Chapter 4

Multiple Random Variables

4.1 Joint and Marginal Distributions

Definition 4.1.1 *An n-dimensional random vector is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean space.*

Suppose, for example, that with each point in a sample space we associate an ordered pair of numbers, that is, a point $(x, y) \in \mathbb{R}^2$, where $\mathbb{R}^2$ denotes the plane. Then we have defined a two -dimensional (or bivariate) random vector (X, Y).

Example 4.1.1 *(Sample space for dice) Consider the experiment of tossing two fair dice. The sample space for this experiment has 36 equally likely points. Let*

$$X = \text{sum of the two dice} \quad \text{and} \quad Y = |\text{difference of two dice}|.$$

In this way we have defined then bivariate random vector (X, Y).

The random vector (X, Y) defined above is called a discrete random vector because it has only a countable (in this case, finite) number of possible values. The probabilities of events defined in terms of X and Y are just defined in terms of the probabilities of the corresponding events in the sample space S. For example,

$$P(X = 5, Y = 3) = P(\{4, 1\}, \{1, 4\}) = \frac{2}{36} = \frac{1}{18}.$$

Definition 4.1.2 *Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ defined by $f(x, y) = P(X = x, Y = y)$ is called the joint probability mass function or joint pmf of (X, Y). If it is necessary to stress the fact that f is the joint pmf of the vector (X, Y) rather than some other vector, the notation $f_{X,Y}(x, y)$ will be used.*

The joint pmf can be used to compute the probability of any event defined in terms of (X, Y). Let A be any subset of $\mathbb{R}^2$. Then

$$P((X, Y) \in A) = \sum_{(x,y)\in A} f(x, y).$$

Expectations of functions of random vectors are computed just as with univariate random variables. Let $g(x, y)$ be a real-valued function defined for all possible values (x, y) of the discrete random vector (X, Y). Then $g(X, Y)$ is itself a random variable and its expected value $Eg(X, Y)$ is given by

$$Eg(X, Y) = \sum_{(x,y)\in\mathbb{R}^2} g(x, y)f(x, y).$$

Example 4.1.2 *(Continuation of Example 4.1.1) For the* (X, Y) *whose joint pmf is given in the following table*

							X					
		2	3	4	5	6	7	8	9	10	11	12
	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$	
Y	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		
	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			
	4					$\frac{1}{18}$		$\frac{1}{18}$				
	5						$\frac{1}{18}$					

Letting $g(x, y) = xy$, *we have*

$$EXY = (2)(0)\frac{1}{36} + \cdots + (7)(5)\frac{1}{18} = 13\frac{11}{18}.$$

The expectation operator continues to have the properties listed in Theorem 2.2.5 (textbook). For example, if $g_1(x, y)$ and $g_2(x, y)$ are two functions and a, b and c are constants, then

$$E(ag_1(X, Y) + bg_2(X, Y) + c) = aEg_1(X, Y) + bEg_2(X, Y) + c.$$

For any (x, y), $f(x, y) \geq 0$ since $f(x, y)$ is a probability. Also, since (X, Y) is certain to be in $\mathbb{R}^2$,

$$\sum_{(x,y)\in\mathbb{R}^2} f(x, y) = P((X, Y) \in \mathbb{R}^2) = 1.$$

Theorem 4.1.1 *Let (X, Y) be a discrete bivariate random vector with joint pmf $f_{XY}(x, y)$. Then the marginal pmfs of X and Y, $f_X(x) = P(X = x)$ and $f_Y(y) = P(Y = y)$, are given by*

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \quad \textit{and} \quad f_Y(y) = \sum_{x \in \mathbb{R}} f_{X,Y}(x, y).$$

PROOF: For any $x \in \mathbb{R}$, let $A_x = \{(x, y) : -\infty < y < \infty\}$. That is, A_x is the line in the plane with first coordinate equal to x. Then, for any $x \in \mathbb{R}$,

$$\begin{aligned}
f_X(x) &= P(X = x) \\
&= P(X = x, -\infty < Y < \infty) \qquad (P(-\infty < Y < \infty) = 1) \\
&= P((X, Y) \in A_x) \qquad (\text{definition of } A_x) \\
&= \sum_{(x,y) \in A_x} f_{X,Y}(x, y) \\
&= \sum_{y \in \mathbb{R}} f_{X,Y}(x, y).
\end{aligned}$$

The proof for $f_Y(y)$ is similar. $\square$

Example 4.1.3 *(Marginal pmf for dice) Using the table given in Example 4.1.2, compute the marginal pmf of* Y*. Using Theorem 4.1.1, we have*

$$f_Y(0) = f_{X,Y}(2,0) + \cdots + f_{X,Y}(12,0) = \frac{1}{6}.$$

Similarly, we obtain

$$f_Y(1) = \frac{5}{18}, \quad f_Y(2) = \frac{2}{9}, \quad f_Y(3) = \frac{1}{6}, \quad f_Y(4) = \frac{1}{9}, \quad f_Y(5) = \frac{1}{18}.$$

Notice that $\sum_{i=0}^{5} f_Y(i) = 1$.

The marginal distributions of X and Y do not completely describe the joint distribution of X and Y. Indeed, there are many different joint distributions that have the same marginal distribution. Thus, it is hopeless to try to determine the joint pmf from the knowledge of only the marginal pmfs. The next example illustrates the point.

Example 4.1.4 *(Same marginals, different joint pmf) Considering the following two joint pmfs,*

$$f(0,0) = \frac{1}{12}, \quad f(1,0) = \frac{5}{12}, \quad , f(0,1) = f(1,1) = \frac{3}{12}, \quad f(x,y) = 0 \textit{ for a}$$

and

$$f(0,0) = f(0,1) = \frac{1}{6}, \quad f(1,0) = f(1,1) = \frac{1}{3}, \quad f(x,y) = 0 \quad \textit{for all other}$$

It is easy to verify that they have the same marginal distributions. The marginal of X is

$$f_X(0) = \frac{1}{3}, \quad f_X(1) = \frac{2}{3}.$$

The marginal of Y is

$$f_Y(0) = \frac{1}{2}, \quad f_Y(1) = \frac{1}{2}.$$

In the following we consider random vectors whose components are continuous random variables.

Definition 4.1.3 *A function $f(x, y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ is called a joint probability density function or joint pdf of the continuous bivariate random vector (X, Y) if, for every $A \subset \mathbb{R}^2$,*

$$P((X, Y) \in A) = \int \int_A f(x, y) dx dy.$$

If $g(x, y)$ is a real-valued function, then the expected value of $g(X, Y)$ is defined to be

$$Eg(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy.$$

The marginal probability density functions of X and Y are defined as

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad -\infty < x < \infty,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx, \quad -\infty < y < \infty.$$

Any function $f(x, y)$ satisfying $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$ and

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

is the joint pdf of some continuous bivariate random vector (X, Y).

Example 4.1.5 *(Calculating joint probabilities-I) Define a joint pdf by*

$$f(x,y)=\begin{cases}6xy^2 & 0<x<1 \text{ and } 0<y<1\\ 0 & \text{otherwise}\end{cases}$$

Now, consider calculating a probability such as $P(X+Y\geq 1)$. *Let* $A=\{(x,y):x+y\geq 1\}$, *we can re-express* A *as*

$$A=\{(x,y):x+y\geq 1, 0<x<1, 0<y<1\}=\{(x,y):1-y\leq x<1, 0<$$

Thus, we have

$$P(X+Y\geq 1)=\int_A\int f(x,y)dxdy=\int_0^1\int_{1-y}^1 6xy^2dxdy=\frac{9}{10}.$$

The joint cdf is the function $F(x,y)$ defined by

$$F(x,y)=P(X\leq x, Y\leq y)=\int_{-\infty}^x\int_{-\infty}^y f(s,t)dtds.$$

4.2 Conditional Distributions and Independence

Definition 4.2.1 *Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the conditional pmf of Y given that $X = x$ is the function of y denoted by $f(y|x)$ and defined by*

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}.$$

For any y such that $P(Y = y) = f_Y(y) > 0$, the conditional pmf of X given that $Y = y$ is the function of x denoted by $f(x|y)$ and defined by

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

It is easy to verify that $f(y|x)$ and $f(x|y)$ are indeed distributions. First, $f(y|x) \geq 0$ for every y since $f(x, y) \geq 0$ and $f_X(x) > 0$. Second,

$$\sum_y f(y|x) = \frac{\sum_y f(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

Example 4.2.1 *(Calculating conditional probabilities) Define the joint pmf of (X, Y) by*

$$f(0,10) = f(0,20) = \frac{2}{18}, \quad f(1,10) = f(1,30) = \frac{3}{18},$$
$$f(1,20) = \frac{4}{18}, \quad f(2,30) = \frac{4}{18}.$$

The conditional probability

$$f_{Y|X}(10|0) = \frac{f(0,10)}{f_X(0)} = \frac{f(0,10)}{f(0,10)+f(0,20)} = \frac{1}{2}.$$

Definition 4.2.2 *Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the conditional pdf of Y given that $X = x$ is the function of y denoted by $f(y|x)$ and defined by*

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

For any y such that $f_Y(y) > 0$, the conditional pdf of X given that $Y = y$ is the function of x denoted by $f(x|y)$ and defined by

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

If $g(Y)$ is a function of Y, then the conditional expected value of $g(Y)$ given that $X = x$ is denoted by $E(g(Y)|x)$ and is given by

$$E(g(Y)|x) = \sum_y g(y)f(y|x) \quad \text{and} \quad E(g(Y)|x) = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

in the discrete and continuous cases, respectively.

Example 4.2.2 *(Calculating conditional pdfs) Let the continuous random vector* (X, Y) *have joint pdf*

$$f(x, y) = e^{-y}, \quad 0 < x < y < \infty.$$

The marginal of X *is*

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_x^{\infty} e^{-y}dy = e6{-}x.$$

Thus, marginally, X *has an exponential distribution. The conditional distribution of* Y *is*

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, & \textit{if } y > x, \\ \frac{0}{e^{-x}} = 0, & \textit{if } y \le x \end{cases}$$

The mean of the conditional distribution is

$$E(Y|X = x) = \int_x^{\infty} ye^{-(y-x)}dy = 1 + x.$$

The variance of the conditional distribution is

$$\begin{aligned} Var(Y|x) &= E(Y^2|x) - (E(Y|x))^2 \\ &= \int_x^{\infty} y^2 e^{-(y-x)}dy - (\int_x^{\infty} ye^{-(y-x)})^2 \\ &= 1 \end{aligned}$$

In all the previous examples, the conditional distribution of Y given $X = x$ was different for different values of x. In some situations, the knowledge that $X = x$ does not give us any more information about Y than we already had. This important relationship between X and Y is called independence.

Definition 4.2.3 *Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called independent random variables if, for* **EVERY** *$x \in \mathbb{R}$ and $y \in mR$,*

$$f(x, y) = f_X(x)f_Y(y).$$

If X and Y are independent, the conditional pdf of Y given $X = x$ is

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

regardless of the value of x.

Lemma 4.2.1 *Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,*

$$f(x, y) = g(x)h(y).$$

PROOF: The "only if" part is proved by defining $g(x) = f_X(x)$ and $h(y) = f_Y(y)$. To proved the "if" part for continuous random variables, suppose that $f(x, y) = g(x)h(y)$. Define

$$\int_{-\infty}^{\infty} g(x)dx = c \quad \text{and} \quad \int_{-\infty}^{\infty} h(y)dy = d,$$

where the constants c and d satisfy

$$\begin{aligned} cd &= (\int_{-\infty}^{\infty} g(x)dx)(\int_{-\infty}^{\infty} h(y)dy) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1 \end{aligned}$$

Furthermore, the marginal pdfs are given by

$$f_X(x) = \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)d$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)c.$$

Thus, we have

$$f(x,y) = g(x)h(y) = g(x)h(y)cd = f_X(x)f_Y(y),$$

showing that X and Y are independent. Replacing integrals with sums proves the lemma for discrete random vectors. $\square$

Example 4.2.3 *(Checking independence) Consider the joint pdf* $f(x,y) = \frac{1}{384}x^2y^2e^{-y-(x/2)}$, $x > 0$ *and* $y > 0$. *If we define*

$$g(x) = \begin{cases} x^2e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

and

$$h(y) = \begin{cases} y^4e^{-y}/384 & y > 0 \\ 0 & y \leq 0 \end{cases}$$

then $f(x,y) = g(x)h(y)$ *for all* $x \in \mathbb{R}$ *and all* $y \in \mathbb{R}$. *By Lemma 4.2.1, we conclude that* X *and* Y *are independent random variables.*

Theorem 4.2.1 *Let X and Y be independent random variables.*

(a) For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$; that is, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent events.

(b) Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then

$$E(g(X)h(Y)) = (Eg(X))(Eh(Y)).$$

PROOF: For continuous random variables, part (b) is proved by noting that

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= (\int_{-\infty}^{\infty} g(x)f_X(x)dx)(\int_{-\infty}^{\infty} h(y)f_Y(y)dy) \\ &= (Eg(X))(Eh(Y)). \end{aligned}$$

The result for discrete random variables is proved bt replacing integrals by sums.

Part (a) can be proved similarly. Let $g(x)$ be the indicator function of the set A. let $h(y)$ be the indicator function of the set B. Note that $g(x)h(y)$ is the indicator function of the set $C \in \mathbb{R}^2$ defined by

$C = \{(x, y) : x \in A, y \in B\}$. Also note that for an indicator function such as $g(x)$, $Eg(X) = P(X \in A)$. Thus,

$$P(X \in A, Y \in B) = P((X, Y) \in C) = E(g(X)h(Y))$$

$$= (Eg(X))(Eh(Y)) = P(X \in A)P(Y \in B).$$

$\square$

Example 4.2.4 *(Expectations of independent variables) Let X and Y be independent exponential(1) random variables. So*

$$P(X \geq 4, Y \leq 3) = P(X \geq 4)P(Y \leq 3) = e^{-4}(1 - e^{-3})/$$

Letting $g(x) = x^2$ and $h(y) = y$, we have

$$E(X^2Y) = E(X^2)E(Y) = (2)(1) = 2.$$

Theorem 4.2.2 *Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by*

$$M_Z(t) = M_X(t)M_Y(t).$$

PROOF:

$$M_Z(t) = Ee^{t(X+Y)} = (Ee^{tX})(Ee^{tY}) = M_X(t)M_Y(t).$$

□

Theorem 4.2.3 *Let* $X \sim N(\mu, \sigma^2)$ *and* $Y \sim N(\gamma, \tau^2)$ *be independent normal random variables. Then the random variable* $Z = X + Y$ *has a* $N(\mu + \gamma, \sigma^2 + \tau^2)$ *distribution.*

PROOF: Using Theorem 4.2.2, we have

$$M_Z(t) = M_X(t)M_Y(t) = \exp\{(\mu + \gamma)t + (\sigma^2 + \tau^2)t^2/2\}.$$

Hence, $Z \sim N(\mu + \gamma, \sigma^2 + \tau^2)$. $\square$

4.3 Bivariate Transformations

Let (X, Y) be a bivariate random vector with a known probability distribution. Let $U = g_1(X, Y)$ and $V = g_2(X, Y)$, where $g_1(x, y)$ and $g_2(x, y)$ are some specified functions. If B is any subset of $\mathbb{R}^2$, then $(U, V) \in B$ if and only if $(X, Y) \in A$, where $A = \{(x, y) : (g_1(x, y), g_2(x, y)) \in B\}$. Thus $P((U, V) \in B) = P((X, Y) \in A)$, and the probability of (U, V) is completely determined by the probability distribution of (X, Y).

If (X, Y) is a discrete bivariate random vector, then

$$f_{U,V}(u, v) = P(U = u, V = v) = P((X, Y) \in A_{u,v}) = \sum_{(x,y)\in A_{uv}} f_{X,Y}(x, y),$$

where $A_{u,v} = \{(x, y) : g_1(x, y) = u, g_2(x, y) = v\}$.

Example 4.3.1 *(Distribution of the sum of Poisson variables) Let X and Y be independent Poisson random variables with parameters θ and λ, respectively. Thus, the joint pmf of (X, Y) is*

$$f_{X,Y}(x, y) = \frac{\theta^x e^{-\theta}}{x!} \frac{\lambda^y e^{-\lambda}}{y!}, \quad x = 0, 1, 2, \ldots, \quad y = 0, 1, 2, \ldots$$

Now define $U = X + Y$ and $V = Y$, thus,

$$f_{U,V}(u, v) = f_{X,V}(u-v, v) = \frac{\theta^{u-v} e^{-\theta}}{(u-v)!} \frac{\lambda^v e^{-\lambda}}{v!}, \quad v = 0, 1, 2, \ldots, \quad u = v, v-$$

The marginal of U is

$$\begin{aligned} f_U(u) &= \sum_{v=0}^{u} \frac{\theta^{u-v} e^{-\theta}}{(u-v)!} \frac{\lambda^v e^{-\lambda}}{v!} = e^{-(\theta+\lambda)} \sum_{v=0}^{u} \frac{\theta^{u-v}}{(u-v)!} \frac{\lambda^v}{v!} \\ &= \frac{e^{-(\theta+\lambda)}}{u!} \sum_{v=0}^{u} \binom{u}{v} \lambda^v \theta^{u-v} = \frac{e^{-(\theta+\lambda)}}{u!} (\theta + \lambda)^u, \quad u = 0, 1, 2, \ldots \end{aligned}$$

This is the pmf of a Poisson random variable with parameter $\theta + \lambda$.

Theorem 4.3.1 *If $X \sim Poisson(\theta)$ and $Y \sim Poisson(\lambda)$ and X and Y are independent, then $X + Y \sim Poisson(\theta + \lambda)$.*

If (X, Y) is a continuous random vector with joint pdf $f_{X,Y}(x, y)$, then the joint pdf of (U, V) can be expressed in terms of $F_{X,Y}(x, y)$ in a similar way. As before, let $A = \{(x, y) : f_{X,Y}(x, y) > 0\}$ and $B = \{(u, v) : u = g_1(x, y) \text{ and } v = g_2(x, y) \text{ for some } (x, y) \in A\}$. For the simplest version of this result, we assume the transformation $u = g_1(x, y)$ and $v = g_2(x, y)$ defines a one-to-one transformation of A to B. For such a one-to-one, onto transformation, we can solve the equations $u = g_1(x, y)$ and $v = g_2(x, y)$ for x and y in terms of u and v. We will denote this inverse transformation by $x = h_1(u, v)$ and $y = h_2(u, v)$. The role played by a derivative in the univariate case is now played by a quantity called the Jacobian of the transformation. It is defined by

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix},$$

where $\frac{\partial x}{\partial u} = \frac{\partial h_1(u,v)}{\partial u}$, $\frac{\partial x}{\partial v} = \frac{\partial h_1(u,v)}{\partial v}$, $\frac{\partial y}{\partial u} = \frac{\partial h_2(u,v)}{\partial u}$, and $\frac{\partial y}{\partial v} = \frac{\partial h_2(u,v)}{\partial v}$.

We assume that J is not identically 0 on B. Then the joint pdf of (U, V) is 0 outside the set B and on the set B is given by

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J|,$$

where $|J|$ is the absolute value of J.

Example 4.3.2 *(Sum and difference of normal variables) Let X and Y be independent, standard normal variables. Consider the transformation $U = X + Y$ and $V = X - Y$. The joint pdf of X and Y is, of course,*

$$f_{X,Y}(x, y) = (2\pi)^{-1} \exp(-x^2/2) \exp(-y^2/2), \quad -\infty < x < \infty, -\infty < y <$$

so the set $A = \mathbb{R}^2$. Solving the following equations

$$u = x + y \quad \textit{and} \quad v = x - y$$

for x and y, we have

$$x = h_1(x, y) = \frac{u + v}{2}, \quad \textit{and} \quad y = h_2(x, y) = \frac{u - v}{2}.$$

Since the solution is unique, we can see that the transformation is one-to-one, onto transformation from A to $B = \mathbb{R}^2$.

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

So the joint pdf of (U, V) is

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J| = \frac{1}{2\pi} e^{-((u+v)/2)^2/2} e^{-((u-v)/2)^2/2} \frac{1}{2}$$

for $-\infty < u < \infty$ *and* $-\infty < v < \infty$*. After some simplification and rearrangement we obtain*

$$f_{U,V}(u,v) = (\frac{1}{\sqrt{2p}\sqrt{2}}e^{-u^2/4})(\frac{1}{\sqrt{2p}\sqrt{2}}e^{-v^2/4}).$$

The joint pdf has factored into a function of u *and a function of* v*. That implies* U *and* V *are independent.*

Theorem 4.3.2 *Let X and Y be independent random variables. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.*

PROOF: We will prove the theorem assuming U and V are continuous random variables. For any $u \in mR$ and $v \in \mathbb{R}$, define

$$A_u = \{x : g(x) \le u\} \quad \text{and} \quad B_u = \{y : h(y) \le v\}.$$

Then the joint cdf of (U, V) is

$$\begin{aligned} F_{U,V}(u,v) &= P(U \le u, V \le v) \\ &= P(X \in A_u, Y \in B_v) \\ & P(X \in A_u)P(Y \in B_v). \end{aligned}$$

The joint pdf of (U, V) is

$$f_{U,V}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{U,V}(u,v) = (\frac{d}{du}P(X \in A_u))(\frac{d}{dv}P(Y \in B_v)),$$

where the first factor is a function only of u and the second factor is a function only of v. Hence, U and V are independent. $\square$

In many situations, the transformation of interest is not one-to-one. Just as Theorem 2.1.8 (textbook) generalized the univariate method to many-to-one functions, the same can be done here. As before, $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$. Suppose $A_0, A_1, \ldots, A_k$ form a partition of $\mathcal{A}$ with these properties. The set A_0, which may be empty, satisfies $P((X, Y) \in A_0) = 0$. The transformation $U = g_1(X, Y)$ and $V = g_2(X, Y)$ is a one-to-one transformation from A_i onto B for each $i = 1, 2, \ldots, k$. Then for each i, the inverse function from B to A_i can be found. Denote the ith inverse by $x = h_{1i}(u, v)$ and $y = h_{2i}(u, v)$. Let J_i denote the Jacobian computed from the ith inverse. Then assuming that these Jacobians do not vanish identically on B, we have

$$f_{U,V}(u, v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v))|J_i|.$$

Example 4.3.3 *(Distribution of the ratio of normal variables) Let X and Y be independent $N(0,1)$ random variable. Consider the transformation $U = X/Y$ and $V = |Y|$. (U and V can be defined to be any value, say (1,1), if $Y = 0$ since $P(Y = 0) = 0$.) This transformation is not one-to-one, since the points (x, y) and $(-x, -y)$ are both mapped into the same (u, v) point. Let*

$$A_1 = \{(x, y) : y > 0\}, \quad A_2 = \{(x, y) : y < 0\}, \quad A_0 = \{(x, y) : y = 0\}.$$

A_0, A_1 and A_2 form a partition of $\mathcal{A} = \mathbb{R}^2$ and $P(A_0) = 0$. The inverse transformations from B to A_1 and B to A_2 are given by

$$x = h_{11}(u, v) = uv, \quad y = h_{21}(u, v) = v,$$

and

$$x = h_{12}(u, v) = -uv, \quad y = h_{22}(u, v) = -v.$$

The Jacobians from the two inverses are $J_1 = J_2 = v$. Using

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

we have

$$\begin{aligned} f_{U,V}(u, v) &= \frac{1}{2\pi} e^{-(uv)^2/2} e^{-v^2/2} |v| + \frac{1}{2\pi} e^{-(-uv)^2/2} e^{-(-v)^2/2} |v| \\ &= \frac{v}{\pi} e^{-(u^2+1)v^2/2}, \quad -\infty < u < \infty, \quad 0 < v < \infty. \end{aligned}$$

From this the marginal pdf of U *can be computed to be*

$$\begin{aligned} f_U(u) &= \int_0^\infty \frac{v}{\pi} e^{-(u^2+1)v^2/2} dv \\ &= \frac{1}{2\pi} \int_0^\infty e^{-(u^2+1)z/2} dz \qquad (z = v^2) \\ &= \frac{1}{\pi(u^2+1)} \end{aligned}$$

So we see that the ratio of two independent standard normal random variable is a Cauchy random variable.

4.4 Hierarchical Models and Mixture Distributions

Example 4.4.1 *(Binomial-Poisson hierarchy) Perhaps the most classic hierarchical model is the following. An insect lays a large number of eggs, each surviving with probability* p*. On the average, how many eggs will survive?*

The large number of eggs laid is a random variable, often taken to be Poisson(λ*). Furthermore, if we assume that each egg's survival is independent, then we have Bernoulli trials. Therefore,, if we let* X*=number of survivors and* Y*=number of eggs laid, we have*

$$X|Y \, binomial(Y, p), \qquad Y \sim Poisson(\lambda),$$

a hierarchical model.

The advantage of the hierarchy is that complicated process may be modeled by a sequence of relatively simple models placed in a hierarchy.

Example 4.4.2 *(Continuation of Example 4.4.1) The random variable X has the distribution given by*

$$
\begin{aligned}
P(X=x) &= \sum_{y=0}^{\infty} P(X=x, Y=y) = \sum_{y=0}^{\infty} P(X=x|Y=y)P(Y=y) \\
&= \sum_{y=x}^{\infty} [\binom{y}{x} p^x (1-p)^{y-x}][\frac{e^{-y}\lambda^y}{y!}] \quad (\textit{conditional probability is 0} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda p},
\end{aligned}
$$

so $X \sim Poisson(\lambda)$. Thus, any marginal inference on X is with respect to a Poisson(λp) distribution, with Y playing no part at all. Introducing Y in the hierarchy was mainly to aid our understanding of the model. On the average,

$$EX = \lambda p$$

eggs will survive.

Sometimes, calculations can be greatly simplified be using the following theorem.

Theorem 4.4.1 *If X and Y are any two random variables, then*

$$EX = E(E(X|Y)),$$

provided that the expectations exist.

PROOF: Let $f(x,y)$ denote the joint pdf of X and Y. By definition, we have

$$EX = \int \inf xf(x,y)dxdy = \int [\int xf(x|y)dx]f_Y(y)dy$$
$$\int E(X|y)f_Y(y)dy = E(E(X|Y))$$

Replacing integrals by sums to prove the discrete case. □

Using Theorem 4.4.1, we have

$$EX = E(E(X|Y)) = E(pY) = p\lambda$$

for Example 4.4.2.

Definition 4.4.1 *A random variable X is said to have a mixture distribution if the distribution of X depends on a quantity that also has a distribution.*

Thus, in Example 4.4.1 the Poisson(λp) distribution is a mixture distribution since it is the result of combining a binomial(Y, p) with $Y \sim$ Poisson(λ).

Theorem 4.4.2 *(Conditional variance identity) For any two random variables X and Y,*

$$VarX = E(Var(X|Y)) + Var(E(X|Y)),$$

provided that the expectations exist.

PROOF: By definition, we have

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

The last term in this expression is equal to 0, however, which can easily be seen by iterating the expectation:

$$E([X-E(X|Y)][E(X|Y)-EX]) = E(E\{[X-E(X|Y)][E(X|Y)-EX]|Y$$

In the conditional distribution $X|Y$, X is the random variable. Conditional on Y, E(X—Y) and EX are constants. Thus,

$$E\{[X-E(X|Y)][E(X|Y)-EX]|Y\} = (E(X|Y)-E(X|Y))(E(X|Y)-E$$

Since

$$E([X - E(X|Y)]^2) = E(E\{[X - E(X|Y)]^2|Y\}) = E(\bar{(}X|Y)).$$

and

$$E([E(X|Y) - EX]^2) = \mathrm{Var}(E(X|Y)),$$

Theorem 4.4.2 is proved. $\square$

Example 4.4.3 *(Beta-binomial hierarchy) One generalization of the binomial distribution is to allow the success probability to vary according to a distribution. A standard model for this situation is*

$$X|P \sim binomial(P), \quad i = 1, \ldots, n,$$
$$P \sim beta(\alpha, \beta).$$

The mean of X is then

$$EX = E[E(X|p)] = E[nP] = \frac{n\alpha}{\alpha+\beta}.$$

Since $P \sim beta(\alpha, \beta)$,

$$Var(E(X|P)) = Var(np) = n^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Also, since $X|P$ is binomial(n, P), $Var(X|P) = nP(1-P)$. We then have

$$E[Var(X|P)] = nE[P(1-P)] = n\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 p(1-p)p^{\alpha-1}(1-p)^{\beta-1}dp$$
$$= n\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} = \frac{n\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}.$$

Adding together the two pieces, we get

$$VarX = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

4.5 Covariance and Correlation

In earlier sections, we have discussed the absence or presence of a relationship between two random variables, Independence or nonindependence. But if there is a relationship, the relationship may be strong or weak. In this section, we discuss two numerical measures of the strength of a relationship between two random variables, the covariance and correlation.

Throughout this section, we will use the notation $EX = \mu_X$, $EY = \mu_Y$, Var$X = \sigma_X^2$, and Var$Y = \sigma_Y^2$.

Definition 4.5.1 *The covariance of X and Y is the number defined by*

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

Definition 4.5.2 *The correlation of X and Y is the number defined by*

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

The value ρ_{XY} is also called the correlation coefficient.

Theorem 4.5.1 *For any random variables X and Y,*

$$Cov(X, Y) = EXY - \mu_X \mu_Y.$$

Theorem 4.5.2 *If X and Y are independent random variables, then $Cov(X,Y)=0$ and $\rho_{XY}=0$.*

Theorem 4.5.3 *If X and Y are any two random variables and a and b are any two constants, then*

$$Var(aX+bY)=a^2\,VarX+b^2\,VarY+2abCov(X,Y).$$

If X and Y are independent random variables, then

$$Var(AX+bY)=a^2\,VarX+b^2\,VarY.$$

Theorem 4.5.4 *For any random variables X and Y,*

a. $-1\le\rho_{XY}\le 1$.

b. $|\rho_{XY}|=1$ *if and only if there exist numbers* $a\neq 0$ *and* b *such that* $P(Y=aX+b)=1$. *If* $\rho_{XY}=1$, *then* $a>0$, *and if* $\rho_{XY}=-1$, *then* $a<0$.

PROOF: Consider the function $h(t)$ defined by

$$\begin{aligned}h(t)&=E((X-\mu_X)t+(Y-\mu_Y))^2\\&=t^2\sigma_X^2+2t\mathrm{Cov}(X,Y)+\sigma_Y^2.\end{aligned}$$

Since $h(t)\ge 0$ and it is quadratic function,

$$(2\mathrm{Cov}(X,Y))^2-4\sigma_X^2\sigma_Y^2\le 0.$$

This is equivalent to

$$-\sigma_X\sigma_Y \leq \mathrm{Cov}(X,Y) \leq \sigma_X\sigma_Y.$$

That is,

$$-1 \leq \rho_{XY} \leq 1.$$

Also, $|\rho_{XY}| = 1$ if and only if the discriminant is equal to 0, that is, if and only if $h(t)$ has a single root. But since $((X-\mu_X)t+(Y-\mu_Y))^2 \geq 0$, $h(t) = 0$ if and only if

$$P((X-\mu_X)t + (Y-\mu_Y) = 0) = 1.$$

This $P(Y = aX + b) = 1$ with $a = -t$ and $b = \mu_X t + \mu_Y$, where t is the root of $h(t)$. Using the quadratic formula, we see that this root is $t = -\mathrm{Cov}(X,Y)/\sigma_X^2$. Thus $a = -t$ has the same sign as ρ_{XY}, proving the final assertion. $\square$

Example 4.5.1 *(Correlation-I) Let X have a uniform(0,1) distribution and Z have a uniform(0,0.1) distribution. Suppose X and Z are independent. Let $Y = X + Z$ and consider the random vector (X, Y). The joint pdf of (X, Y) is*

$$f(x, y) = 10, \quad 0 < x < 1, \quad x < y < x + 0.1$$

Note $f(x, y)$ can be obtained from the relationship $f(x, y) = f(y|x)f(x)$. Then

$$\begin{aligned} Cov(X, Y) &= EXY = -(EX)(EY) \\ &= EX(X + Z) - (EX)(E(X + Z)) \\ &= \sigma_X^2 = \frac{1}{12} \end{aligned}$$

The variance of Y is $\sigma_Y^2 = VarX + VarZ = \frac{1}{12} + \frac{1}{1200}$. Thus

$$\rho_{XY} = \frac{1/12}{\sqrt{1/12}\sqrt{1/12 + 1/1200}} = \sqrt{\frac{100}{101}}.$$

The next example illustrates that there may be a strong relationship between X and Y, but if the relationship is not linear, the correlation may be small.

Example 4.5.2 *(Correlation-II) Let* $X \sim Unif(-1,1)$, $Z \sim Unif(0, 0.$
and X *and* Z *be independent. Let* $Y = X^2 + Z$ *and consider the random vector* (X, Y). *Since given* $X = x$, $Y \sim Unif(x^2, x^2{+}0.1)$. *The joint pdf of* X *and* Y *is*

$$f(x,y) = 5, \quad -1 < x < 1, \quad x^2 < y < x^2 + \frac{1}{10}.$$

$$\begin{aligned} Cov(X,Y) &= E(X(X^2+Z)) - (EX)(E(X^2+Z)) \\ &= EX^3 + EXZ - 0E(X^2+Z) \\ &= 0 \end{aligned}$$

Thus, $\rho_{XY} = Cov(X,Y)/(\sigma_X\sigma_Y) = 0$.

Definition 4.5.3 *Let* $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, *and* $-1 < \rho < 1$ *be five real numbers. The bivariate normal pdf with means* μ_X *and* μ_Y, *variances* σ_X^2 *and* σ_Y^2, *and correlation* ρ *is the bivariate pdf given by*

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_Y\sqrt{1-\rho^2}} \exp\Big\{-\frac{1}{2(1-\rho^2)}\Big((\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-}{\sigma}$$

for $-\infty < x < \infty$ *and* $-\infty < y < \infty$.

The many nice properties of this distribution include these:

a. The marginal distribution of X is $N(\mu_X, \sigma_X^2)$.

b. The marginal distribution of Y is $N(\mu_Y, \sigma_Y^2)$.

c. The correlation between X and Y is $\rho_{XY} = \rho$.

d. For any constants a and b, the distribution of $aX+bY$ is $N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$.

Assuming (a) and (b) are true, we will prove (c). Let

$$s = (\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) \quad \text{and} \quad t = (\frac{x-\mu_X}{\sigma_X}).$$

Then $x = \sigma_X t + \mu_X$, $y = (\sigma_Y s/t) + \mu_Y$, and the Jacobian of the transformation is $J = \sigma_X\sigma_Y/t$. With this change of variables, we obtain

$$\begin{aligned}\rho_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} sf(\sigma_X t + \mu_X, \frac{\sigma_Y s}{t} + \mu_Y)|\frac{\sigma_X\sigma_Y}{t}|dsdt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2})^{-1}\exp\big(-\frac{1}{2(1-\rho)^2}(t^2 - 2\rho s + (\frac{s}{t})^2)\big) \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})dt\int_{-\infty}^{\infty}\frac{s}{\sqrt{2\pi}\sqrt{(1-\rho^2)t^2}}\exp\big(-\frac{(s-\rho t^2)^2}{2(1-\rho^2)t^2}\big)ds\end{aligned}$$

The inner integral is ES, where S is a normal random variable with $ES = \rho t^2$ and $\text{Var}S = (1-\rho^2)t^2$. Thus,

$$\rho_{XY} = \int_{-\infty}^{\infty}\frac{\rho t^2}{\sqrt{2\pi}}\exp\{-t^2/2\}dt = \rho.$$

4.6 Multivariate Distributions

The random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ has a sample space that is a subset of $\mathbb{R}^n$. If $\boldsymbol{X}$ is discrete random vector, then the joint pmf of $\boldsymbol{x}$ is the function defined by $f(\boldsymbol{x}) = f(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n - x_n)$ for each $(x_1, \dots, x_n) \in \mathbb{R}^n$. Then for any $A \subset \mathbb{R}^n$,

$$P(\boldsymbol{X} \in A) = \sum_{\boldsymbol{x} \in A} f(\boldsymbol{x}).$$

If $\boldsymbol{X}$ is a continuous random vector, the joint pdf of $\boldsymbol{X}$ is a function $f(x_1, \dots, x_n)$ that satisfies

$$P(\boldsymbol{X} \in A) = \int \cdots \int_A f(\boldsymbol{x}) d\boldsymbol{x} = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Let $g(\boldsymbol{x}) = g(x_1, \dots, x_n)$ be a real-valued function defined on the sample space of $\boldsymbol{X}$. Then $g(\boldsymbol{X})$ is a random variable and the expected value of $g(\boldsymbol{X})$ is

$$Eg(\boldsymbol{X}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

and

$$Eg(\boldsymbol{X}) = \sum_{\boldsymbol{x} \in \mathbb{R}^n} g(\boldsymbol{x}) f(\boldsymbol{x})$$

in the continuous and discrete cases, respectively.

The marginal distribution of $(X_1, \dots, X_n)$, the first k coordinates of $(X_1, \ldots, X_n)$, is given by the pdf or pmf

$$f(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_{k+1} \cdots dx_n$$

or

$$f(x_1, \ldots, x_k) = \sum_{(x_{k+1}, \ldots, x_n) \in \mathbb{R}^{n-k}} f(x_1, \ldots, x_n)$$

for every $(x_1, \ldots, x_k) \in \mathbb{R}^k$.

If $f(x_1, \ldots, x_k) > 0$, the conditional pdf or pmf of $(X_{k+1}, \ldots, X_n)$ given $X_1 = x_1, \ldots, X_k = x_k$ is the function of $(x_{k+1}, \ldots, x_n)$ defined by

$$f(x_{k=1}, \ldots, x_n | x_1, \ldots, x_k) = \frac{f(x_1, \ldots, x_n)}{f(x_1, \ldots, x_k)}.$$

Example 4.6.1 *(Multivariate pdfs) Let* $n = 4$ *and*

$$f(x_1, x_2, x_3, x_4) = \begin{cases} \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) & 0 < x_i < 1, i = 1, 2, 3, 4 \\ 0 & \textit{otherwise} \end{cases}$$

The joint pdf can be used to compute probabilities such as

$$P(X_1 < \frac{1}{2}, X_2 < \frac{3}{4}, X_4 > \frac{1}{2})$$
$$= \int_{\frac{1}{2}}^{1} \int_0^1 \int_0^{\frac{3}{4}} \int_0^{\frac{1}{2}} \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) dx_1 dx_2 dx_3 dx_4 = \frac{151}{1024}.$$

The marginal pdf of (X_1, X_2) *is*

$$f(x_1, x_2) = \int_0^1 \int_0^1 \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) dx_2 dx_4 = \frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2}$$

for $0 < x_1 < 1$ *and* $0 < x_2 < 1$.

Definition 4.6.1 *Let* n *and* m *be positive integers and let* $p_1, \dots, p_n$ *be numbers satisfying* $0 \leq p_i \leq 1$, $i = 1, \dots, n$, *and* $\sum_{i=1}^n p_i = 1$. *Then the random vector* $(X_1, \dots, X_n)$ *has a multinomial distribution with* m *trials and cell proabilities* $p_1, \dots, p_n$ *if the joint pmf of* $(X_1, \dots, X_n)$ *is*

$$f(x_1, \dots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

on the set of $(x_1, \dots, x_n)$ *such that each* x_i *is a nonnegative integer and* $\sum_{i=1}^n x_i = m$.

Example 4.6.2 *(Multivariate pmf) Consider tossing a six-sided die 10 times. Suppose the die is unbalanced so that the probability of observing an i is $i/21$. Now consider the vector $(X_1, \ldots, X_6)$, where X_i counts the number of times i comes up in the 10 tosses. Then $(X_1, \ldots, X_6)$ has a multinomial distribution with $m = 10$ and cell probabilities $p_1 = \frac{1}{21}, \ldots, p_6 = \frac{6}{21}$. For example, the probability of the vector $(0, 0, 1, 2, 3, 4)$ is*

$$f(0,0,1,2,3,4) = \frac{10!}{0!0!1!2!3!4!}(\frac{1}{21})^0(\frac{2}{21})^0(\frac{3}{21})^1(\frac{4}{21})^2(\frac{5}{21})^3(\frac{6}{21})^4 = 0.0059.$$

The factor $\frac{m!}{x_1! \cdots x_n!}$ is called a multinomial coefficient. It is the number of ways that m objects can be divided into n groups with x_1 in the first group, x_2 in the second group, ..., and x_n in the nth group.

Theorem 4.6.1 *(Multinomial Theorem) Let m and n be positive integers. Let A be the set of vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^n x_i = m$. Then, for any real numbers $p_1, \ldots, p_n$,*

$$(p_1 + \ldots + p_n)^m = \sum_{\boldsymbol{x} \in A} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \ldots p_n^{x_n}.$$

Definition 4.6.2 *Let $X_1, \ldots, X_n$ be random vectors with joint pdf or pmf $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $f_{\boldsymbol{X}_i}(x_i)$ denote the marginal pdf or pmf of $\boldsymbol{X}_i$. Then $X_1, \ldots, X_n$ are called mutually independent random vectors if, for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$,*

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \ldots f_{\boldsymbol{X}_n}(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\boldsymbol{X}_i}(\boldsymbol{x}_i).$$

If the X_i's are all one dimensional, then $X_1, \ldots, X_n$ are called mutually independent random variables.

Mutually independent random variables have many nice properties. The proofs of the following theorems are analogous to the proofs of their counterparts in Sections 4.2 and 4.3.

Theorem 4.6.2 *(Generalization of Theorem 4.2.1) Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be mutually independent random variables. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(x_i)$ is a function only of x_i, $i = 1, \ldots, n$. Then*

$$E(g_1(X_1) \cdots g(X_n)) = (Eg_1(X_1)) \cdots (Eg_n(X_n)).$$

Theorem 4.6.3 *(Generalization of Theorem 4.2.2) Let* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *be mutually independent random variables with mgfs* $M_{X_1}(t), \ldots, M_{X_n}(t)$. *Let* $Z = X_1 + \cdots + X_n$. *Then the mgf of* Z *is*

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t).$$

In particular, if $X_1, \ldots, X_n$ *all have the same distribution with mgf* $M_X(t)$, *then*

$$M_Z(t) = (M_X(t))^n.$$

Example 4.6.3 *(Mgf of a sum of gamma variables) Suppose* $X_1, \ldots, X_n$ *are mutually independent random variables, and the distribution of* X_i *is gamma*(α_i, β). *Thus, if* $Z = X_1 + \ldots + X_n$, *the mgf of* Z *is*

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t) = (1-\beta t)^{-\alpha_1} \cdots (1-\beta t)^{-\alpha_n} = (1-\beta t)^{-(\alpha_1+\cdots+\alpha_n)}$$

This is the mgf of a gamma$(\alpha_1 + \cdots + \alpha_n, \beta)$ *distribution. Thus, the sum of a independent gamma random variables that have a common scale parameter* β *also has a gamma distribution.*

Example 4.6.4 *Let $X_1, \ldots, X_n$ be mutually independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Then*

$$Z = \sum_{i=1}^{n}(a_i X_i + b_i) \sim N(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n} a_i^2\sigma_i^2).$$

Theorem 4.6.4 *(Generalization of Lemma 4.2.1) Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are mutually independent random vectors if and only if there exist functions $g_i(\boldsymbol{x}_i)$, $i = 1, \ldots, n$, such that the joint pdf or pmf of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ can be written as*

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = g_1(\boldsymbol{x}_1) \cdots g_n(\boldsymbol{x}_n).$$

Theorem 4.6.5 *(Generalization of Theorem 4.3.2) Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors. Let $g_i(\boldsymbol{x}_i)$ be a function only of $\boldsymbol{x}_i$, $i = 1, \ldots, n$. Then the random vectors $U_i = g_i(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, are mutually independent.*

Let $(X_1, \ldots, X_n)$ be a random vector with pdf $f_X(x_1, \ldots, x_n)$. Let $\mathcal{A} = \{\boldsymbol{x} : f_X(\boldsymbol{x}) > 0\}$. Consider a new random vector $(U_1, \ldots, U_n)$, defined by $U_1 = g_1(X_1, \ldots, X_n)$, ..., $U_n = g_n(X_1, \ldots, X_n)$. Suppose that $A_0, A_1, \ldots, A_k$ form a partition of $\mathcal{A}$ with these properties. The set A_0, which may be empty, satisfies $P((X_1, \ldots, X_n) \in A_0) = 0$. The transformation $(U_1, \ldots, U_n) = (g_1(\boldsymbol{X}), \ldots, g_n(\boldsymbol{X}))$ is a one-to-one transformation from A_i onto B for each $i = 1, 2, \ldots, k$. Then for each i, the inverse functions from B to A_i can be found. Denote the ith inverse by $x_1 = h_{1i}(u-1, \ldots, u_n), \ldots, x_n = h_{ni}(u_1, \ldots, u_n)$. Let J_i denote the Jacobian computed from the ith inverse. That is,

$$J_i = \begin{vmatrix} \frac{\partial h_{1i}(\boldsymbol{u})}{\partial u_1} & \frac{\partial h_{1i}(\boldsymbol{u})}{\partial u_2} & \cdots & \frac{\partial h_{1i}(\boldsymbol{u})}{\partial u_1} \\ \frac{\partial h_{2i}(\boldsymbol{u})}{\partial u_1} & \frac{\partial h_{2i}(\boldsymbol{u})}{\partial u_2} & \cdots & \frac{\partial h_{2i}(\boldsymbol{u})}{\partial u_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{ni}(\boldsymbol{u})}{\partial u_1} & \frac{\partial h_{ni}(\boldsymbol{u})}{\partial u_2} & \cdots & \frac{\partial h_{ni}(\boldsymbol{u})}{\partial u_1} \end{vmatrix}$$

the determinant of an $n \times n$ matrix. Assuming that these Jacobians do not vanish identically on B, we have the following representation of the joint pdf, $f_U(u_1, \ldots, u_n)$, for $\boldsymbol{u} \in B$:

$$f_{\boldsymbol{u}}(u_1, \ldots, u_n) = \sum_{i=1}^{k} f_X(h_{1i}(u_1, \ldots, u_n), \ldots, h_{ni}(u_1, \ldots, u_n))|J_i|.$$