IV

JOINT AND CONDITIONAL DISTRIBUTIONS, STOCHASTIC INDEPENDENCE, MORE EXPECTATION

1 INTRODUCTION AND SUMMARY

The purpose of this chapter is to introduce the concepts of $k$-dimensional distribution functions, conditional distributions, joint and conditional expectation, and independence of random variables. It, like Chap. II, is primarily a "definitions-and-their-understanding" chapter.

The chapter is divided into four main sections in addition to the present one. In Sec. 2, joint distributions, both in cumulative and density-function form, are introduced. The important $k$-dimensional discrete distribution, called the multinomial, is included as an example. Conditional distributions and independence of random variables are the subject of Sec. 3. Section 4 deals with expectation with respect to $k$-variante distributions. Definitions of covariance, the correlation coefficient, and joint moment generating functions, all of which are special expectations, are given. The important concept of conditional expectation is discussed in Subsec. 4.3. Results relating independence and expectation are presented in Subsec. 4.5, and the famous Cauchy-Schwarz inequality is proved in Subsec. 4.6. The last main section, Sec. 5, is devoted to the important bivariate normal distribution, which gives one unified example of many of the terms defined in the preceding sections.
This chapter is the multidimensional analog of Chap. II. It provides definitions needed to understand distributional-theory results of Chap. V.

2 JOINT DISTRIBUTION FUNCTIONS

In the study of many random experiments, there are, or can be, more than one random variable of interest; hence we are compelled to extend our definitions of the distribution and density function of one random variable to those of several random variables. Such definitions are the essence of this section, which is the multivariate counterpart of Secs. 2 and 3 of Chap. II. As in the univariate case we will first define, in Subsec. 2.1, the cumulative distribution function. Although it is not as convenient to work with as density functions, it does exist for any set of \( k \) random variables. Density functions for jointly discrete and jointly continuous random variables will be given in Subsecs. 2.2 and 2.3, respectively.

2.1 Cumulative Distribution Function

**Definition 1** Joint cumulative distribution function Let \( X_1, X_2, \ldots, X_k \) be \( k \) random variables all defined on the same probability space \((\Omega, \mathcal{A}, P[\cdot])\). The joint cumulative distribution function of \( X_1, \ldots, X_k \), denoted by \( F_{X_1,\ldots,X_k}(\cdot, \ldots, \cdot) \), is defined as \( P[X_1 \leq x_1; \ldots; X_k \leq x_k] \) for all \((x_1, x_2, \ldots, x_k)\).

Thus a joint cumulative distribution function is a function with domain euclidean \( k \) space and countdomain the interval \([0, 1]\). If \( k = 2 \), the joint cumulative distribution function is a function of two variables, and so its domain is just the \( xy \) plane.

**EXAMPLE 1** Consider the experiment of tossing two tetrahedra (regular four-sided polyhedron) each with sides labeled 1 to 4. Let \( X \) denote the number on the downturned face of the first tetrahedron and \( Y \) the larger of the downturned numbers. The goal is to find \( F_{X,Y}(\cdot, \cdot) \), the joint cumulative distribution function of \( X \) and \( Y \). Observe first that the random variables \( X \) and \( Y \) jointly take on only the values

- (1, 1), (1, 2), (1, 3), (1, 4),
- (2, 2), (2, 3), (2, 4),
- (3, 3), (3, 4),
- (4, 4).

(The first component is the value of \( X \), and the second the value of \( Y \).)
The sample space for this experiment is displayed in Fig. 1. The 16 sample points are assumed to be equally likely. Our objective is to find $F_{X,Y}(x, y)$ for each point $(x, y)$. As an example let $(x, y) = (2, 3)$, and find $F_{X,Y}(2, 3) = P[X \leq 2; \ Y \leq 3]$. Now the event \( \{X \leq 2 \text{ and } Y \leq 3\} \) corresponds to the encircled sample points in Fig. 1; hence $F_{X,Y}(2, 3) = \frac{6}{16}$. Similarly, $F_{X,Y}(x, y)$ can be found for other values of $x$ and $y$. $F_{X,Y}(x, y)$ is tabled in Fig. 2.

We saw that the cumulative distribution function of a unidimensional random variable had certain properties; the same is true of a joint cumulative. We shall list these properties for the joint cumulative distribution function of two random variables; the generalization to $k$ dimensions is straightforward.

### TABLE OF VALUES OF $F_{X,Y}(x, y)$

<table>
<thead>
<tr>
<th>$4 \leq y$</th>
<th>0</th>
<th>1/4</th>
<th>1/2</th>
<th>3/4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3 \leq y &lt; 4$</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
</tr>
<tr>
<td>$2 \leq y &lt; 3$</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
<td>3/4</td>
<td>1</td>
</tr>
<tr>
<td>$1 \leq y &lt; 2$</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
</tr>
<tr>
<td>$y &lt; 1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$x &lt; 1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| $1 \leq x < 2$ | $2 \leq x < 3$ | $3 \leq x < 4$ | $4 \leq x$ |

FIGURE 2
Properties of bivariate cumulative distribution function \( F(\cdot, \cdot) \)

(i) \( F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0 \) for all \( y \), \( F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0 \) for all \( x \), and \( \lim_{x \to -\infty} F(x, y) = F(\infty, \infty) = 1 \).

(ii) If \( x_1 < x_2 \) and \( y_1 < y_2 \), then \( P[x_1 < X \leq x_2; y_1 < Y \leq y_2] = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0 \).

(iii) \( F(x, y) \) is right continuous in each argument; that is, \( \lim_{0 < h \to 0} F(x + h, y) = \lim_{0 < h \to 0} F(x, y + h) = F(x, y) \).

We will not prove these properties. Property (ii) is a monotonicity property of sorts; it is not equivalent to \( F(x_1, y_1) \leq F(x_2, y_2) \) for \( x_1 \leq x_2 \) and \( y_1 \leq y_2 \). Consider, for example, the bivariate function \( G(x, y) \) defined as in Fig. 3. Note that \( G(x_1, y_1) \leq G(x_2, y_2) \) for \( x_1 \leq x_2 \) and \( y_1 \leq y_2 \), yet \( G(1 + \varepsilon, 1 + \varepsilon) - G(1 + \varepsilon, 1 - \varepsilon) - G(1 - \varepsilon, 1 + \varepsilon) + G(1 - \varepsilon, 1 - \varepsilon) = 1 - (1 - \varepsilon) - (1 - \varepsilon) - 2\varepsilon + 1 < 0 \) for \( \varepsilon < \frac{1}{2} \); so \( G(x, y) \) does not satisfy property (ii) and consequently is not a bivariate cumulative distribution function.

**Definition 2** Bivariate cumulative distribution function Any function satisfying properties (i) to (iii) is defined to be a bivariate cumulative distribution function without reference to any random variables. ||||

**Definition 3** Marginal cumulative distribution function If \( F_{X,Y}(\cdot, \cdot) \) is the joint cumulative distribution function of \( X \) and \( Y \), then the cumulative distribution functions \( F_X(\cdot) \) and \( F_Y(\cdot) \) are called marginal cumulative distribution functions. ||||

<table>
<thead>
<tr>
<th>Table of ( G(x, y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 \leq y )</td>
</tr>
<tr>
<td>( 0 \leq y &lt; 1 )</td>
</tr>
<tr>
<td>( y &lt; 0 )</td>
</tr>
<tr>
<td>( x &lt; 0 )</td>
</tr>
</tbody>
</table>

**Figure 3**
Remark $F_x(x) = F_{X, Y}(x, \infty)$, and $F_y(y) = F_{X, Y}(\infty, y)$; that is, knowledge of the joint cumulative distribution function of $X$ and $Y$ implies knowledge of the two marginal cumulative distribution functions.

The converse of the above remark is not generally true; in fact, an example (Example 8) will be given in Subsec. 2.3 below that gives an entire family of joint cumulative distribution functions, and each member of the family has the same marginal distributions.

We will conclude this section with a remark that gives an inequality involving the joint cumulative distribution and marginal distributions. The proof is left as an exercise.

Remark $F_x(x) + F_y(y) - 1 \leq F_{X,Y}(x, y) \leq \sqrt{F_x(x)F_y(y)}$ for all $x, y$.

2.2 Joint Density Functions for Discrete Random Variables

If $X_1, X_2, \ldots, X_k$ are random variables defined on the same probability space, then $(X_1, X_2, \ldots, X_k)$ is called a $k$-dimensional random variable.

Definition 4 Joint discrete random variables The $k$-dimensional random variable $(X_1, X_2, \ldots, X_k)$ is defined to be a $k$-dimensional discrete random variable if it can assume values only at a countable number of points $(x_1, x_2, \ldots, x_k)$ in $k$-dimensional real space. We also say that the random variables $X_1, X_2, \ldots, X_k$ are joint discrete random variables.

Definition 5 Joint discrete density function If $(X_1, X_2, \ldots, X_k)$ is a $k$-dimensional discrete random variable, then the joint discrete density function of $(X_1, X_2, \ldots, X_k)$, denoted by $f_{x_1, x_2, \ldots, x_k}(x_1, x_2, \ldots, x_k)$, is defined to be

$$f_{x_1, x_2, \ldots, x_k}(x_1, x_2, \ldots, x_k) = P(X_1 = x_1; X_2 = x_2; \ldots; X_k = x_k)$$

for $(x_1, x_2, \ldots, x_k)$, a value of $(X_1, X_2, \ldots, X_k)$ and is defined to be 0 otherwise.

Remark $\sum f_{x_1, x_2, \ldots, x_k}(x_1, x_2, \ldots, x_k) = 1$, where the summation is over all possible values of $(X_1, \ldots, X_k)$.
EXAMPLE 2 Let $X$ denote the number on the downturned face of the first tetrahedron and $Y$ the larger of the downturned numbers in the experiment of tossing two tetrahedra. The values that $(X, Y)$ can take on are $(1, 1)$, $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 2)$, $(2, 3)$, $(2, 4)$, $(3, 3)$, $(3, 4)$, and $(4, 4)$; hence $X$ and $Y$ are jointly discrete. The joint discrete density function of $X$ and $Y$ is given in Fig. 4.

In tabular form it is given as

$$
\begin{array}{c|ccccc}
(x, y) & (1, 1) & (1, 2) & (1, 3) & (1, 4) & (2, 2) \\
fx, y(x, y) & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\
\end{array}
$$

or in another tabular form as

$$
\begin{array}{c|cccc}
4 & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\
3 & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\
2 & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\
1 & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\
y/ x & 1 & 2 & 3 & 4 \\
\end{array}
$$

**Theorem 1** If $X$ and $Y$ are jointly discrete random variables, then knowledge of $F_{X,Y}(',\cdot')$ is equivalent to knowledge of $f_{X,Y}(',\cdot')$. Also, the statement extends to $k$-dimensional discrete random variables.
PROOF Let \((x_1, y_1), (x_2, y_2), \ldots\) be the possible values of \((X, Y)\).
If \(f_{X,Y}(\cdot, \cdot)\) is given, then \(F_{X,Y}(x, y) = \sum f_{X,Y}(x_i, y_i)\), where the summation is over all \(i\) for which \(x_i \leq x\) and \(y_i \leq y\). Conversely, if \(F_{X,Y}(\cdot, \cdot)\) is given, then for \((x_1, y_1)\), a possible value of \((X, Y)\),
\[
f_{X,Y}(x_1, y_1) = F_{X,Y}(x_1, y_1) - \lim_{0<k\to0} F_{X,Y}(x_1-k, y_1)
- \lim_{0<h\to0} F_{X,Y}(x_1, y_1-h)
+ \lim_{0<k\to0} F_{X,Y}(x_1-h, y_1).
\]

**Definition 6** Marginal discrete density If \(X\) and \(Y\) are jointly discrete random variables, then \(f_X(\cdot)\) and \(f_Y(\cdot)\) are called marginal discrete density functions. More generally, let \(X_{i_1}, \ldots, X_{i_m}\) be any subset of the jointly discrete random variables \(X_1, \ldots, X_k\); then \(f_{X_{i_1},\ldots,X_{i_m}}(x_1,\ldots,x_{i_m})\) is also called a marginal density.

**Remark** If \(X_1, \ldots, X_k\) are jointly discrete random variables, then any marginal discrete density can be found from the joint density, but not conversely. For example, if \(X\) and \(Y\) are jointly discrete with values \((x_1, y_1), (x_2, y_2), \ldots\), then
\[
f_X(x_i) = \sum_{y : x = x_i} f_{X,Y}(x_i, y_i) \quad \text{and} \quad f_Y(y_j) = \sum_{x : y = y_j} f_{X,Y}(x_i, y_i).
\]

Heretofore we have indexed the values of \((X, Y)\) with a single index, namely \(i\). That is, we listed values as \((x_1, y_1), (x_2, y_2), \ldots, (x_1, y_i), \ldots\). The values of \((X, Y)\) could also be indexed by using separate indices for the \(X\) and \(Y\) values. For instance, we could let \(i\) index the possible \(X\) values, say \(x_1, \ldots, x_t, \ldots\), and \(j\) index the possible \(Y\) values, say \(y_1, \ldots, y_j, \ldots\). Then the values of \((X, Y)\) would be a subset of the points \((x_i, y_j)\) for \(i = 1, 2, \ldots\) and \(j = 1, 2, \ldots\). If this latter method of indexing is used, then the marginal density of \(X\) is obtained as follows:
\[
f_X(x_i) = \sum_j f_{X,Y}(x_i, y_j),
\]
where the summation is over all \(y_j\) for the fixed \(x_i\). The marginal density of \(Y\) is analogously obtained. The following example may help to clarify these two different methods of indexing the values of \((X, Y)\).

**EXAMPLE 3** Return to the experiment of tossing two tetrahedra, and define \(X\) as the number on the downturned face of the first tetrahedron and \(Y\) as the larger of the numbers on the two downturned faces. The joint
density of $X$ and $Y$ is given in Fig. 4. The values of $(X, Y)$ can be listed as $(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4),$ and $(4, 4)$, 10 points in all. Or, if we note that $X$ has values 1, 2, 3, and 4; $Y$ has values 1, 2, 3, and 4; and $Y$ is greater than or equal $X$, the values of $(X, Y)$ are $((i, j): i = 1, \ldots, 4; j = 1, \ldots, 4; \text{and } i \leq j)$. Let us use each of these methods of indexing to evaluate $F_{X,Y}(2, 3)$ from the joint density.

Under the first method of indexing,

$$
F_{X,Y}(2, 3) = \sum_{(i, j) \leq (2, 3)} f_{X,Y}(x_i, y_j)
= f_{X,Y}(1, 1) + f_{X,Y}(1, 2) + f_{X,Y}(1, 3) + f_{X,Y}(2, 2) + f_{X,Y}(2, 3) = \frac{5}{8}
$$

Under the second method of indexing,

$$
F_{X,Y}(2, 3) = \sum_{i=1}^{2} \sum_{j=1}^{3} f_{X,Y}(i, j) = \frac{5}{8}.
$$

Similarly, all other values of $F_{X,Y}(\cdot, \cdot)$ could be obtained. Also

$$
f_3(3) = \sum_{(i, j) = (3, 3)} f_{X,Y}(x_i, y_j) = f_{X,Y}(1, 3) + f_{X,Y}(2, 3) + f_{X,Y}(3, 3)
= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}.
$$

Similarly, $f_3(1) = \frac{1}{8}, f_3(2) = \frac{1}{8}$, and $f_3(4) = \frac{1}{8}$, which together with $f_3(3) = \frac{3}{8}$ give the marginal discrete density function of $Y$.

EXAMPLE 4 We mentioned that marginal densities can be obtained from the joint density, but not conversely. The following is an example of a family of joint densities that all have the same marginals, and hence we see that in general the joint density is not uniquely determined from knowledge of the marginals. Consider altering the joint density given in the previous examples as follows:

<table>
<thead>
<tr>
<th>$y$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>$\frac{1}{8} + \varepsilon$</td>
<td>$\frac{1}{8} - \varepsilon$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{8} - \varepsilon$</td>
<td>$\frac{1}{8} + \varepsilon$</td>
<td>$\frac{1}{8}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{8}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{8}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sum_{y} f_{x,y}$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
For each $0 \leq \varepsilon \leq \frac{1}{16}$, the above table defines a joint density. Note that the marginal densities are independent of $\varepsilon$, and hence each of the joint densities (there is a different joint density for each $0 \leq \varepsilon \leq \frac{1}{16}$) has the same marginals.

We saw that the binomial distribution was associated with independent, repeated Bernoulli trials; we shall see in the example below that the multinomial distribution is associated with independent, repeated trials that generalize from Bernoulli trials with two outcomes to more than two outcomes.

**Example 5** Suppose that there are $k + 1$ (distinct) possible outcomes of a trial. Denote these outcomes by $\omega_1, \omega_2, \ldots, \omega_{k+1}$, and let $p_i = P[\omega_i], i = 1, \ldots, k + 1$. Obviously we must have $\sum_{i=1}^{k+1} p_i = 1$, just as $p + q = 1$ in the binomial case. Suppose that we repeat the trial $n$ times. Let $X_i$ denote the number of times outcome $\omega_i$ occurs in the $n$ trials, $i = 1, \ldots, k + 1$. If the trials are repeated and independent, then the discrete density function of the random variables $X_1, \ldots, X_k$ is

$$f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = \frac{n!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} p_i^{x_i}, \quad (1)$$

where $x_i = 0, \ldots, n$ and $\sum_{i=1}^{k+1} x_i = n$. Note that $X_{k+1} = n - \sum_{i=1}^{k} X_i$.

To justify Eq. (1), note that the left-hand side is $P[X_1 = x_1; X_2 = x_2; \ldots; X_{k+1} = x_{k+1}]$; so, we want the probability that the $n$ trials result in exactly $x_1$ outcomes $\omega_1$, exactly $x_2$ outcomes $\omega_2$, ..., exactly $x_{k+1}$ outcomes $\omega_{k+1}$, where $\sum_{i=1}^{k+1} x_i = n$. Any specific ordering of these $n$ outcomes has probability $p_1^{x_1} \cdot p_2^{x_2} \cdots p_{k+1}^{x_{k+1}}$ by the assumption of independent trials, and there are $n! x_1! x_2! \cdots x_{k+1}!$ such orderings.

**Definition 7** Multinomial distribution The joint discrete density function given in Eq. (1) is called the multinomial distribution.

The multinomial distribution is a $(k + 1)$-parameter family of distributions, the parameters being $n$ and $p_1, p_2, \ldots, p_k$. $p_{k+1}$ is, like $q$ in the binomial distribution, exactly determined by $p_{k+1} = 1 - p_1 - p_2 - \cdots - p_k$. A
particular case of a multinomial distribution is obtained by putting, for example, \( n = 3, k = 2, p_1 = .2, \) and \( p_2 = .3, \) to get

\[
f_{x_1, x_2}(x_1, x_2) = f(x_1, x_2) = \frac{3!}{x_1!x_2!(3 - x_1 - x_2)!} \cdot (2)^x_1(3)^x_2(5)^3-x_1-x_2.
\]

This density is plotted in Fig. 5.

We might observe that if \( X_1, X_2, \ldots, X_k \) have the multinomial distribution given in Eq. (1), then the marginal distribution of \( X_i \) is a binomial distribution with parameters \( n \) and \( p_i. \) This observation can be verified by recalling the experiment of repeated, independent trials. Each trial can be thought of as resulting either in outcome \( \omega_i \) or not in outcome \( \omega_j \), in which case the trial is Bernoulli, implying that \( X_i \) has a binomial distribution with parameters \( n \) and \( p_i. \)

### 2.3 Joint Density Functions for Continuous Random Variables

**Definition 8** Joint continuous random variables and density function The \( k \)-dimensional random variable \( (X_1, X_2, \ldots, X_k) \) is defined to be a \( k \)-dimensional continuous random variable if and only if there exists a function \( f_{x_1, \ldots, x_k}(\cdot, \ldots, \cdot) \geq 0 \) such that

\[
F_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f_{x_1, \ldots, x_k}(u_1, \ldots, u_k) \, du_1 \cdots du_k
\]

for all \((x_1, \ldots, x_k). \)

\( f_{x_1, \ldots, x_k}(\cdot, \ldots, \cdot) \) is defined to be a joint probability density function.
As in the unidimensional case, a joint probability density function has two properties:

(i) \( f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) \geq 0 \).

(ii) \( \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) \, dx_1 \cdots dx_k = 1 \).

A unidimensional probability density function was used to find probabilities. For example, for \( X \) a continuous random variable with probability density \( f_X(\cdot) \), \( P[a < X < b] = \int_a^b f_X(x) \, dx \); that is, the area under \( f_X(\cdot) \) over the interval \((a, b)\) gave \( P[a < X < b] \); and, more generally, \( P[X \in B] = \int_B f_X(x) \, dx \); that is, the area under \( f_X(\cdot) \) over the set \( B \) gave \( P[X \in B] \). In the two-dimensional case, volume gives probabilities. For instance, let \((X_1, X_2)\) be jointly continuous random variables with joint probability density function \( f_{X_1, X_2}(x_1, x_2) \), and let \( R \) be some region in the \( x_1x_2 \) plane; then \( P[(X_1, X_2) \in R] = \int_R f_{X_1, X_2}(x_1, x_2) \, dx_1 \, dx_2 \); that is, the probability that \((X_1, X_2)\) falls in the region \( R \) is given by the volume under \( f_{X_1, X_2}(\cdot, \cdot) \) over the region \( R \). In particular if \( R = \{(x_1, x_2) : a_1 < x_1 \leq b_1; a_2 < x_2 \leq b_2\} \), then

\[
P[a_1 < X_1 \leq b_1; a_2 < X_2 \leq b_2] = \int_{a_2}^{b_2} \left[ \int_{a_1}^{b_1} f_{X_1, X_2}(x_1, x_2) \, dx_1 \right] \, dx_2.
\]

A joint probability density function is defined as any nonnegative integrand satisfying Eq. (2) and hence is not uniquely defined.

**EXAMPLE 6** Consider the bivariate function

\[
f(x, y) = K(x + y)I_{0, 1}(x)I_{0, 1}(y) = K(x + y)I_0(x, y),
\]

where \( U = \{(x, y) : 0 < x < 1 \text{ and } 0 < y < 1\} \), a unit square. Can the constant \( K \) be selected so that \( f(x, y) \) will be a joint probability density function? If \( K \) is positive, \( f(x, y) \geq 0 \).

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Kf(x, y) \, dx \, dy = \int_0^1 \int_0^1 K(x + y) \, dx \, dy
\]

\[
= K \int_0^1 \int_0^1 (x + y) \, dx \, dy
\]

\[
= K \int_0^1 (\frac{1}{2} + y) \, dy
\]

\[
= K(\frac{1}{2} + \frac{1}{2})
\]

\[
= 1
\]
for $K = 1$. So $f(x, y) = (x + y)I_{[0,1]}(x)I_{[0,1]}(y)$ is a joint probability density function. It is sketched in Fig. 6.

Probabilities of events defined in terms of the random variables can be obtained by integrating the joint probability density function over the indicated region; for example

$$P[0 < X < \frac{1}{2}; 0 < Y < \frac{1}{2}] = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} (x + y) \, dx \, dy$$

$$= \int_0^{\frac{1}{2}} \left( \frac{y^2}{2} \right) \, dy$$

$$= \frac{1}{8} + \frac{1}{8}$$

$$= \frac{1}{4},$$

which is the volume under the surface $z = x + y$ over the region $(x, y): 0 < x < \frac{1}{2}; 0 < y < \frac{1}{2}$ in the $xy$ plane.

\textbf{Theorem 2} If $X$ and $Y$ are jointly continuous random variables, then knowledge of $F_{X,Y}(\cdot, \cdot)$ is equivalent to knowledge of an $f_{X,Y}(\cdot, \cdot)$. The remark extends to $k$-dimensional continuous random variables.

\textbf{Proof} For a given $f_{X,Y}(\cdot, \cdot)$, $F_{X,Y}(x, y)$ is obtained for any $(x, y)$ by

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) \, du \, dv.$$
For given \( F_{X,Y}(\cdot, \cdot) \), an \( f_{X,Y}(x, y) \) can be obtained by

\[
f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}
\]

for \( x, y \) points, where \( F_{X,Y}(x, y) \) is differentiable.

Definition 9 Marginal probability density functions If \( X \) and \( Y \) are jointly continuous random variables, then \( f_X(\cdot) \) and \( f_Y(\cdot) \) are called marginal probability density functions. More generally, let \( X_{i_1}, \ldots, X_{i_m} \) be any subset of the jointly continuous random variables \( X_1, \ldots, X_k \).

\( f_{X_{i_1}, \ldots, X_{i_m}}(x_{i_1}, \ldots, x_{i_m}) \) is called a marginal density of the \( m \)-dimensional random variable \((X_{i_1}, \ldots, X_{i_m})\).

Remark If \( X_1, \ldots, X_k \) are jointly continuous random variables, then any marginal probability density function can be found. (However, knowledge of all marginal densities does not, in general, imply knowledge of the joint density, as Example 8 below shows.) If \( X \) and \( Y \) are jointly continuous, then

\[
f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \tag{3}
\]

since

\[
f_X(x) = \frac{dF_X(x)}{dx} = \frac{d}{dx} \left[ \int_{-\infty}^{x} \left( \int_{-\infty}^{\infty} f_{X,Y}(u, y) \, dy \right) \, du \right] = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy.
\]

EXAMPLE 7 Consider the joint probability density

\[
f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y).
\]

\[
F_{X,Y}(x, y) = I_{(0,1)}(x)I_{(0,1)}(y) \int_{0}^{x} \int_{0}^{y} (u + v) \, du \, dv
\]

\[
+ I_{(0,1)}(x)I_{[1,\infty)}(y) \int_{0}^{x} \int_{0}^{1} (u + v) \, du \, dv
\]

\[
+ I_{[1,\infty)}(x)I_{(0,1)}(y) \int_{0}^{1} \int_{0}^{y} (u + v) \, du \, dv
\]

\[
+ I_{[1,\infty)}(x)I_{[1,\infty)}(y)
\]

\[
= \frac{1}{2}[(x^2 + y + xy)I_{(0,1)}(x)I_{(0,1)}(y) + (x^2 + x)I_{(0,1)}(x)I_{[1,\infty)}(y)
\]

\[
+ (y + y^2)I_{[1,\infty)}(x)I_{(0,1)}(y) + I_{[1,\infty)}(x)I_{[1,\infty)}(y)].
\]
\[ f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dy \]
\[ = I_{(0,1)}(x) \int_0^1 (x + y) \, dy \]
\[ = (x + \frac{1}{2}) I_{(0,1)}(x); \]

or,
\[ f_X(x) = \frac{\partial F_{X,Y}(x, \infty)}{\partial x} \]
\[ = \frac{\partial F_X(x)}{\partial x} \]
\[ = I_{(0,1)}(x) \frac{\partial}{\partial x} \left( \frac{x^2 + x}{2} \right) \]
\[ = (x + \frac{1}{2}) I_{(0,1)}(x). \]

**EXAMPLE 8** Let \( f_x(x) \) and \( f_y(y) \) be two probability density functions with corresponding cumulative distribution functions \( F_x(x) \) and \( F_y(y) \), respectively. For \(-1 \leq \alpha \leq 1\), define
\[ f_{x,y}(x,y; \alpha) = f_x(x)f_y(y)[1 + \alpha[2F_x(x) - 1][2F_y(y) - 1]]. \]  \hspace{1cm} (4)

We will show (i) that for each \( \alpha \) satisfying \(-1 \leq \alpha \leq 1\), \( f_{x,y}(x,y; \alpha) \) is a joint probability density function and (ii) that the marginals of \( f_{x,y}(x,y; \alpha) \) are \( f_x(x) \) and \( f_y(y) \), respectively. Thus, \( \{f_{x,y}(x,y; \alpha); -1 \leq \alpha \leq 1\} \) will be an infinite family of joint probability density functions, each having the same two given marginals. To verify (i) we must show that \( f_{x,y}(x,y; \alpha) \) is nonnegative and, if integrated over the \( xy \) plane, integrates to 1.

\[ f_x(x)f_y(y)[1 + \alpha[2F_x(x) - 1][2F_y(y) - 1]] \geq 0 \]
\[ \text{if } 1 \geq -\alpha[2F_x(x) - 1][2F_y(y) - 1]; \]

but \( \alpha, 2F_x(x) - 1, \) and \( 2F_y(y) - 1 \) are all between \(-1 \) and \( 1 \), and hence also their product, which implies \( f_{x,y}(x,y; \alpha) \) is nonnegative. Since
\[ 1 = \int_{-\infty}^{\infty} f_x(x) \, dx = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f_{x,y}(x,y; \alpha) \, dy \right) \, dx, \]
it suffices to show that $f_X(x)$ and $f_Y(y)$ are the marginals of $f_{X,Y}(x,y;\alpha)$.

$$
\int_{-\infty}^{\infty} f_{X,Y}(x,y;\alpha) \, dy \\
= \int_{-\infty}^{\infty} f_X(x)f_Y(y)[1 + \alpha[2F_X(x) - 1][2F_Y(y) - 1]] \, dy \\
= f_X(x) \int_{-\infty}^{\infty} f_Y(y) \, dy + \alpha f_X(x)[2F_X(x) - 1] \int_{-\infty}^{\infty} [2F_Y(y) - 1]f_Y(y) \, dy \\
= f_X(x), \quad \text{noting that} \quad \int_{-\infty}^{\infty} [2F_Y(y) - 1]f_Y(y) \, dy \\
= \int_{0}^{1} (2u - 1) \, du = 0 \\
\text{by making the transformation} \quad u = F_Y(y).
$$

3 CONDITIONAL DISTRIBUTIONS AND STOCHASTIC INDEPENDENCE

In the preceding section we defined the joint distribution and joint density functions of several random variables; in this section we define conditional distributions and the related concept of stochastic independence. Most definitions will be given first for only two random variables and later extended to $k$ random variables.

3.1 Conditional Distribution Functions for Discrete Random Variables

**Definition 10** Conditional discrete density function Let $X$ and $Y$ be jointly discrete random variables with joint discrete density function $f_{X,Y}(\cdot,\cdot)$. The conditional discrete density function of $Y$ given $X=x$, denoted by $f_{Y|X}(\cdot|x)$, is defined to be

$$
f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad (5)
$$

if $f_X(x) > 0$, where $f_X(x)$ is the marginal density of $X$ evaluated at $x$. $f_{Y|X}(\cdot|x)$ is undefined for $f_X(x) = 0$. Similarly,

$$
f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad (6)
$$

if $f_Y(y) > 0$. 

\[
\]
Since $X$ and $Y$ are discrete, they have mass points, say $x_1, x_2, \ldots$ for $X$ and $y_1, y_2, \ldots$ for $Y$. If $f_X(x) > 0$, then $x = x_i$ for some $i$, and $f_X(x_i) = P[X = x_i]$. The numerator of the right-hand side of Eq. (5) is $f_{X,Y}(x_i, y_j)$ = $P[X = x_i; Y = y_j]$; so

$$f_{Y|X}(y_j|x) = \frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)} = \frac{P[X = x_i; Y = y_j]}{P[X = x_i]} = \frac{P[Y = y_j|X = x_i]}{f_X(x_i)}$$

for $y_j$ a mass point of $Y$ and $x_i$ a mass point of $X$; hence $f_{Y|X}(\cdot|x)$ is a conditional probability as defined in Subsec. 3.6 of Chap. 1. $f_{Y|X}(\cdot|x)$ is called a conditional discrete density function and hence should possess the properties of a discrete density function. To see that it does, consider $x$ as some fixed mass point of $X$. Then $f_{Y|X}(y|x)$ is a function with argument $y$, and to be a discrete density function must be nonnegative and, if summed over the possible values (mass points) of $Y$, must sum to 1. $f_{Y|X}(y|x)$ is nonnegative since $f_{X,Y}(x, y)$ is nonnegative and $f_X(x)$ is positive.

$$\sum_j f_{Y|X}(y_j|x) = \sum_j \frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)} = \frac{1}{f_X(x)} \sum_j f_{X,Y}(x_i, y_j) = f_X(x) = 1,$$

where the summation is over all the mass points of $Y$. (We used the fact that the marginal discrete density of $X$ is obtained by summing the joint density of $X$ and $Y$ over the possible values of $Y$.) So $f_{Y|X}(\cdot|x)$ is indeed a density; it tells us how the values of $Y$ are distributed for a given value $x$ of $X$.

The conditional cumulative distribution of $Y$ given $X = x$ can be defined for two jointly discrete random variables by recalling the close relationship between discrete density functions and cumulative distribution functions.

**Definition 11** Conditional discrete cumulative distribution If $X$ and $Y$ are jointly discrete random variables, the conditional cumulative distribution of $Y$ given $X = x$, denoted by $F_{Y|X}(\cdot|x)$, is defined to be $F_{Y|X}(y|x) = P[Y \leq y|X = x]$ for $f_X(x) > 0$.

**Remark** $F_{Y|X}(y|x) = \sum_{y \leq y_j < x} f_{Y|X}(y_j|x)$.

**Example 9** Return to the experiment of tossing two tetrahedra. Let $X$ denote the number on the downturned face of the first and $Y$ the larger of the downturned numbers. What is the density of $Y$ given that $X = 2$?
\[
\begin{align*}
    f_{Y \mid X}(2 \mid 2) &= \frac{f_{X \times Y}(2, 2)}{f_X(2)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1 \\
    f_{Y \mid X}(3 \mid 2) &= \frac{f_{X \times Y}(2, 3)}{f_X(2)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1 \\
    f_{Y \mid X}(4 \mid 2) &= \frac{f_{X \times Y}(2, 4)}{f_X(2)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1
\end{align*}
\]

Also,
\[
    f_{Y \mid X}(y \mid 3) = \begin{cases} 
        \frac{1}{2} & \text{for } y = 3 \\
        \frac{1}{4} & \text{for } y = 4.
\end{cases}
\]

**Definition 12** Conditional discrete density function

Let \((X_1, \ldots, X_k)\) be a \(k\)-dimensional discrete random variable, and let \(X_{i_1}, \ldots, X_{i_r}\) and \(X_{j_1}, \ldots, X_{j_s}\) be two disjoint subsets of the random variables \(X_1, \ldots, X_k\).

The **conditional density** of the \(r\)-dimensional random variable \((X_{i_1}, \ldots, X_{i_r})\) given the value \((x_{j_1}, \ldots, x_{j_s})\) of \((X_{j_1}, \ldots, X_{j_s})\) is defined to be
\[
    f_{X_{i_1}, \ldots, X_{i_r} \mid X_{j_1}, \ldots, X_{j_s}}(x_{i_1}, \ldots, x_{i_r} \mid x_{j_1}, \ldots, x_{j_s})
    = \frac{f_{X_{i_1}, \ldots, X_{i_r}, X_{j_1}, \ldots, X_{j_s}}(x_{i_1}, \ldots, x_{i_r}, x_{j_1}, \ldots, x_{j_s})}{f_{X_{j_1}, \ldots, X_{j_s}}(x_{j_1}, \ldots, x_{j_s})}.
\]

**EXAMPLE 10** Let \(X_1, \ldots, X_5\) be jointly discrete random variables. Take \(r = s = 2\), \((X_{i_1}, X_{i_2}) = (X_1, X_2)\), and \((X_{j_1}, X_{j_2}) = (X_3, X_5)\); then
\[
    f_{X_1, X_2 \mid X_3, X_5}(x_1, x_2 \mid x_3, x_5) = \frac{f_{X_1, X_2, X_3, X_5}(x_1, x_2, x_3, x_5)}{f_{X_3, X_5}(x_3, x_5)}.
\]

**EXAMPLE 11** Suppose 12 cards are drawn without replacement from an ordinary deck of playing cards. Let \(X_1\) be the number of aces drawn, \(X_2\) be the number of 2s, \(X_3\) be the number of 3s, and \(X_4\) be the number of 4s. The joint density of these four random variables is given by
\[
    f_{X_1, X_2, X_3, X_4}(x_1, x_2, x_3, x_4)
    = \binom{4}{x_1} \binom{4}{x_2} \binom{4}{x_3} \binom{4}{x_4} \binom{36}{12 - x_1 - x_2 - x_3 - x_4}.
\]

\[
    \binom{52}{12}
\]
where \( x_i = 0, 1, 2, 3, \) or \( 4 \) and \( i = 1, \ldots, 4 \), subject to the restriction that 
\[ \sum x_i \leq 12. \]
There are a large number of conditional densities associated 
with this density; an example is

\[
f_{X_2, X_4|x_1, x_3}(x_2, x_4|x_1, x_3) = \left( \frac{4!}{x_1! (x_2! (x_3! (x_4! (12 - x_1 - x_2 - x_3 - x_4)! / (12)! \right) \frac{36}{44}
\[
= \left( \frac{x_2! (x_4! (12 - x_1 - x_2 - x_3 - x_4)! / (12)! \right) \frac{36}{44}
\]
where \( x_i = 0, 1, \ldots, 4 \) and \( x_2 + x_4 \leq 12 - x_1 - x_3. \)

### 3.2 Conditional Distribution Functions

for Continuous Random Variables

**Definition 13** Conditional probability density function Let \( X \) and \( Y \) 
be jointly continuous random variables with joint probability density function \( f_{X,Y}(x, y) \). The **conditional probability density function** of \( Y \) 
given \( X = x \), denoted by \( f_{Y|X}(\cdot | x) \), is defined to be

\[
f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}
\] 
(7)

if \( f_X(x) > 0 \), where \( f_X(x) \) is the marginal probability density of \( X \), and is 
undefined at points when \( f_X(x) = 0 \).

Similarly,

\[
f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0,
\] 
(8)

and is undefined if \( f_Y(y) = 0 \).

\( f_{Y|X}(\cdot | x) \) is called a (conditional) probability density function and hence 
should possess the properties of a probability density function. \( f_{Y|X}(\cdot | x) \) is 
clearly nonnegative, and

\[
\int_{-\infty}^{\infty} f_{Y|X}(y|x) \, dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_X(x)} \, dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy = \frac{f_X(x)}{f_X(x)} = 1.
\]
The density \( f_{Y|X}(\cdot | x) \) is a density of the random variable \( Y \) given that \( x \) is the value of the random variable \( X \). In the conditional density \( f_{Y|X}(\cdot | x_0) \), \( x \) is fixed and could be thought of as a parameter. Consider \( f_{Y|X}(\cdot | x_0) \), that is, the density of \( Y \) given that \( X \) was observed to be \( x_0 \). Now \( f_{x,y}(x, y) \) plots as a surface over the \( xy \) plane. A plane perpendicular to the \( xy \) plane which intersects the \( xy \) plane on the line \( x = x_0 \) will intersect the surface in the curve \( f_{x,y}(x_0, y) \). The area under this curve is

\[
\int_{-\infty}^{\infty} f_{x,y}(x_0, y) \, dy = f_X(x_0).
\]

Hence, if we divide \( f_{x,y}(x_0, y) \) by \( f_X(x_0) \), we obtain a density which is precisely \( f_{Y|X}(\cdot | x_0) \).

Again, the conditional cumulative distribution can be defined in the natural way.

**Definition 14 Conditional continuous cumulative distribution** If \( X \) and \( Y \) are jointly continuous, then the *conditional cumulative distribution* of \( Y \) given \( X = x \) is defined as

\[
F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(z|x) \, dz
\]

for all \( x \) such that \( f_X(x) > 0 \).

---

**EXAMPLE 12** Suppose \( f_{x,y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y) \).

\[
f_{Y|X}(y|x) = \frac{(x + y)I_{(0,1)}(x)I_{(0,1)}(y)}{(x + \frac{1}{2})I_{(0,1)}(x)} = \frac{x + y}{x + \frac{1}{2}} \cdot I_{(0,1)}(y)
\]

for \( 0 < x < 1 \). Note that

\[
F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(z|x) \, dz
\]

\[
= \int_{0}^{y} \frac{x + z}{x + \frac{1}{2}} \, dz = \frac{1}{x + \frac{1}{2}} \cdot \int_{0}^{y} (x + z) \, dz
\]

\[
= \frac{1}{x + \frac{1}{2}} (xy + y^2/2) \quad \text{for } 0 < y < 1.
\]

---

Conditional probability density functions can be analogously defined for \( k \)-dimensional continuous random variables. For instance,

\[
f_{x_1, x_2, x_4|x_3, x_5}(x_1, x_2, x_4 | x_3, x_5) = \frac{f_{x_1, x_2, x_3, x_4, x_5}(x_1, x_2, x_3, x_4, x_5)}{f_{x_3, x_5}(x_3, x_5)}
\]

for \( f_{x_3, x_5}(x_3, x_5) > 0 \).
3.3 More on Conditional Distribution Functions

We have defined the conditional cumulative distribution \( F_{Y \mid X}(y \mid x) \) for either jointly continuous or jointly discrete random variables. If \( X \) is discrete and \( Y \) is any random variable, then \( F_{Y \mid X}(y \mid x) \) can be defined as \( P(Y \leq y \mid X = x) \) if \( x \) is a mass point of \( X \). We would like to define \( P(Y \leq y \mid X = x) \) and more generally \( P(A \mid X = x) \), where \( A \) is any event, for \( X \) either a discrete or continuous random variable. Thus we seek to define the conditional probability of an event \( A \) given a random variable \( X = x \).

We start by assuming that the event \( A \) and the random variable \( X \) are both defined on the same probability space. We want to define \( P(A \mid X = x) \). If \( X \) is discrete, either \( x \) is a mass point of \( X \), or it is not; and if \( x \) is a mass point of \( X \),

\[
P(A \mid X = x) = \frac{P(A; X = x)}{P(X = x)},
\]

which is well defined; on the other hand, if \( x \) is not a mass point of \( X \), we are not interested in \( P(A \mid X = x) \). Now if \( X \) is continuous, \( P(A \mid X = x) \) cannot be analogously defined since \( P(X = x) = 0 \); however, if \( x \) is such that the events \( \{x - h < X < x + h\} \) have positive probability for every \( h > 0 \), then \( P(A \mid X = x) \) could be defined as

\[
P(A \mid X = x) = \lim_{0 < h \to 0} P(A \mid x - h < X < x + h)
\]

provided that the limit exists. We will take Eq. (9) as our definition of \( P(A \mid X = x) \) if the indicated limit exists, and leave \( P(A \mid X = x) \) undefined otherwise. (It is, in fact, possible to give \( P(A \mid X = x) \) meaning even if \( P(X = x) = 0 \), and such is done in advanced probability theory.)

We will seldom be interested in \( P(A \mid X = x) \) per se, but will be interested in using it to calculate certain probabilities. We note the following formulas:

(i)

\[
P(A) = \sum_{i=1}^{\infty} P(A \mid X = x_i) f_x(x_i)
\]

if \( X \) is discrete with mass points \( x_1, x_2, \ldots \).

(ii)

\[
P(A) = \int_{-\infty}^{\infty} P(A \mid X = x) f_x(x) \, dx
\]

if \( X \) is continuous.

(iii)

\[
P(A; X \in B) = \sum_{\{i : x_i \in B\}} P(A \mid X = x_i) f_x(x_i)
\]
if $X$ is discrete with mass points $x_1, x_2, \ldots$.

(iv) \[ P[A; X \in B] = \int_B P[A | X = x] f_X(x) \, dx \] (13)

if $X$ is continuous.

Although we will not prove the above formulas, we note that Eq. (10) is just the theorem of total probabilities given in Subsec. 3.6 of Chap. I and the others are generalizations of the same. Some problems are of such a nature that it is easy to find $P[A | X = x]$ and difficult to find $P[A]$. If, however, $f_X(\cdot)$ is known, then $P[A]$ can be easily obtained using the appropriate one of the above formulas.

Remark $F_{X,Y}(x, y) = \int_{-\infty}^{x} F_{Y|X}(y | x') f_X(x') \, dx'$ results from Eq. (13) by taking $A = \{ Y \leq y \}$ and $B = (-\infty, x]$; and $F_{Y}(y) = \int_{-\infty}^{\infty} F_{Y|X}(y | x) f_X(x) \, dx$ is obtained from Eq. (11) by taking $A = \{ Y \leq y \}$.

We add one other formula, whose proof is also omitted. Suppose $A = \{ h(X, Y) \leq z \}$, where $h(\cdot, \cdot)$ is some function of two variables; then

(v) \[ P[A | X = x] = P[h(X, Y) \leq z | X = x] = P[h(x, Y) \leq z | X = x]. \] (14)

The following is a classical example that uses Eq. (11); another example utilizing Eqs. (14) and (11) appears at the end of the next subsection.

**EXAMPLE 13** Three points are selected randomly on the circumference of a circle. What is the probability that there will be a semicircle on which all three points will lie? By selecting a point "randomly," we mean that the point is equally likely to be any point on the circumference of the circle; that is, the point is uniformly distributed over the circumference of the circle. Let us use the first point to orient the circle; for example, orient the circle (assumed centered at the origin) so that the first point falls on the positive $x$ axis. Let $X$ denote the position of the second point, and let $A$ denote the event that all three points lie on the same half circle. $X$ is uniformly distributed over the interval $(0, 2\pi)$. According to Eq. (11), $P[A] = \int P[A | X = x] f_X(x) \, dx$. Note that for $0 < x < \pi$, $P[A | X = x] = (\pi - x + \pi)/2\pi$ since, given $X = x$, event $A$ occurs if and only if the third point falls between $x - \pi$ and $\pi$. Similarly, $P[A | X = x] = (x + \pi - \pi)/2\pi$ for $\pi \leq x < 2\pi$. Hence $P[A] = \int_0^\pi P[A | X = x](1/2\pi) \, dx = (1/2\pi) \left( \int_0^\pi \frac{\pi - x + \pi}{2\pi} \, dx + \int_\pi^{2\pi} \frac{x}{2\pi} \, dx \right) = \frac{3}{\pi}$. \(/\!/)
3.4 Independence

When we defined the conditional probability of two events in Chap. 1, we also defined independence of events. We have now defined the conditional distribution of random variables; so we should define independence of random variables as well.

**Definition 15 Stochastic independence** Let \((X_1, X_2, \ldots, X_k)\) be a \(k\)-dimensional random variable. \(X_1, X_2, \ldots, X_k\) are defined to be **stochastically independent** if and only if

\[
F_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} F_{x_i}(x_i)
\]

for all \(x_1, x_2, \ldots, x_k\).

**Definition 16 Stochastic independence** Let \((X_1, X_2, \ldots, X_k)\) be a \(k\)-dimensional discrete random variable with joint discrete density function \(f_{x_1, \ldots, x_k}(\cdot, \ldots, \cdot)\). \(X_1, \ldots, X_k\) are **stochastically independent** if and only if

\[
f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{x_i}(x_i)
\]

for all values \((x_1, \ldots, x_k)\) of \((X_1, \ldots, X_k)\).

**Definition 17 Stochastic independence** Let \((X_1, \ldots, X_k)\) be a \(k\)-dimensional continuous random variable with joint probability density function \(f_{x_1, \ldots, x_k}(\cdot, \ldots, \cdot)\). \(X_1, \ldots, X_k\) are **stochastically independent** if and only if

\[
f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{x_i}(x_i)
\]

for all \(x_1, \ldots, x_k\).

**Remark** Often the word "stochastically" will be omitted.

We saw that independence of events was closely related to conditional probability; likewise independence of random variables is closely related to conditional distributions of random variables. For example, suppose \(X\) and \(Y\) are two independent random variables; then \(f_{X,Y}(x, y) = f_X(x)f_Y(y)\) by definition of independence; however, \(f_{X,Y}(x, y) = f_{Y|x}(y|x) f_X(x)\) by definition of conditional density, which implies that \(f_{Y|x}(y|x) = f_Y(y)\); that is, the conditional
density of \( Y \) given \( x \) is the unconditional density of \( Y \). So to show that two random variables are \textit{not} independent, it suffices to show that \( f_{Y\mid X}(y\mid x) \) depends on \( x \).

**Example 14** Let \( X \) be the number on the downturned face of the first tetrahedron and \( Y \) the larger of the two downturned numbers in the experiment of tossing two tetrahedra. Are \( X \) and \( Y \) independent? Obviously not, since \( f_{Y\mid X}(2\mid 3) = P[Y = 2 \mid X = 3] = 0 \neq f_{Y}(2) = P[Y = 2] = 0.5 \).

**Example 15** Let \( f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y) \). Are therefore \( X \) and \( Y \) independent? No, since \( f_{Y\mid X}(y\mid x) = [(x + y)(x + \frac{1}{2})]I_{(0,1)}(y) \) for \( 0 < x < 1 \). \( f_{Y\mid X}(y\mid x) \) depends on \( x \) and hence cannot equal \( f_{Y}(y) \).

**Example 16** Let \( f_{X,Y}(x, y) = e^{-(x+y)}I_{(0,\infty)}(x)I_{(0,\infty)}(y) \). \( X \) and \( Y \) are independent since

\[
    f_{X,Y}(x, y) = [e^{-x}I_{(0,\infty)}(x)][e^{-y}I_{(0,\infty)}(y)] = f_{X}(x)f_{Y}(y)
\]

for all \((x, y)\).

It can be proved that if \( X_1, \ldots, X_k \) are jointly continuous random variables, then Definitions 15 and 17 are equivalent. Similarly, for jointly discrete random variables, Definitions 15 and 16 are equivalent. It can also be proved that Eq. (15) is equivalent to \( P[X_1 \in B_1; \ldots; X_k \in B_k] = \prod_{i=1}^{k} P[X_i \in B_i] \) for sets \( B_1, \ldots, B_k \). The following important result is easily derived using the above equivalent notions of independence.

**Theorem 3** If \( X_1, \ldots, X_k \) are independent random variables and \( g_1(\cdot), \ldots, g_k(\cdot) \) are \( k \) functions such that \( Y_j = g_j(X_j), j = 1, \ldots, k \) are random variables, then \( Y_1, \ldots, Y_k \) are independent.

**Proof** Note that if \( g_j^{-1}(B_j) = \{z: g_j(z) \in B_j\} \), then the events \( \{Y_j \in B_j\} \) and \( \{X_j \in g_j^{-1}(B_j)\} \) are equivalent; consequently, \( P[Y_1 \in B_1; \ldots; Y_k \in B_k] = P[X_1 \in g_1^{-1}(B_1); \ldots; X_k \in g_k^{-1}(B_k)] = \prod_{j=1}^{k} P[X_j \in g_j^{-1}(B_j)] \).

\[
= \prod_{j=1}^{k} P[Y_j \in B_j].
\]
For $k = 2$, the above theorem states that if two random variables, say $X$ and $Y$, are independent, then a function of $X$ is independent of a function of $Y$. Such a result is certainly intuitively plausible.

We will return to independence of random variables in Subsec. 4.5.

Equation (14) of the previous subsection states that $P[h(X, Y) \leq z \mid X = x] = P[h(x, Y) \leq z \mid X = x]$. Now if $X$ and $Y$ are assumed to be independent, then $P[h(x, Y) \leq z \mid X = x] = P[h(x, Y) \leq z]$, which is a probability that may be easy to calculate for certain problems.

**EXAMPLE 17** Let a random variable $Y$ represent the diameter of a shaft and a random variable $X$ represent the inside diameter of the housing that is intended to support the shaft. By design the shaft is to have diameter 99.5 units and the housing inside diameter 100 units. If the manufacturing process of each of the items is imperfect, so that in fact $Y$ is uniformly distributed over the interval $(98.5, 100.5)$ and $X$ is uniformly distributed over $(99, 101)$, what is the probability that a particular shaft can be successfully paired with a particular housing, when “successfully paired” is taken to mean that $X - h < Y < X$ for some small positive quantity $h$? Assume that $X$ and $Y$ are independent; then

$$P[X - h < Y < X] = \int_{-\infty}^{\infty} P[X - h < Y < X \mid X = x] f_X(x) \, dx$$

$$= \int_{99}^{101} P[x - h < Y < x] \frac{1}{2} \, dx.$$  

Suppose now that $h = 1$; then

$$P[x - 1 < Y < x] = \begin{cases} \frac{x - 98.5}{2} & \text{for } 99 < x \leq 99.5 \\ \frac{1}{2} & \text{for } 99.5 < x < 100.5 \\ \frac{100.5 - (x - 1)}{2} & \text{for } 100.5 < x \leq 101. \end{cases}$$

Hence,

$$P[X - 1 < Y < X] = \int_{99}^{101} P[x - 1 < Y < x] \frac{1}{2} \, dx$$

$$= \int_{99}^{99.5} \frac{1}{2}(x - 98.5) \frac{1}{2} \, dx$$

$$+ \int_{99.5}^{100.5} \frac{1}{2} \, dx + \int_{100.5}^{101} \cdot \frac{1}{2}(100.5 - x + 1) \frac{1}{2} \, dx = \frac{1}{2}.$$
4 EXPECTATION

When we introduced the concept of expectation for univariate random variables in Sec. 4 of Chap. II, we first defined the mean and variance as particular expectations and then defined the expectation of a general function of a random variable. Here, we will commence, in Subsec. 4.1, with the definition of the expectation of a general function of a k-dimensional random variable. The definition will be given for only those k-dimensional random variables which have densities.

4.1 Definition

**Definition 18 Expectation** Let \((X_1, \ldots, X_k)\) be a k-dimensional random variable with density \(f_{x_1, \ldots, x_k}(\cdot, \ldots, \cdot)\). The expected value of a function \(g(\cdot, \ldots, \cdot)\) of the k-dimensional random variable, denoted by \(\mathbb{E}[g(X_1, \ldots, X_k)]\), is defined to be

\[
\mathbb{E}[g(X_1, \ldots, X_k)] = \sum_{x_1} \cdots \sum_{x_k} g(x_1, \ldots, x_k) f_{x_1, \ldots, x_k}(x_1, \ldots, x_k)
\]

(18)

if the random variable \((X_1, \ldots, X_k)\) is discrete where the summation is over all possible values of \((X_1, \ldots, X_k)\), and

\[
\mathbb{E}[g(X_1, \ldots, X_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k) f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) \, dx_1 \cdots dx_k
\]

(19)

if the random variable \((X_1, \ldots, X_k)\) is continuous.

In order for the above to be defined, it is understood that the sum and multiple integral, respectively, exist.

**Theorem 4** In particular, if \(g(x_1, \ldots, x_k) = x_i\), then

\[
\mathbb{E}[g(X_1, \ldots, X_k)] = \mathbb{E}[X_i] = \mu_i.
\]

(20)

**Proof** Assume that \((X_1, \ldots, X_k)\) is continuous. [The proof for \((X_1, \ldots, X_k)\) discrete is similar.]

\[
\mathbb{E}[g(X_1, \ldots, X_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{x_1, \ldots, x_k}(x_1, \ldots, x_k) \, dx_1 \cdots dx_k
\]

\[= \int_{-\infty}^{\infty} x_i f_{x_i}(x_i) \, dx_i = \mathbb{E}[X_i]
\]
using the fact that the marginal density \( f_{X_i}(x_i) \) is obtained from the joint density by

\[
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \, dx_1 \cdots dx_{i-1} \cdot dx_{i+1} \cdots dx_n.
\]

Similarly, the following theorem can be proved.

**Theorem 5** If \( g(x_1, \ldots, x_n) = (x_i - \mathcal{E}[X_i])^2 \), then

\[
\mathcal{E}[(X_i - \mathcal{E}[X_i])^2] = \text{var} \{X_i\}.
\]

We might note that the "expectation" in the notation \( \mathcal{E}[X_i] \) of Eq. (20) has two different interpretations; one is that the expectation is taken over the joint distribution of \( X_1, \ldots, X_n \), and the other is that the expectation is taken over the marginal distribution of \( X_i \). What Theorem 4 really says is that these two expectations are equivalent, and hence we are justified in using the same notation for both.

**Example 18** Consider the experiment of tossing two tetrahedra. Let \( X \) be the number on the first and \( Y \) the larger of the two numbers. We gave the joint discrete density function of \( X \) and \( Y \) in Example 2.

\[
\mathcal{E}[XY] = \sum \sum xy f_{X,Y}(x,y)
\]

\[
= 1 \cdot 1\frac{1}{16} + 1 \cdot 2\frac{1}{16} + 1 \cdot 3\frac{1}{16} + 1 \cdot 4\frac{1}{16}
\]

\[
+ 2 \cdot 2\frac{1}{16} + 2 \cdot 3\frac{1}{16} + 2 \cdot 4\frac{1}{16} + 3 \cdot 3\frac{1}{16}
\]

\[
+ 3 \cdot 4\frac{1}{16} + 4 \cdot 4\frac{1}{16} = \frac{15}{2}.
\]

\[
\mathcal{E}[X + Y] = (1 + 1)\frac{1}{16} + (1 + 2)\frac{1}{16} + (1 + 3)\frac{1}{16} + (1 + 4)\frac{1}{16}
\]

\[
+ (2 + 2)\frac{1}{16} + (2 + 3)\frac{1}{16} + (2 + 4)\frac{1}{16} + (3 + 3)\frac{1}{16}
\]

\[
+ (3 + 4)\frac{1}{16} + (4 + 4)\frac{1}{16} = \frac{9}{2}.
\]

\[
\mathcal{E}[X] = \frac{1}{2}, \text{ and } \mathcal{E}[Y] = \frac{1}{3}; \text{ hence } \mathcal{E}[X + Y] = \mathcal{E}[X] + \mathcal{E}[Y].
\]

**Example 19** Suppose \( f_{X,Y}(x,y) = (x+y)I_{(0,1)}(x)I_{(0,1)}(y) \).

\[
\mathcal{E}[XY] = \int_0^1 \int_0^1 xy(x+y) \, dx \, dy = \frac{1}{2}.
\]

\[
\mathcal{E}[X + Y] = \int_0^1 \int_0^1 (x+y)(x+y) \, dx \, dy = \frac{3}{2}.
\]

\[
\mathcal{E}[X] = \mathcal{E}[Y] = \frac{1}{2}.
\]
EXAMPLE 20  Let the three-dimensional random variable $(X_1, X_2, X_3)$ have the density

$$f_{x_1, x_2, x_3}(x_1, x_2, x_3) = 8x_1x_2x_3 I_{(0, 1)}(x_1)I_{(0, 1)}(x_2)I_{(0, 1)}(x_3).$$

Suppose we want to find (i) $\mathbb{E}[3X_1 + 2X_2 + 6X_3]$, (ii) $\mathbb{E}[X_1X_2X_3]$, and (iii) $\mathbb{E}[X_1X_2]$. For (i) we have $g(x_1, x_2, x_3) = 3x_1 + 2x_2 + 6x_3$ and obtain

$$\mathbb{E}[g(X_1, X_2, X_3)] = \mathbb{E}[3X_1 + 2X_2 + 6X_3] = \int_0^1 \int_0^1 \int_0^1 (3x_1 + 2x_2 + 6x_3) 8x_1x_2x_3 \, dx_1 \, dx_2 \, dx_3 = 42.$$

For (ii), we get

$$\mathbb{E}[X_1X_2X_3] = \int_0^1 \int_0^1 \int_0^1 8x_1x_2x_3^2 \, dx_1 \, dx_2 \, dx_3 = \frac{9}{2},$$

and for (iii) we get $\mathbb{E}[X_1X_2] = \frac{3}{5}$. \textbf{\\}

The following remark, the proof of which is left to the reader, displays a property of joint expectation. It is a generalization of (ii) in Theorem 3 of Chap. II.

\textbf{Remark}  \hspace{1cm} $\mathbb{E} \left[ \sum_{i=1}^m c_i g(X_1, \ldots, X_i) \right] = \sum_{i=1}^m c_i \mathbb{E} [g(X_1, \ldots, X_i)]$ for constants $c_1, c_2, \ldots, c_m$. \textbf{\\}

4.2 Covariance and Correlation Coefficient

\textbf{Definition 19  Covariance}  Let $X$ and $Y$ be any two random variables defined on the same probability space. The \textit{covariance} of $X$ and $Y$, denoted by $\text{cov} [X, Y]$ or $\sigma_{X,Y}$, is defined as

$$\text{cov} [X, Y] = \mathbb{E} [(X - \mu_X)(Y - \mu_Y)] \hspace{1cm} (21)$$

provided that the indicated expectation exists. \textbf{\\}

\textbf{Definition 20  Correlation coefficient}  The \textit{correlation coefficient}, denoted by $\rho(X, Y)$ or $\rho_{X,Y}$, of random variables $X$ and $Y$ is defined to be

$$\rho_{X,Y} = \frac{\text{cov} [X, Y]}{\sigma_X \sigma_Y} \hspace{1cm} (22)$$

provided that $\text{cov} [X, Y], \sigma_X$, and $\sigma_Y$ exist, and $\sigma_X > 0$ and $\sigma_Y > 0$. \textbf{\\}
Both the covariance and the correlation coefficient of random variables \(X\) and \(Y\) are measures of a linear relationship of \(X\) and \(Y\) in the following sense: \(\text{cov} [X, Y]\) will be positive when \(X - \mu_X\) and \(Y - \mu_Y\) tend to have the same sign with high probability, and \(\text{cov} [X, Y]\) will be negative when \(X - \mu_X\) and \(Y - \mu_Y\) tend to have opposite signs with high probability. \(\text{cov} [X, Y]\) tends to measure the linear relationship of \(X\) and \(Y\); however, its actual magnitude does not have much meaning since it depends on the variability of \(X\) and \(Y\). The correlation coefficient removes, in a sense, the individual variability of each \(X\) and \(Y\) by dividing the covariance by the product of the standard deviations, and thus the correlation coefficient is a better measure of the linear relationship of \(X\) and \(Y\) than is the covariance. Also, the correlation coefficient is unitless and, as we shall see in Subsec. 4.6 below, satisfies \(-1 \leq \rho_{X,Y} \leq 1\).

**Remark** \(\text{cov} [X, Y] = \sigma([X - \mu_X][Y - \mu_Y]) = \sigma[XY] - \mu_X \mu_Y.\)

**Proof**
\[
\sigma([X - \mu_X][Y - \mu_Y]) = \sigma[XY] - \mu_X \sigma[Y] - \mu_Y \sigma[X] + \mu_X \mu_Y
= \sigma[XY] - \mu_X \sigma[Y] - \mu_Y \sigma[X] + \mu_X \mu_Y
= \sigma[XY] - \mu_X \mu_Y.
\]

**EXAMPLE 21** Find \(\rho_{X,Y}\) for \(X\), the number on the first, and \(Y\), the larger of the two numbers, in the experiment of tossing two tetrahedra. We would expect that \(\rho_{X,Y}\) is positive since when \(X\) is large, \(Y\) tends to be large too. We calculated \(\sigma[XY]\), \(\sigma[X]\), and \(\sigma[Y]\) in Example 18 and obtained \(\sigma[XY] = \frac{1}{36}\), \(\sigma[X] = \frac{1}{2}\), and \(\sigma[Y] = \frac{1}{36}\). Thus \(\text{cov} [X, Y] = \frac{1}{36} - \frac{1}{2} \cdot \frac{1}{36} = \frac{1}{36}\). Now \(\sigma[X^2] = \frac{3}{4}\) and \(\sigma[Y^2] = \frac{1}{36}\); hence \(\text{var} [X] = \frac{1}{4}\) and \(\text{var} [Y] = \frac{1}{36}\). So,
\[
\rho_{X,Y} = \frac{-\frac{1}{36}}{\sqrt{\frac{1}{4}} \sqrt{\frac{1}{36}}} = \frac{2}{\sqrt{11}}.
\]

**EXAMPLE 22** Find \(\rho_{X,Y}\) for \(X\) and \(Y\) if \(f_{X,Y}(x, y) = (x + y)I_{(0, 1)}(x)I_{(0, 1)}(y)\).

We saw that \(\sigma[XY] = \frac{1}{2}\) and \(\sigma[X] = \sigma[Y] = \frac{1}{2}\) in Example 19. Now \(\sigma[X^2] = \sigma[Y^2] = \frac{1}{2}\); hence \(\text{var} [X] = \text{var} [Y] = \frac{1}{4}\). Finally
\[
\rho_{X,Y} = \frac{1 - \frac{1}{4}}{\frac{1}{4}} = \frac{1}{11}.
\]

Does a negative correlation coefficient seem right?
4.3 Conditional Expectations

In the following chapters we shall have occasion to find the expected value of random variables in conditional distributions, or the expected value of one random variable given the value of another.

**Definition 21** Conditional expectation Let \((X, Y)\) be a two-dimensional random variable and \(g(\cdot, \cdot)\), a function of two variables. The conditional expectation of \(g(X, Y)\) given \(X = x\), denoted by \(\mathcal{E}[g(X, Y) \mid X = x]\), is defined to be

\[
\mathcal{E}[g(X, Y) \mid X = x] = \int_{-\infty}^{\infty} g(x, y) f_{Y|X}(y \mid x) \, dy
\]

(23)

if \((X, Y)\) are jointly continuous, and

\[
\mathcal{E}[g(X, Y) \mid X = x] = \sum y \, g(x, y) f_{Y|X}(y \mid x)
\]

(24)

if \((X, Y)\) are jointly discrete, where the summation is over all possible values of \(Y\).

In particular, if \(g(x, y) = y\), we have defined \(\mathcal{E}[Y \mid X = x] = \mathcal{E}[Y \mid x]\). \(\mathcal{E}[Y \mid x]\) and \(\mathcal{E}[g(X, Y) \mid x]\) are functions of \(x\). Note that this definition can be generalized to more than two dimensions. For example, let \((X_1, \ldots, X_k, Y_1, \ldots, Y_m)\) be a \((k + m)\)-dimensional continuous random variable with density \(f_{X_1, \ldots, X_k, Y_1, \ldots, Y_m}(x_1, \ldots, x_k, y_1, \ldots, y_m)\); then

\[
\mathcal{E}[g(X_1, \ldots, X_k, Y_1, \ldots, Y_m) \mid x_1, \ldots, x_k] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k, y_1, \ldots, y_m) f_{Y_1, \ldots, Y_m \mid X_1, \ldots, X_k}(y_1, \ldots, y_m \mid x_1, \ldots, x_k) \, dy_1 \cdots dy_m.
\]

EXAMPLE 23 In the experiment of tossing two tetrahedra with \(X\), the number on the first, and \(Y\), the larger of the two numbers, we found that

\[
f_{Y|X}(y \mid 2) = \begin{cases} 
\frac{1}{2} & \text{for } y = 2 \\
\frac{1}{4} & \text{for } y = 3 \\
\frac{1}{4} & \text{for } y = 4 
\end{cases}
\]

in Example 9. Hence \(\mathcal{E}[Y \mid X = 2] = \sum y f_{Y|X}(y \mid X = 2) = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{11}{4}\).
EXAMPLE 24 For \( f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y) \), we found that
\[
f_{Y|X}(y|x) = \frac{x + y}{x + \frac{1}{2}} I_{(0,1)}(y)
\]
for \( 0 < x < 1 \) in Example 12. Hence
\[
\mathcal{E}[Y|X = x] = \int_0^1 y \frac{x + y}{x + \frac{1}{2}} dy = \frac{1}{x + \frac{1}{2}} \left( \frac{x}{3} + \frac{1}{3} \right)
\]
for \( 0 < x < 1 \).

As we stated above, \( \mathcal{E}[g(Y)|X] \) is, in general, a function of \( x \). Let us denote it by \( h(x) \); that is, \( h(x) = \mathcal{E}[g(Y)|X] \). Now we can evaluate the expectation of \( h(X) \), a function of \( X \), and will have \( \mathcal{E}[h(X)] = \mathcal{E}[\mathcal{E}[g(Y)|X]] \).

This gives us
\[
\mathcal{E}[\mathcal{E}[g(Y)|X]] = \mathcal{E}[h(X)] = \int_{-\infty}^{\infty} h(x)f_X(x)dx
\]
\[
= \int_{-\infty}^{\infty} \mathcal{E}[g(Y)|X]f_X(x)dx
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)dyf_X(x)dx
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)f_X(x)dydx
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f_{X,Y}(x, y)dydx
\]
\[
= \mathcal{E}[g(Y)].
\]

Thus we have proved for jointly continuous random variables \( X \) and \( Y \) (the proof for \( X \) and \( Y \) jointly discrete is similar) the following simple yet very useful theorem.

**Theorem 6** Let \((X, Y)\) be a two-dimensional random variable; then
\[
\mathcal{E}[g(Y)] = \mathcal{E}[\mathcal{E}[g(Y)|X]],
\]
and in particular
\[
\mathcal{E}[Y] = \mathcal{E}[\mathcal{E}[Y|X]].
\]

**Definition 22** Regression curve \( \mathcal{E}[Y|X = x] \) is called the regression curve of \( Y \) on \( x \). It is also denoted by \( \mu_{Y|X=x} = \mu_{Y|X} \).
Definition 23  Conditional variance  The variance of $Y$ given $X = x$ is defined by \( \text{var} [Y | X = x] = \sigma [Y^2 | X = x] - (\sigma [Y | X = x])^2. \)

Theorem 7  \( \text{var} [Y] = \sigma [\text{var} [Y | X]] + \text{var} [\sigma [Y | X]]. \)

**Proof**

\[
\begin{align*}
\sigma [\text{var} [Y | X]] &= \sigma [\sigma [Y^2 | X]] - \sigma [(\sigma [Y | X])^2] \\
&= \sigma [Y^2] - (\sigma [Y])^2 - \sigma [(\sigma [Y | X])^2] + (\sigma [Y])^2 \\
&= \text{var} [Y] - \sigma [(\sigma [Y | X])^2] + (\sigma [\sigma [Y | X]]).
\end{align*}
\]

Let us note in words what the two theorems say. Equation (26) states that the mean of $Y$ is the mean or expectation of the conditional mean of $Y$, and Theorem 7 states that the variance of $Y$ is the mean or expectation of the conditional variance of $Y$, plus the variance of the conditional mean of $Y$.

We will conclude this subsection with one further theorem. The proof can be routinely obtained from Definition 21 and is left as an exercise. Also, the theorem can be generalized to more than two dimensions.

Theorem 8  Let $(X, Y)$ be a two-dimensional random variable and $g_1(\cdot)$ and $g_2(\cdot)$ functions of one variable. Then

(i) \( \sigma [g_1(Y) + g_2(Y) | X = x] = \sigma [g_1(Y) | X = x] + \sigma [g_2(Y) | X = x]. \)

(ii) \( \sigma [g_1(Y)g_2(X) | X = x] = g_2(x)\sigma [g_1(Y) | X = x]. \)

4.4 Joint Moment Generating Function and Moments

We will use our definition of the expectation of a function of several variables to define joint moments and the joint moment generating function.

Definition 24  Joint moments  The joint raw moments of $X_1, \ldots, X_k$ are defined by \( \sigma [X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k}], \) where the $r_i$'s are 0 or any positive integer: the joint moments about the means are defined by \( \sigma [(X_1 - \mu_{X_1})^{r_1} \cdots (X_k - \mu_{X_k})^{r_k}]. \)

**Remark**  If $r_i = r_j = 1$ and all other $r_m$'s are 0, then that particular joint moment about the means becomes \( \sigma [(X_i - \mu_{X_i})(X_j - \mu_{X_j})], \) which is just the covariance between $X_i$ and $X_j$. 

//
Definition 25  Joint moment generating function  The joint moment generating function of \((X_1, \ldots, X_k)\) is defined by

\[
m_{x_1, \ldots, x_k}(t_1, \ldots, t_k) = \mathcal{E}\left[\exp \sum_{j=1}^{k} t_j X_j\right],
\]

if the expectation exists for all values of \(t_1, \ldots, t_k\) such that \(-h < t_j < h\) for some \(h > 0, j = 1, \ldots, k\).

The \(r\)th moment of \(X_j\) may be obtained from \(m_{x_1, \ldots, x_k}(t_1, \ldots, t_k)\) by differentiating it \(r\) times with respect to \(t_j\) and then taking the limit as all the \(t_i\)'s approach 0. Also \(\mathcal{E}[X_j^r]\) can be obtained by differentiating the joint moment generating function \(r\) times with respect to \(t_i\) and \(s\) times with respect to \(t_j\) and then taking the limit as all the \(t_i\)'s approach 0. Similarly other joint raw moments can be generated.

Remark  \(m_x(t_1) = m_{x, x}(t_1, 0) = \lim_{t_2 \to 0} m_{x, x, x}(t_1, t_2)\), and \(m_{x, x}(t_2) = m_{x, x}(0, t_2)\) be obtained from the joint moment generating function.

An example of a joint moment generating function will appear in Sec. 5 of this chapter.

4.5 Independence and Expectation

We have already defined independence and expectation; in this section we will relate the two concepts.

Theorem 9  If \(X\) and \(Y\) are independent and \(g_1(\cdot)\) and \(g_2(\cdot)\) are two functions, each of a single argument, then

\[
\mathcal{E}[g_1(X)g_2(Y)] = \mathcal{E}[g_1(X)] \cdot \mathcal{E}[g_2(Y)].
\]

Proof  We will give the proof for jointly continuous random variables.

\[
\mathcal{E}[g_1(X)g_2(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_{X,Y}(x,y) \, dx \, dy
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y) \, dx \, dy
\]

\[
= \int_{-\infty}^{\infty} g_1(x)f_X(x) \, dx \cdot \int_{-\infty}^{\infty} g_2(y)f_Y(y) \, dy
\]

\[
= \mathcal{E}[g_1(X)] \cdot \mathcal{E}[g_2(Y)].
\]
Corollary  If $X$ and $Y$ are independent, then $\text{cov} \{X, Y\} = 0$.

Proof  Take $g_1(x) = x - \mu_X$ and $g_2(y) = y - \mu_Y$; by Theorem 9,

$$\text{cov} \{X, Y\} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[g_1(X)g_2(Y)]$$

$$= \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]$$

$$= \mathbb{E}[X - \mu_X] \cdot \mathbb{E}[Y - \mu_Y] = 0 \quad \text{since} \mathbb{E}[X - \mu_X] = 0.$$  

Definition 26 Uncorrelated random variables  Random variables $X$ and $Y$ are defined to be uncorrelated if and only if $\text{cov} \{X, Y\} = 0$.

Remark  The converse of the above corollary is not always true; that is, $\text{cov} \{X, Y\} = 0$ does not always imply that $X$ and $Y$ are independent, as the following example shows.

Example 25  Let $U$ be a random variable which is uniformly distributed over the interval $(0, 1)$. Define $X = \sin 2\pi U$ and $Y = \cos 2\pi U$. $X$ and $Y$ are clearly not independent since if a value of $X$ is known, then $U$ is one of two values, and so $Y$ is also one of two values; hence the conditional distribution of $Y$ is not the same as the marginal distribution. $\mathbb{E}[Y] = \int_0^1 \cos 2\pi u \, du = 0$, and $\mathbb{E}[X] = \int_0^1 \sin 2\pi u \, du = 0$; so $\text{cov} \{X, Y\} = \mathbb{E}[XY] = \int_0^1 \sin 2\pi u \cos 2\pi u \, du = \frac{1}{4} \int_0^1 \sin 4\pi u \, du = 0$.

Theorem 10  Two jointly distributed random variables $X$ and $Y$ are independent if and only if $m_{X,Y}(t_1, t_2) = m_X(t_1)m_Y(t_2)$ for all $t_1, t_2$ for which $-h < t_i < h$, $i = 1, 2$, for some $h > 0$.

Proof  [Recall that $m_X(t_i)$ is the moment generating function of $X$. Also note that $m_Y(t_i) = m_{X,Y}(t_i, 0)$.] $X$ and $Y$ independent imply that the joint moment generating function factors into the product of the marginal moment generating functions by Theorem 9 by taking $g_1(x) = e^{tx}$ and $g_2(y) = e^{ty}$. The proof in the other direction will be omitted.

Remark  Both Theorems 9 and 10 can be generalized from two random variables to $k$ random variables.
4.6 Cauchy-Schwarz Inequality

Theorem 11 Cauchy-Schwarz inequality Let $X$ and $Y$ have finite second moments; then $(\mathcal{E}[XY])^2 = |\mathcal{E}[XY]|^2 \leq \mathcal{E}[X^2]\mathcal{E}[Y^2]$, with equality if and only if $P[Y = cX] = 1$ for some constant $c$.

PROOF The existence of expectations $\mathcal{E}[X]$, $\mathcal{E}[Y]$, and $\mathcal{E}[XY]$ follows from the existence of expectations $\mathcal{E}[X^2]$ and $\mathcal{E}[Y^2]$. Define $0 \leq h(t) = \mathcal{E}[(tX - Y)^2] = \mathcal{E}[X^2]t^2 - 2\mathcal{E}[XY]t + \mathcal{E}[Y^2]$. Now $h(t)$ is a quadratic function in $t$ which is greater than or equal to 0. If $h(t) > 0$, then the roots of $h(t)$ are not real; so $4\mathcal{E}[XY]^2 - 4\mathcal{E}[X^2]\mathcal{E}[Y^2] < 0$, or $(\mathcal{E}[XY])^2 < \mathcal{E}[X^2]\mathcal{E}[Y^2]$. If $h(t) = 0$ for some $t$, say $t_0$, then $\mathcal{E}[(t_0 X - Y)^2] = 0$, which implies $P[t_0 X = Y] = 1$.

Corollary $|\rho_{X,Y}| \leq 1$, with equality if and only if one random variable is a linear function of the other with probability 1.

PROOF Rewrite the Cauchy-Schwarz inequality as $|\mathcal{E}[UV]| \leq \sqrt{\mathcal{E}[U^2]\mathcal{E}[V^2]}$, and set $U = X - \mu_X$ and $V = Y - \mu_Y$.

5 BIVARIATE NORMAL DISTRIBUTION

One of the important multivariate densities is the multivariate normal density, which is a generalization of the normal distribution for a unidimensional random variable. In this section we shall discuss a special case, the case of the bivariate normal. In our discussion we will include the joint density, marginal densities, conditional densities, conditional means and variances, covariance, and the moment generating function. This section, then, will give an example of many of the concepts defined in the preceding sections of this chapter.

5.1 Density Function

Definition 27 Bivariate normal distribution Let the two-dimensional random variable $(X, Y)$ have the joint probability density function

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho^2}} \times \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho \frac{x - \mu_X}{\sigma_X} \frac{y - \mu_Y}{\sigma_Y} + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

(28)
for $-\infty < x < \infty$, $-\infty < y < \infty$, where $\sigma_y$, $\sigma_x$, $\mu_x$, $\mu_y$, and $\rho$ are constants such that $-1 < \rho < 1$, $0 < \sigma_y$, $0 < \sigma_x$, $-\infty < \mu_x < \infty$, and $-\infty < \mu_y < \infty$. Then the random variable $(X, Y)$ is defined to have a \textit{bivariate normal distribution}.

The density in Eq. (28) may be represented by a bell-shaped surface $z = f(x, y)$ as in Fig. 7. Any plane parallel to the $xy$ plane which cuts the surface will intersect it in an elliptic curve, while any plane perpendicular to the $xy$ plane will cut the surface in a curve of the normal form. The probability that a point $(X, Y)$ will lie in any region $R$ of the $xy$ plane is obtained by integrating the density over that region:

$$P[(X, Y) \text{ is in } R] = \int_{R} f(x, y) \, dy \, dx.$$  \hfill (29)

The density might, for example, represent the distribution of hits on a vertical target, where $x$ and $y$ represent the horizontal and vertical deviations from the central lines. And in fact the distribution closely approximates the distribution of this as well as many other bivariate populations encountered in practice.

We must first show that the function actually represents a density by showing that its integral over the whole plane is 1; that is,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx = 1.$$  \hfill (30)

The density is, of course, positive. To simplify the integral, we shall substitute

$$u = \frac{x - \mu_x}{\sigma_x} \quad \text{and} \quad v = \frac{y - \mu_y}{\sigma_y},$$  \hfill (31)
so that it becomes
\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\left\{ \frac{1}{2(1 - \rho^2)} u^2 - 2\rho uv + v^2 \right\}} \, dv \, du. \]

On completing the square on \( u \) in the exponent, we have
\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\left\{ \frac{1}{2(1 - \rho^2)} (u - \rho v)^2 + (1 - \rho^2)v^2 \right\}} \, dv \, du, \]

and if we substitute
\[ w = \frac{u - \rho v}{\sqrt{1 - \rho^2}} \quad \text{and} \quad dw = \frac{du}{\sqrt{1 - \rho^2}}, \]

the integral may be written as the product of two simple integrals
\[ \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-w^2/2} \, dw \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-v^2/2} \, dv, \quad (32) \]

both of which are 1, as we have seen in studying the univariate normal distribution. Equation (30) is thus verified.

**Remark** The cumulative bivariate normal distribution
\[ F(x, y) = \int_{-\infty}^{x} \left( \int_{-\infty}^{y} f(x', y') \, dx' \right) \, dy' \]

may be reduced to a form involving only the parameter \( \rho \) by making the substitution in Eq. (31).

### 5.2 Moment Generating Function and Moments

To obtain the moments of \( X \) and \( Y \), we shall find their joint moment generating function, which is given by
\[ m_{X,Y}(t_1, t_2) = m(t_1, t_2) = \sigma^2 \left\{ e^{t_1 X + t_2 Y} \right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x, y) \, dy \, dx. \]

**Theorem 12** The moment generating function of the bivariate normal distribution is
\[ m(t_1, t_2) = \exp \left( t_1 \mu_X + t_2 \mu_Y + \frac{1}{2} \left( t_1^2 \sigma_X^2 + 2 \rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2 \right) \right). \quad (33) \]
PROOF Let us again substitute for \( x \) and \( y \) in terms of \( u \) and \( v \) to obtain
\[
m(t_1, t_2) = e^{i\mu_x + i\mu_y} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(\sigma_x u + \sigma_y v)} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\left\{ \frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) \right\}} \, dv \, du.
\]
(34)

The combined exponents in the integrand may be written
\[
-\frac{1}{2(1-\rho^2)} [u^2 - 2\rho uv + v^2 - 2(1 - \rho^2)t_1 \sigma_x u - 2(1 - \rho^2)t_2 \sigma_y v],
\]
and on completing the square first on \( u \) and then on \( v \), we find this expression becomes
\[
-\frac{1}{2(1-\rho^2)} \left[ (u - \rho v - (1-\rho^2)t_1 \sigma_x)^2 + (v - \rho t_1 \sigma_x - t_2 \sigma_y)^2 \right.
\]
\[
\left. - (1-\rho^2)(t_1^2 \sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2 \sigma_y^2) \right],
\]
which, if we substitute
\[
w = \frac{u - \rho v - (1-\rho^2)t_1 \sigma_x}{\sqrt{1-\rho^2}} \quad \text{and} \quad z = v - \rho t_1 \sigma_x - t_2 \sigma_y,
\]
becomes
\[
-\frac{1}{2} w^2 - \frac{1}{2} z^2 + \frac{1}{4}(t_1^2 \sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2 \sigma_y^2),
\]
and the integral in Eq. (34) may be written
\[
m(t_1, t_2) = e^{i\mu_x + i\mu_y} \exp\left[ \frac{1}{4}(t_1^2 \sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2 \sigma_y^2) \right]
\]
\[
\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-w^2/2 - z^2/2} \, dw \, dz
\]
\[
= \exp[t_1 \mu_x + t_2 \mu_y + \frac{1}{4}(t_1^2 \sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2 \sigma_y^2)]
\]
since the double integral is equal to unity.

Theorem 13 If \((X, Y)\) has bivariate normal distribution, then
\[
\mathcal{E}[X] = \mu_x,
\]
\[
\mathcal{E}[Y] = \mu_y,
\]
\[
\text{var}[X] = \sigma_X^2,
\]
\[
\text{var}[Y] = \sigma_Y^2,
\]
\[
\text{cov}[X, Y] = \rho \sigma_X \sigma_Y,
\]
and

$$\rho_{X,Y} = \rho.$$  

**Proof** The moments may be obtained by evaluating the appropriate derivative of \( m(t_1, t_2) \) at \( t_1 = 0, t_2 = 0 \). Thus,

$$\sigma[X] = \left. \frac{\partial m}{\partial t_1} \right|_{t_1, t_2 = 0} = \mu_X$$

$$\sigma[X^2] = \left. \frac{\partial^2 m}{\partial t_1^2} \right|_{t_1, t_2 = 0} = \mu_X^2 + \sigma_X^2.$$  

Hence the variance of \( X \) is

$$\sigma[(X - \mu_X)^2] = \sigma[X^2] - \mu_X^2 = \sigma_X^2.$$  

Similarly, on differentiating with respect to \( t_2 \), one finds the mean and variance of \( Y \) to be \( \mu_Y \) and \( \sigma_Y^2 \). We can also obtain joint moments

$$\sigma[XY]$$

by differentiating \( m(t_1, t_2) \) \( r \) times with respect to \( t_1 \) and \( s \) times with respect to \( t_2 \) and then putting \( t_1 \) and \( t_2 \) equal to 0. The covariance of \( X \) and \( Y \) is

$$\sigma[(X - \mu_X)(Y - \mu_Y)] = \sigma[XY - X\mu_Y - Y\mu_X + \mu_X \mu_Y] = \sigma[XY] - \mu_X \mu_Y$$

$$= \left. \frac{\partial^2}{\partial t_1 \partial t_2} m(t_1, t_2) \right|_{t_1, t_2 = 0} - \mu_X \mu_Y$$

$$= \rho \sigma_X \sigma_Y.$$  

Hence, the parameter \( \rho \) is the correlation coefficient of \( X \) and \( Y \). 

**Theorem 14** If \( (X, Y) \) has a bivariate normal distribution, then \( X \) and \( Y \) are independent if and only if \( X \) and \( Y \) are uncorrelated. 

**Proof** \( X \) and \( Y \) are uncorrelated if and only if \( \text{cov} [X, Y] = 0 \) or, equivalently, if and only if \( \rho_{X,Y} = \rho = 0 \). It can be observed that if \( \rho = 0 \), the joint density \( f(x, y) \) becomes the product of two univariate normal distributions; so that \( \rho = 0 \) implies \( X \) and \( Y \) are independent. We know that, in general, independence of \( X \) and \( Y \) implies that \( X \) and \( Y \) are uncorrelated.
5.3 Marginal and Conditional Densities

Theorem 15 If \((X, Y)\) has a bivariate normal distribution, then the marginal distributions of \(X\) and \(Y\) are univariate normal distributions; that is, \(X\) is normally distributed with mean \(\mu_X\) and variance \(\sigma_X^2\), and \(Y\) is normally distributed with mean \(\mu_Y\) and variance \(\sigma_Y^2\).

Proof The marginal density of one of the variables \(X\), for example, is by definition

\[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy; \]

and again substituting

\[ v = \frac{y - \mu_Y}{\sigma_Y} \]

and completing the square on \(v\), one finds that

\[ f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sqrt{1 - \rho^2}} \exp\left[ -\frac{1}{2} \left( \frac{x - \mu_X}{\sigma_X} \right)^2 - \frac{1}{2(1 - \rho^2)} \left( v - \rho \frac{x - \mu_X}{\sigma_X} \right)^2 \right] \, dv. \]

Then the substitutions

\[ w = \frac{v - \rho(x - \mu_X)/\sigma_X}{\sqrt{1 - \rho^2}} \quad \text{and} \quad dw = \frac{dv}{\sqrt{1 - \rho^2}} \]

show at once that

\[ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X^2} \exp\left[ -\frac{1}{2} \left( \frac{x - \mu_X}{\sigma_X} \right)^2 \right], \]

the univariate normal density. Similarly the marginal density of \(Y\) may be found to be

\[ f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y^2} \exp\left[ -\frac{1}{2} \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right]. \]

Theorem 16 If \((X, Y)\) has a bivariate normal distribution, then the conditional distribution of \(X\) given \(Y = y\) is normal with mean \(\mu_X + (\rho\sigma_X/\sigma_Y)(y - \mu_Y)\) and variance \(\sigma_X^2(1 - \rho^2)\). Also, the conditional distribution of \(Y\) given \(X = x\) is normal with mean \(\mu_Y + (\rho\sigma_Y/\sigma_X)(x - \mu_X)\) and variance \(\sigma_Y^2(1 - \rho^2)\).
PROOF The conditional distributions are obtained from the joint and marginal distributions. Thus, the conditional density of $X$ for fixed values of $Y$ is

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

and, after substituting, the expression may be put in the form

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi\sigma_X\sqrt{1 - \rho^2}}} \exp\left\{-\frac{1}{2\sigma_X^2(1 - \rho^2)} \left[ x - \mu_x - \frac{\rho \sigma_x}{\sigma_Y} (y - \mu_Y) \right]^2 \right\}. \tag{35}$$

which is a univariate normal density with mean $\mu_x + (\rho \sigma_x/\sigma_Y)(y - \mu_Y)$ and with variance $\sigma_X^2(1 - \rho^2)$. The conditional distribution of $Y$ may be obtained by interchanging $x$ and $y$ throughout Eq. (35) to get

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y\sqrt{1 - \rho^2}}} \exp\left\{-\frac{1}{2\sigma_Y^2(1 - \rho^2)} \left[ y - \mu_y - \frac{\rho \sigma_Y}{\sigma_X} (x - \mu_X) \right]^2 \right\}. \tag{36}$$

As we already noted, the mean value of a random variable in a conditional distribution is called a regression curve when regarded as a function of the fixed variable in the conditional distribution. Thus the regression for $X$ on $Y = y$ in Eq. (35) is $\mu_x + (\rho \sigma_x/\sigma_Y)(y - \mu_Y)$, which is a linear function of $y$ in the present case. For bivariate distributions in general, the mean of $X$ in the conditional density of $X$ given $Y = y$ will be some function of $y$, say $g(y)$, and the equation

$$x = g(y)$$

when plotted in the $xy$ plane gives the regression curve for $X$. It is simply a curve which gives the location of the mean of $X$ for various values of $Y$ in the conditional density of $X$ given $Y = y$.

For the bivariate normal distribution, the regression curve is the straight line obtained by plotting

$$x = \mu_x + \frac{\rho \sigma_x}{\sigma_y} (y - \mu_Y),$$

as shown in Fig. 8. The conditional density of $X$ given $Y = y$, $f_{X|Y}(x|y)$, is also plotted in Fig. 8 for two particular values $y_0$ and $y_1$ of $Y$. 
PROBLEMS

1 Prove or disprove:
   (a) If \( P[X > Y] = 1 \), then \( \delta[X] > \delta[Y] \).
   (b) If \( \delta[X] > \delta[Y] \), then \( P[X > Y] = 1 \).
   (c) If \( \delta[X] > \delta[Y] \), then \( P[X > Y] > 0 \).

2 Prove or disprove:
   (a) If \( F_X(z) > F_Y(z) \) for all \( z \), then \( \delta[Y] > \delta[X] \).
   (b) If \( \delta[Y] > \delta[X] \), then \( F_X(z) > F_Y(z) \) for all \( z \).
   (c) If \( \delta[Y] > \delta[X] \), then \( F_X(z) > F_Y(z) \) for some \( z \).
   (d) If \( F_X(z) = F_Y(z) \) for all \( z \), then \( P[X = Y] = 1 \).
   (e) If \( F_X(z) > F_Y(z) \) for all \( z \), then \( P[X < Y] > 0 \).
   (f) If \( Y = X + 1 \), then \( F_X(z) = F_Y(z + 1) \) for all \( z \).

3 If \( X_1 \) and \( X_2 \) are independent random variables with distribution given by
   \( P[X_i = -1] = P[X_i = 1] = \frac{1}{2} \) for \( i = 1, 2 \), then are \( X_1 \) and \( X_1X_2 \) independent?

4 A penny and a dime are tossed. Let \( X \) denote the number of heads up. Then
   the penny is tossed again. Let \( Y \) denote the number of heads up on the dime
   (from the first toss) and the penny from the second toss.
   (a) Find the conditional distribution of \( Y \) given \( X = 1 \).
   (b) Find the covariance of \( X \) and \( Y \).

5 If \( X \) and \( Y \) have joint distribution given by
   \[ f_{x, y}(x, y) = 2I_{(0, 1)}(x)I_{(0, 1)}(y). \]
   (a) Find \( \text{cov} [X, Y] \).
   (b) Find the conditional distribution of \( Y \) given \( X = x \).

6 Consider a sample of size 2 drawn without replacement from an urn containing
   three balls, numbered 1, 2, and 3. Let \( X \) be the number on the first ball drawn
   and \( Y \) the larger of the two numbers drawn.
   (a) Find the joint discrete density function of \( X \) and \( Y \).
   (b) Find \( P[X = 1 | Y = 3] \).
   (c) Find \( \text{cov} [X, Y] \).
7 Consider two random variables $X$ and $Y$ having a joint probability density function
\[ f_{X,Y}(x, y) = \frac{1}{2}xyI_{0,0}(y)I_{0,1}(x). \]
(a) Find the marginal distributions of $X$ and $Y$.
(b) Are $X$ and $Y$ independent?
8 If $X$ has a Bernoulli distribution with parameter $p$ (that is, $P[X = 1] = p = 1 - P[X = 0]$), $P[Y | X = 0] = 1$, and $P[Y | X = 1] = 2$, what is $P[Y]$?
9 Consider a sample of size 2 drawn without replacement from an urn containing three balls, numbered 1, 2, and 3. Let $X$ be the smaller of the two numbers drawn and $Y$ the larger.
(a) Find the joint discrete density function of $X$ and $Y$.
(b) Find the conditional distribution of $Y$ given $X = 1$.
(c) Find cov $[X, Y]$.
10 Let $X$ and $Y$ be independent random variables, each having the same geometric distribution. Find $P[X = Y]$.
11 If $F(\cdot)$ is a cumulative distribution function:
(a) Is $F(x, y) = F(x) + F(y)$ a joint cumulative distribution function?
(b) Is $F(x, y) = F(x)F(y)$ a joint cumulative distribution function?
(c) Is $F(x, y) = \max\{F(x), F(y)\}$ a joint cumulative distribution function?
(d) Is $F(x, y) = \min\{F(x), F(y)\}$ a joint cumulative distribution function?
12 Prove
\[ F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y) \leq \sqrt{F_X(x)F_Y(y)} \quad \text{for all } x, y. \]
13 Three fair coins are tossed. Let $X$ denote the number of heads on the first two coins, and let $Y$ denote the number of tails on the last two coins.
(a) Find the joint distribution of $X$ and $Y$.
(b) Find the conditional distribution of $Y$ given that $X = 1$.
(c) Find cov $[X, Y]$.
14 Let random variable $X$ have a density function $f(\cdot)$, cumulative distribution function $F(\cdot)$, mean $\mu$, and variance $\sigma^2$. Define $Y = \alpha + \beta X$, where $\alpha$ and $\beta$ are constants satisfying $-\infty < \alpha < \infty$ and $\beta > 0$.
(a) Select $\alpha$ and $\beta$ so that $Y$ has mean 0 and variance 1.
(b) What is the correlation coefficient between $X$ and $Y$?
(c) Find the cumulative distribution function of $Y$ in terms of $\alpha$, $\beta$, and $F(\cdot)$.
(d) If $X$ is symmetrically distributed about $\mu$, is $Y$ necessarily symmetrically distributed about its mean? (Hint: $Z$ is symmetrically distributed about constant $C$ if $Z - C$ and $-(Z - C)$ have the same distribution.)
15 Suppose that random variable $X$ is uniformly distributed over the interval $(0, 1)$; that is, $f_X(x) = I_{0,1}(x)$. Assume that the conditional distribution of $Y$ given $X = x$ has a binomial distribution with parameters $n$ and $p = x$; i.e.,
\[ P[Y = y | X = x] = \binom{n}{y} x^y(1 - x)^{n-y} \quad \text{for } y = 0, 1, \ldots, n. \]
(a) Find $\delta[Y]$.
(b) Find the distribution of $Y$.

16 Suppose that the joint probability density function of $(X, Y)$ is given by

$$f_{x, y}(x, y) = [1 - x(1 - 2x)(1 - 2y)]\mathbb{I}_{[0, 1]}(x)\mathbb{I}_{[0, 1]}(y),$$

where the parameter $\alpha$ satisfies $-1 \leq \alpha \leq 1$.

(a) Prove or disprove: $X$ and $Y$ are independent if and only if $X$ and $Y$ are uncorrelated.

An isosceles triangle is formed as indicated in the sketch.

(b) If $(X, Y)$ has the joint density given above, pick $\alpha$ to maximize the expected area of the triangle.

(c) What is the probability that the triangle falls within the unit square with corners at $(0, 0), (1, 0), (1, 1)$, and $(0, 1)$?

* (d) Find the expected length of the perimeter of the triangle.

17 Consider tossing two tetrahedra with sides numbered 1 to 4. Let $Y_i$ denote the smaller of the two downturned numbers and $Y_2$ the larger.

(a) Find the joint density function of $Y_1$ and $Y_2$.
(b) Find $P[Y_1 \geq 2, Y_2 \geq 2]$.
(c) Find the mean and variance of $Y_1$ and $Y_2$.
(d) Find the conditional distribution of $Y_2$ given $Y_1$ for each of the possible values of $Y_1$.
(e) Find the correlation coefficient of $Y_1$ and $Y_2$.

18 Let $f_{x, y}(x, y) = e^{-x-y}I_{[0, \infty]}(x)I_{[0, \infty]}(y)$

(a) Find $P[X > 1]$.
(b) Find $P[X + Y < 2]$.
(c) Find $P[X < Y | X < 2]$. 
(d) Find $m$ such that $P[X + Y < m] = \frac{1}{2}$.
(e) Find $P[0 < X < 2 | Y = 2]$.
(f) Find the correlation coefficient of $X$ and $Y$.

* 19 Let $f_{x, y}(x, y) = e^{-\gamma(1 - e^{-\gamma})}I_{[0, \infty]}(x)I_{[0, \infty]}(y) + e^{-\gamma(1 - e^{-\gamma})}I_{[0, \infty]}(x)I_{[0, \infty]}(y)$.

(a) Show that $f_{x, y}(\cdot, \cdot)$ is a probability density function.
(b) Find the marginal distributions of $X$ and $Y$.
(c) Find $\delta[Y | X = x]$ for $0 < x$.
(d) Find $P[X \leq 2, Y \leq 2]$.
(e) Find the correlation coefficient of $X$ and $Y$.
(f) Find another joint probability density function having the same marginals.
*20 Suppose \( X \) and \( Y \) are independent and identically distributed random variables with probability density function \( f(\cdot) \) that is symmetrical about 0.

(a) Prove that \( P(\left| X + Y \right| \leq 2|X|) > \frac{1}{2} \).

(b) Select some symmetrical probability density function \( f(\cdot) \), and evaluate \( P(\left| X + Y \right| \leq 2|X|) \).

*21 Prove or disprove: If \( \delta(Y|X) = X, \delta(X|Y) = Y \), and both \( \delta(X^2) \) and \( \delta(Y^2) \) are finite, then \( P(X = Y) = 1 \). (Possible Hint: \( P(X = Y) = 1 \) if \( \text{var} \left( X - Y \right) = 0 \).)

22 A multivariate Chebyshev inequality: Let \( (X_1, \ldots, X_n) \) be jointly distributed with \( \delta[X_j] = \mu_j \) and \( \text{var}[X_j] = \sigma_j^2 \) for \( j = 1, \ldots, m \). Define \( A_i = \{ |X_i - \mu_i| \leq \sqrt{m}\sigma_i \} \). Show that \( P(\bigcap_{i=1}^{m} A_i) \geq 1 - t^{-2} \), for \( t > 0 \).

23 Let \( f_2(\cdot) \) be a probability density function with corresponding cumulative distribution function \( F_2(\cdot) \). In terms of \( f_2(\cdot) \) and/or \( F_2(\cdot) \):

(a) Find \( P(X > x_0 + \Delta x | X > x_0) \).

(b) Find \( P(x_0 < X < x_0 + \Delta x | X > x_0) \).

(c) Find the limit of the above divided by \( \Delta x \) as \( \Delta x \) goes to 0.

(d) Evaluate the quantities in parts (a) to (c) for \( f_2(x) = \lambda e^{-\lambda x} I_{[0, \infty)}(x) \).

24 Let \( N \) equal the number of times a certain device may be used before it breaks. The probability is \( p \) that it will break on any one try given that it did not break on any of the previous tries.

(a) Express this in terms of conditional probabilities.

(b) Express it in terms of a density function, and find the density function.

25 Player \( A \) tosses a coin with sides numbered 1 and 2. \( B \) spins a spinner evenly graduated from 0 to 3. \( B \)'s spinner is fair, but \( A \)'s coin is not; it comes up 1 with a probability \( p \), not necessarily equal to \( \frac{1}{2} \). The payoff \( X \) of this game is the difference in their numbers (\( A \)'s number minus \( B \)'s).

Find the cumulative distribution function of \( X \).

26 An urn contains four balls; two of the balls are numbered with a 1, and the other two are numbered with a 2. Two balls are drawn from the urn without replacement. Let \( X \) denote the smaller of the numbers on the drawn balls and \( Y \) the larger.

(a) Find the joint density of \( X \) and \( Y \).

(b) Find the marginal distribution of \( Y \).

(c) Find the cov \((X, Y)\).

27 The joint probability density function of \( X \) and \( Y \) is given by

\[ f_{x, y}(x, y) = 3(x + y)I_{[0, 1]}(x + y)I_{[0, 1]}(x)I_{[0, 1]}(y). \]

(Note the symmetry in \( x \) and \( y \).)

(a) Find the marginal density of \( X \).

(b) Find \( P(X + Y < .5) \).

(c) Find \( \delta(Y|X = x) \).

(d) Find cov \((X, Y)\).
28 The discrete density of $X$ is given by $f_x(x) = x/3$ for $x = 1, 2$, and $f_{Y|X}(y|x)$ is binomial with parameters $x$ and $1/3$; that is,

$$f_{Y|X}(y|x) = P[Y = y|X = x] = \binom{x}{y}(1/3)^y$$

for $y = 0, \ldots, x$ and $x = 1, 2$.

(a) Find $d[Y]$ and $\text{var}[Y]$.

(b) Find $d[Y]$.

(c) Find the joint distribution of $X$ and $Y$.

29 Let the joint density function of $X$ and $Y$ be given by $f_{x, y}(x, y) = 8xy$ for $0 < x < y < 1$ and be 0 elsewhere.

(a) Find $d[Y|X = x]$.

(b) Find $d[Y|X = x]$.

(c) Find $\text{var}[Y|X = x]$.

30 Let $Y$ be a random variable having a Poisson distribution with parameter $\lambda$. Assume that the conditional distribution of $X$ given $Y = y$ is binomially distributed with parameters $y$ and $p$. Find the distribution of $X$, if $X = 0$ when $Y = 0$.

31 Assume that $X$ and $Y$ are independent random variables and $X$ ($Y$) has binomial distribution with parameters $3$ and $1/2$ ($2$ and $1/2$). Find $P[X = Y]$.

32 Let $X$ and $Y$ have bivariate normal distribution with parameters $\mu_x = 5$, $\mu_y = 10$, $\sigma_x = 1$, and $\sigma_y = 25$.

(a) If $\rho > 0$, find $r$ when $P[4 < Y < 16|X = 5] = .954$.

(b) If $\rho = 0$, find $P[X + Y \leq 16]$.

33 Two dice are cast 10 times. Let $X$ be the number of times no 1s appear, and let $Y$ be the number of times two 1s appear.

(a) What is the probability that $X$ and $Y$ each be less than 3?

(b) What is the probability that $X + Y$ will be 4?

34 Three coins are tossed $n$ times.

(a) Find the joint density of $X$, the number of times no heads appear; $Y$, the number of times one head appears; and $Z$, the number of times two heads appear.

(b) Find the conditional density of $X$ and $Z$ given $Y$.

35 Six cards are drawn without replacement from an ordinary deck.

(a) Find the joint density of the number of aces $X$ and the number of kings $Y$.

(b) Find the conditional density of $X$ given $Y$.

36 Let the two-dimensional random variable $(X, Y)$ have the joint density

$$f_{X, Y}(x, y) = \frac{1}{4}((x-y)I_{(0, 2)}(x)I_{(0, 2)}(y)).$$

(a) Find $d[Y|X = x]$.

(b) Find $d[Y|X = x]$.

(c) Find $\text{var}[Y|X = x]$.


37 The trinomial distribution (multinomial with $k + 1 = 3$) of two random variables $X$ and $Y$ is given by

$$f_{x, y}(x, y) = \frac{n!}{x!y!(n-x-y)!} p^x q^y (1-p-q)^{n-x-y}$$

for $x, y = 0, 1, \ldots, n$ and $x + y \leq n$, where $0 \leq p, 0 \leq q$, and $p + q \leq 1$. 
(a) Find the marginal distribution of \( Y \).
(b) Find the conditional distribution of \( X \) given \( Y \), and obtain its expected value.
(c) Find \( p[X, Y] \).

Let \( (X, Y) \) have probability density function \( f_{x, y}(x, y) \), and let \( u(X) \) and \( v(Y) \) be functions of \( X \) and \( Y \), respectively. Show that
\[
\delta[u(X)v(Y) \mid X = x] = u(x)\delta[v(Y) \mid X = x].
\]

If \( X \) and \( Y \) are two random variables and \( \delta[Y \mid X = x] = \mu \), where \( \mu \) does not depend on \( x \), show that \( \text{var}[Y] = \delta[\text{var}[Y \mid X]] \).

If \( X \) and \( Y \) are two independent random variables, does \( \delta[Y \mid X = x] \) depend on \( x \)?

If the joint moment generating function of \( (X, Y) \) is given by \( m_{x, y}(t_1, t_2) = \exp\{t_1^2 + t_2\} \), what is the distribution of \( Y \)?

Define the moment generating function of \( Y \mid X = x \). Does \( m_Y(t) = \delta[m_{Y \mid X}(t)] \)?

Toss three coins. Let \( X \) denote the number of heads on the first two and \( Y \) denote the number of heads on the last two.

(a) Find the joint distribution of \( X \) and \( Y \).
(b) Find \( \delta[Y \mid X = 1] \).
(c) Find \( \rho_{X, Y} \).
(d) Give a joint distribution that is not the joint distribution given in part (a) yet has the same marginal distributions as the joint distribution given in part (a).

Suppose that \( X \) and \( Y \) are jointly continuous random variables, \( f_{x, y}(x, y) = I_{(0,1)}(x, y) \), and \( f_x(x) = I_{(0,1)}(x) \).

(a) Find \( \delta[Y] \).
(b) Find \( \text{cov}[X, Y] \).
(c) Find \( P[X + Y < 1] \).
(d) Find \( f_{x, y}(x, y) \).

Let \( (X, Y) \) have a joint discrete density function
\[
f_{x, y}(x, y) = p(1 - p_1)^{x-\alpha}(1 - p_2)^{y-\beta}[1 + \alpha(x - p_1)(y - p_2)]I_{(0,1)}(x)I_{(0,1)}(y),
\]
where \( 0 < p_1, 0 < p_2 < 1 \), and \(-1 \leq \alpha \leq 1 \). Prove or disprove: \( X \) and \( Y \) are independent if and only if they are uncorrelated.

Let \( (X, Y) \) be jointly discrete random variables such that each \( X \) and \( Y \) have at most two mass points. Prove or disprove: \( X \) and \( Y \) are independent if and only if they are uncorrelated.