SOME THEORY AND PRACTICE OF STATISTICS
by Howard G. Tucker

CHAPTER 5. DEPENDENT AND INDEPENDENT RANDOM VARIABLES AND LIMIT THEOREMS

5.1 Multivariate Distribution Functions. We considered earlier the concept of independent events. We now extend this to the concept of independent random variables. The extension will appear natural. A large part of statistical inference deals with independent random variables, and so this must be gone into first. We begin the study of independent random variables by studying multivariate distribution functions, and we introduce these by simply looking into the special case of bivariate distribution functions, namely, the joint distribution function of two random variables.

Definition. If X and Y are random variables defined over the same fundamental probability set, then their joint (or bivariate) distribution function $F_{X,Y}(x,y)$ is defined by

$$F_{X,Y}(x,y) = P([X \le x] \cap [Y \le y]) \ .$$

Let us consider a simple example of a bivariate distribution function. Suppose we have an urn that contains three red balls, two white balls and one blue ball. Now suppose we sample twice *without replacement.* Let X denote the number of red balls in the sample, and let Y denote the number of white balls in the sample. We shall determine now their joint distribution function, $F_{X,Y}(x,y)$, for all pairs of real numbers, (x,y). It is easy to verify that

$$\begin{aligned}
P([X=0]\cap[Y=1]) &= \frac{\binom{2}{1}}{\binom{6}{2}} = \frac{2}{15} \ , \\
P([X=1]\cap[Y=0]) &= \frac{\binom{3}{1}}{\binom{6}{2}} = \frac{3}{15} \ , \\
P([X=1]\cap[Y=1]) &= \frac{\binom{3}{1}\binom{2}{1}}{\binom{6}{2}} = \frac{6}{15} \ , \\
P([X=0]\cap[Y=2]) &= \frac{\binom{2}{2}}{\binom{6}{2}} = \frac{1}{15}, \text{ and} \\
P([X=2]\cap[Y=0]) &= \frac{\binom{3}{2}}{\binom{6}{2}} = \frac{3}{15} \ .
\end{aligned}$$

(Notice that events like $[X=0]\cap[Y=0]$ or $]X=2]\cap[Y=1]$ are impossible events.) Now, if one uses the convention that an empty sum is equal to zero,

one obtains after a small amount of computation that, for every pair of real numbers (x, y),

$$F_{X,Y}(x,y) = \sum \{P([X=k] \cap [Y=l]) : 0 \le k \le x, 0 \le l \le y, k+l = 1 \text{ or } 2\} .$$

The values of this distribution function and the regions over which it has those values are as follows:

Value	Region over which $F_{X,Y}(x,y)$ has the value
0	$\min\{x,y\} < 0$ or $\max\{x,y\} < 1$
$\frac{2}{15}$	$0 \le x < 1$ and $1 \le y < 2$
$\frac{3}{15}$	$0 \le x < 1$ and $y \ge 2$
$\frac{3}{15}$	$1 \le x < 2$ and $0 \le y < 1$
$\frac{11}{15}$	$1 \le x < 2$ and $1 \le y < 2$
$\frac{12}{15}$	$1 \le x < 2$ and $y \ge 2$
$\frac{6}{15}$	$x \ge 2$ and $0 \le y < 1$
$\frac{14}{15}$	$x \ge 2$ and $1 \le y < 2$
1	$x \ge 2$ and $y \ge 2$.

In figure ? below there is a graph of this function, the values of the function being recorded over the domains over which the function takes these values.

FIGURE 1

Joint distribution functions have much the same properties as distribution functions of one variable (or, the so-called univariate distribution function). The proofs of the following three theorems are so much like the proofs of the three theorems in section 2.1 that only hints of the proofs will be supplied. Formal proofs are left to the reader.

Theorem 1. *If X and Y are random variables, if $x_1 \le x_2$ and if $y_1 \le y_2$, then*

$$F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2) .$$

Hint of proof: Use the easily proved fact that

$$[X \le x_1] \cap [Y \le y_1] \subset [X \le x_2] \cap [Y \le y_2].$$

Theorem 2. *If X and Y are random variables, then*
(i) $\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y)$ *for every real y, and*
(ii) $\lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x)$ *for every real x.*

Hint of proof: In order to prove part (i), one should use the fact that, for every value of $y \geq 0$,

$$[Y \leq y] = ([X \leq 0] \cap [Y \leq y]) \cup \bigcup_{k=1}^{\infty} [k-1 < X \leq k] \cap [Y \leq y]$$

and then closely folow the steps in the proof of theorem 17 in section 2.1. In order to prove part (ii), one should use the result from part (i) and the fact that $F_{X,Y}(x,y) = F_{Y,X}(y,x)$.

Theorem 3. *If X and Y are random variables, then*

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

and

$$(ii)\ \lim_{y \to -\infty} F_{X,Y}(x_1,y_1) = 0\ .$$

The proof of this theorem is very much like the proof of theorem 17 in section 2.1.

At this time we need not present a formal development of the general notion of multivariate distributions. After carefully going through the development this far, the reader should be able to supply a definition of the joint distribution function of three random variables X_1, X_2, X_3 and should be able to prove theorems similar to theorems 1, 2 and 3. We next introduce the notion of independent random variables.

Definition. Let $X_1, X_2, \cdots$ denote a finite or infinite sequence of random variables. These random variables are said to be **independent** if for all possible selections of pairs of real numbers

$$(a_1, b_1), (a_2, b_2), \cdots$$

where $a_i < b_i$ for all i, and where the values of $\pm\infty$ are allowed, the events

$$[a_1 < X_1 \leq b_1],\ [a_2 < X_2 \leq b_2],\ \cdots$$

are independent.

Theorem 4. *Let X_1, X_2, $\cdots$, X_m denote m random variables. A necessary and sufficient condition that these random variables be independent is that for every m-tuple of real numbers, $(x_1, x_2, \cdots, x_m)$, the equality*

$$F_{X_1,X_2,\cdots,X_m}(x_1, x_2, \cdots, x_m) = \prod_{j=1}^{m} F_{X_j}(x_j)$$

holds.

Proof: We first prove that the condition is necessary. Accordingly, we assume that the random variables are independent and prove that the condition, namely the above equation holds for any point $(x_1, x_2, \cdots, x_m)$. In this case, let $-\infty = a_1 = a_2 = \cdots = a_m$ and $x_i = b_i$ for $1 \leq i \leq m$. Then by the definition of independence and joint distribution we have

$$\begin{aligned} F_{X_1,X_2,\cdots,X_m}(x_1, x_2, \cdots, x_m) &= P\left(\bigcap_{k=1}^{m} [-\infty < X_k \leq x_k]\right) \\ &= \prod_{k=1}^{m} P\left([-\infty < X_k \leq x_k]\right) \\ &= \prod_{k=1}^{m} P([X_k \leq x_k]) \\ &= \prod_{j=1}^{m} F_{X_j}(x_j) \ , \end{aligned}$$

which establishes the condition. We now prove that the equation in the theorem implies independence of the random variables.We first remark that if $(i_1, i_2, \cdots, i_k)$ is any subset of the integers $1, 2, , \cdots, m, \ ,$ where $1 \leq i_1 < \cdots < i_k \leq m$, then

$$F_{X_{i_1},\cdots,X_{i_m}}(x_{i_1}, \cdots, x_{i_m}) = \prod_{j=1}^{k} F_{X_{i_j}}(x_{i_j}) \ .$$

This follows by taking the limit of both sides as $x_r \to \infty$ for all $r \notin \{i_1, \cdots, i_k\}$ and by using theorems 1 and 2. We next prove that

$$P([a_1 < X_1 \leq b_1] \cap [a_2 < X_2 \leq b_2]) = \prod_{j=1}^{2} P([a_j < X_j \leq b_j]) \ .$$

Accordingly, we obtain from the above remark that

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)$$

for all real number pairs (x_1, x_2). We first note that the following equations hold:

$$\begin{aligned} F_{X_1,X_2}(b_1, b_2) &= F_{X_1}(b_1) F_{X_2}(b_2) \\ F_{X_1,X_2}(a_1, b_2) &= F_{X_1}(a_1) F_{X_2}(b_2) \\ F_{X_1,X_2}(b_1, a_2) &= F_{X_1}(b_1) F_{X_2}(a_2) \\ F_{X_1,X_2}(a_1, a_2) &= F_{X_1}(a_1) F_{X_2}(a_2). \end{aligned}$$

Also it is easy to verify that

$$\begin{aligned}[X_1 \le b_1] \cap [X_2 \le b_2] \quad &= [a_1 < X_1 \le b_1] \cap [a_2 < X_2 \le b_2] \\ &\cup [a_1 < X_1 \le b_1] \cap [X_2 \le a_2] \\ &\cup [X_1 \le a_1] \cap [a_2 < X_2 \le b_2] \\ &\cup [X_1 \le a_1] \cap [X_2 \le a_2] \ .\end{aligned}$$

Since

$$[X_1 \le b_1] \cap [X_2 \le a_2] = ([X_1 \le a_1] \cap [X_2 \le a_2]) \cup ([a_1 < X_1 \le b_1] \cap [X_2 \le a_2])$$

we obtain, after taking probabilities of both sides, that

$$P([a_1 < X_1 \le b_1] \cap [X_2 \le a_2]) = F_{X_1,X_2}(b_1, a_2) - F_{X_1,X_2}(a_1, a_2) \ .$$

In a similar manner, we obtain

$$P([X_1 \le a_1] \cap [a_2 < X_2 \le b_2]) = F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(a_1, a_2) \ .$$

Now, making use of all of the above, we finally obtain

$$\begin{aligned}&P([a_1 < X_1 \le b_1] \cap [a_2 < X_2 \le b_2]) \\ &= F_{X_1}(b_1)F_{X_2}(b_2) - F_{X_1}(a_1)F_{X_2}(b_2) - F_{X_1}(b_1)F_{X_2}(a_2) + F_{X_1}(a_1)F_{X_2}(a_2) \\ &= (F_{X_1}(b_1) - F_{X_1}(a_1))(F_{X_2}(b_2) - F_{X_2}(a_2)) \\ &= P([a_1 < X_1 \le b_1])P([a_2 < X_2 \le b_2]) \ .\end{aligned}$$

If, in the above proof, one replaces X_1 by X_{i_1} and X_2 by X_{i_2} where $1 \le i_1 < i_2 \le n$, then one proves in the very same way that

$$P(\bigcap_{j=1}^{2} [a_{i_j} < X_{i_j} \le b_{i_j}]) = \prod_{j=1}^{2} P([a_{i_j} < X_{i_j} \le b_{i_j}]) \ .$$

The proof that

$$P(\bigcap_{j=1}^{k} [a_{i_j} < X_{i_j} \le b_{i_j}]) = \prod_{j=1}^{k} P([a_{i_j} < X_{i_j} \le b_{i_j}])$$

for $3 \le k \le m$ is the same but is more complicated and is more a problem of complicated notation than one of mathematics. Consequently, a proof of the above equation for $k \ge 3$ will not be given here. However, the reader is urged to try his or her hand at proving this in the case when $k = 3$.

EXERCISES

1-3. Write out formal proofs of theorems 1,2 and 3.

4. State the definition of the joint distribution function of random variables X, Y and Z.

5-7. State and prove theorems analogous to theorems 1, 2 and 3 for three random variables X, Y and Z.

8. Prove: If X, Y and Z are random variables that satisfy

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

for all real x, y and z, then X, Y and Z are independent.

5.2 Multivariate Densities. In sections 2.2 and 2.3 we discussed univariate discrete and absolutely continuous distribution functions. In this section we discuss discreteness and absolutely continuity of multivariate distribution functions and determine their densities. First we look at the multivariate discrete distribution.

Definition. If $\mathbf{X} = (X_1, \cdots, X_n)$ are random variables, then they are said to have a discrete joint distribution if there exists a countable set of distinct points in $\mathbf{R}^n$, $\{\mathbf{x}_1, \mathbf{x}_2, \cdots\}$, such that

$$\sum\nolimits_{j\geq 1} P([\mathbf{X} = \mathbf{x}_j]) = 1.$$

The countable set of points $\{\mathbf{x}_1, \mathbf{x}_2, \cdots\} \in \mathbf{R}^n$ that satisfy this requirement is referred to as the *range* of $\mathbf{X}$ and is sometimes denoted by $range(\mathbf{X})$. The *discrete density* of $\mathbf{X}$ is then defined as

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P\left(\bigcap_{j=1}^{n} [X_j = x_j]\right).$$

An important multivariate discrete density in statistics is that of the multinomial distribution. It arises in the following manner. Suppose that a game consists of $k+1$ disjoint outcomes denoted by $A_0, A_1, \cdots, A_k$. Suppose that this game is played n times under identical conditions, so that the outcome of one play does not influence the outcome of any other play. Let X_i denote the number of times in the n plays that the outcome is A_i, $0 \leq i \leq k$. One should first notice that $X_0 + X_1 + \cdots + X_k = n$. Thus if we can determine

the joint density of $X_1, \cdots, X_k$, we are able to determine the joint density of all $k+1$ discrete random variables. So let us determine this joint density. It is easy to see that the fundamental probability set, Ω, is the set of all $(k+1)^n$ ordered n-tuples of the outcomes $A_0, A_1, \cdots, A_k$. (Here is another case where the individual outcomes are not necessarily equally likely.) Let p_i denote the probability that A_i occurs at a particular trial. For $1 \leq i \leq k$, If we let x_i be a non-negative integer, $1 \leq i \leq k$, such that $0 \leq x_1 + \cdots + x_k \leq n$, then the probability that, for one particular elementary event, x_1 of the outcomes are A_1's, x_2 of the outcomes are A_2's, $\cdots$, and x_k of the outcomes are A_k's is

$$p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} (1 - p_1 - \cdots - p_k)^{n - x_1 - \cdots - x_k}.$$

We must now determine how many such elementary events there are in $\bigcap_{j=1}^{k} [X_j = x_j]$. There are $\binom{n}{x_1}$ ways in which x_1 out of the n trials can be selected for A_1 to occur. For each way that x_1 trials can be selected for A_1 to occur, there are $\binom{n-x_1}{x_2}$ ways in which one can select x_2 trials out of the remaining $n - x_1$ trials for the outcome to be A_2. After selecting $x_1 + \cdots + x_k$ trial numbers for x_1 of the A_1's to occur on up to x_k of the A_k's there is exactly one way in which to select the remaining trial numbers for A_0 only to occur. Thus the joint density of $X_1, \cdots, X_k$ is

$$f_{X_1, \cdots, X_k}(x_1, \cdots, x_k) = \binom{n}{x_1} \binom{n - x_1}{x_2} \cdots \binom{n - x_1 - \cdots - x_{k-1}}{x_k} \prod_{i=0}^{k} p_i^{x_i} \ ,$$

where $x_0 = n - x_1 - \cdots - x_{k-1}$. Upon suitable cancelling, one observes that

$$\binom{n}{x_1} \binom{n - x_1}{x_2} \cdots \binom{n - x_1 - \cdots - x_{k-1}}{x_k} = \frac{n!}{x_0! x_1! \cdots x_k!} \ .$$

Consequenntly

$$f_{X_1, \cdots, X_k}(x_1, \cdots, x_k) = \frac{n!}{x_0! x_1! \cdots x_k!} \prod_{i=0}^{k} p_i^{x_i}$$

for integers $x_0, x_1, \cdots, x_k$, where each $x_i \geq 0$ and $x_0 + x_1 + \cdots + x_k = n$. This joint distribution arises in many applied statistical problems and is called the *multinomial distribution.*

Theorem 1. *Random variables $X_1, \cdots, X_n$ have a joint discrete distribution if and only if each X_i has a discrete distribution.*

Proof: Suppose $X_1, \cdots, X_n$ have a joint discrete distribution. We shall prove that each random variable has a discrete distribution only in the case $n = 2$. Suppose that

$$range(X_1, X_2) = \{(x_{11}, x_{12}), (x_{21}, x_{22}), \cdots, (x_{m1}, x_{m2})\} \ .$$

Now let $F = \{x : (x, x') \in range(X_1, X_2)$ for some $x'\}$, i.e., F is the set of all first coordinates of points in $range(X_1, X_2)$. For every $x \in F$, let $S_x = \{y : (x, y) \in range(X_1, X_2)\}$. Then, $range(X_1, X_2) = \cup\{\{(x, y) : y \in S_x\} : x \in F\}$. For every $x \in F$,

$$\begin{aligned} P([X_1 = x]) &= P\left(\bigcup\{[X_1 = x] \cap [X_2 = y] : y \in S_x\}\right) \\ &= \sum\{P[X_1 = x] \cap [X_2 = y]) : y \in S_x\} > 0 \ . \end{aligned}$$

Hence $F \subset range(X_1)$. But

$$\sum_{x \in F} P([X_1 = x]) = \sum_{x \in F} \sum_{y \in S_x} P([X_1 = x] \cap [X_2 = y]) = 1,$$

and thus F is the range of X_1, thus proving that X_1 has a discrete distribution. A proof that X_2 has a discrete distribution is similar to the above proof. Conversely, suppose that each of the individual random variables has a discrete distribution. Then,

$$X_1 = \sum\nolimits_{n \geq 1} x_n I_{[X_1 = x_n]} \text{ and } X_2 = \sum\nolimits_{n \geq 1} y_n I_{[X_2 = y_n]}$$

for countable sets of distinct numbers $\{x_n\}$ and $\{y_n\}$. Now the set, S, of all possible pairs $\{(x_i, y_j) : i \geq 1, j \geq 1\}$ is countable. For each $i \geq 1$,

$$[X_1 = x_i] = \bigcup\nolimits_{j \geq 1} ([X_1 = x_i] \cap [X_2 = y_j]),$$

and hence the set S_i defined by $S_i = \{j \geq 1 : P([X_1 = x_i] \cap [X_2 = y_j]) > 0\}$ is nonempty, and

$$\sum\{P([X_1 = x_i] \cap [X_2 = y_j]) : j \in S_i\} = P([X_1 = x_i]) > 0.$$

Thus,

$$\sum\nolimits_{i \geq 1} \sum\{P([X_1 = x_i] \cap [X_2 = y_j]) : j \in S_i\} = \sum\nolimits_{i \geq 1} P([X_1 = x_i]\} = 1,$$

and we have shown that $range(X_1, X_2) \subset \cup_{i \geq 1}\{(i,j) : j \in S_i\}$, which is a countable set. This shows that the joint distribution of (X_1, X_2) is discrete.

Theorem 2. *If X_1, X_2 have a joint discrete distribution, then so does X_1, and*

$$f_{X_1}(x_1) = \sum \{f_{X_1,X_2}(x_1, x_2) : x_2 \in S_{x_1}\}.$$

Proof: This was actually proved in the proof of theorem 1.

Theorem 3. *If random variables $X_1, \cdots, X_n$ have a joint discrete distribution, then they are independent if and only if*

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j)$$

for all $(x_1, \cdots, x_n) \in range(X_1, \cdots, X_n)$.

Proof: We first prove that the condition is sufficient. Assuming that the above factorization of the joint density is true, we obtain

$$\begin{aligned} F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) &= \sum f_{X_1,X_2,\cdots,X_n}(t_1, t_2, \cdots, t_n) \\ &= \sum \prod_{j=1}^{n} f_{X_j}(t_j) \ , \end{aligned}$$

where these sums are taken over the set

$$\{(t_1, t_2, \cdots, t_n) \in range(X_1, X_2, \cdots, X_n) : t_i \leq x_i, 1 \leq i \leq n\} \ .$$

But this last sum above is equal to the product of the sums, and thus

$$\begin{aligned} F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) &= \prod_{j=1}^{n} \sum \{f_{X_j}(t_j) : t_j \in range(X_j), t_j \leq x_j\} \\ &= \prod_{j=1}^{n} F_{X_j}(x_j). \end{aligned}$$

Independence of the random variables follows from theorem 4 in section 5.1. We now prove that the condition is necessary. Assuming the random variables to be independent, we find that for every $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$,

$$\begin{aligned} F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) &= \prod_{j=1}^{n} F_{X_j}(x_j). \\ &= \prod_{j=1}^{n} \sum \{f_{X_j}(t_j) : t_j \in range(X_j), t_j \leq x_j\} \\ &= \sum \prod_{j=1}^{n} f_{X_j}(t_j) \ , \end{aligned}$$

where this last sum is taken over the set

$$\{(t_1, t_2, \cdots, t_n) \in range(X_1, X_2, \cdots, X_n) : t_i \leq x_i, 1 \leq i \leq n\} \ .$$

By the uniqueness of the discrete density, this proves that the condition is true.

Definition. Random variables $X_1, \cdots, X_n$ are said to have a joint absolutely continuous distribution (function) if there exists a nonnegative function $f_{X_1,\cdots,X_k}(x_1, \cdots, x_k)$ defined over n-dimensional euclidean space, $\mathbf{R}^n$, such that

$$F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1,\cdots,X_n}(t_1, \cdots, t_n) dt_1 \cdots dt_n$$

for every n-tuple of real numbers $(x_1, \cdots, x_n)$.

The function $f_{X_1,\cdots,X_k}(x_1, \cdots, x_k)$ is called a joint density of the random variables $X_1, \cdots, X_n$. At this point we recall that the fundamental theorem of calculus tells us that: *if a real-valued function g is Riemann-integrable over the interval $[a, b]$, then the function G defined by $G(x) = \int_a^x g(t)dt$ for all $x \in [a, b]$ is continuous over the interval $[a, b]$,and if g is continuous at $x_0 \in [a, b]$, then is G differentiable at x_0 and $G'(x_0) = g(x_0)$.* Using this theorem, one easily sees that at every value of x at which a univariate density $f_X(x)$ is continuous, then $F'_X(x) = f_X(x)$. In the multivariate case, at every point $(x_1, \cdots, x_k)$ in n-dimensional euclidean space, $\mathbf{R}^n$, at which $f_{X_1,\cdots,X_k}(x_1, \cdots, x_k)$ is continuous, we have

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = f_{X_1,\cdots,X_k}(x_1, \cdots, x_k) \ .$$

The following theorem will not be stated and proved in all its generality because the general statement and proof require too much complicated notation. We instead state and prove a special case; the reader should be able to extend the proof to more general cases.

Theorem 4. *If X_1, X_2, X_3 have a joint absolutely continuous distribution function, then so does X_1, X_3, and a joint density of X_1, X_3 may be written as*

$$f_{X_1,X_3}(x_1, x_3) = \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1, x_2, x_3) dx_2.$$

Proof: By theorem 2 in section 5.1, the definition of density and the usual properties of integrals (i.e., that we can change the order of integration in multiple integrals)

$$\begin{aligned} F_{X_1,X_3}(x_1,x_3) &= P([X_1 \le x_1] \cap [X_3 \le x_3]) \\ &= P([X_1 \le x_1] \cap [X_2 \le \infty] \cap [X_3 \le x_3]) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \int_{-\infty}^{x_3} f_{X_1,X_2,X_3}(t_1,t_2,t_3) dt_3 dt_2 dt_1 \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_3} \left\{ \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(t_1,t_2,t_3) dt_2 \right\} dt_1 dt_3 \ . \end{aligned}$$

Thus by the definition of joint density function of X_1 and X_3, we may take

$$f_{X_1,X_3}(x_1,x_3) = \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(t_1,t_2,t_3) dt_2 \ ,$$

which proves the theorem.

The joint density $f_{X_1,X_3}(x_1,x_3)$ just obtained is called a "marginal" or "marginal density" of the joint density $f_{X_1,\cdots,X_k}(x_1,\cdots,x_k)$ .In general, we see that when we want to find the joint density of a certain finite set of random variables, and if we are given a joint density of a finite set of random varaibles that includes them, we integrate out the extraneous variables. As an example, suppose X, Y are random variables whose joint density is given by

$$f_{X,Y}(x,y) = \begin{cases} 1 \text{ if } x \ge 0, y \ge 0, y + 2x \le 2 \\ 0 \text{ otherwise.} \end{cases}$$

One can easily check that this is a density, namely that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1.$$

Applying the above theorem to the taking of marginal densities, we have

$$f_X(x) = \begin{cases} \int_0^{2-2x} 1 dy = 2 - 2x \text{ if } 0 \le x \le 1 \\ 0 \text{ otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} \int_0^{1-\frac{1}{2}y} 1 dx = 1 - \frac{1}{2}y \text{ if } 0 \le y \le 2 \\ 0 \text{ otherwise.} \end{cases}$$

If random variables have a joint absolutely continuous distribution, then it is easy to tell if they are independent. This is accomplished by the following theorem.

Theorem 5. *If random variables $X_1, X_2, \cdots, X_n$ have a joint absolutely continuous distribution, then a necessay and sufficient condition that they are independent is that their joint density factors into the product of their univariate marginals, i.e.,*

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j)$$

for all $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$.

Proof: We first prove that the condition is sufficient. Assuming the above factorization to be true, we obtain

$$\begin{aligned} F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\cdots,X_n}(t_1, \cdots, t_n) dt_n \cdots dt_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left\{ \prod_{j=1}^{n} f_{X_j}(t_j) \right\} dt_n \cdots dt_1 \\ &= \prod_{j=1}^{n} \int_{-\infty}^{x_j} f_{X_j}(t_j) dt_j = \prod_{j=1}^{n} F_{X_j}(x_j) \ , \end{aligned}$$

from which independence follows by theorem 4 in section 5.1. We now prove that the condition is necessary. Assuming the random variables to be independent, we find that, for every $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$,

$$\begin{aligned} F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) &= \prod_{j=1}^{n} F_{X_j}(x_j) \\ &= \prod_{j=1}^{n} \int_{-\infty}^{x_j} f_{X_j}(t_j) dt_j \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left\{ \prod_{j=1}^{n} f_{X_j}(t_j) \right\} dt_n \cdots dt_1 \ . \end{aligned}$$

By the very definition of a joint density, the integrand of the last multiple integral above is a joint density, and the theorem is proved.

Aside to sophisticates: For an undergraduate course where one is limited to the Riemann integral, much is left unsaid. Certain statements made above are justified only by appealing to the a.e. uniquenss of the Radon-Nikodym theorem and also to Tonelli's theorem.

EXERCISES

1. Let X_1 and X_2 be random variables with a joint discrete distribution, define

$$F = \{x : (x, x') \in range(X_1, X_2) \text{ for some } x'\},$$

i.e., F is the set of all first coordinates of points in $range(X_1, X_2)$, and for every $x \in F$, let $S_x = \{y : (x, y) \in F\}$. Prove that

$$[X_1 = x] = [(X_1, X_2) \in \{(x, y) : y \in S_x\}] \text{ for every } x \in F,$$

and

$$P([X_1 = x]) = \sum \{P([X_1 = x] \cap [X_2 = y]) : y \in S_x\} \text{ for every } x \in F.$$

2. Three tags, numbered 1, 2 and 3, are placed in a box. One samples two times with replacement. Let X denote the number on the first tag selected, and let Y denote the number on the second tag selected. Find the joint discrete density function and and the joint distribution function of (X, Y).

3. Use theorem 4 in this section to prove that X and Y in problem 2 are independent.

4. An urn contains two red balls, two white balls and one blue ball. Let X denote the number of red balls in a sample of size three taken without replacement, and let Y denote the number of white balls in the same sample. Find the joint density function of (X, Y), and show that X and Y are not independent.

5. Let X and Y be random variables with joint absolutely continuous distribution whose joint density is given by

$$f_{X,Y}(x.y) = \begin{cases} \frac{3}{4} \text{ if } \mid x \mid \leq 1, x^2 \leq y \leq 1 \\ 0 \text{ otherwise.} \end{cases}$$

Find the two univariate marginal densities, $f_X(x)$ and $f_Y(y)$.

6. Let X, Y and Z be independent random variables, each of which has an absolutely continuous distribution function.

(i) Prove that the joint distribution fuction of X, Y,Z is absolutely continuous.

(ii) If

$$f_X(x) = \begin{cases} e^{-x} \text{ for } x > 0 \\ 0 \text{ for } x \leq 0, \end{cases}$$

if

$$f_Y(y) = \begin{cases} 1 \text{ if } \ 0 < y < 1 \\ 0 \text{ if } \ \mid y - \frac{1}{2} \mid \geq \frac{1}{2} \end{cases}$$

and if

$$f_Z(z) = \frac{1}{2}e^{-|z|} \text{ for all real } z,$$

determine the joint density of X, Y,Z.

7. Let U and V be random variables with joint absolutely continuous distribution function whose density is given by

$$f_{U,V}(u,v) = \begin{cases} \frac{2}{\pi} \text{ if } \mid v \mid \leq 1, 0 \leq u \leq \sqrt{1-v^2} \\ 0 \text{ otherwise.} \end{cases}$$

(i) Find the marginal densities $f_U(u)$ and $f_V(v)$.

(ii) Show that U and V are not independent.

8. At the beginning of this section where we were dixcussing the multinomial distribution, there is the statement, "Thus if we can determine the joint density of $X_1, \cdots, X_k$, we are able to determine the joint density of all $k+1$ discrete random variables." Render a precise meaning of this statement.

5.3 Variance and Covariance. In this section we assemble some basic results about variance and covariance and their relationships with independence.

Theorem 1. *If X and Y are independent random variables with finite first moments, then the random variable XY has a finite first moment, and $E(XY) = E(X)E(Y)$.*

Proof: We shall first prove this theorem when both X and Y are positive random variables, i.e., when $P([X \geq 0]) = 1$ and $P([Y \geq 0]) = 1$. Case (i): Suppose Y is discrete and takes only a finite number of non-negative values, $0 < z_1 < \cdots < z_n$, with positive probabilities. Then for $z > 0$,

$$P([XY > z]) = \sum_{j=1}^{n} P([XY > z] \cap [Y = z_j]) \ .$$

Now, by the theorem of total probabilities,

$$\begin{aligned} P([XY > z]) &= \sum_{j=1}^{n} P([XY > z] \mid [Y = z_j])P([Y = z_j]) \\ &= \sum_{j=1}^{n} P([X > \tfrac{z}{z_j}] \mid [Y = z_j])P([Y = z_j]) \ , \end{aligned}$$

and since X and Y are independent, we obtain

$$P([XY > z]) = \sum_{j=1}^{n} P([X > \frac{z}{z_j}])P([Y = z_j]) \ .$$

Hence

$$\int_0^\infty P([XY > z])dz = \sum_{j=1}^n P([Y = z_j]) \int_0^\infty P([X > \frac{z}{z_j}])dz \ ,$$

and after a change of variable in each integral on the right hand side, we obtain

$$\begin{aligned} \int_0^\infty P([XY > z])dz &= \sum_{j=1}^n P([Y = z_j])z_j \int_0^\infty P([X > x])dx \\ &= E(Y) \int_0^\infty P([X > x])dx \\ &= E(X)E(Y) \end{aligned} \quad ,$$

since X and Y are positive random variables. Case (ii) : Now suppose X is as before in case (i) and Y is a bounded non-negative random variable, i.e., there exists a positive integer K such that $0 \leq Y(\omega) \leq K$ for all $\omega \in \Omega$. As in the proof of theorem 4 in section 3.2, we define for every positive integer n,

$$\begin{aligned} Y_n^- &= \sum_{j=1}^{K2^n} \tfrac{j-1}{2^n} I_{[(j-1)/2^n < Y \leq j/2^n]}, \text{ and} \\ Y_n^+ &= \sum_{j=1}^{K2^n} \tfrac{j}{2^n} I_{[(j-1)/2^n < Y \leq j/2^n]} \ . \end{aligned}$$

As was established in the proof of theorem 4 in section 3.2, $Y_n^-(\omega) \leq Y(\omega) \leq Y_n^+(\omega)$ for all $\omega \in \Omega$ and

$$\begin{aligned} &0 \leq E(Y_n^-) \leq E(Y) \leq E(Y_n^+) \ , \text{ plus} \\ &0 \leq E(Y) - E(Y_n^-) \leq \tfrac{1}{2^n} \text{ and} \\ &0 \leq E(Y_n^+) - E(Y) \leq \tfrac{1}{2^n} \ . \end{aligned}$$

Thus

$$\lim_{n\to\infty} E(Y_n^+) = E(Y) \text{ and } \lim_{n\to\infty} E(Y_n^-) = E(Y) \ .$$

Also note that for each n, X and Y_n^+ are independent, and X and Y_n^- are independent. To prove this, one observes that for every $y \in (0, K)$, there exists exactly one value of k such that $\frac{k}{2^n} \leq y < \frac{k+1}{2^n}$. Thus

$$[Y_n^+ \leq y] = [Y_n^+ \leq \frac{k}{2^n}] = [Y \leq \frac{k}{2^n}].$$

Using the independence of X and Y, we obtain

$$\begin{aligned} P([X \leq x] \cap [Y_n^+ \leq y]) &= P([X \leq x] \cap [Y \leq \tfrac{k}{2^n}]) \\ &= P([X \leq x])P([Y \leq \tfrac{k}{2^n}]) \\ &= P([X \leq x])P([Y_n^+ \leq y]), \end{aligned}$$

which proves the independence of X and Y_n^+. The proof of independence of X and Y_n^- is the same where each k is replaced by $k-1$. Because $XY_n^- \leq XY \leq XY_n^-$ over Ω, then by theorem 3 in section 3.2 and case (i),

$$E(X)E(Y_n^-) = E(XY_n^-) \leq E(XY) \leq E(XY_n^+) = E(X)E(Y_n^+) \ .$$

Letting $n \to \infty$, we obtain the conclusion of the theorem for case (ii). Case (iii) We now consider the case where X and Y are any independent, positve random variables whose expectations exist. First, we show that if $K > 0$ is any positive integer, then X and $YI_{[Y<K]}$ are independent. Indeed, we first observe that, for any $y > 0$,

$$[YI_{[Y \leq K]} \leq y] = [Y \leq \min\{K, y\}] \cup [Y > K],$$

the union being a disjoint union. Then, because of independence and disjointness, we have, for every $x > 0$ and every $y > 0$,

$$\begin{aligned} &P([X \leq x] \cap [YI_{[Y \leq K]} \leq y]) \\ &= P([X \leq x] \cap ([Y \leq \min\{K, y\}]) + P([X \leq x] \cap [Y > K]) \\ &= P([X \leq x])P([Y \leq \min\{K, y\}]) + P([X \leq x])P([Y > K]) \\ &= P([X \leq x])P([Y \leq \min\{K, y\}] \cup [Y > K]) \\ &= P([X \leq x])P([YI_{[Y \leq K]} \leq y]), \end{aligned}$$

which proves that X and $YI_{[Y<K]}$ are independent. Since $YI_{[Y<K]}$ is a bounded, positive random variable, we have by case (ii) that

$$E(XYI_{[Y<K]}) = E(X)E(YI_{[Y<K]}).$$

This and lemma 1 in section 3.2 and its corollary imply

$$\begin{aligned} \int_0^\infty P([XY > x])dx &= lim_{K\to\infty} \int_0^\infty P([XY > x] \cap [| Y |< K]))dx \\ &= lim_{K\to\infty} \int_0^\infty P([XYI_{[Y<K]} > x])dx \\ &= lim_{K\to\infty} E(XYI_{[Y<K]}) \\ &= lim_{K\to\infty} E(X)E(YI_{[Y<K]}) \\ &= E(X)E(Y) \ , \end{aligned}$$

which is finite. Thus the expectation of XY exists and equals $E(X)E(Y)$. We now establish the theorem in the general case. Let us denote

$$\begin{array}{ll} X^+ = XI_{[X\geq 0]} \ , & X^- = -XI_{[X<0]} \ , \\ Y^+ = YI_{[Y\geq 0]} \ , & Y^- = -YI_{[Y<0]} \ . \end{array}$$

Clearly, each of the random variables X^+, X^-, Y^+ and Y^- is a non-negative random variable, and $X = X^+ - X^-$ and $Y = Y^+ - Y^-$. Now

$$XY = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+ \ .$$

Claim: X^+ and Y^+ are independent, X^+ and Y^- are independent, X^- and Y^+ are independent and X^- and Y^- are independent. We prove this claim in only one case and prove that X^+ and Y^- are independent. We first note that for $y \geq 0$,

$$[Y^- \leq y] = [-YI_{[Y<0]} \leq y] = [Y \geq -y],$$

and thus, for $x \geq 0$ and $y \geq 0$ and by the definition of independent random variables,

$$\begin{aligned} P([X^+ \leq x] \cap [Y^- \leq y]) &= P([X \leq x] \cap [Y \geq -y]) \\ &= P([X \leq x])P([Y \geq -y]) \\ &= P([X^+ \leq x])P([Y^- \leq y]), \end{aligned}$$

which proves the claim. Also, one can show that the existence of $E(X)$ and $E(Y)$ imply that of $E(X^+)$, $E(X^-)$, $E(Y^+)$ and $E(Y^-)$ and that $E(X) = E(X^+) - E(X^-)$ and $E(Y) = E(Y^+) - E(Y^-)$. Thus, by all this and the last displayed equation, we have by the fact that we have shown the theorem to to be true for all non-negative random variables:

$$\begin{aligned} E(XY) &= E(X^+)E(Y^+) + E(X^-)E(Y^-) - E(X^+)E(Y^-) - E(X^-)E(Y^+) \\ &= ((E(X^+) - E(X^-))(E(Y^+) - E(Y^-)) = E(X)E(Y) \ , \end{aligned}$$

which proves the theorem.

Definition. The covariance of two random variables, X and Y, is defined by

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))),$$

provided that the expectaions involved exist.

Theorem 2. *If X and Y are random variables, then $Cov(X,Y) = E(XY) - E(X)E(Y)$, provided the expectations exist.*

Proof: By previous results on expectations and the definition above, we have

$$\begin{aligned} Cov(X,Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - XE(Y) - E(X)Y + E(X)E(Y)) \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y)\ . \end{aligned}$$

Lemma 1. *If X is a random variable such that $E(X^2) = 0$,* then $P([X = 0]) = 1$.

Proof: Let us suppose that the conclusion is not true. Since $P([X^2 \geq 0]) = 1$, there exists $x_0 > 0$ such that $P([X^2 \leq x_0]) < 1$ or $1 - P([X^2 \leq x_0]) > 0$. Let $\epsilon = 1 - P([X^2 \leq x_0]) > 0$. Then for all $x \in [0, \epsilon)$, $P([X^2 > x]) \geq \epsilon$. Thus

$$E(X^2) = \int_0^\infty P([X^2 > x])dx \geq \int_0^{x_0} P([X^2 > x])dx \geq \epsilon x_0 > 0,$$

which contradicts the hypothesis and establishes the lemma.

Lemma 2. *If a and b are real numbers, then $(a + b)^2 \leq 4a^2 + 4b^2$.*

Proof: Easily, if $|a| \leq |b|$, then

$$\begin{aligned} (a + b)^2 &\leq a^2 + 2|a||b| + b^2 \\ &\leq a^2 + 3b^2 \\ &\leq 3a^2 + 3b^2. \end{aligned}$$

The same conclusion is established when $|a| \geq |b|$.

Lemma 3. *If X and Y are random variables with finite second moments, then $X + Y$ and XY have finite second moments.*

Proof: Since by hypothesis, X^2 and Y^2 have finite expectations, and since by lemma 2, $0 \leq (X + Y)^2 \leq 3X^2 + 3Y^2$, then by theorem 3 in section 3.2, $0 \leq E((X + Y)^2) \leq 3E(X^2) + 3E(Y^2)$. Thus $X + Y$ has a finite second moment. On the other hand, since $XY = \frac{1}{2}((X + Y)^2 - X^2 - Y^2)$, and since all three random variables on the right hand side have finite expectations, then the expectation of the left hand side exists.

Theorem 3. (*Schwarz' Inequality*) *If X and Y are random variables with finite second moments, then the product XY has a finite second moment, and*

$$(E(XY))^2 \leq E(X^2)E(Y^2);$$

in addition, *equality holds if and only if there exists a real number t such that*

$$P([tX + Y = 0]) = 1 \text{ or } P([X + tY = 0]) = 1.$$

Proof: It should be recalled that the quadratic equation $ax^2 + 2bx + c = 0$ has two distinct roots if and only if $b^2 - ac > 0$, one double root if and only if $b^2 - ac = 0$ and no real roots if and on;y if $b^2 - ac < 0$. Since, by hypothesis, both X and Y have finite second moments, then by lemma 3 and for any real number t, it follows that $tX + Y$ has a finite second moment. Now, for any real t, $P([(tX + Y)^2 \geq 0]) = 1$, and this implies that

$$E((tX + Y)^2) = t^2E(X^2) + 2tE(XY) + E(Y^2) \geq 0.$$

This inequality tells us that the quadratic equation

$$t^2E(X^2) + 2tE(XY) + E(Y^2) = 0$$

either has no real roots or one double root, and hence

$$(E(XY))^2 - E(X^2)E(Y^2) \leq 0,$$

which gives the first conclusion of the theorem. Strict equality holds if and only if there exists a value of t such that $E((tX + Y)^2) = 0$. By lemma 1, this implies that $P([tX + Y = 0]) = 1$. Conversely, if $P([tX + Y = 0]) = 1$ for some t, then $E((tX + Y)^2) = 0$ for that value of t, and thus the proof is complete.

Definition. If a random variable X has a finite second moment, then the standard deviation of X, $s.d.(X)$, is defined by $s.d.(X) = \sqrt{Var(X)}$.

Definition. Let X and Y be random variables with finite second moments. The correlation coefficient $\rho(X, Y)$ of X and Y is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{s.d.(X)s.d.(Y)} ,$$

provided $s.d.(X) > 0$ and $s.d.(Y) > 0$. If the standard deviation of any of these random variables is equal to zero, then the correlation coefficient is not defined.

Theorem 4. *If X and Y are random variables with positive standard deviations, then $-1 \leq \rho(X, Y) \leq 1$. The equality $\rho(X, Y) = 1$ holds if and*

only if there exists a positive number β such that $Y - E(Y) = \beta(X - E(X))$. The equality $\rho(X, Y) = -1$ holds if and only if there exists a negative number γ such that $Y - E(Y) = \gamma(X - E(X))$.

Proof: By Schwarz' inequality

$$(E(((X - E(X))(Y - E(Y))))^2 \leq (E(X - E(X)))^2(E(X - E(X)))^2,$$

which implies that $(\rho(X, Y))^2 \leq 1$, which in turn implies that $-1 \leq \rho(X, Y) \leq 1$. By Schwarz' inequality, one of the two equal signs holds if and only if there exists a constant t such that $Y - E(Y) = t(X - E(X))$. Suppose $t < 0$. Then, multiplying both sides by $X - E(X)$ and taking expectations, we get $Cov(X, Y) = tVar(X)$. Hence $Cov(X, Y) = -1$. Similarly, one can show that when $t > 0$, then $\rho(X, Y) = 1$. This concludes the proof.

Corollary. *If X and Y are independent random variables with finite second moments and positive standard deviations, then $Cov(X, Y) = 0$ and $\rho(X, Y) = 0$.*

Proof: By theorems 1 and 2 of this section, we have

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0,$$

from which the second conclusion trivially follows.

At this point it should be mentioned that the converse to the above corollary is not necessarily true. Two random variables can be uncorrelated and yet be highly dependent. A problem in the set of exercises at the end of this section illustrates this.

Theorem 5. *If $X_1, \cdots, X_n$ are independent random variables with finite second moments, then*

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k).$$

Proof: First note that $E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$. Then

$$\begin{aligned}
Var\left(\sum_{k=1}^{n} X_k\right) &= E\left((\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k))^2\right) \\
&= E\left((\sum_{k=1}^{n} (X_k - E(X_k)))^2\right) \\
&= E\left\{\sum_{k=1}^{n} (X_k - E(X_k))^2 + \sum_{i \neq j} (X_i - E(X_i))(X_j - E(X_j))\right\} \\
&= \sum_{k=1}^{n} Var(X_k) + \sum_{i \neq j} Cov(X_i, X_j) \ .
\end{aligned}$$

For $i \neq j$, X_i and X_j are independent, and so by the above corollary, $Cov(X_i, X_j) = 0$, thus proving the theorem.

EXERCISES

1. Prove: if Z is a random variable whose second moment exists, and if a random variable W is defined by $W = \frac{1}{s.d.(Z)} Z$, then $Var(W) = 1$.

2. An urn contains three red balls and two black balls.One samples four times with replacement, selecting one ball at random each time and noting its color. Let X denote the number of red balls selected in the first two trials, and let Y denote the number of black balls selected in the third and fourth trials.

(i) Find the discrete density of X, of Y and of XY.

(ii) Compute the correlation coefficient of X and Y, $\rho(X, Y)$.

3. An urn contains three red balls and two black balls. One samples without replacement two times. Let X denote the number of red balls selected, and let Y denote the number of white balls selected.

(i) Find the discrete density of X, then of Y and finally of XY.

(ii) Compute the correlation coefficient, $\rho(X, Y)$.

(iii) Show how proper application of a certain theorem in this section would have shown you that $\rho(X, Y) = -1$.

4. A fair coin is tossed three times. Let X denote the number of heads that appear in the first trial, let Y denote the number of times a head comes up in the first two tosses, and let Z denote the number of times the coin comes up heads in all three trials.

(i) Find the densities of X, of Y, of Z, of XZ and of YZ.

(ii) Compute $\rho(X, Z)$ and $\rho(X, Z)$.

5. In the proof of theorem 1, we noted that since X and Y are independent, then for each n, X and Y_n^+ are independent, and X and Y_n^- are independent. Prove these two statements.

5.4 Limit Theorems. Limit theorems are among the most treasured results of probability and statistics. We have already encountered our first limit theorem, Bernoulli's theorem, which provided us with an interpretation of probability of an event and justified the method used in permutation tests. In this section we obtain a generalization of Bernoulli's theorem, the law of large numbers, which is crucial in statistical inference and at the very basis of the insurance industry.

Theorem 1. (Law of Large Numbers) *If* $\{X_1, X_2, \cdots, X_n, \cdots\}$ *is a sequence of independent and identically distributed random variables with finite common expectation* μ and *finite common second moment, then for every* $\epsilon > 0$*, the following is true:*

$$P\left(\left[\mid \frac{1}{n}\sum_{j=1}^{n} X_j - \mu) \mid \geq \epsilon\right]\right) \to 0 \text{ as } n \to \infty \ .$$

Proof: Let us denote

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \ ,$$

and let $\sigma^2 > 0$ denote the common variance of the random variables. It is easy to verify that $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Thus by Chebishev's inequality proved in section 4.1,

$$0 \leq P\left(\left[\mid \frac{1}{n}\sum_{j=1}^{n} X_j - \mu \mid \geq \epsilon\right]\right) = P\left(\left[\mid \overline{X}_n - E(X) \mid \geq \epsilon\right]\right) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$$

as $n \to \infty$, which proves the theorem.

In popular discourse, the Law of Large Numbers is usually referred to as the "law of averages".

We shall include here two further limit theorems and a practical addendum. The first is called the *Central Limit Theorem*, whose proof is beyond the scope of this course. The second theorem is the Laplace DeMoivre theorem that is most frequently used in connection with binomial probabilities.

Theorem 2. *(Central Limit Theorem) If $\{X_1, X_2, \cdots, X_n, \cdots\}$ is a sequence of independent and identically distributed random variables with finite common second moment, then*

$$lim_{n\to\infty} P\left(\left[\frac{\sum_{j=1}^{n}(X_j - E(X_j))}{\sqrt{nVar(X_1)}} \leq x\right]\right) = \Phi(x)$$

for all real x, where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

Theorem 3. (Laplace-DeMoivre theorem) *If $\{S_n\}$ is a sequence of random variables, and if, for some $p \in (0,1)$, S_n is $Bin(n,p)$* for all *n, then*

$$lim_{n\to\infty} P\left(\left[\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right]\right) = \Phi(x),$$

where $\Phi(x)$ is the same as in theorem 2.

Proof: If we let $\{X_n\}$ denote the outcomes of a sequence of independent plays of a game where $X_n = 1$ if a certain event of probability p occurs and $= 0$ if the event does not occur, then S_n has the same distribution as $\sum_{j=1}^{n} X_j$, from which we obtain $E(S_n) = np$ and $Var(S_n) = np(1-p)$. The conclusion of this theorem follows as a corollary to theorem 2.

We may insert a practical addendum here. Suppose one has on hand the number of successes, S_n, of n Bernoulli trials, for which the value of n is known and the value of S_n is observed, and one wishes to use these data to estimate the unknown value of p. If n is reasonably large, the Laplace-DeMoivre theorem may be used as follows. Find from the tables the value of x, call it x_0, such that $\Phi(x_0) = .975$. (To save you the trouble, let me give it to you; it is 1.96.) Conditions of symmetry imply that $\Phi(-x_0) = .025$, and thus $\Phi(x_0) - \Phi(-x_0) = .95$. By the Laplace-DeMoivre theorem, for large values of n, the probability of the event

$$\left[-x_0 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq x_0\right]$$

is very close in value to .95. This event is the same as the event

$$\left[\frac{S_n}{n} - x_0\sqrt{p(1-p)/n} \leq p \leq \frac{S_n}{n} + x_0\sqrt{p(1-p)/n}\right].$$

Since $\sqrt{p(1-p)/n} \leq \sqrt{\frac{1}{4n}}$, this last event implies the event

$$\left[\frac{S_n}{n} - x_0\sqrt{\frac{1}{4n}} \leq p \leq \frac{S_n}{n} + x_0\sqrt{\frac{1}{4n}}\right].$$

Thus we know that for large values of n, this last inequality supplies us with what is called a 95% *confidence interval* for p, since this last event has a probability that is approximately equal to or greater than .95.

Let us nail down an interpretation of what has been done above. If someone has the number of successes S_n in n trials and wishes to know what are the possible values of p, the statistician will use the above formulas and will tell the person that the value of p is any number between $\frac{S_n}{n} - x_0\sqrt{\frac{1}{4n}}$ and $\frac{S_n}{n} + x_0\sqrt{\frac{1}{4n}}$. What this means, according to Bernoulli's theorem given above, is that among the many independent projects in your statistical career that you use these formulas for left and right endpoints for possible values of p, in roughly 95% or more of the cases you will be correct and in approximately 5% or fewer of the cases the values of the unknown p will not be included in the interval. (And you may never know in which cases you were wrong.) What it *does not* mean is that you have 95% confidence in a particular situation that the interval you supply contains the unknown value of p, unless you define this expression as being the interpretation given before it.

EXERCISES

1. In a hotly fought election campaign, a polling agency took a random sample of voters of size $1,600$. The results were that 830 were for candidate A. Find a 95% confidence interval for the proportion of voters who intended to vote for candidate A.

2. Another polling agency looked at the data in problem 1. They were concerned with this problem. Should they announce that candidate A was in a dead heat with his opponent, or should they predict that candidate A will win? Thus they wanted to test the null hypothesis that the proportion p of voters for candidate A is 0.5 against the alternative that $p > 0.5$. So they would like to determine $P([S_{1600} \geq 830])$ when $p = 0.5$. If this probability is small, say, less than 0.01, then they are willing to state that A will win. What will they do?

3. Yet another polling agency selects their random sample in such a way that they encounter considerable expense. So they wish to minimize the size of the sample. If the voter sentiment is in a dead heat, they do not wish to say that candidate A will win, so they wish to limit the probability of making this error to be no greater than 0.05. However, if the true proportion of voters for candidate A is at least 0.55, they wish to be able to ascertain this with large probability, say, 0.90. So how large a sample of voters should they take, and under what circumstance will they state that candidate A is leading?