CHAPTER 4: DISTRIBUTIONS IN THE PRESENCE OF MULTI-DIMENSIONALITY AND FUNCTIONS OF RANDOM VARIABLES

PART 1: Multi-dimensional distributions

1 Introduction

2 Joint distribution functions

2.1 Cumulative distribution functions

2.2 Joint density function for discrete random variables

2.3 Joint density function for continuous random variables

3 Conditional distributions and stochastic independence

3.1 Conditional distribution for discrete random variables

3.2 Conditional distribution for continuous random variables

3.3 Conditional probabilities of an event given a random variable

4 A multi-dimensional world of expectations

4.1 Unconditional expectations

4.2 Covariances and correlations

4.3 Conditional expectations

4.4 Joint moment generating functions and moments

4.5 Independence and expectations

5 Highlight: The bivariate normal distribution

5.1 Density function

5.2 Moment generating function and moments

5.3 Marginal and conditional densities

PART 2: Distributions for functions of random variables

1 Introduction

2 Expectations of functions of random variables

2.1 Expectation two ways

2.2 Sums of random variables

2.3 Product and quotient

3 Cumulative-distribution-function technique

3.1 Description of the technique

3.2 Distribution of minimum and maximum

3.3 Distribution of sum and difference of two random variables

3.4 Distribution of product and quotient

4 Moment-generating-function technique

4.1 Description of the technique

4.2 Distribution of sums of independent random variables

Central limit theorem

5 The transformation *Y=g(X)*

5.1 Distribution of *Y=g(X)*

5.2 Probability integral transform

5.3 The transformation $Y_1 = g_1(X_1, ..., X_n), ..., Y_k = g_k(X_1, ..., X_n)$

PART 1: Multi-dimensional distributions

1 Introduction

- It is sometimes of interest to draw conclusions about multiple random variables at once
- As in the univariate case, we will start with cumulative distribution functions
- These exist for any set of k random variables
- Joint density functions will be introduced afterwards

2 Joint distribution functions

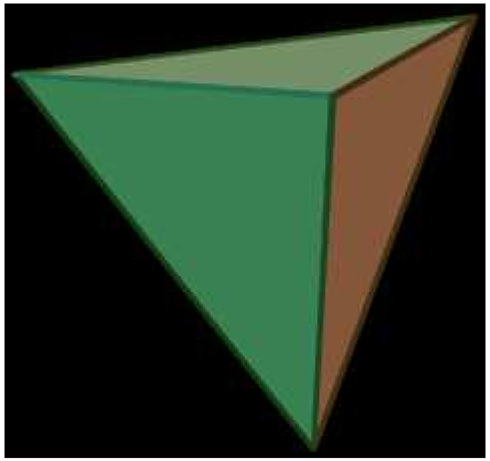
2.1 Cumulative distribution functions

Definition **Joint cumulative distribution function** Let $X_1, X_2, \ldots, X_k$ be k random variables all defined on the same probability space $(\Omega, \mathscr{A}, P[\cdot])$. 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 $(x_1, x_2, \ldots, x_k)$. ////

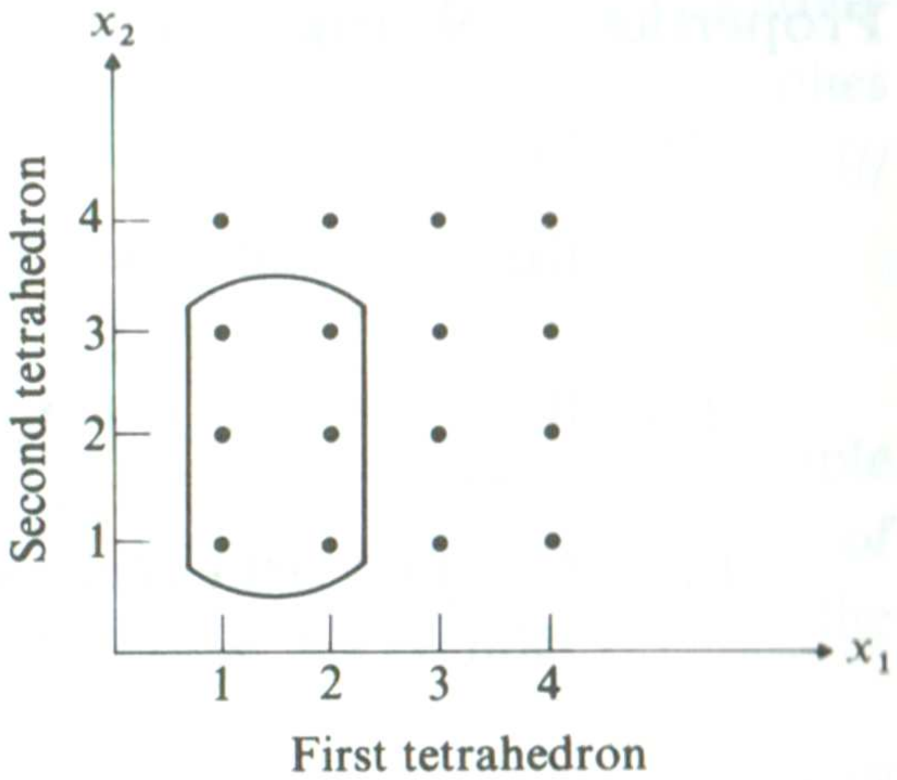
- A joint cumulative distribution function is a function with domain Euclidean *k* space and counterdomain the interval [0,1].
- If k=2, then the joint cumulative distribution function is a function of two variables, and so its domain is simply the *xy* plane.

EXAMPLE 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 sample space for this experiment is displayed . 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. ; 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 ////



x_2
Second tetrahedron
4
3
2
1
1
2
3
4
x_1
First tetrahedron

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

$4 \le y$	0	$\frac{4}{16}$	$\frac{8}{16}$	$\frac{12}{16}$	1
$3 \le y < 4$	0	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{9}{16}$
$2 \le y < 3$	0	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$
$1 \le y < 2$	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
$y < 1$	0	0	0	0	0
	$x < 1$	$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x$

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 $\lim_{\substack{x \to \infty \\ y \to \infty}} F(x, y) = F(\infty, \infty) = 1$.

(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 $F(x_1, y_1) \le F(x_2, y_2)$ for $x_1 \le x_2$ and $y_1 \le y_2$. Consider, for example, the bivariate function $G(x, y)$ defined

Note that $G(x_1, y_1) \le G(x_2, y_2)$ for $x_1 \le x_2$ and $y_1 \le y_2$, 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.

TABLE OF $G(x, y)$

$1 \le y$	0	x	1
$0 \le y < 1$	0	0	y
$y < 0$	0	0	0
	$x < 0$	$0 \le x < 1$	$1 \le x$

Definition 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*. ////

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 function 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 $f_{X_1, X_2, \ldots, X_k}(\cdot, \cdot, \ldots, \cdot)$, 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 $\sum f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) = 1$, where the summation is over all possible values of $(X_1, \ldots, X_k)$. ////

EXAMPLE 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

(x, y)	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(2, 2)	(2, 3)	(2, 4)	(3, 3)	(3, 4)	(4, 4)
$f_{X,Y}(x,y)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

or in another tabular form as

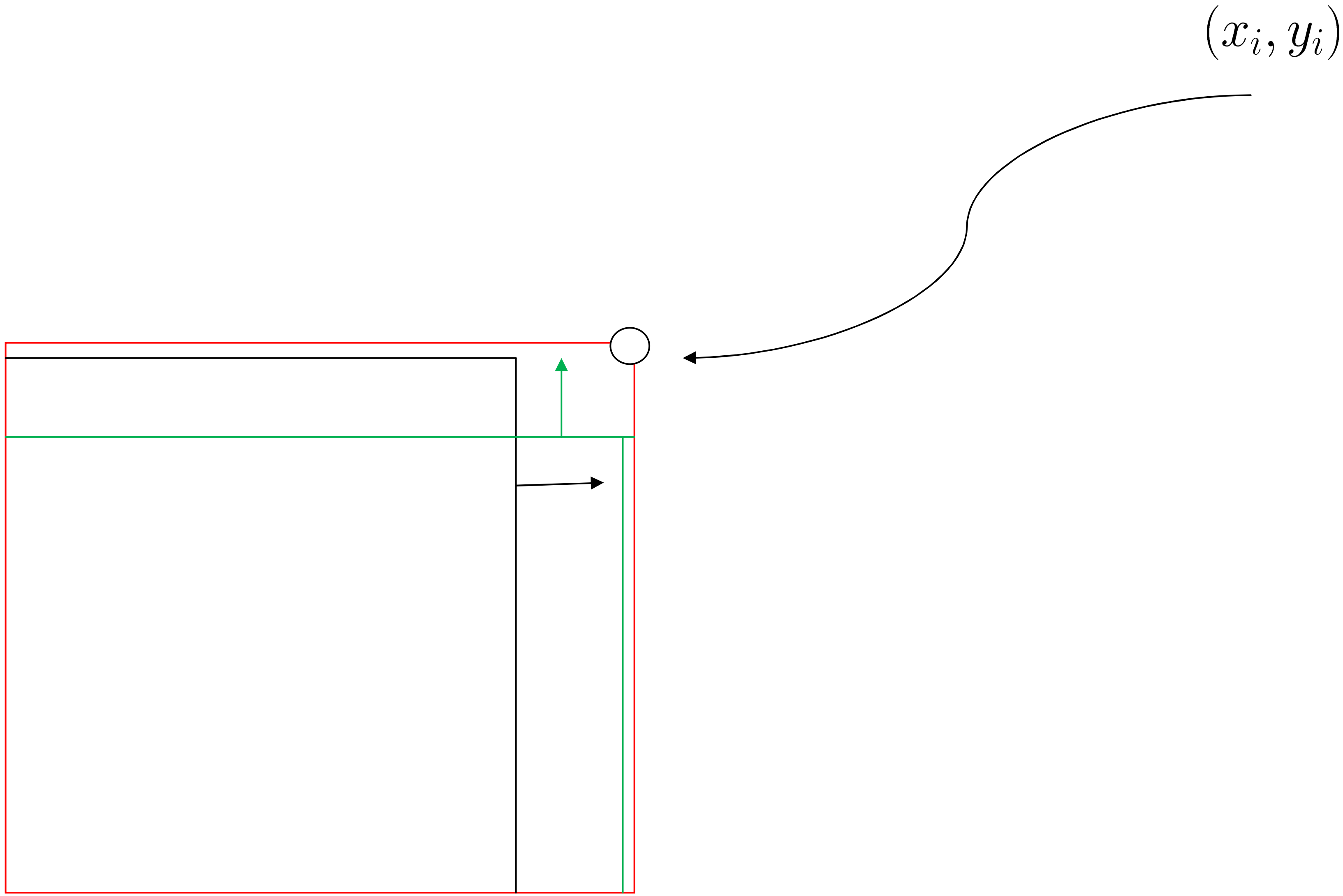
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	
2	$\frac{1}{16}$	$\frac{2}{16}$		
1	$\frac{1}{16}$			
$y \diagup x$	1	2	3	4

////

Theorem If X and Y are jointly discrete random variables, then knowledge of $F_{X,Y}(\cdot,\cdot)$ is equivalent to knowledge of $f_{X,Y}(\cdot,\cdot)$.

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 $x_i \le x$ and $y_i \le y$. Conversely, if $F_{X,Y}(\cdot,\cdot)$ is given, then for (x_i, y_i), a possible value of (X, Y),

$$\begin{aligned} f_{X,Y}(x_i, y_i) = F_{X,Y}(x_i, y_i) &- \lim_{0<h\to 0} F_{X,Y}(x_i - h, y_i) \\ &- \lim_{0<h\to 0} F_{X,Y}(x_i, y_i - h) \\ &+ \lim_{0<h\to 0} F_{X,Y}(x_i - h, y_i - h). \end{aligned}$$ ////


(x_i, y_i)

Definition **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 $f_{X_{i_1}, \ldots, X_{i_m}}(x_{i_1}, \ldots, x_{i_m})$ 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

$$f_X(x_k) = \sum_{\{i: x_i = x_k\}} f_{X,Y}(x_i, y_i) \qquad \text{and} \qquad f_Y(y_k) = \sum_{\{i: y_i = y_k\}} f_{X,Y}(x_i, y_i). \qquad ////$$

EXAMPLE . 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:

4	$\frac{1}{16}+\varepsilon$	$\frac{1}{16}-\varepsilon$	$\frac{1}{16}$	$\frac{4}{16}$
3	$\frac{1}{16}-\varepsilon$	$\frac{1}{16}+\varepsilon$	$\frac{3}{16}$	
2	$\frac{1}{16}$	$\frac{2}{16}$		
1	$\frac{1}{16}$			
y / x	1	2	3	4

For each $0 \leq \varepsilon \leq \frac{1}{16}$, the above table defines a joint density. Note that the marginal densities are independent of ε, 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 Suppose that there are $k+1$ (distinct) possible outcomes of a trial. Denote these outcomes by $\mathscr{o}_1, \mathscr{o}_2, \ldots, \mathscr{o}_{k+1}$, and let $p_i = P[\mathscr{o}_i]$, $i = 1, \ldots, k+1$. 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 $\mathscr{o}_i$ 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

$$f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) = \frac{n!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} p_i^{x_i}, \qquad (1)$$

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 $\mathscr{J}_1$, exactly x_2 outcomes $\mathscr{J}_2$, ..., exactly x_{k+1} outcomes $\mathscr{J}_{k+1}$, where $\sum_1^{k+1} x_i = n$. 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 **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 $p_{k+1} = 1 - p_1 - p_2 - \cdots - p_k$.

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 $\mathscr{o}_i$ or not in outcome $\mathscr{o}_i$, 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 function for continuous random variables

Definition 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

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f_{X_1,\ldots,X_k}(u_1, \ldots, u_k)\, du_1 \ldots du_k \quad (2)$$

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.* ////

- Note that a joint probability density function is defined as any non-negative integrand satisfying the definition statements above. Hence, it is NOT UNIQUE!

Properties

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) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)\, dx_1 \ldots dx_k = 1.$

About areas and volums

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 $x_1 x_2$ plane; then $P[(X_1, X_2) \in R] = \iint_R f_{X_1, X_2}(x_1, x_2)\,dx_1\,dx_2$; 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 $R = \{(x_1, x_2): a_1 < x_1 \le b_1;\, a_2 < x_2 \le b_2\}$, then

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

EXAMPLE Consider the bivariate function

$$f(x, y) = K(x + y)I_{(0,1)}(x)I_{(0,1)}(y) = K(x + y)I_U(x, y),$$

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, $f(x, y) \geq 0$.

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Kf(x, y)\,dx\,dy &= \int_0^1\int_0^1 K(x + y)\,dx\,dy\\ &= K\int_0^1\int_0^1 (x + y)\,dx\,dy\\ &= K\int_0^1 (\tfrac{1}{2} + y)\,dy\\ &= K(\tfrac{1}{2} + \tfrac{1}{2})\\ &= 1\end{aligned}$$

for $K = 1$. So $f(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y)$ is a joint probability density function.

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

$$
\begin{aligned}
P[0 < X < \tfrac{1}{2}; 0 < Y < \tfrac{1}{4}] &= \int_0^{\frac{1}{4}} \int_0^{\frac{1}{2}} (x + y)\, dx\, dy \\
&= \int_0^{\frac{1}{4}} \left(\tfrac{1}{8} + \frac{y}{2} \right) dy \\
&= \tfrac{1}{32} + \tfrac{1}{64} \\
&= \tfrac{3}{64},
\end{aligned}
$$

which is the volume under the surface $z = x + y$ over the region $\{(x, y): 0 < x < \frac{1}{2}; 0 < y < \frac{1}{4}\}$ in the xy plane. ////

Theorem 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.

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

$$F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\, du\, dv.$$

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 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 below shows.) If X and Y are jointly continuous, then

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

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 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

$$f_{X,Y}(x, y; \alpha) = f_X(x)f_Y(y)\{1 + \alpha[2F_X(x) - 1][2F_Y(y) - 1]\}. \tag{4}$$

We will show (i) that for each α 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 α, $2F_X(x) - 1$, 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 $f_{X,Y}(x, y; \alpha)$.

$$\int_{-\infty}^{\infty} f_{X,Y}(x, y; \alpha)\, dy$$

$$= \int_{-\infty}^{\infty} f_X(x) f_Y(y)\{1 + \alpha[2F_X(x) - 1][2F_Y(y) - 1]\}\, dy$$

$$= f_X(x) \int_{-\infty}^{\infty} f_Y(y)\, dy + \alpha f_X(x)[2F_X(x) - 1] \int_{-\infty}^{\infty} [2F_Y(y) - 1] f_Y(y)\, dy$$

$$= f_X(x), \qquad \text{noting that} \quad \int_{-\infty}^{\infty} [2F_Y(y) - 1] f_Y(y)\, dy$$

$$= \int_0^1 (2u - 1)\, du = 0$$

by making the transformation $u = F_Y(y)$. ////

3 Conditional distributions and stochastic independence

3.1 Conditional distribution for discrete random variables

Definition **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 $f_{Y|X}(\cdot \mid x)$, is defined to be

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

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

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

if $f_Y(y) > 0$. ////

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

Remark $F_{Y|X}(y \mid x) = \sum_{\{j : y_j \leq y\}} f_{Y|X}(y_j \mid x)$. ////

EXAMPLE 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 $X = 2$?

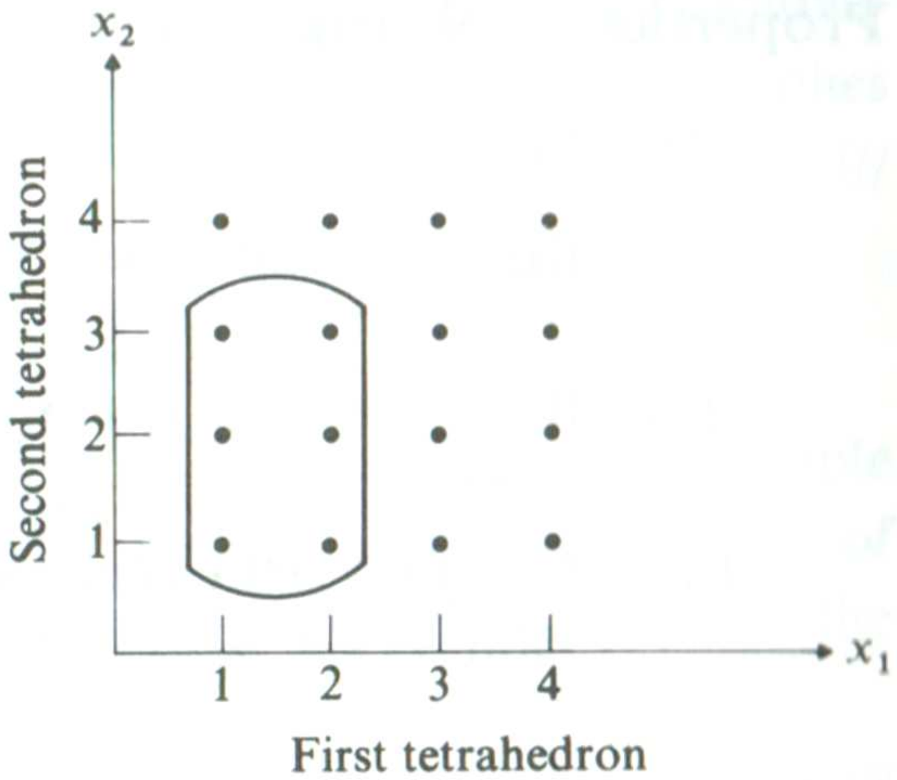
$$f_{Y|X}(2|2) = \frac{f_{X,Y}(2,2)}{f_X(2)} = \frac{\frac{2}{16}}{\frac{4}{16}} = \frac{1}{2}$$

$$f_{Y|X}(3|2) = \frac{f_{X,Y}(2,3)}{f_X(2)} = \frac{\frac{1}{16}}{\frac{4}{16}} = \frac{1}{4}$$

$$f_{Y|X}(4|2) = \frac{f_{X,Y}(2,4)}{f_X(2)} = \frac{\frac{1}{16}}{\frac{4}{16}} = \frac{1}{4}.$$

Also,

$$f_{Y|X}(y|3) = \begin{cases} \frac{3}{4} & \text{for } y = 3 \\ \frac{1}{4} & \text{for } y = 4. \end{cases}$$ ////


x_2
Second tetrahedron
4
3
2
1
1
2
3
4
x_1
First tetrahedron

Definition 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

$$f_{X_{i_1},\ldots,X_{i_r} | X_{j_1},\ldots,X_{j_s}}(x_{i_1}, \ldots, x_{i_r} | x_{j_1}, \ldots, x_{j_s})$$

$$= \frac{f_{X_{i_1},\ldots,X_{i_r},X_{j_1},\ldots,X_{j_s}}(x_{i_1}, \ldots, x_{i_r}, x_{j_1}, \ldots, x_{j_s})}{f_{X_{j_1},\ldots,X_{j_s}}(x_{j_1}, \ldots, x_{j_s})}. \qquad ////$$

3.2 Conditional distribution for continuous random variables

Definition : 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

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

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,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \qquad \text{if } f_Y(y) > 0,$$

and is undefined if $f_Y(y) = 0$ ////

Definition **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 Suppose $f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y)$.

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

for $0 < x < 1$. Note that

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

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

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

3.3 Conditional probabilities of an event given a random variable

We have defined the conditional cumulative distribution $F_{Y|X}(y|x)$ for either jointly continuous or jointly discrete random variables. If X is discrete and Y is any random variable, then $F_{Y|X}(y|x)$ can be defined as $P[Y \leq y | X = x]$ if x is a mass point of X. We would like to define $P[Y \leq y | X = x]$ and more generally $P[A | X = x]$, 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 $P[A | X = x]$.

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. ↑ as our definition of $P[A \mid X = x]$ if the indicated limit exists, and leave $P[A \mid X = x]$ undefined otherwise.

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.

$$(iii) \quad P[A : X \in B] = \sum_{x_i \in B} P[A|X = x_i] f_X(x)$$

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

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

if X is continuous.

3.4 Independence

- When we defined conditional probabilities early on, we were able to introduce the concepts of “independence” and “dependence” of two events
- We have now defined the conditional distribution of random variables. Hence, similarly as in the “probability world” we should now be able to define “independence” and “dependence” or random variables
- Although in this context one can / should talk about “stochastic” independence, often the term “stochastic” is omitted

Definition **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 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 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 $x_1, \ldots, x_k$. ////

EXAMPLE 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 $f_{Y|X}(2|3) = P[Y = 2 | X = 3] = 0 \neq f_Y(2) = P[Y = 2] = \frac{3}{16}$.
////

Theorem 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, $P[Y_1 \in B_1; \ldots; Y_k \in B_k] = P[X_1 \in g_1^{-1}(B_1); \ldots ; X_k \in g_k^{-1}(B_k)] = \prod_{j=1}^{k} P[X_j \in g_j^{-1}(B_j)]$

$$= \prod_{j=1}^{k} P[Y_j \in B_j].$$ ////

Application

EXAMPLE 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

$$P[X-h<Y<X]=\int_{-\infty}^{\infty}P[X-h<Y<X\,|\,X=x]f_X(x)\,dx$$
$$=\int_{99}^{101}P[x-h<Y<x]\tfrac{1}{2}\,dx.$$

Suppose now that $h=1$; then

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

Hence,

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

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

$$+\int_{99.5}^{100.5}\tfrac{1}{2}(\tfrac{1}{2})\,dx+\int_{100.5}^{101}(\tfrac{1}{2})(100.5-x+1)\tfrac{1}{2}\,dx=\tfrac{7}{16}.$$

////

4 A multi-dimensional world of expectations

4.1 Unconditional expectations

Definition **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 $\mathscr{E}[g(X_1, \ldots, X_k)]$, is defined to be

$$\mathscr{E}[g(X_1, \ldots, X_k)] = \sum g(x_1, \ldots, x_k) f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) \tag{18}$$

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

$$\mathscr{E}[g(X_1, \ldots, X_k)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k) f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)\, dx_1 \ldots dx_k \tag{19}$$

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 In particular, if $g(x_1, \ldots, x_k) = x_i$, then

$$\mathscr{E}[g(X_1, \ldots, X_k)] = \mathscr{E}[X_i] = \mu_{X_i}.$$

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

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

$$= \int^{\infty} x_i f_{X_i}(x_i)\, dx_i = \mathscr{E}[X_i]$$

Theorem If $g(x_1, \ldots, x_k) = (x_i - \mathscr{E}[X_i])^2$, then

$$\mathscr{E}[g(X_1, \ldots, X_k)] = \mathscr{E}[(X_i - \mathscr{E}[X_i])^2] = \operatorname{var}[X_i]. \qquad ////$$

Remark $\mathscr{E}\left[\sum_{1}^{m} c_i g_i(X_1, \ldots, X_k)\right] = \sum_{1}^{m} c_i \mathscr{E}[g_i(X_1, \ldots, X_k)]$ for constants $c_1, c_2, \ldots, c_m$. ////

EXAMPLE 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

$$\begin{aligned}
\mathscr{E}[XY] &= \sum xy f_{X,Y}(x, y) \\
&= 1 \cdot 1(\tfrac{1}{16}) + 1 \cdot 2(\tfrac{1}{16}) + 1 \cdot 3(\tfrac{1}{16}) + 1 \cdot 4(\tfrac{1}{16}) \\
&\quad + 2 \cdot 2(\tfrac{2}{16}) + 2 \cdot 3(\tfrac{1}{16}) + 2 \cdot 4(\tfrac{1}{16}) + 3 \cdot 3(\tfrac{3}{16}) \\
&\quad + 3 \cdot 4(\tfrac{1}{16}) + 4 \cdot 4(\tfrac{4}{16}) = \tfrac{135}{16}.
\end{aligned}$$

$$\begin{aligned}
\mathscr{E}[X + Y] &= (1 + 1)\tfrac{1}{16} + (1 + 2)\tfrac{1}{16} + (1 + 3)\tfrac{1}{16} + (1 + 4)\tfrac{1}{16} \\
&\quad + (2 + 2)\tfrac{2}{16} + (2 + 3)\tfrac{1}{16} + (2 + 4)\tfrac{1}{16} + (3 + 3)\tfrac{3}{16} \\
&\quad + (3 + 4)\tfrac{1}{16} + (4 + 4)\tfrac{4}{16} = \tfrac{90}{16}.
\end{aligned}$$

$\mathscr{E}[X] = \frac{5}{2}$, and $\mathscr{E}[Y] = \frac{50}{16}$; hence $\mathscr{E}[X + Y] = \mathscr{E}[X] + \mathscr{E}[Y]$. ////

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

$$\mathscr{E}[g(X_1, X_2, X_3)] = \mathscr{E}[3X_1 + 2X_2 + 6X_3]$$
$$= \int_0^1 \int_0^1 \int_0^1 (3x_1 + 2x_2 + 6x_3)8x_1x_2x_3 \, dx_1 \, dx_2 \, dx_3 = \tfrac{22}{3}.$$

For (ii), we get

$$\mathscr{E}[X_1X_2X_3] = \int_0^1 \int_0^1 \int_0^1 8x_1^2x_2^2x_3^2 \, dx_1 \, dx_2 \, dx_3 = \tfrac{8}{27},$$

and for (iii) we get $\mathscr{E}[X_1X_2] = \frac{4}{9}$. ////

4.2 Covariances and correlations

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

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

provided that the indicated expectation exists. ////

Definition **Correlation coefficient** The *correlation coefficient*, denoted by $\rho[X, Y]$ or $\rho_{X,Y}$, of random variables X and Y is defined to be

$$\rho_{X,Y} = \frac{\text{cov}[X, Y]}{\sigma_X \sigma_Y}$$

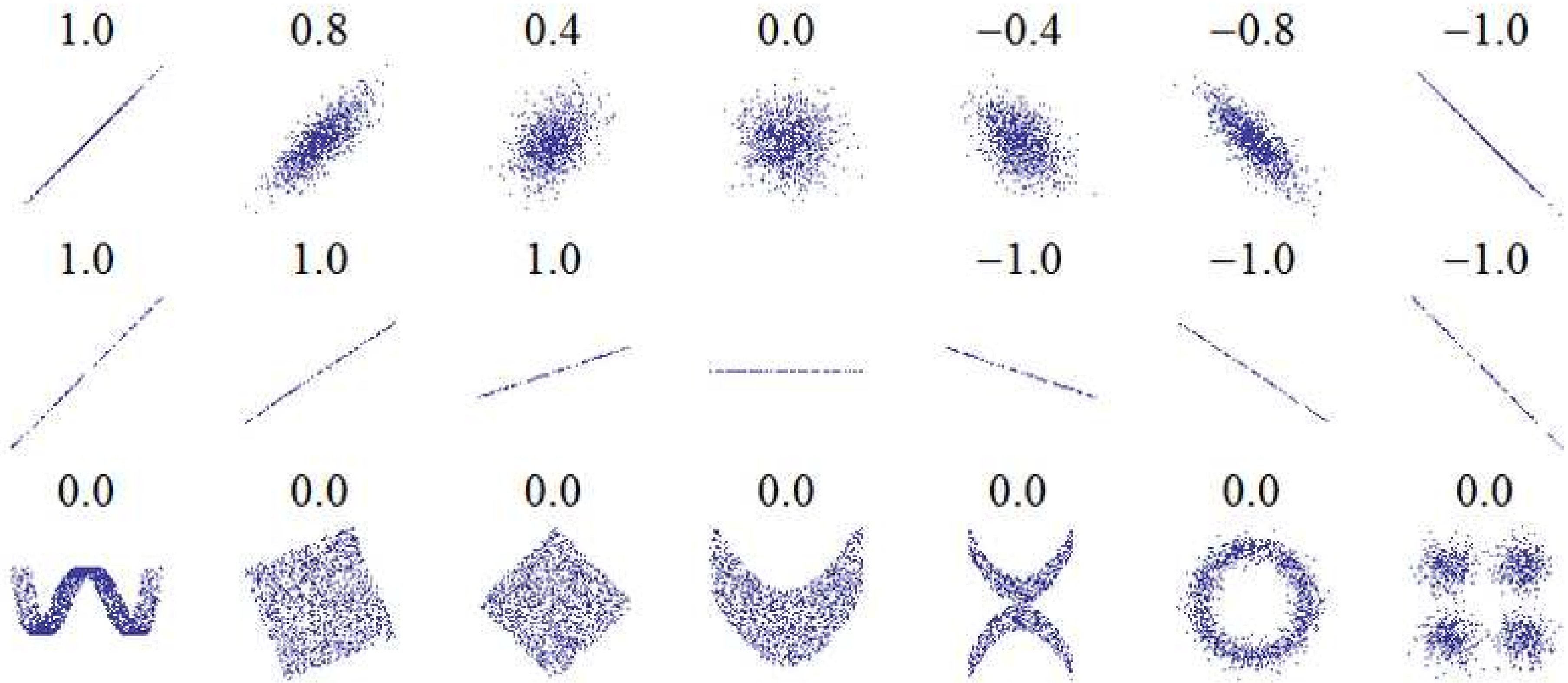
provided that cov $[X, Y]$, σ_X, and σ_Y exist, and $\sigma_X > 0$ and $\sigma_Y > 0$. ////

Remark $\text{cov}[X, Y] = \mathscr{E}[(X - \mu_X)(Y - \mu_Y)] = \mathscr{E}[XY] - \mu_X \mu_Y$.

PROOF
$$\begin{aligned} \mathscr{E}[(X - \mu_X)(Y - \mu_Y)] &= \mathscr{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &= \mathscr{E}[XY] - \mu_X \mathscr{E}[Y] - \mu_Y \mathscr{E}[X] + \mu_X \mu_Y \\ &= \mathscr{E}[XY] - \mu_X \mu_Y . \end{aligned}$$
////

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: cov $[X, Y]$ will be positive when $X - \mu_X$ and $Y - \mu_Y$ tend to have the same sign with high probability, and cov $[X, Y]$ will be negative when $X - \mu_X$ and $Y - \mu_Y$ tend to have opposite signs with high probability. 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, $-1 \leq \rho_{X,Y} \leq 1$.

- Several sets of (x, y) points, with the **correlation coefficient of *x* and *y*** for each set, are shown in the following plot. Note that the correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom).
- Remark: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero


1.0
0.8
0.4
0.0
−0.4
−0.8
−1.0
1.0
1.0
1.0
−1.0
−1.0
−1.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0

EXAMPLE 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 $\mathscr{E}[XY]$, $\mathscr{E}[X]$, and $\mathscr{E}[Y]$ and obtained $\mathscr{E}[XY] = \frac{135}{16}$, $\mathscr{E}[X] = \frac{5}{2}$, and $\mathscr{E}[Y] = \frac{50}{16}$. Thus cov $[X, Y] = \frac{135}{16} - \frac{5}{2} \cdot \frac{50}{16} = \frac{10}{16}$. Now $\mathscr{E}[X^2] = \frac{30}{4}$ and $\mathscr{E}[Y^2] = \frac{170}{16}$; hence var $[X] = \frac{5}{4}$ and var $[Y] = \frac{55}{64}$. So,

$$\rho_{X,Y} = \frac{\frac{10}{16}}{\sqrt{\frac{5}{4}}\sqrt{\frac{55}{64}}} = \frac{2}{\sqrt{11}}. \qquad ////$$

4.3 Conditional expectations

Definition **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 $\mathscr{E}[g(X,Y)|X = x]$, is defined to be

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

if (X, Y) are jointly continuous, and

$$\mathscr{E}[g(X,Y)|X = x] = \sum g(x, y_j) f_{Y|X}(y_j|x)$$

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

Theorem Let (X, Y) be a two-dimensional random variable; then

$$\mathscr{E}[g(Y)] = \mathscr{E}[\mathscr{E}[g(Y) | X]],$$

and in particular

$$\mathscr{E}[Y] = \mathscr{E}[\mathscr{E}[Y | X]].$$

////

Partial proof

$$
\begin{aligned}
\mathscr{E}[\mathscr{E}[g(Y) \mid X]] = \mathscr{E}[h(X)] &= \int_{-\infty}^{\infty} h(x) f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} \mathscr{E}[g(Y) \mid x] f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(y) f_{Y|X}(y \mid x)\, dy \right] f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y) f_{Y|X}(y \mid x) f_X(x)\, dy\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y) f_{X,Y}(x, y)\, dy\, dx \\
&= \mathscr{E}[g(Y)].
\end{aligned}
$$

Definition **Conditional variance** The *variance* of Y given $X = x$ is defined by $\text{var}[Y|X = x] = \mathscr{E}[Y^2|X = x] - (\mathscr{E}[Y|X = x])^2$. ////

Theorem $\text{var}[Y] = \mathscr{E}[\text{var}[Y|X]] + \text{var}[\mathscr{E}[Y|X]]$.

PROOF

$$\begin{aligned}\mathscr{E}[\text{var}[Y|X]] &= \mathscr{E}[\mathscr{E}[Y^2|X]] - \mathscr{E}[(\mathscr{E}[Y|X])^2] \\ &= \mathscr{E}[Y^2] - (\mathscr{E}[Y])^2 - \mathscr{E}[(\mathscr{E}[Y|X])^2] + (\mathscr{E}[Y])^2 \\ &= \text{var}[Y] - \mathscr{E}[(\mathscr{E}[Y|X])^2] + (\mathscr{E}[\mathscr{E}[Y|X]])^2 \\ &= \text{var}[Y] - \text{var}[\mathscr{E}[Y|X]].\end{aligned}$$

////

Additional useful "to-knows"

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

(i) $\mathscr{E}[g_1(Y) + g_2(Y) | X = x] = \mathscr{E}[g_1(Y) | X = x] + \mathscr{E}[g_2(Y) | X = x]$.

(ii) $\mathscr{E}[g_1(Y)g_2(X) | X = x] = g_2(x)\mathscr{E}[g_1(Y) | X = x]$. ////

Definition : **Regression curve** $\mathscr{E}[Y | X = x]$ is called the *regression curve* of Y on x. It is also denoted by $\mu_{Y|X=x} = \mu_{Y|x}$. ////

- We will see more about regression in a subsequent chapter

4.4 Joint moment generating functions and moments

Definition **Joint moments** The *joint raw moments* of $X_1, \ldots, X_k$ are defined by $\mathscr{E}[X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k}]$, where the r_i's are 0 or any positive integer; the *joint moments* about the means are defined by

$$\mathscr{E}[(X_1 - \mu_{X_1})^{r_1} \cdots (X_k - \mu_{X_k})^{r_k}]. \qquad ////$$

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

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

$$m_{X_1, \ldots, X_k}(t_1, \ldots, t_k) = \mathscr{E}\left[\exp \sum_{j=1}^{k} t_j X_j\right],$$

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

Remark $m_X(t_1) = m_{X,Y}(t_1, 0) = \lim_{t_2 \to 0} m_{X,Y}(t_1, t_2)$, and $m_Y(t_2) = m_{X,Y}(0, t_2) = \lim_{t_1 \to 0} m_{X,Y}(t_1, t_2)$; that is, the marginal moment generating functions can be obtained from the joint moment generating function. ////

The rth 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's approach 0. Also $\mathscr{E}[X_i^r X_j^s]$ 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's approach 0. Similarly other joint raw moments can be generated.

4.5 Independence and expectations

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

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

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

$$\begin{aligned}
\mathscr{E}[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_{X,Y}(x, y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} g_1(x)f_X(x)\,dx \cdot \int_{-\infty}^{\infty} g_2(y)f_Y(y)\,dy \\
&= \mathscr{E}[g_1(X)] \cdot \mathscr{E}[g_2(Y)].
\end{aligned}$$

////

Corollary If X and Y are independent, then $\text{cov}[X, Y] = 0$.

PROOF Take $g_1(x) = x - \mu_X$ and $g_2(y) = y - \mu_Y$;

$$\begin{aligned}\text{cov}[X, Y] &= \mathscr{E}[(X - \mu_X)(Y - \mu_Y)] = \mathscr{E}[g_1(X)g_2(Y)] \\ &= \mathscr{E}[g_1(X)]\mathscr{E}[g_2(Y)] \\ &= \mathscr{E}[X - \mu_X] \cdot \mathscr{E}[Y - \mu_Y] = 0 \qquad \text{since } \mathscr{E}[X - \mu_X] = 0.\end{aligned}$$ ////

Definition **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.

EXAMPLE 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. $\mathscr{E}[Y] = \int_0^1 \cos 2\pi u \, du = 0$, and $\mathscr{E}[X] = \int_0^1 \sin 2\pi u \, du = 0$; so $\text{cov}[X, Y] = \mathscr{E}[XY] = \int_0^1 \sin 2\pi u \cos 2\pi u \, du = \frac{1}{2} \int_0^1 \sin 4\pi u \, du = 0$. ////

Theorem 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 $m_X(t_1)$ is the moment generating function of X. Also note that $m_X(t_1) = m_{X,Y}(t_1, 0)$.] X and Y independent imply that the joint moment generating function factors into the product of the marginal moment generating functions by taking $g_1(x) = e^{t_1 x}$ and $g_2(y) = e^{t_2 y}$. The proof in the other direction will be omitted.

////

Cauchy-Schwarz inequality

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

PROOF The existence of expectations $\mathscr{E}[X]$, $\mathscr{E}[Y]$, and $\mathscr{E}[XY]$ follows from the existence of expectations $\mathscr{E}[X^2]$ and $\mathscr{E}[Y^2]$. Define $0 \leq h(t) = \mathscr{E}[(tX - Y)^2] = \mathscr{E}[X^2]t^2 - 2\mathscr{E}[XY]t + \mathscr{E}[Y^2]$. 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 $4(\mathscr{E}[XY])^2 - 4\mathscr{E}[X^2]\mathscr{E}[Y^2] < 0$, or $(\mathscr{E}[XY])^2 < \mathscr{E}[X^2]\mathscr{E}[Y^2]$. If $h(t) = 0$ for some t, say t_0, then $\mathscr{E}[(t_0 X - Y)^2] = 0$, 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 $|\mathscr{E}[UV]| \leq \sqrt{\mathscr{E}[U^2]\mathscr{E}[V^2]}$, and set $U = X - \mu_X$ and $V = Y - \mu_Y$. ////

5 Highlight: The bivariate normal distribution

5.1 Density function

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

$$f_{X,Y}(x, y) = f(x, y) = \frac{1}{2\pi\sigma_X \sigma_Y \sqrt{1-\rho^2}}$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho \frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, where σ_Y, σ_X, μ_X, μ_Y, and ρ 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 *bivariate normal distribution*. ////