Chapter 2 Random Variables

Contents

2 Random Variables

2.1 Definition of a Random Variable

An experiment E is specified by the three tuple $(S, F, P(\cdot))$ where S is a finite, countable, or noncountable set called the sample space, F is a Borel field specifying a set of events, and $P(\cdot)$ is a probability measure allowing calculation of probabilities of all events.

Using an underlying experiment **a random variable $X(e)$ is defined as a real-valued function on S that satisfies the following: (a) $\{e: X(e) \le x\}$ is a member of F for all x, which guarantees the existence of the cumulative distribution function, and (b) the probabilities of the events $\{e: X(e) = +\infty\}$ and $\{e: X(e) = -\infty\}$ are both zero.** This means that the function is not allowed to be + or – infinite with a zero probability.

Example 2.1

E is specified by $(S, F, P(\cdot))$ where $S = \{a, b, c, d\}$, F is the power set of S, and $P(\cdot)$ is defined by $P\{a\} = 0.4$, $P\{b\} = 0.3$, $P\{c\} = 0.2$, and $P\{d\} = 0.1$. Define the following mapping $X(a) = 1$, $X(b) = 0$, $X(c) = 1$, $X(d) = 4$. Is the function $X(.)$ a random variable?

Solution

To show that the function is a random variable, it must be shown that conditions (a) and (b), specified above, are satisfied. The sets $\{e: X(e) \le x\}$ as x varies from $-\infty$ to ∞ are as follows:

$$\text{For}\quad \begin{array}{cc} x < 0 & \{e: X(e) \le x\} = \phi \\ 0 \le x < 1 & \{e: X(e) \le x\} = \{b\} \\ 1 \le x < 4 & \{e: X(e) \le x\} = \{a, b, c\} \\ x \ge 4 & \{e: X(e) \le x\} = \{a, b, c, d\} = S \end{array}$$

Since the sets described are subsets of S, they are members of the power of S, and thus condition (a) is satisfied. If S has a finite number of elements and F is the power set of S, condition (a) will always be satisfied.

If the power set is not used for F, it is possible to construct a function that is not a random variable. It is easily seen that

$$P\{e: X(e) = +\infty\} = P(\phi) = 0$$
$$P\{e: X(e) = -\infty\} = P(\phi) = 0$$

Thus conditions (b) is satisfied and since (a) is also satisfied, $X(.)$ is a random variable.

Most common functions defined on a given experiment are random variables, however, condition (a) can be easily violated if F is not the power set of S or condition (b) cannot be satisfied if X is defined to be ∞ for a set with finite probability.

Random variables are said to be **totally characterized** or described with relation to calculating probabilities of acceptable events (i.e., events that are a member of F) by their cumulative distribution function or probability density function. Weaker characterizations, called "**partial characterizations**" would include specifying higher-order moments, variance, mean, and the like. Knowing just the mean of a random variable is certainly less information about the random variable than knowing the probability density function, yet it still provides some idea about values of the random variable.

2.1.1 Cumulative Distribution Function (CDF)

The **cumulative distribution function** $F_X(x)$ for a random variable $X(e)$, or when convenient represented by just X, is defined for all x as

$$F_x(x)\overset{\Delta}{=}P\{e: X(e)\le x\}=P\{X\le x\} \tag{2.1}$$

It is sufficient information to calculate the probabilities of all allowable events, and as a result is called a **total characterization**. Because of the properties of the probability measure $P(\cdot)$ described in the overall experiment, the cumulative distribution function (CDF) can be shown to have a number of important properties:

(1) $F_X(x)$ is **bounded** from above and below,

$$0\le F_X(x)\le 1 \quad \text{for} \quad \text{all} \quad x \tag{2.2}$$

(2) $F_X(x)$ is a **nondecreasing** function of x,

$$F_X(x_2)\ge F_X(x_1) \quad \text{for all} \quad x_2>x_1 \quad \text{and all} \quad x_1 \tag{2.3}$$

(3) $F_X(x)$ is **continuous** from the right,

$$\lim_{a\to 0^+} F_X(x+a)=F_X(x) \tag{2.4}$$

(4) $F_X(x)$ can be used to **calculate probabilities** of events

$$\begin{aligned} P\{x_1<X\le x_2\}&=F_X(x_2)-F_X(x_1)\\ P\{x_1<X\le x_2\}&=F_X(x_2)-F_X(x_1^-)\\ P\{x_1<X\le x_2\}&=F_X(x_2^-)-F_X(x_1^-)\\ P\{x_1<X\le x_2\}&=F_X(x_2^-)-F_X(x_1)\end{aligned} \tag{2.5}$$

Where $x^-=\lim_{\varepsilon\to 0^+} F_X(x-\varepsilon)$ (left-hand limit)

(5) Relation to the probability density function $f_X(x)$ (to be defined later) is written as

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx \tag{2.6}$$

Example 2.3

For the random variable defined in Example 2.1 determine (a) the cumulative distribution function $F_X(x)$ and calculate the probabilities of the following events. Using this distribution function, determine (b) $P\{1 < X \le 2\}$, (c) $P\{1 < X \le 4\}$, (d) $P\{0 \le x \le 1\}$,(e) $P\{0 \le x < 1\}$,(f) $P\{0 < x < 1\}$,(g) $P\{X > 1\}$,(h) $P\{X \le 1\}$,(i) $P\{X = 1\}$, and (j) $P\{X = 3\}$.

Solution

(a) The CDF $F_X(x)$ for the random variable X is obtained by determining the probabilities of the events $\{e : X(e) \le x\}$ for all $x \in (-\infty, \infty)$. The results from Example 2.1 help us determine $F_X(x)$ as follows for the following regions.

$$\begin{array}{cc}
x < 0 & \{e : X(e) \le x\} = P(\phi) = 0 \\
0 \le x < 1 & \{e : X(e) \le x\} = P(b) = 0.3 \\
1 \le x < 4 & \{e : X(e) \le x\} = P(a,b,c) = P(a) + P(b) + P(c) = 0.9 \\
x \ge 4 & \{e : X(e) \le x\} = P(a,b,c,d) = P(S) = 1
\end{array}$$

These results can be plotted as shown.

The CDF along with the probabilities of the intervals given in Eq.(2.5) will be used to determine the probabilities of the events listed.

(a) $P\{1 < X \le 2\} = F_X(2) - F_X(1) = 0.9 - 0.9 = 0$

(b) $P\{1 < X < 4\} = F_X(4) - F_X(1) = 1 - 0.9 = 0.1$

(c) $P\{0 \le X \le 1\} = F_X(1) - F_X(0^-) = 0.9 - 0 = 0.9$

(d) $P\{0 \le X < 1\} = F_X(1^-) - F_X(0^-) = 0.3 - 0 = 0.3$

(e) $P\{0 < X < 1\} = F_X(1^-) - F_X(0) = 0.3 - 0.3 = 0$

(f) $P\{X > 1\} = F_X(\infty) - F_X(1) = 1 - 0.9 = 0.1$

(g) $P\{X \le 0\} = F_X(0) = 0.3$

(h) $P\{X = 1\} = F_X(1) - F_X(1^-) = 0.9 - 0.3 = 0.6$

(i) $P\{X = 3\} = F_X(3) - F_X(3^-) = 0.9 - 0.9 = 0$

2.1.2 Probability Density Function (PDF)

The probability density function $f_X(x)$ for a random X is a total characterization and is defined

as the derivative of the cumulative distribution function

$$f_X(x) \overset{\Delta}{=} \frac{d}{dx} F_X(x) \qquad \textbf{(2.7)}$$

If $F_X(x)$ has jump discontinuities, it is convenient to use delta functions so that a probability density function (PDF) will always be defined. Therefore a probability density function will contain delta functions at the points of discontinuities of $F_X(x)$ with weights equal to the size of the jump at those points. Important properties of the probability density function for a random variable X are as follows:

(1) **positivity**

$$f_X(x) \geq 0 \qquad \text{for all } x \qquad (2.8)$$

(2) Integral over all x (**unit area**)

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad (2.9)$$

(3) $f_X(x)$ used to **calculate probability of events**,

$$\begin{aligned}
p\{x_1 < X \leq x_2\} &= \int_{x_1+\in}^{x_2+\in} f_X(x)dx \overset{\Delta}{=} \int_{x_1^+}^{x_2^+} f_X(x)dx \\
p\{x_1 \leq X < x_2\} &= \int_{x_1-\in}^{x_2-\in} f_X(x)dx \overset{\Delta}{=} \int_{x_1^-}^{x_2^-} f_X(x)dx \\
p\{x_1 \leq X \leq x_2\} &= \int_{x_1-\in}^{x_2+\in} f_X(x)dx \overset{\Delta}{=} \int_{x_1^-}^{x_2^+} f_X(x)dx \\
p\{x_1 < X < x_2\} &= \int_{x_1+\in}^{x_2-\in} f_X(x)dx \overset{\Delta}{=} \int_{x_1^+}^{x_2^-} f_X(x)dx
\end{aligned} \qquad (2.10)$$

(4) **Relationship to cumulative distribution function**

$$f_X(x) = \frac{dF_X(x)}{dx} \qquad (2.11)$$

A random variable is called **a discrete random variable** **if its probability density function $f_X(x)$ is a sum of delta function only, or correspondingly if its cumulative distribution function $F_X(x)$ is a staircase function**. **A random variable is called** **a continuous random variable** **if its cumulative distribution function has no finite discontinuities or equivalently its probability density function $f_X(x)$ has no delta functions**. **If a random is neither a continuous or discrete random variable we call it** **a mixed random variable**.

Random variable
- **Discrete random variable**
- **Continuous random variable**
- **Mixed random variable**

Examples of these three types follow.

Discrete random variable,

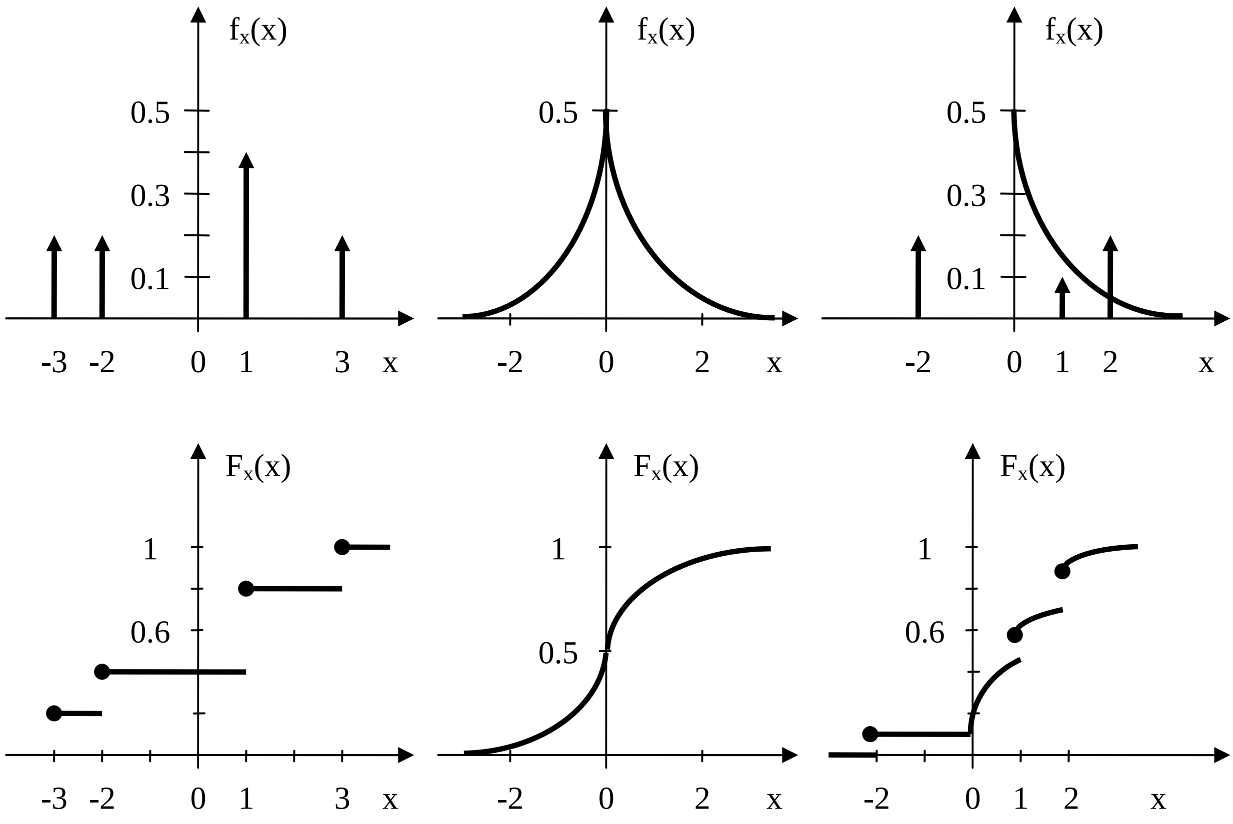
$$f_X(x) = 0.2\delta(x+3) + 0.2\delta(x+2) + 0.4\delta(x-1) + 0.2\delta(x-3)$$

Continuous random variable,

$$f_X(x) = \frac{1}{2}e^{-|x|}$$

Mixed random variable,

$$f_X(x) = 0.5e^{-x}\mu(x) + 0.2\delta(x+2) + 0.1\delta(x-1) + 0.2\delta(x-2)$$



(a) Discrete RV (b) Continuous RV (c) Mixed RV

Figure 2.2 Examples of (a) discrete (b) continuous and (c) mixed random variables.

Example 2.5

Given a mixed random variable with probability density function

$$f_X(x)=\frac{1}{2}e^{-x}\mu(x)+\frac{1}{8}\delta(x+1)+\frac{1}{4}\delta(x)+\frac{1}{8}\delta(x-1)$$

Determine the cumulative distribution function $F_X(x)$.

Solution

The cumulative distribution function $F_X(x)$ can be determined by the relationship given in (2.6).

$$\begin{aligned}F_X(x)&=\int_{-\infty}^{x} f_x(x)dx\\&=\int_{-\infty}^{x}\frac{1}{2}e^{-x}\mu(x)dx+\int_{-\infty}^{x}\left[\frac{1}{8}\delta(x+1)+\frac{1}{4}\delta(x)+\frac{1}{8}\delta(x-1)\right]dx\\&=\left[\frac{1}{2}-\frac{1}{2}e^{-x}\right]\mu(x)+\frac{1}{8}\mu(x+1)+\frac{1}{4}\mu(x)+\frac{1}{8}\mu(x-1)\end{aligned}$$

The integrals of the delta function as x goes through the values gives unit steps,causing discontinuities or jumps, at the points $x=-1,0,1$ of sizes $\frac{1}{8},\frac{1}{4}$, and $\frac{1}{8}$ respectively.

2.1.3 partial characterizations

Important **partial characterizations** of random variables are the mean, variance, higher-order moments and central moments, and conditional distribution and density functions. Definitions for these partial characterizations follow.

The **mean** η_X of a random variable X or equivalently the expected value $E[X]$ of a random variable X is defined by

$$\eta_X=\int_{-\infty}^{\infty} xf_X(x)dx\overset{\Delta}{=}E[X] \tag{2.12}$$

The mean of a random variable is also referred to as the average of the random variable as seen from the integral above. Knowing the mean of a random variable is not sufficient information to calculate probabilities of given events. It is just barely enough information to bound probabilities of some special events as seen in the following inequality for the special case when the random variable takes on positive values:

$$P\{X\geq a\}\leq\frac{E[X]}{a}\quad \text{for}\quad a>0 \tag{2.13}$$

The **variance** σ_X^2 of a random variable X is defined by

$$\sigma_X^2=\int_{-\infty}^{\infty}(x-\eta_X)^2 f_X(x)dx\overset{\Delta}{=}E[(X-\eta_X)^2] \tag{2.14}$$

The **standard deviation** is defined as the square root of the variance and denoted as σ_X .The variance gives a sense of spread of the random variable as larger values would indicate a wider density function and as with the mean it is not sufficient information for calculation of probabilities of events. However, like the mean, bounds can be placed on the calculation of some special events.

Two of the most important bounds are described in the Tchebycheff inequalities for the events $\{|X-\eta_X| \geq k\sigma_X\}$ and $\{|X-\eta_X| < k\sigma_X\}$. Notice that these events represent the regions corresponding to the mean $\pm k$ standard deviations about the mean and the complement of that region.

The **Tchebycheff inequalities** corresponding to bounds on the probabilities of the tails and core of the density function, respectively, for any density function and $k \geq 1$ are

$$P\{|X-\eta_X| \geq k\sigma_X\} \leq \frac{1}{k^2}$$
$$P\{|X-\eta_X| < k\sigma_X\} \geq 1-\frac{1}{k^2} \tag{2.15}$$

For a given positive ε we can also bound the $P\{\eta_X-\varepsilon < X < \eta_X+\varepsilon\}$ which is the probability that X is in the region $\pm\varepsilon$ around the mean by

$$P\{\eta_X-\varepsilon < X < \eta_X+\varepsilon\} \geq 1-\frac{\sigma_X^2}{\varepsilon^2} \tag{2.16}$$

Eq. (2.15) can be obtained from (2.14) by letting $k\sigma_X = \varepsilon$.

Proof : $$P\{|X-\eta_X| \geq \varepsilon\} = \int_{|X-\eta_X|\geq\varepsilon} f_X(x)dx$$

$$\leq \int_{|X-\eta_X|\geq\varepsilon} \frac{|X-\eta_X|^2}{\varepsilon^2} f_X(x)dx \leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(X-\eta_X)^2 f_X(x)dx = \frac{\sigma_X^2}{\varepsilon^2}$$

Tchebycheff inequalities: $$P\{|X-\eta_X| \geq \varepsilon\} \leq \frac{\sigma_X^2}{\varepsilon^2} \quad or \quad P\{|X-\eta_X| \leq \varepsilon\} \leq 1-\frac{\sigma_X^2}{\varepsilon^2}$$

The **higher-order moments** m_k and **higher-order central moments** μ_k of a random variable X are partial characterizations and provide additional information about the statistics properties of random variables. They are defined as follow for $k > 2$:

$$m_k = \int_{-\infty}^{\infty} x^k f_X(x)dx \overset{\Delta}{=} E[X^k] \tag{2.17}$$

$$\mu_k = \int_{-\infty}^{\infty} (x-\eta_X)^k f_X(x)dx \overset{\Delta}{=} E[(X-\eta_X)^k] \tag{2.18}$$

A generalization of the Tchebycheff inequality called the **inequality of Bienayme** gives a bound on the event $\{|X-a| \geq \varepsilon\}$ in higher-order moments about an arbitrary value a, not necessarily the mean, as

$$P\{|X-a|\geq \varepsilon\}\leq \frac{E[|X-a|^n]}{\varepsilon^n} \tag{2.19}$$

If $\varepsilon = k\varepsilon$, $a=\eta_X$, and $n=2$, the inequality above reduces to Tchebycheff inequality.

2.1.4 Conditional Cumulative Distribution Function

Another partial characterization of a random variable X is **the conditional distribution function** defined for conditional event C by

$$F_X(x|C)\overset{\Delta}{=}P\{X\leq x|C\} \tag{2.20}$$

The determination of $F_X(x|C)$ for a given event C and a given $F_X(x)$ for a random variable X uses the definition of conditional probability as follow:

$$F_X(x|C)=\frac{P\{X\leq x, C\}}{P(C)} \tag{2.21}$$

Similarly **the conditional probability density function** for a random variable conditioned on the event C is defined in the terms of the derivative of the conditional distribution function as

$$f_X(x|C)\overset{\Delta}{=}\frac{dF_X(x|C)}{dx} \tag{2.22}$$

If C is given in terms of elements of the sample space S, then the $P\{X\leq x, C\}$ can be determined by working with the original sample space. In many problems the conditioning event is given in terms of the values of the random variable.

Example 2.6

A random variable X is characterized by its cumulative distribution function $F_X(x)=\left(1-e^{-x}\right)\mu(x)$ and an event C is defined by $C=\{0.5<X\leq 1\}$. Determine the conditional cumulative distribution function $F_X(x|C)$ and conditional probability density function $f_X(x|C)$ for the random variable X conditioned on C.

Solution

By Eq.(2.21) the conditional distribution can be written as

$$F_X(x|C)=\frac{P\{X\leq x, 0.5<X\leq 1\}}{P(0.5<X\leq 1)}$$

The numerator will be different depending on where x is in the infinite interval as

$$\begin{array}{lll} for & x<0.5 & P\{X \le x, 0.5<X \le 1\} = P(\phi) = 0 \\ for & 0 \le x \le 1 & P\{X \le x, 0.5<X \le 1\} = P(0.5<X \le x) = F_X(x) - F_X(0.5) \\ & & = (1-e^{-x}) - (1-e^{-0.5}) = e^{-0.5} - e^{-x} \\ for & x>1 & P\{X \le x, 0.5<X \le 1\} = P(0.5<X \le 1) = F_X(1) - F_X(0.5) \\ & & = (1-e^{-1}) - (1-e^{-0.5}) = e^{-0.5} - e^{-1} = 0.2386512 \end{array}$$

But we know that the $P(C)$ is determined as

$$\begin{aligned} P(C) = P\{0.5<X \le 1\} = P(0.5<X \le x) &= F_X(1) - F_X(0.5) \\ &= (1-e^{-1}) - (1-e^{-0.5}) = 0.2386512 \end{aligned}$$

Dividing the probabilities by P(C) gives the final answer for the conditional distribution as

$$\begin{array}{lll} for & x<0.5 & F_X(x|C) = 0 \\ for & 0.5 \le x \le 1 & F_X(x|C) = (e^{-0.5} - e^{-x})/0.2386512 \\ for & x>1 & F_X(x|C) = 1 \end{array}$$

Plots of the $F_X(x)$ and $F_X(x|C)$ are shown for this example in Figure 2.6.

The conditional probability density function can be obtained by taking the derivative of the conditional distribution throughout the region of the conditioning event, and the result is

$$f_X(x|C) = \begin{cases} 4.1902157e^{-x} & 0.5<x \le 1 \\ 0 & elsewhere \end{cases}$$

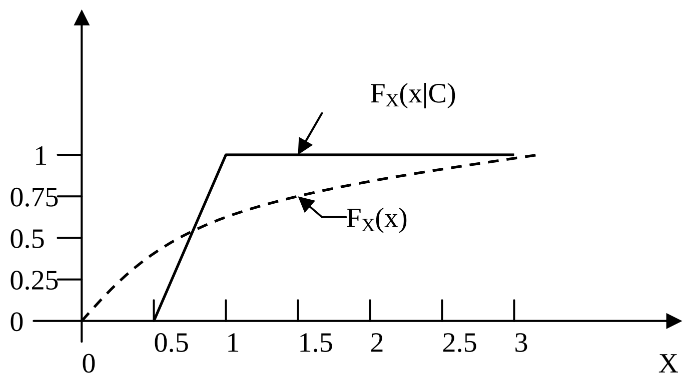


Figure 2.6 $F_X(x)$ and $F_X(x|C)$ for Example 2.6

2.1.5 Characteristic function

The characterization of a random variable can also be given by specifying its **characteristic function** $\phi_X(\omega)$, which is the modified Fourier transform of $f_X(x)$ and defined as

$$\phi_X(\omega) \overset{\Delta}{=} \int_{-\infty}^{\infty} f_X(x) e^{+j\omega x} dx = E[e^{j\omega X}] = F(f_X(x))\Big|_{\omega=-\omega} \tag{2.23}$$

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(\omega) e^{-j\omega x} d\lambda$$

Specification of characteristic function is equivalent to a total characterization, since its inverse transform yields the probability density function.

Example 2.7

Determine the characteristic function for a Gaussian random variable with known mean n_X and variance δ_X^2.

Solution

By the definition above, the characteristic function becomes

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{(x-\eta_X)^2}{2\sigma_X^2}\right\} e^{j\omega X} dx$$

After expanding out the term in the brackets and combining it with the other exponent over a common denominator, the characteristic function becomes

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{\left(x^2 - \eta_X x + \eta_X^2 - 2j\omega x\sigma_X^2\right)}{2\sigma_X^2}\right\} dx$$

Next complete the square in the brackets and multiply by the proper correction so that the characteristic function becomes

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{\left(x^2 - (\eta_X x + j\omega\sigma_X^2)\right)^2}{2\sigma_X^2}\right\} \exp\left\{\frac{-\eta_X^2 + (n_X^2 + j\omega\sigma_X^2\eta_X)^2}{2\sigma_X^2}\right\} dx$$

The exponential term on the right can be taken through the integral sign, since it is not a function of x and since the integral is of a Gaussian density it gives a value of 1.

We can finally write, after simplifying the second exponent, the characteristic function for a Gaussian random variable as

$$\Phi_X(\omega) = \exp\left\{ j\omega\eta_X - \frac{\omega^2\sigma_X^2}{2} \right\} \tag{2.24}$$

The higher-order moments can also be determined from the characteristic function as

$$E\left[X^k\right] = (-j)^k \frac{\partial^k \phi_X(\omega)}{\partial^k \omega}\Bigg|_{\omega=0} \tag{2.25}$$

The above equation can be used to determine the moments of any random variable by obtaining the characteristic function or modified Fourier transform and performing the differentiation and evaluation at zero.

Example 2.8

A random variable X has a probability density function $f_X(x)$ given by $f_X(x) = ae^{-ax}\mu(x)$. Determine the moments by using the characteristic function.

Solution

The characteristic function can be obtained by taking the modified Fourier transform of the $f_X(x)$ as

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)e^{+j\omega x}dx = F\left(ae^{-ax}\mu(x)\right)\Big|_{\omega=-\omega}$$

$$= \frac{a}{j\omega + a}\Big|_{\omega=-\omega} = \frac{a}{j\omega + a}$$

By Eq.(2.25) the moments are determined as follows:

$$E\left[X^1\right] = (-j)^1 \frac{\partial\Phi(\omega)}{\partial\omega}\Big|_{\omega=0} = (-j)^1 \frac{\partial\left[a/(-j\omega + a)\right]}{\partial\omega}\Big|_{\omega=0}$$

$$= (-j)^1 a(-j)^1/(-j\omega + a)^2\Big|_{\omega=0} = \frac{1}{a}$$

$$E\left[X^2\right] = (-j)^2 \frac{\partial^2\Phi(\omega)}{\partial^2\omega}\Big|_{\omega=0} = (-j)^2 \frac{\partial^2\left[a/(-j\omega + a)\right]}{\partial^2\omega}\Big|_{\omega=0}$$

$$= (-j)^2 a(-j)^2/(-j\omega + a)^3\Big|_{\omega=0} = \frac{1}{a^2}$$

On continuing this process, it is easy to see that

$$E\left[X^k\right] = (-j)^k \frac{\partial^k\Phi(\omega)}{\partial^k\omega}\Big|_{\omega=0} = (-j)^k a \frac{\partial^k\left[1/(-j\omega + a)\right]}{\partial^k\omega}\Big|_{\omega=0}$$

$$= (-j)^k a(-j)^k/(-j\omega + a)^{k+1}\Big|_{\omega=0} = \frac{1}{a^k}$$

Another function that can be used to determine the moments for a random variable X is **the moment-generating function** defined as follows:

$$M_X(t) \overset{\Delta}{=} \int_{-\infty}^{\infty} f_X(x)e^{tx}dx = E[e^{tx}] \qquad (2.26)$$

As the same implies, the moment-generating function can be used to calculate the moments for any random variable for which the derivatives exist and relationship can be shown to be

$$E[X^k] = \frac{\partial^k M_X(t)}{\partial^k t}\Big|_{t=0} \qquad (2.27)$$

The moment-generating function does not always exist, but when it does, it provides an alternative to using the characteristic function.

2.1.6 Higher-Order Moments for Gaussian Random Variable

Let X be a Gaussian random variable with known mean η_X and variance σ_X^2. The higher-order moments, $m_k = E[X^k]$, can be written in terms of the probability density function as

$$E[X^m] = \int_{-\infty}^{\infty} x^m f_X(x)dx = \int_{-\infty}^{\infty} \frac{x^m}{\sqrt{2\pi}\sigma_X} \exp\{-\frac{(x-\eta_X)^2}{2\sigma_X^2}\}dx \tag{2.28}$$

If $x-\eta_X$ is replaced by y, the integral can be written as

$$E[X^m] = \int_{-\infty}^{\infty} \frac{(y+\eta_X)^m}{\sqrt{2\pi}\sigma_X} \exp\{-\frac{y^2}{2\sigma_X^2}\}dy \tag{2.29}$$

After expending the $(y+\eta_X)^m$, the higher-order moments are written as

$$\begin{aligned} E[X^m] &= \int_{-\infty}^{\infty} \sum_{i=0}^{m} \binom{m}{i} y^{m-i}\eta_X^i \frac{1}{\sqrt{2\pi}\sigma_X} \exp\{-\frac{y^2}{2\sigma_X^2}\}dy \\ &= \sum_{i=0}^{m} \binom{m}{i} E[Y^{m-i}]\eta_X^i \end{aligned} \tag{2.30}$$

But Y, defined by $X-\eta_X$, is a Gaussian random variable with zero mean and variance equal to the variance of X. By symmetry of the zero mean Gaussian density, all higher-order odd moments of Y are zero and (2.30) can be written as

$$E[X^m] = \begin{cases} \sum_{j=0}^{\frac{m}{2}-1} \binom{m}{2j} E[Y^{m-2j}]\eta_X^{2j} + \eta_X^m & m(even) \\ \sum_{j=1}^{\frac{m-1}{2}} \binom{m}{2j-1} E[Y^{m-2j+1}]\eta_X^{2j-1} + \eta_X^m & m(odd) \end{cases} \tag{2.31}$$

The higher-order moments up to order six are easily determined from (2.31) as follows:

$$\begin{aligned} E[X] &= \eta_X \\ E[X^2] &= \sigma_X^2 + \eta_X^2 \\ E[X^3] &= 3\eta_X\sigma_X^2 + \eta_X^3 \\ E[X^4] &= 3\sigma_X^4 + 6\sigma_X^2\eta_X^2 + \eta_X^4 \\ E[X^5] &= 15\sigma_X^4\eta_X + 10\sigma_X^2\eta_X^3 + \eta_X^5 \\ E[X^6] &= 15\sigma_X^6 + 45\sigma_X^4\eta_X^2 + 15\sigma_X^2\eta_X^4 + \eta_X^6 \end{aligned} \tag{2.32}$$

If $\eta_X = 0$, then it is seen from the above that $E[X^{(2k-1)}] = 0$ for $k = 1,2,\ldots$, as expected.

2.2 Common Continuous Random Variable

Gaussian Random Variable. A random variable X is called a Gaussian or normal random if it has a probability density function $f_X(x)$ given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{(x-\eta_X)^2}{2\sigma_X^2}\right\} \overset{\Delta}{=} g(x;\eta_X,\sigma_X^2) \tag{2.33}$$

It can be seen that η_X and σ_X^2 are the mean and variance of the Gaussian random variable. The **standard deviation** is defined as σ_X .

Rather than write the expression (2.33), it is common practice to use the notation $X \sim N(\eta_X,\sigma_X^2)$, Which means that X is a Gaussian random variable with mean η_X and variance σ_X^2 .

The higher-order central moments can be determined for the Gaussian density in terms of the σ_X^2 as follows:

$$E[(X-\eta_X)^n] = \begin{cases} 1\cdot 3\cdots(n-1)\sigma_X^2 & for\, n\, even \\ 0 & for\, n\, odd \end{cases} \tag{2.34}$$

Uniform Random Variable. A random is uniform if its probability density on an interval is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases} \tag{2.35}$$

The mean and variance of a uniformly distributed random variable can be shown to be

$$\eta_X = \frac{a+b}{2}, \qquad \sigma_X^2 = \frac{(b-a)^2}{12} \tag{2.36}$$

Exponential Random Variable. An exponential random variable has the following probability density function for $a > 0$:

$$f_X(x) = \begin{cases} ae^{-ax} & x \ge 0 \\ 0 & otherwise \end{cases} \tag{2.37}$$

The mean and variance for an exponential distributed random variable are

$$\eta_X = \frac{1}{a}, \qquad \sigma_X^2 = \frac{1}{a^2} \tag{2.38}$$

Rayleigh Random Variable. A Rayleigh distributed random variable has for $b > 0$ the following probability density function:

$$f_X(x) = \begin{cases} \dfrac{x}{\alpha^2}\exp(-\dfrac{x^2}{2\alpha^2}), & x \ge 0 \\ 0, & otherwise \end{cases} \tag{2.39}$$

The mean and variance for a Rayleigh distributed random variable are

$$\eta_X = \sqrt{\pi/2\alpha}, \qquad \sigma_X^2 = (2 - \frac{\pi}{2})\alpha^2 \tag{2.40}$$

Gamma Random Variable. A Gamma distributed random has for $\alpha > 0$ and $\beta > 0$ the following probability density function:

$$f_X(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \ge 0 \\ 0, & otherwise \end{cases} \tag{2.41}$$

Where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$

If α is a positive integer, then it is well known that $\Gamma(\alpha) = (\alpha - 1)!$.

The mean and variance for a Gamma distributed random variable are

$$\eta_X = \beta\alpha, \qquad \sigma_X^2 = \beta^2\alpha \tag{2.42}$$

Cauchy Random Variable. A random is Cauchy distributed random if its probability density takes the following form for $a > 0$:

$$f_X(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2} \tag{2.43}$$

The mean and variance for a Cauchy distributed random variable are

$$\eta_X = \text{ does not exist}, \qquad \sigma_X^2 = \text{ does not exist} \tag{2.44}$$

Chi–square Random Variable. A random variable is chi-squared distributed with degree of freedom N if

$$f_X(x) = \frac{x^{(N/2)-1}}{2^{N/2}\Gamma(N/2)} \exp(-\frac{x}{2})\mu(x) \tag{2.45}$$

The mean and variance for a Chi-square random variable are

$$\eta_X = N, \qquad \sigma_X^2 = 2N \tag{2.46}$$

Log Normal Random Variable. A random variable is log normally distributed if its probability density function is of the form

$$f_X(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\dfrac{1}{2\sigma^2} (\log_e x - m) \right\}, & x \ge 0 \\ 0, & otherwise \end{cases} \tag{2.47}$$

The mean and variance in terms of the positive parameters n and σ are

$$\eta_X = \exp[m+\sigma^2/2], \qquad \sigma_X^2 = \exp[2m+2\sigma^2]-\exp[2m+\sigma^2] \tag{2.48}$$

Beta Random Variable. A Beta-distributed random variable has for $\alpha > 0$ and $\beta > 0$ the probability density function

$$f_X(x) = \begin{cases} kx^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1 \\ 0, & otherwise \end{cases} \tag{2.49}$$

$$k = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \quad \text{and} \quad \Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$$

If α, β, and $\alpha+\beta$ are positive integers then $\Gamma(\alpha)=(\alpha-1)!$, $\Gamma(\beta)=(\beta-1)!$, and $\Gamma(\alpha+\beta)=(\alpha+\beta-1)!$.

The mean and variance for a beta-distributed random variable can be shown to be

$$\eta_X = \frac{\alpha}{\alpha+\beta}, \qquad \sigma_X^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{2.50}$$

2.3 Common Discrete Random Variables

Bernoulli Random Variable. A discrete random variable is Bernoulli if its probability density is given by

$$f_X(x) = p\delta(x) + (1-p)\delta(x-1) \tag{2.51}$$

The mean and variance for a Bernoulli distributed random variable can be shown to be

$$\eta_X = 1-p, \qquad \sigma_X^2 = p - p^2 \tag{2.52}$$

Discrete Uniform Random Variable. A discrete random variable is uniform if its probability density on a range is given by

$$f_X(x) = \sum_{i=0}^{n} \frac{1}{n+1}\delta\left(x - \left(a + i\frac{(b-a)}{n}\right)\right) \tag{2.53}$$

The mean and variance for a uniformly distributed random variable can be shown to be

$$\eta_X = \frac{a+b}{2}, \qquad \sigma_X^2 = \frac{(n+2)(b-a)^2}{12n} \tag{2.54}$$

Poisson Distribution. A random variable X is Poisson distributed if its probability density function is given by

$$f_X(x) = \sum_{k=0}^{\infty} \lambda^k \frac{e^{-\lambda}}{k!}\delta(x-k) \tag{2.55}$$

The mean and variance for a Poisson distributed random variable can be shown to be

$$\eta_X = \lambda, \qquad \sigma_X^2 = \lambda \tag{2.56}$$

Binomial Random Variable. A random variable X is binomially distributed if its probability density function is given by

$$f_X(x) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k) \tag{2.57}$$

The mean and variance for a binomially distributed random variable can be shown to be

$$\eta_X = np, \qquad \sigma_X^2 = np(1-p) \tag{2.58}$$

X could be interpreted as the number of successes in n repeated trials of the Bernoulli type.

Geometric Random Variable. A random variable X is geometrically distributed if its probability density function is given by

$$f_X(x) = \sum_{k=1}^{\infty} p^k (1-p)^{k-1} \delta(x-k) \tag{2.59}$$

The mean and variance for a geometric random variable can be shown to be

$$E[X] = \frac{1}{P}, \qquad \sigma_X^2 = \frac{1-P}{P^2} \tag{2.60}$$

X could be interpreted as the number of the trial for which the first success occurs for repeated trials of the Bernoulli type.

Negative Binomial Random Variable. A random variable X is Negative binomially distributed if its probability density function is given by

$$f_X(x) = \sum_{j=k}^{\infty} \binom{j-1}{k-1} p^k (1-p)^{j-k} \delta(x-j) \tag{2.61}$$

The mean and variance for a negative binomially distributed random variable can be shown to be

$$\eta_X = \frac{k}{P}, \qquad \sigma_X^2 = \frac{k}{P}\left(\frac{1}{p}-1\right) \tag{2.62}$$

X could be interpreted as the number of the trial on which the kth success occurs for repeated trials of the Bernoulli type.

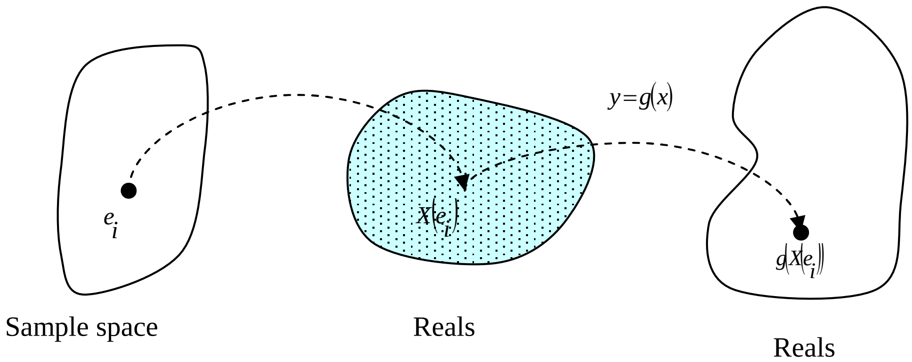
2.4 Transformations of One Random Variable

Let $X(e)$ be a random variable defined on a sample space S. Define a real-valued function $g(x)$ on the real numbers. If $X(e)$ for all e a member of S is the domain of the $g(x)$, then the range becomes the set of all $Y(e)$ such that $Y(e)=g(X(e))$. This can be described by

$$e_i \Rightarrow X(e_i) \Rightarrow g(X(e_i)) \stackrel{\Delta}{=} Y(e_i) \quad \text{for all } e_i \text{ member of } S \tag{2.63}$$

As we usually drop the e from the random variable $X(e)$ and use just X, it is expedient to drop the index from $Y(e)$ and use just Y. Therefore Eq.(2.63) is usually represented as

$$Y=g(X) \tag{2.64}$$



$\zeta \in S$	$X(\zeta)$	$Y=X(\zeta)$
ζ_1	$X(\zeta_1)$	$Y=g(X(\zeta_1))$
ζ_2	$X(\zeta_2)$	$Y=g(X(\zeta_2))$
$\vdots$	$\vdots$	$\vdots$
ζ_i	$X(\zeta_i)$	$Y=g(X(\zeta_i))$
$\vdots$	$\vdots$	$\vdots$

Figure 2.7 Transformation of random variables as a mapping and a tabular

2.4.1 Transformation of One Random Variable

In general, unless $g(x)$ is a linear function of x described by $g(x)=ax+b$, with a and b constants, or other special information regarding the type probability and type of function are specified, the expected value of Y cannot be determined. In general, for $Y=g(x)$, the mean of Y is not the function $g(.)$ of the mean, that is,

$$E[Y] \neq g(E[x]) \tag{2.65}$$

Probability Density Function (Discrete Random Variable).

If X is a real discrete random variable it takes on a finite or countable set S_X of possible values x_i.

Therefore X is totally characterized by its probability density function $f_X(x)$ consisting of a sum of weighted impulse function at the x_i as follows:

$$f_X(x) = \sum_{x_i \in S_X} p(x_i)\delta(x - x_i) \tag{2.66}$$

In the expression above $p(x_i)$ is the probability that $X = x_i$ or more precisely is denoted by $P\{e : X(e) = x_i\}$. If X takes on only the values x_i it is reasonable that Y takes only on the values $y_i = g(x_i)$. Thus Y is a discrete random variable whose probability density function $f_Y(y)$ can be written as

$$f_Y(y) = \sum_{x_i \in S_X} p(x_i)\delta\big(y - g(x_i)\big) \tag{2.67}$$

In some cases the nonlinear function $g(x)$ is such that $g(x_i)$ of several different x_i give the same value y_j, thus allowing us to write the density as

$$f_Y(y) = \sum_{y_j \in S_Y} p(y_j)\delta(y - y_j) \tag{2.68}$$

Where S_Y is the set of all unique $g(x_i)$ and $p(y_j)$ is the sum of the probabilities.

Example 2.9

Let X be a discrete random variable characterized by its probability density function $f_X(x)$ as follows:

$$f_X(x) = 0.1\delta(x+3) + 0.2\delta(x+1) + 0.25\delta(x) + 0.15\delta(x-1) + 0.05\delta(x-2) + 0.3\delta(x-3)$$

Define the function $y = g(x)$ as shown in Figure 2.8.

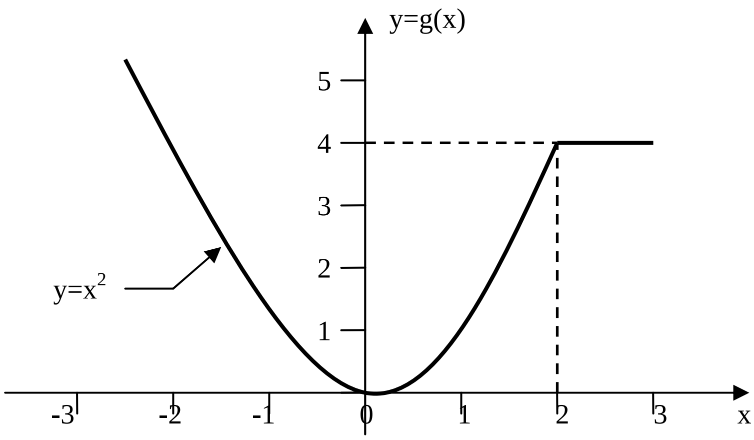


Figure 2.8 Function y=g(x) for Example 2.9

Solution

Each point is taken through the transformation, and like values are collected to give

$$f_y(y) = 0.1\delta(y-9) + 0.2\delta(y-1) + 0.25\delta(y) + 0.15\delta(y-1) + 0.05\delta(y-4) + 0.3\delta(y-4)$$
$$= 0.1\delta(y-9) + 0.35\delta(y-1) + 0.25\delta(y) + 0.35\delta(y-4)$$

Probability Density Function (Continuous Random Variable).

Theorem. If X is a continuous random variable characterized by its probability density function $f_X(x)$ and $g(x)$ is a continuous function with no flat spots (finite lengths of constant value), then the probability density function $f_Y(y)$, which characterized the random variable $Y = g(x)$, is given for each value of y as follows

$$f_Y(y) = \sum_{x_i \in S_X} \frac{f_X(x_i)}{\left|dg(x)/dx\right|\Big|_{x=x_i}} \tag{2.69}$$

Where S_X is the set of all real solutions of $y = g(x)$

$f_Y(y) = 0$ if no real solution of $y = g(x)$ exists

The theorem above can be used to find the probability density function $f_Y(y)$ for $Y = g(x)$, and it requires solving $Y = g(x)$ for each value of y from $-\infty$ to $+\infty$.

Example 2.10

Given a random variable X with probability density function $f_X(x)$, find the probability density function $f_Y(y)$ for a random variable Y defined by

$$Y = X^2$$

as shown in Figure 2.9. Work the problem for the following cases:

$$(a)\ f_X(x) = 0.5\exp\{-|x|\}.$$
$$(b)\ f_X(x) = \exp\{-x\}u(x).$$

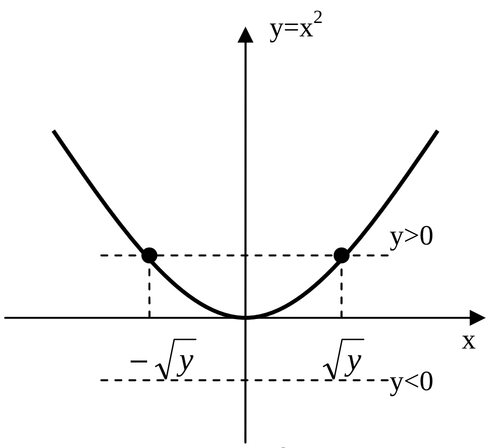


Figure 2.9 Transformation y=x² for Example 2.10

Solution

(a) Conditions are satisfied to use the fundamental theorem directly as $g(x)$ is a continuous function and X is a continuous random variable. For any $y < 0$, $y = x^2$ has no real solution (the dotted line for an arbitrary $y < 0$ doesn't intersect the curve $y = x^2$). Therefore $f_Y(y) = 0$ for those values of y.

If $y > 0$,then, $y = x^2$ has two real roots, $x_1 = +\sqrt{y}$ and $x_2 = -\sqrt{y}$. Using the fundamental theorem gives for $y > 0$.

$$f_Y(y) = \left.\frac{f_X(x)}{|2x|}\right|_{x=x_1=\sqrt{y}} + \left.\frac{f_X(x)}{|2x|}\right|_{x=x_1=-\sqrt{y}}$$

Using the two-sided-exponential density for $f_X(x)$, $f_Y(y)$ can be simplified to

$$f_Y(y) = \left(\frac{0.5e^{-|\sqrt{y}|}}{|2\sqrt{y}|} + \frac{0.5e^{-|-\sqrt{y}|}}{|-2\sqrt{y}|}\right)\mu(y) = \frac{1}{2\sqrt{y}}e^{-\sqrt{y}}\mu(y)$$

(b) Using $f_X(x) = \exp\{-x\}\mu(x)$, it is seen that for $y < 0$, $f_Y(y) = 0$, and that for $y > 0$,

$$f_Y(y) = \left.\frac{e^{-x}\mu(x)}{|2x|}\right|_{x=\sqrt{y}} + \left.\frac{e^{-x}\mu(x)}{|2x|}\right|_{x=-\sqrt{y}} = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}\mu(y)$$

If X is a continuous random variable and $g(x)$ is a continuous function except for a finite or countable number of flat spots, then the probability density for $Y = g(X)$ can be found by a slight modification of the fundamental theorem. The probability density function $f_Y(y)$ is composed of two parts. For all values of y such that there are no flat spots, find the density as given by the fundamental theorem, and then add to this sum of delta functions, one for each flat spot, at y_j with weight equal to the probability that the random variable X produces y_j after passing through $g(x)$. In Figure 2.10 a typical continuous $g(x)$ is shown that contains flat spots at y_j for values of x on the intervals $x_{j1} \le x \le x_{j2}$, respectively. In this notation the probability density function $f_Y(y)$ can be written as

$$f_Y(y) = \begin{cases} \displaystyle\sum_{x_j \in m} \frac{f_X(x_i)}{\left|dg(x)/dx\right|_{x_i}} + p(x_{j1} \le x \le x_{j2})\delta(y - y_j) \\ 0, \quad if\ no\quad real\quad solution\quad of\quad y = g(x)\quad exist \end{cases} \tag{2.70}$$

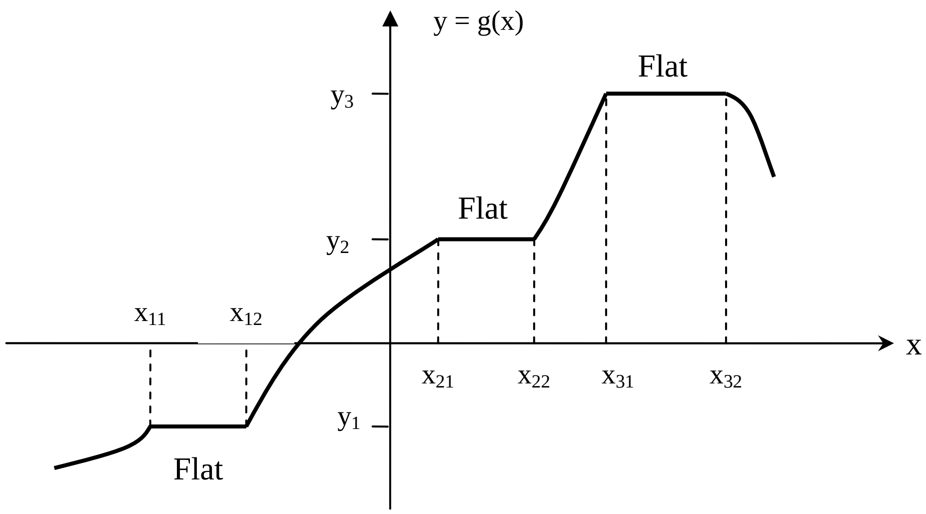


Figure 2.10 Function y = g(x) that contains flat spots

Example 2.11

Let X be a Gaussian random variable with zero mean and unity variance. The function $y = g(x)$ is defined in Figure 2.11, and the random variable Y is specified by the transformation $Y = g(X)$. Find the probability density function $f_Y(y)$ for Y.

Solution

In this example the function $y = g(x)$ has two flat spots, ones at $y = 9$ and the other at $y = 4$. From the graph of $g(x)$ it is seen that probability that $\{Y = 9\}$ is the probability that $\{X \le -3\}$, while the probability that $\{Y = 4\}$ equals the probability that $\{X \ge 0.2\}$. Using the Gaussian density and the parameters given, these probabilities become

$$P(Y = 9) = P(X \le -3) = \int_{-\infty}^{-3} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = \Phi(-3) = 0.0013$$

$$P(Y = 4) = P(X \ge 2) = \int_{2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 1 - \Phi(3) = 0.0228$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx$.

Treating the continuous part of the transformation separately it is seen that there are four different situations as y goes from $-\infty$ to ∞.

For $y < 0$, $y = g(x)$ has no real roots, therefore $f_Y(y) = 0$.

For $y > 9$, again no real roots so $f_Y(y) = 0$.

For $0 < y < 4$, there are two real roots at $+\sqrt{y}$ and $-\sqrt{y}$. Therefore $f_Y^{(c)}(y)$ is given by

$$f_Y^{(c)}(y) = \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-x^2/2\right)}{|2x|}\Bigg|_{x=\sqrt{y}} + \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-x^2/2\right)}{|2x|}\Bigg|_{x=-\sqrt{y}}$$
$$= \frac{\exp(-y/2)}{\sqrt{2\pi y}}$$

For $4 < y < 9$, $y = g(x)$ has omly one real root at $x = \sqrt{y}$, and by the fundamental theorem $f_Y^{(c)}(y)$ becomes

$$f_Y^{(c)}(y) = \frac{\exp(-y/2)}{2\sqrt{2\pi y}}$$

Using unit step functions, the continuous and discrete parts can be combined to give the final answer as

$$f_Y(y) = \frac{\exp(-y/2)}{2\sqrt{2\pi y}}[\mu(y) - \mu(y-4)] + \frac{\exp(-y/2)}{2\sqrt{2\pi y}}[\mu(y-4) - \mu(y-9)]$$
$$+ 0.0228\delta(y-4) + 0.0013\delta(y-9)$$

Probability Density Function (Mixed Random Variable).

If X is a mixed continuous variable, its probability density function $f_X(x)$ can be written as

$$f_X(x) = f_X^{(c)}(x) + \sum_i p[X = x_i]\delta(x - x_i) \tag{2.71}$$

where $f_X^{(c)}(x)$ represents the density without delta functions. If $g(x)$ is a continuous function, then $f_Y(y)$ can be obtained by adding the results from using the fundamental theorem on $f_X^{(c)}(x)$ to the results obtained by handling the impulses as previously described for transformation of discrete random variable. If $g(x)$ has flat spots, then the modification of the fundamental theorem can be used on $f_X^{(c)}(x)$ with the results added to the discrete part.

2.4.2 Cumulative Distribution Function

Assume that a random variable X is characterized by its probability density function $f_X(x)$ and that a random variable Y is defined by $Y = g(X)$, where $g(x)$ is a real-valued function. It is desired to find the cumulative distribution function $F_Y(y)$ directly, and then use it to obtain the $f_Y(y)$. Using the basic definition of a cumulative distribution function gives

$$F_Y(y) \stackrel{\Delta}{=} P\{Y \le y\} = P\{g(X) \le y\} = P\{x : g(x) \le y\} = \int_{I_y} f_X(x)dx \tag{2.72}$$

where $I_y = \{x : g(x) \le y\}$. This region is illustrated in Figure 2.12.

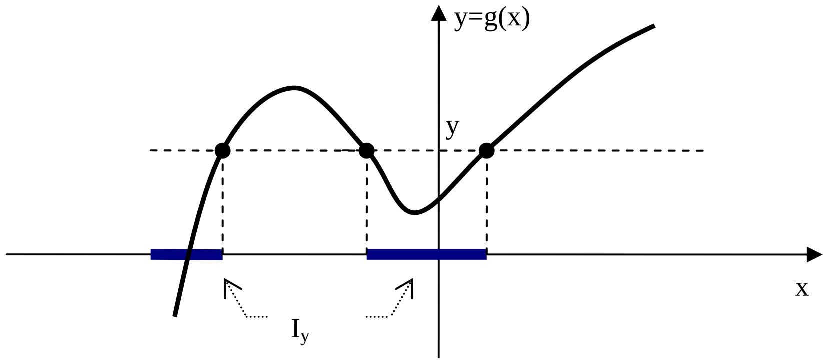


Figure 2.12 Region of integration I_y to obtain the cumulative distribution function.

Example 2.12

Assume that X is a random variable with probability density function $f_X(x)$ given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$$

Define the radom variable Y by $Y = g(X)$, where $g(x)$ is shown in Figure 2.15. Find the cumulative distribution function $F_Y(y)$ for the random variable Y.

Solution

To find $F_Y(y)$ using Eq.(2.72) it is necessary to identify the regions I_y for all values of y. for this problem it turns out that there are four distinct regions for different ranges of y as shown in Figure 2.13.

Region1. For $y < 0$, $\{x : g(x) \le y\}$ is the null set ϕ which results in $F_Y(y) = 0$.

Region2. For $0 < y < 4$, $I_y = \{x : g(x) < y\} = \{x : -\sqrt{y} \le x \le \sqrt{y}\}$. Therefore $F_Y(y)$ becomes

$$F_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

Region3. For $4 < y < 9$, $I_y = \{x : x \ge -\sqrt{y}\}$ and $F_Y(y)$ is

$$F_Y(y) = \int_{-\sqrt{y}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 1 - \Phi(-\sqrt{y})$$

Region4 .For $y > 9$, $I_y = \{all \quad x\}$; therefore $F_Y(y) = \int_{-\infty}^{\infty} F_x(x) dx = 1$.

Summarizing the results for the four regions above the cumulative distribution function $F_Y(y)$ becomes

$$F_Y(y)=\begin{cases}0, & y\le 0\\ \Phi(\sqrt{y})-\Phi(-\sqrt{y}), & 0<y<4\\ 1-\Phi(-\sqrt{y}), & 4<y<9\\ 1, & y\ge 9\end{cases}$$

which is illustrated in Figure 2.14.

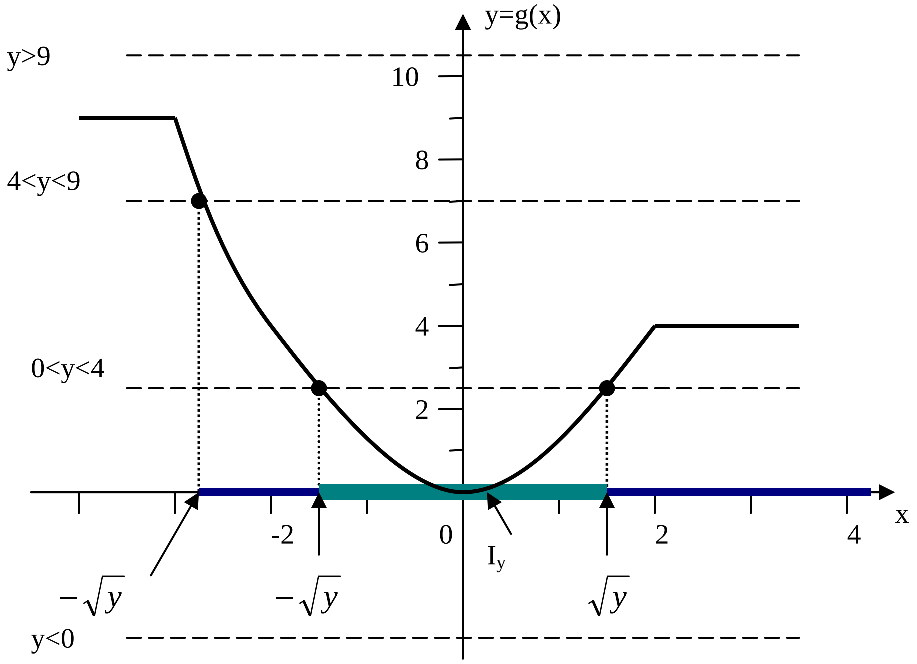


Figure 2.13 Region for Example 2-12.

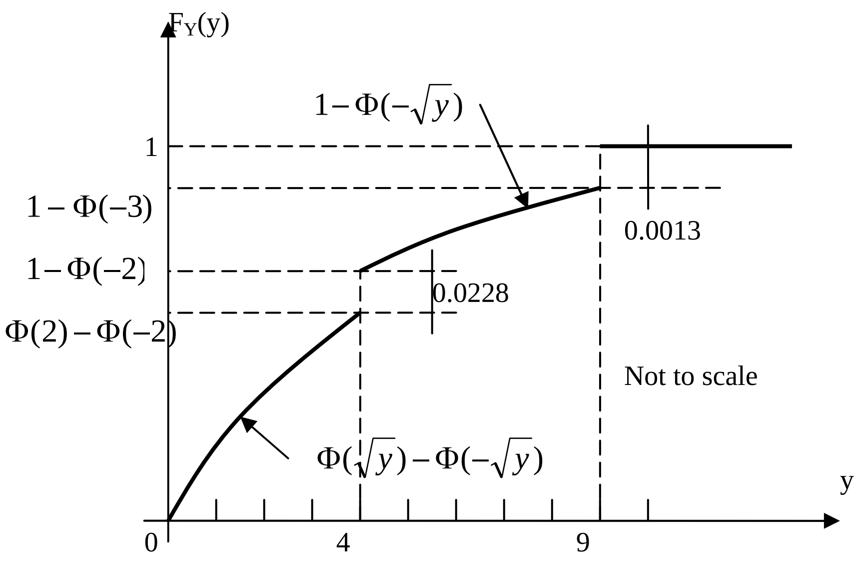


Figure 2.14 Cumulative distribution function for Example 2.12

When taking the derivation of an integral with respect to a variable that appears in the limits of the integral, it is actually expedient to use the **Liebnitz rule,** which in its most general form is

$$\frac{d}{dx}\left[\int_{\alpha(x)}^{\beta(x)} f(x,\gamma)d\gamma\right] = f\left(x,\beta(x)\right)\frac{d\beta(x)}{dx} - f\left(x,\alpha(x)\right)\frac{d\alpha(x)}{dx} + \int_{\alpha(x)}^{\beta(x)} \frac{d}{dx} f(x,\gamma)d\gamma \tag{2.73}$$

2.5 Computation of Expected Values

For a random variable X with probability density function $f_X(x)$ and a real-valued function $g(x)$, a random variable Y can be defined by $Y = g(x)$. There are three common ways to compute the expected value of $Y = g(x)$.

Method 1. By using $f_Y(y)$, which yields

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy \tag{2.74}$$

Method 2. By using $f_X(x)$, which yields

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{2.75}$$

Method 3. By using the Monte Carlo technique, which essentially synthetic sampling.

Method 3 represents an alternative to method1 and method 2 and it is not an analytical method. Monte Carlo Sampling is useful in problems where the integrals are of such a nature that analytical solutions are not easily obtained or do not exist. Roughly speaking, it is an experimental method using synthetic sampling from the density on X to generate samples x, computing samples $y = g(x)$, and then computing the average of those results to get the approximation for $E[Y]$.

Example 2.13

In Example 2.10 a random variable Y was defined as $Y = g(X) = X^2$ and $f_X(x)$ was given by

$$f_X(x) = \frac{1}{2}\exp\left(-|x|\right)$$

For this problem compute the $E[Y]$ using methods 1 and 2.

Solution

Method1. The density $f_Y(y)$ determined in Example 2.9 is used to compute the integral

$$E[Y]=\int_{-\infty}^{\infty} y f_Y(y)dy=\int_0^{\infty}\frac{ye^{-\sqrt{y}}}{2\sqrt{y}}dy$$

By making a change of variables $x=(y)^{\frac{1}{2}}$, $E[Y]$ becomes

$$E[Y]=\int_0^{\infty}x^2e^{-x}dx=\frac{e^{-x}}{-1}\left(x^2+2x+2\right)\Big|_0^{\infty}=2$$

Method2. the original density for X and $g(x)$ is used:

$$E[Y]=E[g(X)]=\int_{-\infty}^{\infty}g(x)f_X(x)dx=\int_{-\infty}^{\infty}x^2\frac{e^{-|x|}}{2}dx=\int_0^{\infty}x^2e^{-x}dx=2$$

As this example shows, the relationship between the two methods is a change of variables within the intergration. Clearly, if the probability density function for Y is not needed, its calculation is an unnecessary step in the process of computing the expected value and the second method would be preferred.

2.6 Two Random Variables

An experiment E is specified by the three-tuple $(S,F,P(\cdot))$ where S is a finite, countable, or noncountable set called the sample space, F is a Borel field specifying a set of events, and $P(\cdot)$ is a probability measure allowing calculation of probabilities of all events. Based on this underlying experiment, two random variables $X(e)$ and $Y(e)$ are defined as real-valued functions on the same S that satisfies the following conditions:

(a) $\{e: X(e)\le x \quad and \quad Y(e)\le y\}$ is a member of F for all x and y. This guarantees the existence of the cumulative distribution function.

(b) The probabilities of the events $\{e: X(e)=-\infty\}$, $\{e: Y(e)=+\infty\}$, $\{e: X(e)=+\infty\}$, and $\{e: Y(e)=-\infty\}$ are all zero. This means that the function is not allowed to be + or – infinity with a nonzero probability.

2.6.1 Joint Cumulative Distribution Function

The **joint cumulative distribution function** $F_{XY}(x,y)$ for a random variable $X(e)$ and $Y(e)$, represented by X and Y is defined for all x and y as

$$F_{XY}(x,y) = P\{e : X(e) \le x \cap Y(e) \le y\} \overset{\Delta}{=} P\{X \le x, Y \le y\} \quad (2.76)$$

It is sufficient information to calculate the probabilities of all allowable events, and thus is called a **total characterization.** Because of the properties of the probability measure $P(\cdot)$ described in the overall experiment, the joint cumulative distribution function can be shown to have a number of properties.

(1) $F_{XY}(x,y)$ is **bounded** from above and below,

$$0 \le F_{XY}(x,y) \le 1 \quad \text{for all } x \text{ and } y \quad (2.77)$$

(2) $F_{XY}(x,y)$ is a **nondecreasing** function of x and y,

$$F_{XY}(x_2,y) \ge F_{XY}(x_1,y) \quad \text{for all } x_2 > x_1\text{, all } x_1\text{, and all } y$$

$$F_{XY}(x,y_2) \ge F_{XY}(x,y_1) \quad \text{for all } y_2 > y_1\text{, all } y_1\text{, and all } x \quad (2.78)$$

(3) $F_{XY}(x,y)$ is **continuous from the right** in both x and y

$$\lim_{\substack{\varepsilon \to 0^+ \\ \delta \to 0^+}} F_{XY}(x+\varepsilon, y+\delta) = F_{XY}(x,y) \quad (2.79)$$

(4) $F_{XY}(x,y)$ can be used to **calculate probabilities** of rectangular events as

$$P\{x_1 < X \le x_2, y_1 < Y \le y_2\} = F_{XY}(x_2,y_2) - F_{XY}(x_1,y_1) \quad (2.80)$$

(5) $F_{XY}(x,y)$ is **related to the joint probability density function** by

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dxdy \quad (2.81)$$

Example 2.14

Random variables X and Y are defined in Figure 2.16. In this example X and Y are both discrete random variables. Find the joint cumulative distribution function $F_X(x)$, $F_Y(y)$, and $F_{XY}(x,y)$ for X and Y.

Solution

From the table it observed that X takes on values 0 and 1 with finite probabilities of 0.6 and 0.4, respectively, while Y takes on values 0, 1, and 2 with probabilities of 0.3, 0.4, and 0.3, respectively. Therefore $F_X(x)$ and $F_Y(y)$ are

$$F_X(x) = 0.6\mu(x) + 0.4\mu(x-1),\ F_Y(y) = 0.3\mu(x) + 0.4\mu(x-1) + 0.3\mu(x-2)$$

To obtain the joint distribution it is necessary to compute probability of the event $\{X \le x, Y \le y\}$ as x and y are both varied from $-\infty$ to ∞. This will be done by fixing x in different regions and determining the probability of $\{X \le x, Y \le y\}$ as y varies

ζ	$X(\zeta)$	$Y(\zeta)$	$P(\zeta)$
ζ_1	0	0	0.2
ζ_2	0	1	0.3
ζ_3	0	0	0.1
ζ_4	1	2	0.2
ζ_5	1	2	0.1
ζ_6	1	1	0.1

Figure 2.16 Random Variables X and Y for Example 2-14

For $x < 0$ and all y the set $\{X \le x, Y \le y\} = \phi$ so $F_{XY}(x,y) = 0$.

For $0 \le x \le 1$, there will be three different regions on the y=axis for which different values of the distribution are obtained as

$$
\begin{array}{ll}
y < 0 & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\phi) = 0 \\
0 \le y < 1 & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\zeta_1, \zeta_3) = 0.2 + 0.1 = 0.3 \\
1 \le y & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\zeta_1, \zeta_3, \zeta_2) = 0.2 + 0.1 + 0.1 = 0.4
\end{array}
$$

For $x \ge 1$, the results are similar to above but there are four different regions that give the joint cumulative distribution function as

$$
\begin{array}{ll}
y < 0 & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\phi) = 0 \\
0 \le y < 1 & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\zeta_1, \zeta_3) = 0.2 + 0.1 = 0.3 \\
1 \le y < 2 & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\zeta_1, \zeta_3, \zeta_2, \zeta_6) = 0.2 + 0.1 + 0.3 + 0.1 = 0.7 \\
2 \le y & F_{XY}(x,y) = P\{X \le x, Y \le y\} = P(\zeta_1, \zeta_3, \zeta_2, \zeta_6, \zeta_4, \zeta_5) = P(S) = 1
\end{array}
$$

The joint cumulative distribution function is shown in Figure 2.17

2.6.2 Joint Probability Density Function

The **joint probability density function** $f_{XY}(x,y)$ for the random variable X and Y is a total characterization and is defined as the derivative of the joint cumulative distribution function.

$$f_{XY}(x,y) \stackrel{\Delta}{=} \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) \tag{2.82}$$

If $F_{XY}(x,y)$ has jump discontinuities, it is convenient to use delta functions so that a joint probability density function will always be defined. Therefore a probability density function will contain delta functions at the points of discontinuities of $F_{XY}(x,y)$ with weights equal to the size of the jumps at those points.

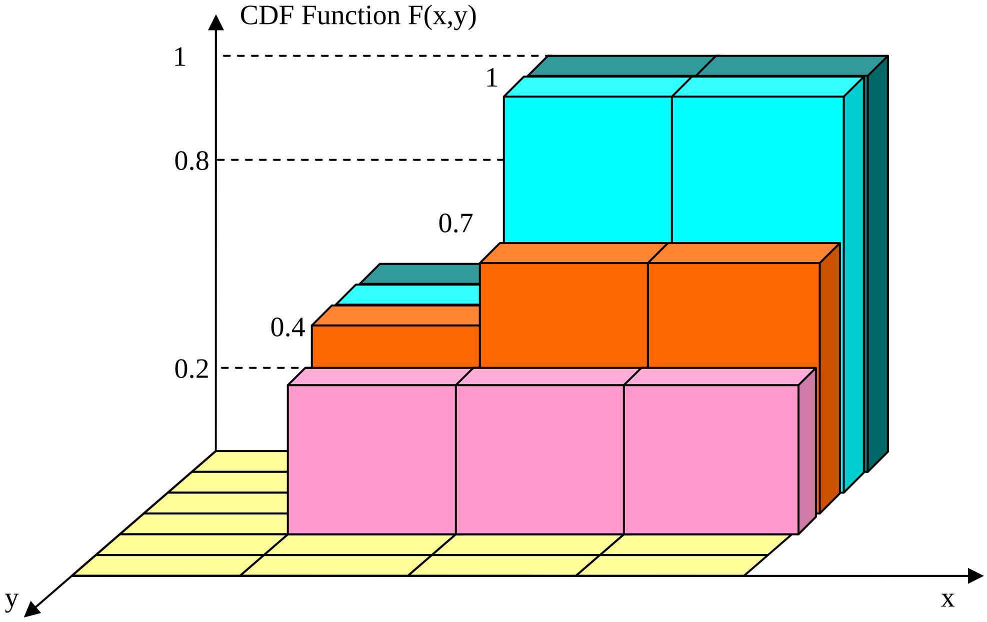


Figure 2.17 Cumulative distribution function $F_{XY}(x,y)$ for Example 2-14

Important properties of the probability density function are as follow:

(1) **Positivity**,

$$f_{XY}(x,y) \geq 0 \quad \text{for all } x \text{ and } y \tag{2.83}$$

(2) **Integral over all x and y**,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) = 1 \tag{2.84}$$

(3) $f_{XY}(x,y)$ can be used to **calculate probability** of rectangular events as

$$P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = \int_{x_1^+}^{x_2^+}\int_{y_1^+}^{y_2^+} f_{XY}(x,y)dxdy \tag{2.85}$$

Where x_1^+, x_2^+ , y_1^+ and y_2^+ are limits from the positive sides, or any event A as

$$P(\{X,Y\} \in A) = \iint_A f_{XY}(x,y)dxdy \tag{2.86}$$

(4)**Relationship to joint cumulative distribution function**,

$$\int_{-\infty}^{x^+}\int_{-\infty}^{y^+} f_{XY}(x,y)dxdy = F_{XY}(x,y) \tag{2.87}$$

Random variables are called **jointly discrete random variables** if their probability density function $f_{XY}(x,y)$ is a sum of two dimensional delta functions only, and correspondingly if its cumulative distribution function $F_{XY}(x,y)$ is a box staircase type function. Random variables are called **joitly continuous random variables** if their cumulative distribution function has no finite discontinuities or equivalently its probability density function $f_{XY}(x,y)$ has no delta functions. If random variables are neither jointly continuous or jointly discrete random variables, they are **jointly mixed random variables**.

Example 2.15

The random variables described in Example 2.14 are jointly discrete random variables. Give the joint probability density function $f_{XY}(x,y)$ for these random variables X and Y.

Solution

The joint density is seen to be

$$f_{XY}(x,y) = 0.3\delta(x,y) + 0.3\delta(x,y-1) + 0.3\delta(x-1,y-1) + 0.3\delta(x-1,y-2)$$

2.6.3 Partial Characterizations

Important partial characterizations for two random variables are the marginal densities, means, variances, covariances, higher-order joint moments, and joint central moments.

For two random variables X and Y, the **marginal densities** for X and Y are defined as the densities of X and Y by themselves and will be denoted by $f_X(x)$ and $f_Y(y)$ as before. These marginal densities can be obtained from the joint probability density function for the two random variances as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \tag{2.88}$$

The **conditional probability density functions** for X and Y are defined as

$$f_X(x \mid y) \overset{\Delta}{=} \frac{f_{XY}(x,y)}{f_Y(y)}$$
$$f_Y(y \mid x) \overset{\Delta}{=} \frac{f_{XY}(x,y)}{f_X(x)} \tag{2.89}$$

The means η_X and η_Y for two random variables are X and Y, using the conditional densities define the **conditional means** as

$$E[X \mid y] = \int_{-\infty}^{\infty} x f_X(x \mid y)dx \overset{\Delta}{=} E[X \mid Y = y]$$
$$E[Y \mid x] = \int_{-\infty}^{\infty} y f_X(y \mid x)dy \overset{\Delta}{=} E[Y \mid X = x] \tag{2.90}$$

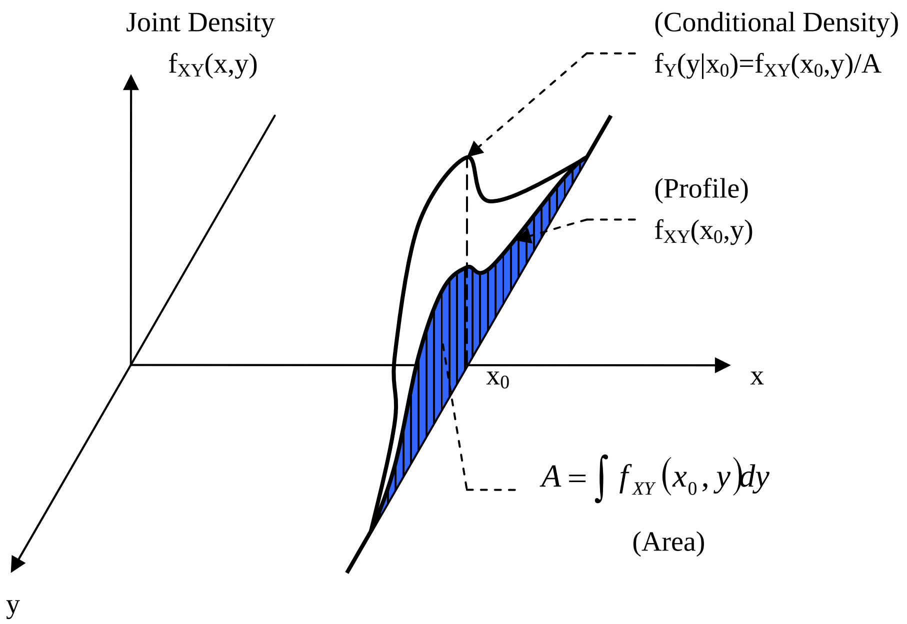


Figure 2.18 Conditional probability dendity functions

By rearranging Eq.(2.89), we can rewrite the joint probability density function in terms of the conditional and marginal densities as

$$f_{XY}(x,y)=f_Y(y|x)f_X(x)$$
$$f_{XY}(x,y)=f_X(x|y)f_Y(y) \qquad (2.91)$$

By substituting (2.91) into (2.88), alternative formulas for determining the marginal densities for X and Y can be found as follows:

$$f_X(x)=\int_{-\infty}^{\infty} f_X(x|y)f_Y(y)dy$$
$$f_Y(y)=\int_{-\infty}^{\infty} f_Y(y|x)f_X(x)dx \qquad (2.92)$$

Examining (2.91), we see that it is also possible to write the relationships between the two conditional densities as

$$f_X(x|y)=f_Y(y|x)f_X(x)/f_Y(y)$$
$$f_Y(y|x)=f_X(x|y)f_Y(y)/f_X(x) \qquad (2.93)$$

These formulas are comparable to **Bayes's rule**, which expresses the relationship between conditional probabilities, except that they represent the relationships between the conditional probability density functions and not probabilities.

Should either X or Y be discrete random variables, these results would need to be re-written. For example, let X be a discrete random variable and Y a continuous random variable with known $f(y|\{X=x_i\})$ for all x_i and known $f_X(x)$:

$$f_X(x)=\sum_{i=1}^{N}P\{X=x_i\}\delta(x-x_i) \tag{2.94}$$

Then it is easily shown, for all $j=1$ to N, that

$$P(\{X=x_j\}\,|\,y)=\frac{f(y\,|\,\{X=x_j\})P\{X=x_j\}}{\sum_{i=1}^{N}f(y\,|\,\{X=x_j\})P\{X=x_j\}} \tag{2.95}$$

Example 2.16

Given the joint probability density function $f_{XY}(x,y)$

$$f_{XY}(x,y)=\begin{cases}8xy, & 0<y<1, 0<x<y\\ 0, & elsewhere\end{cases}$$

Determine (a) $f_X(x)$, (b) $f_Y(y)$, and (c) $f_Y(y|x)$.

Sulution

(a) The $f_X(x)$ is obtained by integtating over all y at each value of x. for $x<0$ and $x>1$, $f_X(x)=0$ because $f_{XY}(x,y)$ being zero leads to a zero integral. For $0<x<1$, $f_X(x)$ is determined as

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)dy=\int_0^x 8xydy=4x^3$$

In summary $f_X(x)$ can be written as

$$f_X(x)=4x^3[\mu(x)-\mu(x-1)]$$

(b) $f_Y(y)$ is obtained by integtating over all x at each value of y. For $y<0$ and $y>1$, the $f_Y(y)=0$ because $f_{XY}(x,y)$ is zero leads to a zero integral. For $0<y<1$, $f_Y(y)$ is determined as

$$f_Y(y)=\int_{-\infty}^{\infty}f_{XY}(x,y)dx=\int_0^y 8xydx=4y^3$$

Thus $f_Y(y)$ can be written as

$$f_Y(y)=4y^3[\mu(y)-\mu(y-1)]$$

(c) The conditional density $f_Y(y|x)$ from (2.106) for $0<x<y$ and $0<x<1$ is

$$f_Y(y|x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{8xy}{4x^3}=\frac{2y}{x^2}$$

For $y>x$ the joint density is zero, and thus $f_Y(y|x)$ can be summarized as

$$f_Y(y|x)=\begin{cases}\dfrac{2y}{x^2}, & 0<y<x\\ 0, & elsewhere\end{cases}$$

Several plots of $f_Y(y|x)$ for various values of x are shown in Figure 2.19. Notice the different domains for nonzero values.

Example 2.17

Find $E[Y|x]$ for the random variables X and Y that have the joint probability density function given in Example 2.16.

Solution

Using the conditional density determined in Example 2.16 and (2.90), the conditional expected value is determined for $0 < x < 1$ as

$$E[Y|x] = \int_{-\infty}^{\infty} y f_Y(y|x) dy = \int_0^x y \frac{2y}{x^2} dy = \frac{2}{3} \frac{y^3}{x^2}\Big]_0^x = \frac{2}{3} x$$

For x outside the interval [0,1], the density is zero as is the expected value.

Two random variables X and Y are defined to be **statistically independent** if the joint probability density function can be written as a product of the two marginal densities for all x and y as shown:

$$f_{XY}(x,y) = f_X(X) f_Y(y) \tag{2.96}$$

The **correlation** between two random variables X and Y is defined as the expected value of their product;

$$R_{XY} = E[XY] \tag{2.97}$$

Two random variables are defined as **uncorrelated** if $R_{XY} = E[X]E[Y]$,which is the same as writing

$$E[XY] = E[X]E[Y] \tag{2.98}$$

Using Eq.(2.96), it is simple to show that if two random variables are statistically independent, then they are uncorrelated. The steps in the proof follow.

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(xy) dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dxdy = \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy = E[X]E[Y] \tag{2.99}$$

Two random variables X and Y are called **orthogonal** if $E[XY] = 0$.

If the random variables X and Y are independent and both of the means are zero, or one of the means is zero and the other is finite, then they are orthogonal since by independence the $E[XY] = E[X] \cdot E[Y]$ and thus $E[X] \cdot E[Y] = 0$. These relationships between independence, orthogonality and uncorrelated random variables are show in Figure2.20. In the figure the single solid arrow indicates always true, whereas a dotted arrow indicates always true if the conditions written along side are true. The arrow back from uncorrelated to independence will be shown later to be true for jointly Gaussian random variables.

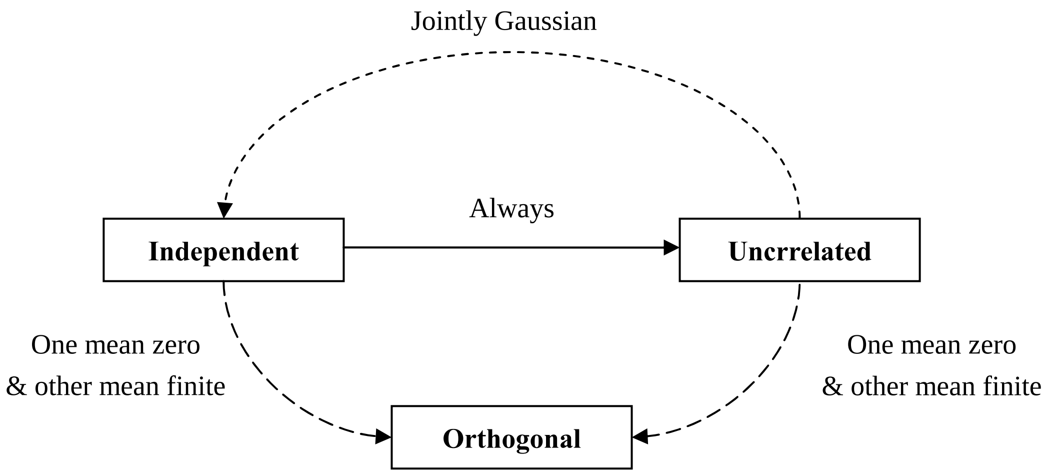


Figure 2.20 Relationship between the definitions of independent, uncorrelated, and orthogonal.

With two random variables the individual variances are as before for each random variable, but more information considering the relationship between X and Y is available through the **covariance** σ_{XY} defined as

$$\sigma_{XY} \stackrel{\Delta}{=} E[(X-\eta_X)(Y-\eta_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\eta_X)(y-\eta_Y)f_{XY}(x,y)dxdy \tag{2.100}$$

The **correlation coefficient** ρ_{XY} between the two random variables X and Y is defined as the normalized covariance by

$$\rho_{XY} \stackrel{\Delta}{=} \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.101}$$

It provides a measure of linear relationship between the two random variables X and Y. the correlation coefficient for two random variables X and Y can be easily shown to be bounded by $+1$ and -1 as

$$-1 \le |\rho_{XY}| \le 1 \tag{2.102}$$

The closer ρ_{XY} is to -1 or $+1$ the more the random variables X and Y are said to be linearly related. It is noticed that the closer ρ_{XY} is to -1 or $+1$, the more ridgelike the probability density becomes, which indicates that one random variable can be written as almost a scaler multiple of the other.

The **higher-order moments** m_{jk} and the **central moments** μ_{jk} for two random variables X and Y are defined by

$$m_{jk} \stackrel{\Delta}{=} E[X^jY^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^j y^k f_{XY}(x,y)dxdy \tag{2.103}$$

$$\mu_{jk} \stackrel{\Delta}{=} E[(X-\eta_X)^j(Y-\eta_Y)^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\eta_X)^j(y-\eta_Y)^k f_{XY}(x,y)dxdy \tag{2.104}$$

2.6.4 Jointly Normal Random Variables

The random variables X and Y are defined as **jointly Gaussian or jointly normal** if their probability density function $f_{XY}(x,y)$ has the following form:

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y r}\exp\left\{-\frac{1}{2r}\left[\frac{(x-\eta_X)^2}{\sigma_X^2}-\frac{2\rho_{XY}(x-\eta_X)(y-\eta_Y)}{\sigma_X\sigma_Y}+\frac{(y-\eta_Y)^2}{\sigma_Y^2}\right]\right\} \qquad (2.105)$$

Where η_X, η_Y, σ_X^2, σ_Y^2, and ρ_{XY} are, respectively, the mean of X, mean of Y, variance of X, variance of Y, and correlation coefficient of X and Y, and $r=\sqrt{1-\rho_{XY}^2}$. The marginal densities or densities of the random variables X and Y can be obtained by integrating out the proper variable to give

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\{-\frac{(x-\eta_X)^2}{2\sigma_X^2}\}, \qquad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\{-\frac{(y-\eta_Y)^2}{2\sigma_Y^2}\} \qquad (2.106)$$

By dividing the joint density by the marginal density and simplifying the conditional probability density function for Y conditional on $\{X=x\}$ and X conditional on $\{Y=y\}$, the conditional probability density function can be determined as

$$\begin{aligned} f_Y(y\mid x)&=\frac{1}{\sqrt{2\pi}\sigma_{Y|x}}\exp\{-\frac{(x-\eta_{Y|x})^2}{2\sigma_{Y|x}^2}\}, & \eta_{Y|x}&=\eta_Y+\frac{\rho_{XY}\sigma_Y}{\sigma_X}(x-\eta_X)\\ & & \sigma_{Y|x}^2&=\sigma_Y^2(1-\rho_{XY}^2)\\ f_X(x\mid y)&=\frac{1}{\sqrt{2\pi}\sigma_{X|y}}\exp\{-\frac{(x-\eta_{X|y})^2}{2\sigma_{X|y}^2}\}, & \eta_{X|y}&=\eta_X+\frac{\rho_{XY}\sigma_X}{\sigma_Y}(y-\eta_Y)\\ & & \sigma_{X|y}^2&=\sigma_X^2(1-\rho_{XY}^2)\end{aligned} \qquad (2.107)$$

When the means are zero, the conditional means and variance are easily obtained from the above by setting $\eta_X=\eta_Y=0$ to give

$$\begin{aligned} f_Y(y\mid x)&=\frac{1}{\sqrt{2\pi}\sigma_{Y|x}}\exp\{-\frac{(x-\eta_{Y|x})^2}{2\sigma_{Y|x}^2}\}, & \eta_{Y|x}&=\frac{\rho_{XY}\sigma_Y}{\sigma_X}x\\ & & \sigma_{Y|x}^2&=\sigma_Y^2(1-\rho_{XY}^2)\\ f_X(x\mid y)&=\frac{1}{\sqrt{2\pi}\sigma_{X|y}}\exp\{-\frac{(x-\eta_{X|y})^2}{2\sigma_{X|y}^2}\}, & \eta_{X|y}&=\frac{\rho_{XY}\sigma_X}{\sigma_Y}y\\ & & \sigma_{X|y}^2&=\sigma_X^2(1-\rho_{XY}^2)\end{aligned} \qquad (2.108)$$

Note that although X and Y have zero means, the conditional means are not equal to zeros.

A number of useful formulas for various moments of jointly normal zero mean random variable in terms of their given parameters σ_X, σ_Y, and ρ_{XY} follow:

$$
\begin{aligned}
&E[XY] = \rho_{XY}\sigma_X\sigma_Y \\
&E[X^2Y^2] = \sigma_X^2\sigma_Y^2 + 2\rho_{XY}\sigma_X^2 \\
&E[X^2] = \sigma_X^2, \qquad E[Y^2] = \sigma_Y^2, \\
&E[X^k] = E[Y^k] = 0, \quad k \quad odd \\
&E[X^4] = 3\sigma_X^4, \qquad E[Y^4] = 3\sigma_Y^4, \\
&E[|X| \cdot |Y|] = \frac{2\sigma_X\sigma_Y}{\pi}(\cos\alpha + \alpha\sin\alpha)
\end{aligned} \tag{2.109}
$$

Where $\sin\alpha = \rho_{XY}, -\frac{\pi}{2} < \alpha < \frac{\pi}{2}$

2.7 Two Function of Two Random Variables

The basic problem is to obtain the statistical characterization of two random variables Z and W that are functions of two other random variables X and Y. The expressions for Z and W are specified by

$$W = g(X,Y) \qquad \text{and} \qquad Z = h(X,Y) \tag{2.110}$$

2.7.1 Probability Density Function (Discrete Random Variables)

If X and Y are discrete random variables, their joint probability density function $f_{XY}(x,y)$ is composed of two-dimensional impulses:

$$f_{XY}(x,y) = \sum_{(x_i,y_i)} p(x_i,y_i)\delta(x-x_i, y-y_i) \tag{2.111}$$

Then the joint probability density function for $W = g(X,Y)$ and $Z = h(X,Y)$ will also contain only impulses with locations obtained by taking (x_i, y_i) through the function $w = g(x,y)$ and $z = h(x,y)$, and with weights equal to the probabilities of each (x_i, y_i). The $f_{WZ}(w,z)$ is therefore seen to be

$$f_{WZ}(w,z) = \sum_{(x_i,y_i)} p(x_i,y_i)\delta(w-g(x_i,y_i), z-h(x_iy_i)) \tag{2.112}$$

Example 2.18

Given the following probability density function for the discrete random variables X and Y :

$$f_{XY}(x,y) = \frac{1}{4}\delta(x-1,y-1) + \frac{1}{2}\delta(x,y) + \frac{1}{6}\delta(x-1,y) + \frac{1}{12}\delta(x,y+1)$$

Consider two new random variables W and Z that are defined by the following transformations:

$$W = X^2 + Y^2, Z = X^2$$

Find the joint probability density function $f_{WZ}(w,z)$.

Solution

Taking each delta function through the transformation gives the joint density $f_{WZ}(w,z)$ as

$$f_{WZ}(w,z)=\frac{1}{4}\delta(w-2,z-1)+\frac{1}{2}\delta(w,z)+\frac{1}{6}\delta(w-1,z)+\frac{1}{12}\delta(w-1,z)$$
$$=\frac{1}{4}\delta(w-2,z-1)+\frac{1}{2}\delta(w,z)+\frac{1}{4}\delta(w-1,z)$$

2.7.2 Probability Density Function (Continuous Random Variables and Continuous Functions)

If X and Y are continuous random variables, and $g(x,y)$ and $h(x,y)$ are continuous functions of x and y with no flat spots, the counterpart to the one function of one random variable result can be written as follows for random variables $W=g(X,Y)$ and $Z=h(X,Y)$

$$f_{WZ}(w,z)=\sum_{(x_i,y_i)}\left.\frac{f_{XY}(x,y)}{|J(x,y)|}\right|_{(x,y)=(x_i,y_i)} \tag{2.113}$$

Where (x_i,y_i) are all real solutions of $w=g(x,y)$ and $z=h(x,y)$. The $J(x,y)$ is the Jacobian of the transformation, which is expressed as the following determinant:

$$J(X,Y)=\begin{vmatrix}\frac{\partial g(x,y)}{\partial x} & \frac{\partial g(x,y)}{\partial y}\\ \frac{\partial h(x,y)}{\partial x} & \frac{\partial h(x,y)}{\partial y}\end{vmatrix} \tag{2.114}$$

If no real solution of $w=g(x,y)$ and $z=h(x,y)$ exists for a given w and z, then the joint probability density $f_{WZ}(w,z)=0$ for those values of w and z.

Example 2.19

Let W and Z be two random variables defined by the following functions of two other random variables X and Y:

$$W=X^2+Y^2 andZ=X^2$$

Let X and Y be independent Gaussian random variable described by $X\sim N(0,1)$ and $Y\sim N(0,1)$.

(a) Find the joint probability density function for the random variables X and Y.

(b) Are X and Y statistically independent?

Solution

(a) The solution is obtained by using Eqs.(2.113) and (2.114). The Jacobian for the transformation defined is

$$J(x,y)=\begin{vmatrix}\dfrac{\partial(x^2+y^2)}{\partial x} & \dfrac{\partial(x^2+y^2)}{\partial y}\\ \dfrac{\partial(x^2)}{\partial x} & \dfrac{\partial(x^2)}{\partial y}\end{vmatrix}=\begin{vmatrix}2x & 2y\\ 2x & 0\end{vmatrix}=4|xy|$$

We must find all real solutions (x,y) of the equations $w=x^2+y^2$ and $z=x^2$ for all w and z.

For $w>0$ and $z>0$ the second equation can be solved for x in terms of z as $x=\pm\sqrt{z}$. Substituting this result into the first equation allows us to solve for y as $y=\pm\sqrt{(w-z)}$. Since y will not be real if $(w-z)<0$, there is no real solution for $z>w$, so by the transformation theorem the joint density $f_{WZ}(w,z)=0$.

For there are no real solutions for any $w<0$ or $z<0$, since $w=x^2+y^2$ and $z=x^2$ can never be negative, thus $f_{WZ}(w,z)=0$ for those regions of w and z. In summary, there are only real roots for $w>0$, $z>0$ and $z<w$, and for that case there are four of them as combinations of the following: $x=\pm\sqrt{z}$, $y=\pm\sqrt{(w-z)}$.

By Eq.(2.113), the joint density for w and z becomes

$$f_{WZ}(w,z)=\left.\frac{f_{XY}(x,y)}{4|xy|}\right|_{\substack{x=\sqrt{z}\\ y=\sqrt{(w-z)}}}+\left.\frac{f_{XY}(x,y)}{4|xy|}\right|_{\substack{x=\sqrt{z}\\ y=-\sqrt{(w-z)}}}+\left.\frac{f_{XY}(x,y)}{4|xy|}\right|_{\substack{x=-\sqrt{z}\\ y=\sqrt{(w-z)}}}+\left.\frac{f_{XY}(x,y)}{4|xy|}\right|_{\substack{x=-\sqrt{z}\\ y=-\sqrt{(w-z)}}}$$

Substituting these roots into the following density function

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

and collecting like terms (they are all the same because of the squaring operation in the numerator and the Gaussian density), the $f_{WZ}(w,z)$ is easily seen to be

$$f_{WZ}(w,z)=\frac{e^{-w/2}}{2\pi\sqrt{z(w-z)}}\mu(w-Z)\mu(w)\mu(z)$$

(b) The random variables W and Z are not statistically independent because the joint density is not a product of the marginal densities. This can be easily seen by finding the marginal densities and multiplying them. However, it is easier to see that the product cannot be the same as the product would be nonzero for all $w>0$ and $z>0$ whereas the joint density is zero for $z>w$ and $w>0$.

Example 2.20

Let Z and W be random variables defined by the following functions of two other random variables X and Y:

$$W = X + Y$$
$$Z = 2(X + Y)$$

Assume that $f_{XY}(x, y)$ is known and that it is required to find the joint density funciton $f_{WZ}(w, z)$.

Solution

The transformation theorem cannot be used since the Jacobian is given by

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 0$$

regardless of the roots of $w = x + y$ and $z = 2(x + y)$. Certainly W and Z are random variables, so they must have a probability density function. Notice that $Z = 2W$, that is, Z can be written as a linear function of W. Thus, as W takes on values w, Z takes on only the values $z = 2w$. In the $w - z$ two-dimensional space, the joint density must be zero for (z, w), not on the line, and have meaning along the line $z = 2w$ as shown in Figure 2.22. This is what is meant by a line mass. Using the distribution function method, the derivative yields the joint density function as

$$f_{WZ}(w, z) = f_W(w)\delta(z - 2w)$$

where $f_W(w)$ is obtained by using the one function of two random variables approach.

2.7.3 Distribution Function (Continuous, Discrete, or Mixed)

The problem of finding the joint cumulative distribution function for the random variables W and Z defined by

$$W = g(X, Y) \quad Z = h(X, Y) \tag{2.117}$$

$F_{WZ}(w, z)$ can be written as

$$F_{WZ}(w, z) = P\{W \le w, Z \le z\} = P\{g(X, Y) \le w, h(X, Y) \le z\} \tag{2.118}$$

The $P\{g(x, y) \le w, h(x, y) \le z\}$ can be obtained by integrating the probability density function over the region I_{WZ} defined in the x - y space as

$$I_{WZ} = \{(x, y) : g(x, y) \le w \quad and \quad h(x, y) \le z\} \tag{2.119}$$

Using this I_{WZ} and $F_{WZ}(w, z)$ can now be written as

$$F_{WZ}(w, z) = \int\int_{I_{WZ}} f_{XY}(x, y) dxdy \tag{2.120}$$

This approach can be used regardless whether the probability density function is from a discrete continuous or mixed random variable. If $f_{WZ}(w, z)$ is desired, it can be obtained by taking partial derivatives as follows:

$$f_{WZ}(w,z)=\frac{\partial}{\partial w}\frac{\partial}{\partial z}F_{WZ}(w,z) \tag{2.121}$$

Special care is required if $F_{WZ}(w,z)$ is discontinuous, since various types of impulse functions will be generated.

Example 2.23

Define two random variables W and Z by the following transformations of two random variables X and Y : $W = X^2 + Y^2$ and $Z = X^2$. Assume that X and Y are independent random variables characterized by their joint probability density function $f_{XY}(x,y)$, which is given by $f_{XY} = e^{-(x+y)}\mu(x)\mu(y)$. Find the joint cumulative distribution function $F_{WZ}(w,z)$ for W and Z, and then take partial derivitives to obtain the joint probability density function $f_{WZ}(w,z)$.

Solution

Using Eq.(2.120) to obtain the cumulative distribution function depends on first obtaining the region I_{WZ}, which for our problem is defined by

$$I_{WZ}=\{(x,y): x^2+y^2 \le w and x^2 \le w\}$$

and is shown in Figure 2.26.

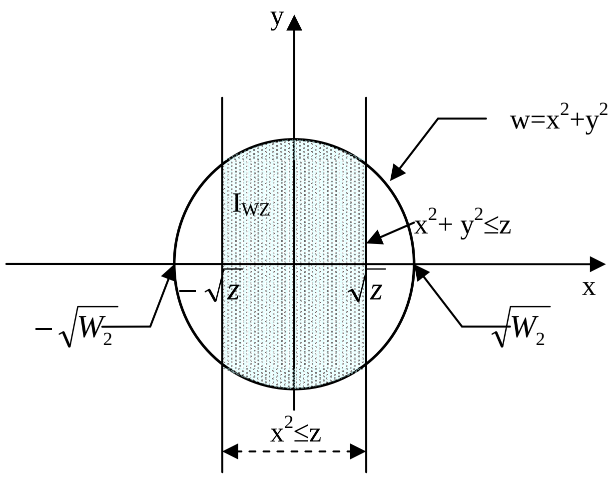


Figure 2.26 The region of integration I_{WZ} for Example 2.23

For $w<0$ and $z<0$, $I_{WZ}=\phi$, the null set. Clearly, the integral over that region gives zero, and so $F_{WZ}(w,z)=0$.

For $w \ge 0$ and $z \ge 0$, the region is shown as the shaded area in Figure 2.26. Thus the $F_{WZ}(w,z)$ can be written in terms of the integral of the joint probability density function for the random variables X and

Y as follows:

$$F_{WZ}(w,z) = \int_{-\sqrt{z}}^{\sqrt{z}} \left[\int_{-\sqrt{w-x^2}}^{\sqrt{w-x^2}} \exp-(x+y)\mu(x)\mu(y)dy \right] dx$$

$$= \int_0^{\sqrt{z}} e^{-x} \left[\int_0^{\sqrt{w-x^2}} e^{-y} dy \right] dx$$

This expression can be simplified to give

$$F_{WZ}(w,z) == \int_0^{\sqrt{z}} e^{-x} \left(1 - e^{-\sqrt{w-x^2}}\right) dx \mu(w)\mu(x)$$

The joint probability density function $f_{WZ}(w,z)$ is then obtained by taking the partial derivatives of $F_{WZ}(w,z)$ with respect to w and z as follows:

$$f_{WZ}(w,z) = \frac{\partial^2}{\partial w \partial z} F_{WZ}(w,z) = \frac{\partial}{\partial w} \left(\frac{\partial}{\partial z} \int_0^{\sqrt{z}} e^{-x} \left(1 - e^{-\sqrt{w-x^2}}\right) dx \right) \mu(w)\mu(x)$$

$$= \frac{\partial}{\partial w} \left(e^{-\sqrt{z}} \left(1 - e^{-\sqrt{w-z}}\right) \frac{1}{2} z^{-\frac{1}{2}} \right) \mu(w)\mu(x)$$

In the equation above the partial with respect to z was obtained by Liebnitz's rule. Continuing and taking the partial with respect to w and simplifying gives

$$f_{WZ}(w,z) = \frac{e^{-\sqrt{w-z}} e^{-\sqrt{z}}}{4\sqrt{w-z}\sqrt{z}} \mu(w)\mu(x)$$

This result can be easily checked by using the transformation theorem for two functions of two random variables.

2.8 One Function of Two Random Variables

Define a random variable Z by $Z = g(X,Y)$, where X and Y are random variables. The basic problem is that of characterizing Z, knowing either partial or tital characterizations of X and Y.

2.8.1 Probability Density Function (Discrete Random Variables)

If X and Y are discrete random variables, their joint probability density function $f_{XY}(x,y)$ is composed of two-dimensional delta functions as

$$f_{XY}(x,y) = \sum_{(x_i,y_i)} P(x_i,y_i)\delta(x - x_i, y - y_i) \qquad (2.122)$$

Where $P(x_i,y_i) = P\{x = x_i, y = y_i\}$ and S is the set of all pairs (x_i,y_i) considered. If $g(x,y)$

represents a function defined on real x and y, then the random variable $Z = g(X,Y)$ is a discrete random variable with probability density function $f_Z(z)$ given by

$$f_Z(z) = \sum_{(x_i, y_i)\in S} P(x_i, y_i)\delta(z - g(x_i, y_i)) \tag{2,123}$$

Some of the $g(x_i, y_i)$ will be the same, so adding those probabilities the $f_Z(z)$ can be written in terms of a reduced number of terms. If z_k is an element from the set of unique values of $g(x_i, y_i)$, where $(x_i, y_i) \in S$, then $f_Z(z)$ can be rewritten as

$$f_Z(z) = \sum_{z_k \in K} P(z_k)\delta(z - z_k) \tag{2.124}$$

Where $P(z_k) = P\{(x_i, y_i) : g(x_i, y_i) = z_k\}$ and K is the set of all unique value of $g(x_i, y_i)$.

2.8.2 Probability Density Function (Continuous Random Variables)

Let X and Y be continuous random variables with known joint probability density function $f_{XY}(x, y)$ and $g(x, y)$ be a known continuous real-valued function of x and y. The basic problem is that of determining the probability density function $f_Z(z)$ for the random variable defined by $Z = g(X,Y)$. Four important basic methods for finding the density are (1) the cumulative distribution approach, (2) the auxiliary random variable approach, (3) the incremental probability approach and (4) the synthetic sampling or Monte Carlo approach. Of these only the synthetic sampling approach would be considered a direct approach.

Cumulative Distribution approach. In this approach the cumulative distribution function $F_Z(z)$ is found and then differentiated to obtain $f_Z(z)$. The distribution function $F_Z(z)$ is determined as follow:

$$F_Z(z) = P(Z \le z) = P\{g(X,Y) \le z\} = \iint_{I_z} f_{XY}(x, y)dxdy \tag{2.125}$$

Where $I_z = \{(x, y) : g(x, y) \le z\}$. To obtain $F_Z(z)$ for z from $-\infty$ to ∞, $f_{XY}(x, y)$ must be integrated over changing regions of integration. The regions are composed of all (x, y) such that $g(x, y) \le z$ with boundaries equal to the intersection of a plane at z and the function as shown in Figure 2.27. The probability density function $f_Z(z)$ is then obtained by differentiating $F_Z(z)$ to give

$$f_z(z) = \frac{d}{dz}F(z) \tag{2.126}$$

For certain problems, whose regions can be expressed easily in the limits if the two-dimensional integral, Liebnitz's rule can be applied so that the integral may never need to be evaluated. However, for other problems the method is a tour de force using two-dimensional integration.

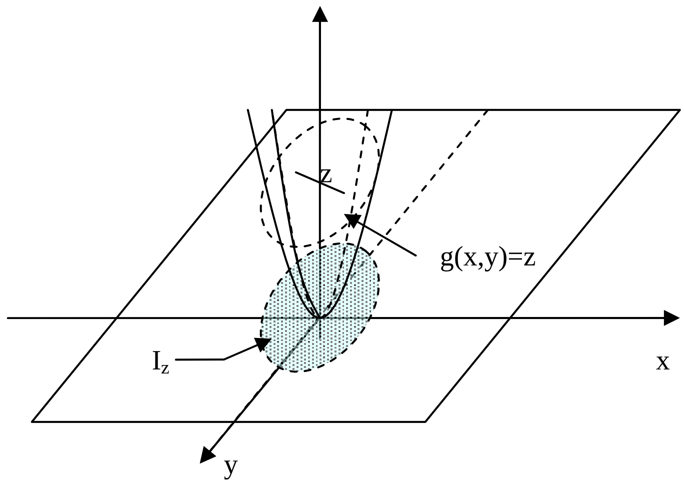


Figure 2.27 Obtaining the region of integration for the distribution method.

Example 2.24

Let X and Y be a pair of random variables characterized by their joint probability density function $f_{XY}(x,y)$. Define a new random variable Z by $Z = X + Y$. Find the probability density function $f_Z(z)$ for Z by using the distribution method and determine the answer for the special case where X and Y are independent random variables.

Solution

To find the distribution function $F_Z(z)$ the regions I_z must be identified for all z and are given by

$$I_Z = \{(x,y): x + y \le z\}$$

Thus the region is seen to be all points (x,y) below and including the boundary line $x + y = z$, or in standard form $y = z - x$ as seen in Figure 2.28.

Thus we can write $F_Z(z)$ for an arbitrary z as the following integral:

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{XY}(x,y)dxdy$$

The probability density function is then given by

$$f_z(z) = \frac{d}{dz}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{XY}(x,y)dxdy\right]$$

Interchanging the first integral and the derivative and applying Liebnitz's rule to the second integral gives the answer as

$$f_z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)dx \tag{2.127}$$

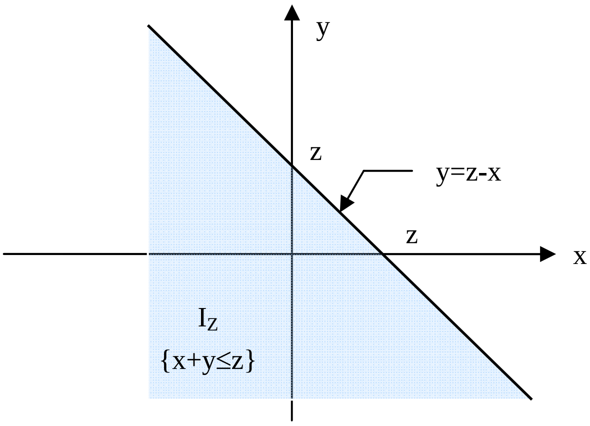


Figure 2.28 Region I_Z for Example 2.24

If X and Y are independent random variables, then $f(x,y)=f_X(x)f_Y(y)$. For this important case the $f_Z(z)$, for the random variable $Z=X+Y$, is determined from Eq.(2.127) to be the convolution integral as follows:

$$f_z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = f_X(z)*f_Y(z) \tag{2.128}$$

Auxiliary Random Variable Approach. The auxiliary random variable approach is a method for obtaining the probability density function $f_Z(z)$ for the random variable $Z=h(x,y)$. The approach consists three basic steps: defining an auxiliary random variable, using the thansformational theorem to obtain a joint probability density function, and an integration to obtain the marginal density for Z and thus the desired density.

In the first step a random variable W is defined by $W=g(X,Y)$ this is a fictitious random variable and selected as a vehicle for using the two-dimensional transformation theorem. Usually $g(X,Y)$ is selected as just X, just Y, or a simple function of X and Y that might be suggested by the form of the function $h(X,Y)$ which defines Z.

The second step is the application of the two-dimensional transformation theorem to the following:

$$\begin{aligned} W &= g(X,Y) \\ Z &= h(X,Y) \end{aligned} \tag{2.129}$$

To find the joint probability density $f_{WZ}(w,z)$ which is given by

$$f_{WZ}(w,z)=\sum_{(x_i,y_i)} \left.\frac{f_{XY}(x,y)}{|J(x,y)|}\right|_{(x,y)=(x_i,y_i)} \tag{2.130}$$

Where (x_i,y_i) are all real solution of $w=g(x,y)$ and $z=h(x,y)$.

The final step in the process of obtaining the probability density function $f_Z(z)$ is to integrate the joint density function $f_{WZ}(w,z)$ as follows:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{WZ}(w,z)dw \qquad (2.131)$$

Example 2.25

Let X and Y be a pair of random variables characterized by their joint probability density function $f_{XY} = (x,y)$. Define a new random variable Z by $Z = X + Y$ (same as in Example 2.24).Find the probability density function $f_Z(z)$ for z by using the auxialary random variable method.

Solution

Define the auxiliary random $W = X$, which now gives us two functions of two random variables $W = X$ and $Z = X + Y$. Using the two functions of two random variables theorem the joint probability density $f_{WZ}(w,z)$ is given by Eq.(2.130). There is only one solution to the set of equations, $w = x$ and $z = x + y$, which are easily seen to be $x = w$ and $y = z - w$. The Jacobian of the transformations is determined as

$$J(x,y) = \begin{vmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\ \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{vmatrix} = \begin{vmatrix} \dfrac{\partial x}{\partial x} & \dfrac{\partial x}{\partial y} \\ \dfrac{\partial (x+y)}{\partial x} & \dfrac{\partial (x+y)}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

Therefore the joint density can be written as

$$f_{WZ}(w,z) = f_{XY}(x,y)\big|_{\substack{x=w \\ y=z-w}} = f_{XY}(w, z-w)$$

The desired probability density function, $f_W(w)$, is obtained by integrating the joint density $f_{WZ}(w,z)$ to give the result

$$f(z) = \int_{-\infty}^{\infty} f_{WZ}(w,z)dw = \int_{-\infty}^{\infty} f_{XY}(w, z-w)dw$$

Incremental Approach. The incremental approach can be used for certain problems of finding the probability density function $f_Z(z)$ for random variable Z defined by $Z = h(X,Y)$. The method uses the fact that $f_Z(z)\Delta z$ can be approximated by

$$f_Z(z)\Delta z \approx P\{z < Z \le z + \Delta z\} \qquad (2.132)$$

Where $P\{z < Z \le z + \Delta z\}$ is equivalent to the probability that (x,y) is a member of $I_{\Delta z} = \{(x,y) : z < h(x,y) \le z + \Delta z\}$. Thus the approximation above can be written as

$$f_Z(z)\Delta z \approx \iint_{I_{\Delta z}} f_{XY}(x,y)dxdy \qquad (2.133)$$

The approach relies on the assumption that as $\Delta z \to 0$ the integral over $I_{\Delta z}$ becomes a linear function of Δz. When this happens, the Δz is canceled from both sides of the approximation thus leaving $f_Z(z)$ in the limit as $\Delta z \to 0$.

Examples 2.26

Suppose we are given random variables X and Y that are characterized by their joint probability density function $f_{XY} = (x, y)$ and that $f_Y(y) = 0$ for $y < 0$. A new random variable Z is defined by $Z = XY$, the product of the two random variables. Find the probability density function $f_Z(z)$ for Z by the incremental method.

Solution

Using the incremental method we must first determine the probability that Z is in the interval $(z, z+\Delta z)$, which is given by

$$P\{z < Z \le z+\Delta z\} = P\{z < XY \le z+\Delta z\} = P\left\{\frac{z}{Y} < X < \frac{z+\Delta z}{Y}\right\}$$

$$= \int_{-\infty}^{\infty}\int_{z/y}^{(z+\Delta z)/y} f_{XY}(x,y)dxdy \approx \int_{-\infty}^{\infty} f_{XY}(\frac{z}{y}, y)\frac{\Delta z}{y}dy$$

Using (2.132) gives

$$f_Z(z)\Delta z \approx \int_{-\infty}^{\infty} f_{XY}(\frac{z}{y}, y)\frac{\Delta z}{y}dy$$

Cancelling the Δz, the equation above in the limit reduces to the result

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{y} f_{XY}(\frac{z}{y}, y)dy$$

This proof can be modified to include the case where $f_Y(y) \ne 0$ for $y < 0$ to give the following result for the probability density $f_Z(z)$ for the product of any two random variables $Z = XY$:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{XY}(\frac{z}{y}, y)dy \tag{2.134}$$

2.9 Computation of E[h(X,Y)]

As for the case of one random variable there are several basic methods for obtaining $E[Z]$, where $Z = h(X, Y)$. The first three methods described are comparable to those described for one function of one random variable. The forth technique is very useful for functions of more than one random variable and does not have a counterpart in the function of one random variable case.

Method 1. Using $f_{XY}(x, y)$,

$$E[Z]=E[h(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(x,y)f_{XY}(x,y)dxdy \tag{2.135}$$

Method 2. Using $f_Z(z)$,

$$E[h(X,Y)]=E[Z]=\int_{-\infty}^{\infty}zf_Z(z)dz \tag{2.136}$$

Method 3. Using a Monte Carlo technique or synthetic sampling.

Method 4. Using iterated expected value,

$$E[Z]=E[h(X,Y)]=E[E[h(x,Y)]|_{x=X}]=E[E[h(X,y)]|_{y=Y}] \tag{2.137}$$

Method 4 has no parallel for the case of one function of one random variable but can be a powerful method for computing expected values of functions of two random variables. The meaning of **iterated expected value** operation described above is easily understood and represents conditional expected values as follows:

$$E[h(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)h(x,y)dxdy \tag{2.138}$$

Breaking the joint probability density function into the product of its conditional and marginal densities gives.

$$E[h(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_Y(y\,|\,x)f_X(x)h(x,y)dxdy \tag{2.139}$$

Rearranging the integral on the right side of the equation above, and taking the $f_X(x)$ through the integral with respect to y (it is not a function of y) gives

$$E[h(X,Y)]=\int_{-\infty}^{\infty}f_X(x)[\int_{-\infty}^{\infty}h(x,y)f_Y(y\,|\,x)dy]dx \tag{2.140}$$

The term in brackets is recognized as the conditional expected value of $g(x,Y)$ or

$$E[h(x,Y)]=\int_{-\infty}^{\infty}h(x,y)f_Y(y\,|\,x)dy \tag{2.141}$$

Therefore $E[h(X,Y)]$ becomes

$$E[h(X,Y)]=\int_{=\infty}^{\infty}f_X(x)E[h(x,Y)]dx \tag{2.142}$$

In a similar fashion $E[h(X,Y)]$ can be written in terms of the conditional expected value with respect to y as follows

$$E[Z]=E[h(X,Y)]=E[E[h(X,y)]|_{y=Y}] \tag{2.143}$$

Example 2.27

Given $Z=\cos(X+Y)$ with X a uniform random variable on $(-\pi,\pi)$ and Y a Gaussian random variable with zero mean and variance equal to 1. Assume that X and Y are independent, and compute $E[Z]$ the expected value of Z using method 4.

Solution

Using (2.137), we have the expected value of Z as

$$E[Z]=E[\cos(X+Y)]=E[E[\cos(X+y)]|_{y=Y}]$$

where

$$E[\cos(X+y)]=\int_{-\pi}^{\pi}\cos(x+y)f_X(x|y)dx$$

For this problem X and Y are independent, therefore the conditional density is the same as the marginal density, and the equation above can be written as

$$E[\cos(X+y)]|_{y=Y}=\int_{-\pi}^{\pi}\cos(x+y)\frac{1}{2\pi}dx\Big|_{y=Y}=0|_{y=Y}$$

The expected value of Z can be found

$$E[Z]=E[0|_{y=Y}]=\int_{-\infty}^{\infty}0\,f_Y(y)dy=0$$

The following example calculates some higher-order moments for jointly Gaussian random variable also using the concept of iterated expected values.

Example 2.28

Let X and Y be jointly normal random variables with parameters $\eta_X=\eta_Y=0$ with σ_X, σ_Y, and ρ_{XY} known. Find $\mathrm{E}[XY]$ and $\mathrm{E}[X^2Y^2]$.

Solution

Using the iterated expected value formula just presented, the $\mathrm{E}[XY]$ can be written as

$$\mathrm{E}[XY]=\int_{-\infty}^{\infty}f_X(x)\left[\int_{-\infty}^{\infty}xyf_Y(y|x)dy\right]dx$$

where

$$f_Y(y|x)=\frac{1}{\sqrt{2\pi(1-\rho_{XY}^2)}\sigma_Y}\exp\left\{-\frac{(y-\rho_{XY}\sigma_Y\sigma_X^{-1}x)^2}{2\sigma_Y^2(1-\rho_{XY}^2)}\right\}$$

The term inside the square brackets can be simplified by taking x through the integral and using the result for the conditional mean Eq. (2.108) to give

$$\int_{-\infty}^{\infty}xyf_Y(y|x)dy=x\int_{-\infty}^{\infty}yf_Y(y|x)dy=x(\rho_{XY}\sigma_Y\sigma_X^{-1}x)$$

Substituting this result for the bracketed term and using second moment for the random variable X gives the $\mathrm{E}[XY]$ as

$$\mathrm{E}[XY] = \int_{-\infty}^{\infty} f_X(x)[x^2 \rho_{XY}\sigma_Y \sigma_X^{-1}]dx$$
$$= \rho_{XY}\sigma_Y\sigma_X^{-1}\int_{-\infty}^{\infty} f_X(x)x^2 dx = \rho_{XY}\sigma_X\sigma_Y$$

Similarly $\mathrm{E}[X^2Y^2]$ can be found in the following fashion:

$$\mathrm{E}[X^2Y^2] = \int_{-\infty}^{\infty} f_X(x)\left(\int_{-\infty}^{\infty} x^2y^2 f_Y(y|x))dy\right)dx$$
$$= \int_{-\infty}^{\infty} f_X(x)\left(x^2\int_{-\infty}^{\infty} y^2 f_Y(y|x))dy\right)dx$$

The integral in the parentheses above represents the conditional second moment, and it can be written in terms of the conditional variance and the conditional mean form Eq. (2.107) as

$$\mathrm{E}[X^2Y^2] = \int_{-\infty}^{\infty} f_X(x)x^2[\sigma_{Y|x}^2 + \eta_{Y|x}^2]dx$$
$$= \sigma_Y^2(1-\rho_{XY}^2)\int_{-\infty}^{\infty} f_X(x)x^2dx + \frac{\rho_{XY}^2\sigma_Y^2}{\sigma_X^2}\int_{-\infty}^{\infty} f_X(x)x^4dx$$

The first integral is the variance of X, since X has a zero mean, and the second is the fourth-order moment for the one-dimensional Gaussian random variable, which is known from Eq. (2.23). Substituting these second-, and fourth-order moments gives

$$\mathrm{E}\left[X^2Y^2\right] = \sigma_Y^2(1-\rho_{XY}^2)[\sigma_X^2] + \frac{\rho_{XY}^2\sigma_Y^2}{\sigma_X^2}[3\sigma_X^4]$$
$$= \sigma_X^2\sigma_Y^2 + 2\rho_{XY}^2\sigma_X^2\sigma_Y^2$$
$$= \sigma_X^2\sigma_Y^2(1+2\rho_{XY}^2)$$

2.10 Multiple Random Variables

2.10.1 Total Characterizations

The real-valued functions for random variables $X_1, X_2 \cdots, X_n$ and Y are defined on a sample space and are said to be **totally characterized** or described with relation to calculating probabilities of acceptable events (i.e., events a member of F) by their cumulative distribution function or probability density function.

The joint cumulative distribution function $F_{X_1X_2\cdots X_n}(x_1, x_2, \cdots, x_n)$ for random variables $X_1, X_2 \cdots, X_n$ is defined by

$$F_{X_1X_2\cdots X_n}(x_1, x_2, \cdots, x_n) \overset{\Delta}{=} P\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\} \tag{2.144}$$

It satisfies the extensions of the bounded, nondecreasing, and continuity properties given in 2.6.1 and can be used to determine probabilities of given events. The **joint probability density function** $f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)$ is related to the joint cumulative distribution function through partial derivatives, and vice versa in terms of integrals, as

$$f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=\frac{\partial^n F_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)}{\partial x_1\partial x_2\cdots\partial x_n}$$
$$F_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n \tag{2.145}$$

For a given vector $\mathbf{X}$ of random variables $X_1,X_2\cdots,X_n$, the characteristic function $\phi_{\mathbf{X}}(\omega_1,\omega_2,\cdots,\omega_n)$ is defined as the "modified" Fourier transform of the joint probability density function

$$\begin{aligned}\phi_{\mathbf{X}}(\omega_1,\omega_2,\cdots,\omega_n)&=E[e^{j(\omega_1X_1+\omega_2X_2+\cdots\omega_nX_n)}]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x_1,x_2,\cdots,x_n)e^{j(\omega_1X_1+\omega_2X_2+\cdots\omega_nX_n)}dx_1dx_2\cdots dx_n\\&=F(f_{\mathbf{X}}(x_1,x_2,\cdots,x_n))\big|_{\boldsymbol{\omega}\to-\boldsymbol{\omega}}\end{aligned} \tag{2.146}$$

Where $\boldsymbol{\omega}=[\omega_1,\omega_2,\cdots,\omega_n]^T$. This integral does not necessarily exist for all probability density functions and has a given region of convergence.

The **moment-generating function** is defined as follows:

$$\begin{aligned}M_{\mathbf{X}}(t_1,t_2,\cdots,t_n)&=E[e^{t_1X_1+t_2X_2+\cdots+t_nX_n}]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2,\cdots,x_n)e^{(t_1X_1+t_2X_2+\cdots t_nX_n)}dx_1dx_2\cdots dx_n\end{aligned} \tag{2.147}$$

2.10.2 Partial Characterizations

Partial characterizations include the various marginal and conditional densities and the higher-order moments. The higher-order moments can be determined from the definition of the moments. However, in many problems the characteristic function or moment-generating function is useful for that purpose.

The **rth-order higher-order moments** for $k_1+k_2+\cdots+k_n=r$ are defined by

$$\begin{aligned}m_{k_1k_2\cdots k_n}&\overset{\Delta}{=}E[X_1^{k_1}X_2^{k_2}\cdots X_n^{k_n}]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}f(x_1x_2\cdots x_n)dx_1dx_2\cdots dx_n\end{aligned} \tag{2.148}$$

Where $k_1+k_2+\cdots+k_n=r$

A well-known theorem gives the higher rth-order moments in terms of the characteristic function $\phi_{\mathbf{X}}(\omega_1,\omega_2,\cdots,\omega_n)$ as follows:

$$m_{k_1k_2\cdots k_n} \overset{\Delta}{=} (-j)^r \left.\frac{\partial^r \phi_{\mathbf{X}}(\omega_1,\omega_2,\cdots,\omega_n)}{\partial^{k_1}\omega_1\partial^{k_2}\omega_2\cdots\partial^{k_n}\omega_n}\right]_{\omega_1=0,\omega_2=0,\cdots,\omega_n=0} \tag{2.149}$$

Where $k_1+k_2+\cdots+k_n=r$ and $r\le n$

Similarly the moments can be generated by using the moment-generating function through a corresponding theorem, from Davenport, that gives the higher-order moments in terns of the moment-generating function as follows:

$$m_{k_1k_2\cdots k_n} = \left.\frac{\partial^r M_{\mathbf{X}}(t_1,t_2,\cdots,t_n)}{\partial^{k_1}t_1\partial^{k_2}t_2\cdots\partial^{k_n}t_n}\right]_{t_1=0,t_2=0,\cdots,t_n=0} \tag{2.150}$$

Where $k_1+k_2+\cdots+k_n=r$ and $r\le n$

The moment-generating function is simpler to obtain in that it is determined as a real integra. However, the use of the existing Fourier transform table to obtain the characteristic function, and the fact the characteristic function exists for cases where the moment-generating function does not exist, makes the characteristic function more usable.

Example 2.29

The random variables X_1, X_2, and X_3 are characterized by their joint probability density function $f_{X_1X_2X_3}(x_1,x_2,x_3)$,which is

$$f_{X_1X_2X_3}(x_1,x_2,x_3)=6e^{-(x_1+2x_2+3x_3)}\mu(x_1)\mu(x_2)\mu(x_3)$$

Determine, by using Eq. (2.149), the following third-order moments: (a) m_{001}, (b) m_{111}, (c) m_{030},(d) m_{210}.

Solution

To obtain the moments from (2.149), we first need to calculate the characteristic function, which from (2.146) is as follows:

$$\Phi(\omega_1,\omega_2,\omega_3)=\int_0^\infty\int_0^\infty\int_0^\infty 6e^{-(x_1+2x_2+3x_3)}e^{j(\omega_1x_1+\omega_2x_2+\omega_3x_3)}dx_1dx_2dx_3$$
$$=6\frac{1}{(-j\omega_1+1)}\frac{1}{(-j\omega_2+2)}\frac{1}{(-j\omega_1+3)}$$

(a) The first-order moment m_{001} is determined from (2.149) as

$$m_{001}=(-j)^1\left.\frac{\partial\Phi}{\partial\omega_3}\right]_{\substack{\omega_1=0\\\omega_2=0\\\omega_3=0}}=\left.\frac{6(-j)(-1)(-j)}{(-j\omega_1+1)(-j\omega_2+2)(-j\omega_1+3)^2}\right]_{\substack{\omega_1=0\\\omega_2=0\\\omega_3=0}}=\frac{1}{3}$$

(b) The third-order moment m_{111} is determined from (2.149) as

$$m_{111}=(-j)^3\left.\frac{\partial^3\Phi}{\partial\omega_1\partial\omega_2\partial\omega_3}\right]_{\substack{\omega_1=0\\\omega_2=0\\\omega_3=0}}=\left.\frac{(-j)^3 6(-1)^3(-j)^3}{(-j\omega_1+1)^2(-j\omega_2+2)^2(-j\omega_1+3)^2}\right]_{\substack{\omega_1=0\\\omega_2=0\\\omega_3=0}}=\frac{1}{6}$$

(c) The third-order moment m_{030} is determined from (2.149) as

$$m_{030} = (-j)^3 \frac{\partial^3 \Phi}{\partial^3 \omega_3}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0\\ \omega_3=0}} = \frac{(-j)^3 6(-1)(-2)(-3)(-j)^3}{(-j\omega_1+1)(-j\omega_2+2)^4(-j\omega_1+3)}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0\\ \omega_3=0}} = \frac{1}{8}$$

(c) The third-order moment m_{210} is determined from (2.149) as

$$m_{210} = (-j)^3 \frac{\partial^3 \Phi}{\partial^2 \omega_1 \partial \omega_2}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0\\ \omega_3=0}} = \frac{(-j)^3 6(-1)(-2)(-1)(-j)^3}{(-j\omega_1+1)^3(-j\omega_2+2)^2(-j\omega_1+3)}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0\\ \omega_3=0}} = 1$$

Since the random variables for this problem are independent, it is easy to check these results by computing the moments shown as the product of various moments of the random variables taken by themselves. When the random variables are not independent, then the higher-order moments cannot be calculated as products of the individual moments, so the formula above becomes more useful for their evaluation.

Another useful characterization for random variables is the cumulants that will play a role in the analysis of nonlinear systems with random process excitations. The cumulants are defined in terms of the joint characteristic function as follows:

$$c_{k_1 k_2 \cdots k_n} = (-j)^r \frac{\partial^r \ln \phi(\omega_1, \omega_2, \cdots, \omega_n)}{\partial^{k_1} \omega_1 \partial^{k_2} \omega_2 \cdots \partial^{k_n} \omega_n}\Bigg]_{\omega_1=0, \omega_2=0, \cdots, \omega_n=0} \tag{2.151}$$

Where $k_1 + k_2 + \cdots + k_n = r$ and $r \le n$

The only difference in the preceding formulas for moments and cumulants is that the natural *log(ln)* of the characteristic function is the object of the partial derivatives for the cumulants, whereas the characteristic function is used for determining the moments.

Example 2.30

The random variables X_1 and X_2 are characterized by their joint probability density function $f_{X_1 X_2}(x_1, x_2)$, which is

$$f_{X_1 X_2}(x_1, x_2) = 2e^{-(x_1 + 2x_2)} \mu(x_1) \mu(x_2)$$

Determine, by using Eq. (2.151), the following cumulants: (a) c_{01}, (b) c_{10}, (c) c_{11} and (d) the *r*th order cumulant c_{jk} for $j > 1, k > 1$, and $j + k = r$.

Solution

To obtain the cumulants from (2.151), we first need to calculate the characteristic function, which from (2.146) is as follows

$$\Phi(\omega_1,\omega_2)=\int_0^{\infty}\int_0^{\infty}2e^{-(x_1+2x_2)}e^{j(\omega_1x_1+\omega_2x_2)}dx_1dx_2$$
$$=2\frac{1}{(-j\omega_1+1)}\frac{1}{(-j\omega_2+2)}$$

(a) The first-order cumulant c_{10} is determined from (2.151) as

$$c_{10}=(-j)^1\frac{\partial(-\ln(-j\omega_1+1)-\ln(-j\omega_2+2)+\ln 2)}{\partial\omega_1}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}$$
$$=\frac{-(-j)(-j)1}{(-j\omega_1+1)}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}=1$$

(b) The first-order cumulant c_{01} is determined from (2.151) as

$$c_{10}=(-j)^1\frac{\partial(-\ln(-j\omega_1+1)-\ln(-j\omega_2+2)+\ln 2)}{\partial\omega_2}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}$$
$$=\frac{-(-j)(-j)1}{(-j\omega_1+2)}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}=\frac{1}{2}$$

(c) The first-order cumulant c_{11} is determined from (2.151) as

$$c_{11}=(-j)^1\frac{\partial(-\ln(-j\omega_1+1)-\ln(-j\omega_2+2)+\ln 2)}{\partial\omega_1\partial\omega_2}\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}$$
$$=\frac{\partial}{\partial\omega_2}\left(\frac{-(-j)(-j)1}{(-j\omega_1+1)}\right)\Bigg]_{\substack{\omega_1=0\\ \omega_2=0}}=\frac{1}{2}$$

(d) The rth order cumulant c_{jk} for $j>1,k>1$, and $j+k=r$ is seen from the equation above to be 0.

A theorem presented by Nikias and Petropulu originally presented by Brillinger gives the nth-order cumulant in terms of all the nth-order moments as follows:

Theorem. If $X_1,X_2\cdots,X_n$ are jointly Gaussian random variables specified by their various order moments $m_{ij}=E[X_iX_j]$, then the nth cumulant $c_{11\cdots1}=cum[X_1X_2\cdots X_n]$ can be written in terms of the $E[X_{i_1}X_{i_2}\cdots,X_{i_n}]$ as follows:

$$c_{11\cdots1}=\sum_{all\ \ partitions}(-1)^{p-1}(p-1)!E[\Pi_{i\in s_1}X_i]\cdot E[\Pi_{i\in s_2}X_i]\cdots E[\Pi_{i\in s_p}X_i] \tag{2.152}$$

Where the summation is taken over all partitions $\{s_1,s_2,\cdots,s_n\}$, $p=1,2,\ldots,n$ of the set of integers $\{1,2,\ldots,n\}$.

Example (2.31)

Use the theorem above to determine the first, second, and third cumulants in terms of various third-order moments (from Nikias and Petropulu).

Solution

The $c_1(X_1)$ and $c_{11}(X_1, X_1)$ are easily shown to be

$$\begin{aligned} c_1 &= c_1(X_1) = E[X_1] \\ c_{11} &= c_{11}(X_1, X_1) = E[X_1 X_2] - E[X_1]E[X_2] \end{aligned} \tag{2.153}$$

In order to determine the cumulant $c_{111} = c(X_1 X_2 X_3)$, it is necessary to determine all partitions of $\{1,2,3\}$. For $p = 1$, just one set we get $s_1 = \{1,2,3\}$. For $p = 2$, we have the following ways to partition: $s_1 = \{1\}$ and $s_2 = \{2,3\}$, or $s_1 = \{2\}$ and $s_2 = \{1,3\}$ or $s_1 = \{3\}$ and $s_2 = \{1,2\}$. And finally for $p = 3$, there is only one way: $s_1 = \{1\}$, $s_2 = \{2\}$ and $s_3 = \{3\}$.

From (2.152) we can now write $c(X_1 X_2 X_3)$ as

$$\begin{aligned} c_{111} &\overset{\Delta}{=} c(X_1, X_2, X_3) \\ &= (-1)^{1-1}(1-1)!E\left[X_1 X_2 X_3\right] \\ &\quad + (-1)^{2-1}(2-1)!(E[X_1]E[X_2 X_3] + E[X_2]E[X_1 X_3] + E[X_3]E[X_1 X_2]) \\ &\quad + (-1)^{3-1}(3-1)!(E[X_1]E[X_2]E[X_3] \\ &= E\left[X_1 X_2 X_3\right] - (E[X_1]E[X_2 X_3] + E[X_2]E[X_1 X_3] + E[X_3]E[X_1 X_2]) \\ &\quad + 2E[X_1]E[X_2]E[X_3] \end{aligned} \tag{2.154}$$

Similarly it can be shown that c_{1111} is as follows:

$$\begin{aligned} c_{1111} &\overset{\Delta}{=} c(X_1, X_2, X_3, X_4) \\ &= E\left[X_1 X_2 X_3 X_4\right] - E\left[X_1 X_2\right]E\left[X_3 X_4\right] - E\left[X_1 X_3\right]E\left[X_2 X_4\right] - E\left[X_1 X_4\right]E\left[X_2 X_3\right] \\ &\quad - E[X_1]E[X_2 X_3 X_4] - E[X_2]E[X_1 X_3 X_4] - E[X_3]E[X_1 X_2 X_4] - E[X_4]E[X_1 X_2 X_3] \\ &\quad + 2(E[X_1 X_2]E[X_3]E[X_4] + E[X_1 X_3]E[X_2]E[X_4] + E[X_1 X_4]E[X_2]E[X_3] \\ &\quad + E[X_2 X_3]E[X_1]E[X_4] + E[X_2 X_4]E[X_1]E[X_3] + E[X_3 X_4]E[X_1]E[X_2]) \\ &\quad - 6E[X_1]E[X_2]E[X_3]E[X_4] \end{aligned}$$

(2.155)

For zero mean random variables the first, second, third, and fourth-order cumulants reduce to the following:

$$
\begin{aligned}
&c_1(X_1) = E[X_1] = 0 \\
&c_{11}(X_1, X_2) = E[X_1 X_2] \\
&c_{111}(X_1, X_2, X_3) = E[X_1 X_2 X_3] \\
&c_{1111}(X_1, X_2, X_3, X_4) = E[X_1 X_2 X_3 X_4] - E[X_1 X_2]E[X_3 X_4] \\
&\qquad\qquad - E[X_1 X_3]E[X_2 X_4] - E[X_1 X_4]E[X_2 X_3]
\end{aligned} \tag{2.156}
$$

2.10.3 Gaussian Random Vectors

When more than two jointly normal random variables are considered, it becomes convenient to use a vector formulation. Random variables $X_1, X_2 \cdots, X_n$ are **jointly Gaussian** if their joint probability function can be written in the form

$$
f_{\mathbf{X}}(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}|\mathbf{K}|^{1/2}} \exp\left(\frac{-(\mathbf{x}-\mathbf{m})^T \mathbf{K}^{-1}(\mathbf{x}-\mathbf{m})}{2}\right)
$$

Where

$$
\begin{aligned}
&\mathbf{X} = [X_1, X_2, \cdots, X_n]^T \\
&\mathbf{m} = \mathrm{E}[\mathbf{X}] = [m_1, m_2, \cdots, m_n]^{\mathrm{T}} \\
&\mathbf{K} = E[(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T] \\
&\mathbf{x} = [x_1, x_2, \cdots, x_n]^T
\end{aligned} \tag{2.157}
$$

For the random vector $\mathbf{X}$, $\mathbf{m}$ and $\mathbf{K}$ are the mean vector and covariance matrix, respectively. The $E[.]$ is the expected value operator. The vector $\mathbf{m}$, called the mean vector of the random vector $\mathbf{X}$, has the means of each random variable as its components. The argument of the exponential is a vector-matrix-vector product that yields a scalar result. The matrix $\mathbf{K}$ is called the covariance matrix. It can be easily shown that $\mathbf{K}$ is a square matrix with variances of the individual random variables down the main diagonal and covariances of the respective random variables off the main diagonal. It can also be shown that $\mathbf{K}$ is symmetric, positive definite, and has all its eigen values greater than 0. instead of writing the long form given above, **it is convenient to use a short-hand notation and write that $\mathbf{X} \sim N[\mathbf{m}, \mathbf{K}]$, which will mean that $\mathbf{X}$ is a Gaussian random vector with mean vector $\mathbf{m}$, covariance matrix $\mathbf{K}$**, and have the joint density given in (2.157).

Moments and cumulants for Gaussian random variables. Let $X_1, X_2 \cdots, X_n$ be jointly Gaussian random variables characterized by their mean vector $\mathbf{m}_{\mathbf{X}}$ and covariance matrix $\mathbf{K}$. The cumulants and moments for the jointly Gaussian random variables will now be presented.

The **joint characteristic function for a Gaussian random vector** can be shown

$$
\phi(\omega_1, \omega_2, \cdots, \omega_n) = \exp[j\boldsymbol{\omega}^T \mathbf{m}_X - \frac{1}{2}\boldsymbol{\omega}^T \mathbf{K}^{-1}\boldsymbol{\omega}]
$$

$$\text{Where } \boldsymbol{\omega}^T = [\omega_1, \omega_2, \cdots, \omega_n] \text{ and} \tag{2.158}$$

$$\mathbf{m}_X^T = [E[X_1], E[X_2], \cdots, E[X_n]]$$

Theorem. If X_1, X_2, X_3, and X_4 are zero mean jointly Gaussian random variables specified by their second-order moments $m_{ij} = E[\mathrm{X_i X_j}]$, then the fourth-moment $E[X_1 X_2 X_3 X_4]$ can be written in terms of the $E[\mathrm{X_i X_j}]$ as follows:

$$E[X_1 X_2 X_3 X_4] = E[\mathrm{X_1 X_2}] \cdot E[\mathrm{X_3 X_4}] + E[\mathrm{X_1 X_3}] \cdot E[\mathrm{X_2 X_4}] + E[\mathrm{X_1 X_4}] \cdot E[\mathrm{X_2 X_3}] \tag{2.159}$$

2.11 M Functions of N Random Variables

In general, we would wish to find the joint probability density function $f(y_1, y_2, \cdots, y_N)$ for random variables $Y_1, Y_2, \cdots, Y_N$ defined by

$$Y_i = g_i(X_1, X_2, \cdots, X_N) \qquad i = 1, 2, \cdots, M$$

When M and N are equal, it is possible to find the joint density of the transformed variables by extending the transformation theorem given in Eq.(2.113). When M and N are not equal auxiliary random variables can be defined to make the orders equal then use the transformation theorem followed by integrating out with respect to the auxiliary variables.

The extension of the distribution function approach is possible, but usually the high-dimensional space makes it difficult to describe the regions over which we integrate, and the following partial derivatives of the distribution are certainly not a picnic either.

We are then led to obtaining the joint density by use of statistical sampling, where again the higher dimensions make the number of the samples necessary to get a reasonable estimate of the probability density by a histogram becomes exceedingly large. This problem can be alleviated if the densities involved are of a certain parametric form.

2.12 Summary

In this chapter the basic definitions for a single random variable and several joint random variables were presented along with the various forms of their characterizations, which included the mean, variance, covariance, and higher-order moments as partial characterization. Total characterizations for random variables were given in terms of the probability density function and the corresponding probability distribution functions and joint probability distribution functions.

Common probability density functions for both continuous and discrete random variables were then discussed including the Gaussian, uniform, exponential Raleigh, chi-squared, beta, and Cauchy densities for the continuous case and uniform, Bernoulli, Poisson, binomial, geometric, and negative binomial densities for discrete case. The mean, moments, and variance of these common random values were given to form a small catalog of densities and partial characterizations of the associated random variables.

For a single function of a single random variable, the main techniques discussed for obtaining the density of the transformed random variable in terms of the given random variable were the transformation theorem and the distributional approach. The distributional approach involved first obtaining the distribution function and then obtaining the density by taking its derivative. These same basic techniques were used for a single function of multiple random variables by using an auxiliary random variable approach. In this way the n-dimensional transformation could be used first to obtain a joint probability density function followed by an integrating out of those auxiliary random variables from the joint density to give the required density.

----------This is the end of Chapter02----------