random Variables). *Two numbers are selected independently from the positive real numbers, each according to an exponential distribution with parameter λ. We wish to find the PDF of their sum. Let X and Y represent these two numbers, and denote this sum by $U = X + Y$. The random variables X and Y have PDFs*

$$
f_X(\xi) = f_Y(\xi) = \begin{cases} \lambda e^{-\lambda \xi} & \xi \geq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

When $u \geq 0$, we can use the convolution formula and write

$$
\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} f_X(u - \xi) f_Y(\xi) d\xi \\
&= \int_0^u \lambda e^{-\lambda(u - \xi)} \lambda e^{-\lambda \xi} d\xi \\
&= \int_0^u \lambda^2 e^{-\lambda u} d\xi = \lambda^2 u e^{-\lambda u}.
\end{aligned}
$$

On the other hand, if $u < 0$ then we get $f_U(u) = 0$. The PDF of U is given by

$$
f_U(u) = \begin{cases} \lambda^2 u e^{-\lambda u} & u \geq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

This is an Erlang distribution with parameter $m = 2$ and $\lambda > 0$.

Example 103 (Sum of Gaussian Random Variables). *It is an interesting and important fact that the sum of two independent Gaussian random variables is itself a Gaussian random variable. Suppose X is Gaussian with mean m_1 and variance σ_1^2, and similarly Y is Gaussian with mean m_2 and variance σ_2^2, then $U = X + Y$ has a Gaussian density with mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$. We will show this property in the special case where both random variables*

are standard normal random variable. The general case can be attained in a similar manner, but the computations are somewhat tedious.

Suppose X and Y are two independent Gaussian random variables with PDFs

$$f_X(\xi) = f_Y(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}.$$

Then, the PDF of $U = X + Y$ is given by

$$f_U(u) = (f_X * f_Y)(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(u-\xi)^2}{2}} e^{-\frac{\xi^2}{2}} d\xi$$

$$= \frac{1}{2\pi} e^{-\frac{u^2}{4}} \int_{-\infty}^{\infty} e^{-\left(\xi - \frac{u}{2}\right)^2} d\xi$$

$$= \frac{1}{2\sqrt{\pi}} e^{-\frac{u^2}{4}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\left(\xi - \frac{u}{2}\right)^2} d\xi \right).$$

The expression within the parentheses is equal to one since the integrant is a Gaussian PDF with $m = u/2$ and $\sigma^2 = 1/2$. Thus, we obtain

$$f_U(u) = \frac{1}{\sqrt{4\pi}} e^{-\frac{u^2}{4}},$$

which verifies that U is indeed Gaussian.

Let X and Y be independent random variables. Consider the random variable $U = X + Y$. The moment generating function of U is given by

$$\begin{aligned}
M_U(s) &= \mathrm{E}\left[e^{sU}\right] = \mathrm{E}\left[e^{s(X+Y)}\right] \\
&= \mathrm{E}\left[e^{sX}e^{sY}\right] = \mathrm{E}\left[e^{sX}\right]\mathrm{E}\left[e^{sY}\right] \\
&= M_X(s)M_Y(s).
\end{aligned}$$

That is, the moment generating function of the sum of two independent random variables is the product of the individual moment generating functions.

Example 104 (Sum of Gaussian Random Variables). *In this example, we revisit the problem of adding independent Gaussian variables using moment generating functions. Again, let X and Y denote the two independent Gaussian variables. We denote the mean and variance of X by m_1 and σ_1^2. Likewise, we represent the mean and variance of Y by m_2 and σ_2^2. We wish to show that the sum $U = X + Y$ is Gaussian with parameters $m_1 + m_2$ and $\sigma_1^2 + \sigma_2^2$.*

The moment generating functions of X and Y are

$$M_X(s) = e^{m_1 s + \frac{\sigma_1^2 s^2}{2}}$$

$$M_Y(s) = e^{m_2 s + \frac{\sigma_2^2 s^2}{2}}$$

The moment generating function of $U = X + Y$ is therefore equal to

$$M_U(s) = M_X(s)M_Y(s) = \exp\left((m_1 + m_2)s + \frac{(\sigma_1^2 + \sigma_2^2)s^2}{2}\right),$$

which demonstrates that U is a Gaussian random variable with mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$, as anticipated.

Further Reading

1. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Section 3.5.

2. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Chapter 6, Section 7.5.

3. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Chapter 5, Sections 6.1–6.3.

4. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Chapters 7–9.