CHAPTER TWO

RANDOM VARIABLES AND THEIR PROBABILITY DISTRIBUTION

2.1 Random Variables

Suppose we consider a coin flip. P(heads) = 1/2; P(tails) = 1/2. We have assigned a number (1/2) to each of two points (heads and tails) of a sample space. A random variable can take on finitely or countably infinitely many values or even a continuous set of values. If a set X is countably infinite, this means that a 1 to 1 correspondence can be made between the elements of (1, 2, 3,...) and the elements of X. For example, consider X = (1/2, 1/4, 1/8, ...)

$$1 \to (1/2)^1$$

$$2 \to (1/2)^2 = 1/4$$

$$3 \to (1/2)^3 = 1/8$$

We can easily see that there is a 1-1 correspondence between the two sets.

Sets that are either finite or countably infinite are called ***discrete***.

Example: A = {3, 4, 6} is a discrete set. It has three elements. B = {2, 4, 6, ...} is discrete. It can be paired with the elements of X = {1, 2, 3, ...} .

If X is a ***continuous*** set, e.g., the set of all numbers between 0 and 1, we can't establish a 1-1 correspondence between X and any set that is countably infinite.

Sets such as A = {1, 2, 3,...}, B = {2, 4, 6, ...} , and C = {1/2, 2/2, 3/2, ...} are countably infinite since their elements can be paired with the elements of (1, 2, 3,...). No countably infinite set can ever be paired with any continuous set, even such a seemingly simple set as the set of real numbers X between 1/10 and 1/5. Sets that are continuous require calculus to define associated measures of probability. To illustrate, consider your chance of getting to the office precisely on time - 8 A.M. Have you ever arrived at exactly 8 o'clock? It is very unlikely. Even if the clock's second hand hit 8 A.M. as you walked through the portals of your office, you would have to rely on your faulty vision and an office clock - both lacking in exactness. Also, what level of accuracy are we talking about? Exactness means 100% accuracy, and no measurement that we mortals take qualifies. Consequently, the chances of arriving at a point in a continuous probability space is 0. But with the calculus we can shift measurement to an interval, e.g., 7:55 A.M. to 8:05 A.M. Since you are probably a conscientious worker, the odds are 100% or close that you arrive for work within this interval. Whenever we are measuring the probability that X lies between two values, we are in the domain of ***continuous*** sets and their associated probability distributions.

EXERCISES 2.1

1. Give an example of a discrete set with four elements.

2. Are the following sets discrete or continuous? Explain.

   a) {a, b, e, 1, 2, 3}

   b) {5, 10, 15, ...}

   c) {1/3, 2/3, 3/3, 4/3, ...}

   d) the set of integers between 1 and 100

   e) the set of numbers between 6 and 7

3. Prove that the set of numbers between 0 an 1 is continuous, i.e., cannot be placed in a 1-1 correspondence with 1, 2, 3, ...

   Hint: This is a famous modern proof that is due to the genius of Cantor. First assume that such a correspondence is possible:

   $1 \rightarrow .\, a_1\, a_2\, a_3$

   $2 \rightarrow .\, b_1\, b_2\, b_3$

   $3 \rightarrow .\, c_1\, c_2\, c_3$

   .
   .
   .

   Form a number $.x_1\, x_2\, x_3$ where $x_1 \neq a_1$, $x_2 \neq b_2$, $x_3 \neq c_3$, , etc. Now you can complete this ingenious proof.

2.2 Probability Distribution

We want to assign a probability to event X. If X is a point in a continuous sample space, P(X) = 0. Therefore, we need to use calculus to develop probability through summing P(X) on an interval. In the discrete case, we can simply sum the probabilities of the points that we wish to consider.

The distribution function accumulates the probability associated with random variable X and is defined:

$$F(X) = P(X \leq x). \; -\infty < x < \infty$$

F(X) is the probability that X will take on values less than or equal to x.

Example (Discrete): Suppose we spin a wheel that has numbers from 1-100. The probability that X lands on a number less than or equal to 75 is 75/100 = 3/4. In the language of probability, $F(75) = P(X \leq 75) = 3/4$.

Example (Continuous): Consider a bank that opens at 8 A.M. and closes at 4 P.M. An efficiency expert will pop in at a random time to check on the functioning of the bank. Find the probability that he/she arrives on or before noon. Suppose the chance that the expert coming between any equal interval of time is the same throughout the day. This means that the chance of her arriving between 8:30 A.M. and 9 A.M. is the same as her arriving between 3 P.M. and 3:30 P.M. Also consider the probability of her arriving between 8 A.M. and 12 noon. This would be four times the probability of her arriving between 8 A.M. and 9 A.M.

These common and specific assumptions lead us to a probability distribution called the ***uniform density***. To find the probability associated with continuous problems we use the calculus. We define the distribution function F(x) as follows:

$$F(x) = P(X \le x) = P(-\infty < X \le x)$$

$$\int_{-\infty}^{x} f(u)\ du$$

For the example of the efficiency expert, we need f(u) to enable us to calculate the required probabilities. Since there is no chance of her arriving earlier than 8 A.M. or later than 4 P.M., we define f(x) = 0 at any point corresponding to a time she cannot arrive.

For convenience, let 8 A.M. = 0(hrs.); let 4 P.M. = 8(hrs.). The ***uniform density*** is defined as follows:

$$f(x) = \frac{1}{b-a}\ \text{, where } a = 0,\ b = 8.$$

$$f(x) = \int_{0}^{x} \frac{i}{8}\ du = P(x \le x) = \frac{x}{8}$$

We define $f(x) = 0$ if $x > 8$

$f(x) = 0$ if $x < 0$.

Now we can determine the probability that the efficiency expert will arrive on any interval. To illustrate, consider determining the probability that she will arrive before noon. The associated distribution function is:

$$F(4) = \int_{-\infty}^{4} \frac{1}{8}\ dx = \int_{-\infty}^{0} 0\ dx + \int_{0}^{4} \frac{1}{8}\ du$$

$$= 0 + \frac{1}{8}\ x\ \Big|_{0}^{4} = 1/2$$

Therefore, F(4) = 1/2. Of course, you could intuit the answer by common sense, but common sense is sometimes wrong in mathematics.

In the previous example, we have shown the relationship that exists between the density function f(x) (sometimes called the ***probability mass function***) and the distribution function F(x).

Discrete Case $F(x) = P(X \le x) = \Sigma f(u) \qquad u \le x$

Continuous Case $F(x) = \int_{-\infty}^{x} f(u)\, du$

Let us look at two more examples of probability distributions.

Example: Discrete Case - Binomial Distribution

Although technically this should be called the ***binomial density function***, consider applications of the binomial distribution:

$$P(x) = \binom{n}{x} P^x (1 - P)^{n-x} \quad x = 0, 1, 2, \ldots n$$

where P = probability of success on a single trial, x is the number of successes in n independent trials, and

$$\binom{n}{x} = \frac{n!}{(n-x)!\, x!}$$

We assume that P stays the same for each trial. With a deck of cards, this means that each card that is picked would be replaced.

To find the probability of getting exactly three heads in four tosses of a fair coin, we proceed as follows:

P = probability of a head in a single toss

Therefore, P = 1/2. The number of successes you are asking for, x, is 3. The total number of trials, n, is 4. By substituting in the above formula,

$$P(3) = \binom{4}{3} (1/2)^3 (1 - 1/2)^2 = \frac{4!}{(4-3)!\,3!} (1/8)(1/2)$$

$$P(3) = 4 \bullet 1/16 = 1/4$$

Example: <u>Continuous Case - Exponential Distribution</u>

We will postpone discussion of how we select a distribution to fit data from the real world until a later chapter. However, the exponential distribution is a very useful tool in giving us measures such as the times that customers or planes enter systems such as banks or flight queues.

The ***<u>exponential density function</u>*** is defined by:

$$f(x) = \frac{1}{B} e^{-x/B} \quad \text{if } x \geq 0$$

$$0 \text{ elsewhere, where } B > 0.$$

The corresponding distribution function:

$$F(x) = P(X \leq x) = \int_0^x \frac{1}{B} e^{-x/B}\, dt = 1 - e^{-t/B}$$

Verify this for exercise #1.

To illustrate, let B = 3. We will discuss how to calculate B in the next chapter.

$$F(x) = \int_0^x (1/3)\ e^{-t/3}\ dt$$

$$F(2) = \int_0^2 (1/3)\ e^{-t/3}\ dt = e^{-t/3} \Big|_2^0$$

$$= e^{-2/3} - [-1]$$

$$F(2) = 1 - e^{-2/3}$$

F(2) would represent the probability that an event, like a customer entering a bank, would occur between the time corresponding to t = 0 and the time corresponding to t = 2.

$F(3) \;= 1 - e^{-3/3} = 1 - e^{-1}$

$F(3) - F(2) = (1 - e^{-1}) - (1 - e^{-2/3}) = e^{-2/3} - e^{-1}$

F(3) - F(2) represents the probability that an event would occur between the time corresponding to t = 2 and time t = 3.

<u>Properties</u>

If f(x) is a density function:

<u>Discrete Case</u> 1) $f(x) \geq 0$

2) $\Sigma f(x) = 1$

<u>Continuous Case</u> 1) $f(x) \geq 0$

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

3) $P(a < x < b) = \int_{a}^{b} f(x)\, dx$

Example: <u>Discrete</u>

Let us verify that the probabilities f(x) associated with a die throw x is a density function.

$f(1) = f(2) = f(3) = f(4) = f(5) = f(6) = 1/6$

Therefore, $f(x) \geq 0$ and $\sum_{1,2,3,\ldots 6} x = \sum_{1,2,3,\ldots 6} 1/6 = 1$

Example: <u>Continuous</u>

To show that $f(x) = 1/2\ e^{-x/2}$ for $x \geq 0$

$= 0$ for $x < 0$

is a density function for a continuous random variable:

$$f(x) = (1/2)\ e^{-x/2} \geq 0 \quad \forall\ x \geq 0$$

$$\int_{-\infty}^{\infty} (1/2)\ e^{-x/2}\ dx = \int_{0}^{\infty} (1/2)\, e^{-x/2}\ dx$$

$$= -\ \int e^{-x/2}\,(-1/2)\,dx = -\ e^{-x/2} \Big|_{0}^{\infty}$$

$$= 0 - (-1) = 1$$

EXERCISES 2.2

1. a) Verify that $f(x) = 1/8 \quad 0 < x < 8$

$= 0$ elsewhere

is a continuous probability density function.

b) Verify that $\int_0^{\infty} (1/B)^{-t/B}\, dt = 1 - e^{-x/B}$

2. Find the value of the constant c so that

$$f(x) = \frac{cx^2}{2} \quad 0 < x < 5$$

$$= 0 \quad \text{elsewhere}$$

is a density function. Then calculate $P(1 < x < 4)$.

3. a) Let $f(x) = \dfrac{1}{\pi(x^2+1)}$ $\quad -\infty < x < \infty$. Show that f(x) is a density function

This means that one should show that $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

b) Find the distribution function corresponding to the previous density function.

4. In an "ESP" Zeno deck there are five different symbols - five cards of each. Calculate the probability of obtaining four correct guesses out of 210 trials [with replacement after each trial].

2.3 Joint Distributions

The previous ideas are naturally generalizable to two or more variables. Consider the case of two variables where we define the ***joint probability function*** of x and y as follows:

Discrete Case: If x, y are discrete random variables, the joint probability function of x and y is defined by:

$$f(x,y) = P(X = x, Y = y)$$

We know: 1) $f(x,y) \geq 0$

2) $\sum_x \sum_y f(x,y) = 1$

Consider the joint probability function associated with rolling two dice. Let $x_1 = 1$, $x_2 = 2$, $x_3 = 3$. Let $y_1 = 1$, $y_2 = 2$, $y_3 = 3$. The probability that $x_1 = 2$ and $y_1 = 1$ equals 1/36. This would be written $f(1, 1) = 1/36$. Our table below would have 36 probabilities f(x, y), each equal to 1/36.

x \ y	y_1	y_2	y_3
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	$f(x_1,y_3)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$f(x_2,y_3)$
x_3	$f(x_3,y_1)$	$f(x_3,y_2)$	$f(x_3,y_3)$

For the joint probability table above $P(X = x_1, Y = y_1) = f(x_1, y_1) \bullet P(X = x_3, Y = y_3) = f(x_3, y_3)$. Of course, x and y could assume a countably infinite or finite set of values.

The ***marginal probability function*** of x_i means that you are summing the row probabilities for the ith row. This gives you the probability that $x = x_i$. The formula is

$$P(X = x_i) = f_x(x_i) = \sum_{k=1}^{n} f(x_i, y_k) \quad i = 1,2,\ldots m; \quad k = 1,2,\ldots n.$$

In the above table there are three rows so i = 1,2,3. There are three columns so k = 1,2,3. If we created a joint probability table for the rolling of two dice, we would have six rows (i = 1, 2, 3, 4, 5, 6) and six columns (k = 1, 2, 3, 4, 5, 6).

The corresponding formula for the marginal probability function for $Y = y_k$ is obtained by adding the probabilities in the kth column. The formula is:

$$P(Y = y_k) = f_y(y_k) = \sum_{i=1}^{n} f(x_i, y_k)$$

Consider the following specific example:

x \ y	$y_1 = 1$	$y_2 = 2$	$y_3 = 3$	Total
$x_1 = 1$	1/40	1/20	1/40	1/10
$x_2 = 2$	1/40	1/10	1/40	3/20
$x_3 = 3$	1/20	1/40	27/40	3/4
Total	1/40	7/40	29/40	1

$$P(X = x_2) = \sum_{k=1}^{3} f(x_2, y_k) = 1/40 + 1/10 + 1/40 = 3/20$$

$$P(Y = y_3) = \sum_{i=1}^{3} f(x_i, y_3) = 1/40 + 1/40 + 27/40 = 29/40$$

Note that the totals of the respective marginal probability functions must equal 1. We write this as:

The ***<u>Marginal Probability Function of x</u>*** $= f_x(x) = \sum_{\text{all } y} f(x,y) = 1.$

The ***<u>Marginal Probability Funciton of y</u>*** $= f_y(y) = \sum_{\text{all } x} f(x,y) = 1.$

The ***joint distribution function*** $F(x,y) = P(X \le x, Y \le y)$. For example, $F(2, 2) = 1/40 + 1/20 + 1/40 + 1/10$. We are summing all entries for which $x_i \le 2$ and $y_k \le 2$.

Continuous Case

A natural extension can be made to the continuous case.

$$P(c < x < d,\ e < y < f) = \int_e^f \int_c^d f(x, y)\ dx\,dy$$

$f(x, y)$ is called the **joint density function** of x and y.

The corresponding properties also hold for continuous random variables:

1) $f(x,y) \ge 0$

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

The **joint distribution function** of x and y is defined by:

$$F(x, y) = P(X \le x, Y \le y) = \int_{u=-\infty}^{y} \int_{v=-\infty}^{x} f(u, v)\ du\ dv$$

It follows that:

$$\frac{\partial^2 F}{\partial x\, \partial y} = f(x, y)$$

The **marginal distribution functions** of x and y are defined as:

(1) $$P(X \le x) = F_X(x) = \int_{u=-\infty}^{x} \left[\int_{v=-\infty}^{\infty} f(u,v)\, dv \right] du$$

Integrate the innermost integral first - then the outermost. If one has constant limits, one can integrate in either order, but be careful with having the limits correspond to the variable of integration.

Similarly,

$$\text{(2)} \qquad P(Y \le y) = F_y(y) = \int_{u=-\infty}^{y} \left[\int_{v=-\infty}^{\infty} f(u,v)\, du \right] dv$$

Sometimes, $F_x(x)$ and $F_y(y)$ are simply called the distribution functions of x and y respectively.

We can take the derivative of (1) with respect to x and (2) with respect to y to obtain the ***marginal density function of x and y***. These are written as:

$$f_x(x) = \int_{v=-\infty}^{\infty} f(x,v)\, dv \text{ and } f_y(y) = \int_{u=-\infty}^{\infty} f(u,y)\, du$$

To illustrate these concepts, let us examine the joint density.

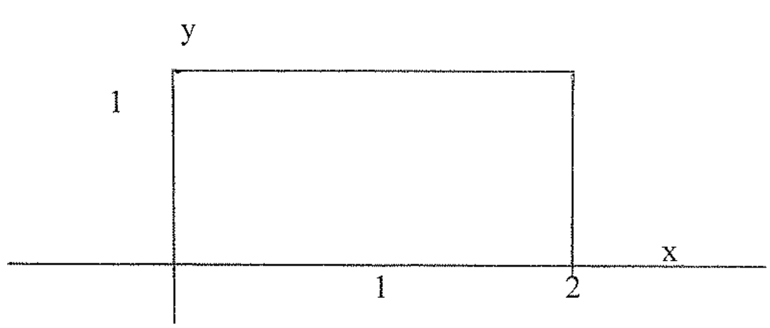
$$f(x,y) = c[x + y] \quad 0 < x < 2, \ 0 < y < 1$$
$$= 0 \ \text{ elsewhere}$$

To find the value of c, use the property of joint density functions

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

Since $f(x, y) = 0$, outside of the rectangle defined by $0 < x < 2, \ 0 < y < 1$,

we know $\int_0^1 \int_0^2 c(x + y)\, dx\, dy = 1$



$$c \int_0^1 \int_0^2 (x + y)\ dx\, dy = 1$$

$$c \int_0^1 \left. \frac{x^2}{2} + xy \right|_0^2 dy = 1$$

$$c \int_0^1 (2 + 2y)\, dy = 1$$

$$c[2y + y^2 \Big|_0^1] = 1$$

$$c[2 + 1] = 1 \ \rightarrow c = 1/3$$

Therefore, $f(x,y) = 1/3\ (x + y)\quad 0 < x < 2,\ 0 < y < 1$

$= 0$ elsewhere

$$f_X(x) = \int_{-\infty}^{\infty} (x,y)\, dy$$

For our example, $f_X(x) = \int_{-\infty}^{\infty} 1/3\ (x + y)\ dy$

$$f_X(x) = \int_0^1 1/3\ (x + y)\, dy$$

$$f_X(x) = 1/3\ [xy + \frac{y^2}{2} \Big|_0^1]$$

$$f_X(x) = 1/3\ [x + 1/2]$$

EXERCISES 2.3

1. Is the following a discrete joint density function?

$x \backslash y$	y_1	y_2
x_1	1/2	1/2
x_2	3/4	1/4

2. Verify that $f(x,y) = 1/3\ (x + y)\ 0 < x < 2,\ 0 < y < 1$ is a continuous joint density function.

3. a) Compute $f_y(y)$ for

$$f(x,y) = 1/3\ (x + y);\quad 0 < x < 2,\quad 0 < y < 1$$
$$= 0 \quad \text{elsewhere}$$

b) Compute the probability that $0 < x < 1,\ 0 < y < 1/2$ for the above probability density function.

4. a) Compute c for the joint density function

$$f(x,y) = c\, x^2y^2;\quad 0 < x < 1, 0 < y < 1$$

b) Compute $f_x(x),\ f_y(y)$.

2.4 Independence

Independence is one of the thorniest problems in mathematical statistics. Most of statistics, including the ***central limit theorem***, is based upon independence assumptions.

The ***central limit theorem***, which you will cover in depth in a full year sequence of probability and statistics with calculus, is the basis for the confidence intervals and hypothesis tests that were part of Chapter One. An elementary statement of this extraordinary and broadly generalizable theorem is:

In selecting random samples of size n from a population with mean μ and standard deviation σ, the sampling distribution of x approaches a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ as $n \to \infty$. (If $n \geq 30$, the sampling distribution can be approximated by a normal probability distribution.)

Using calculus notation, we state the central limit as follows:

Let $x_1, x_2, \ldots$ be independent random variables which are identically distributed with finite mean μ and variance σ^2. Let $x_1 + x_2 + \ldots + x_n = S_n$.

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma/\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dz$$

The computation of the integral requires numerical analysis - another highly important advanced mathematics course. The assumption of identical distribution means that there is an underlying formula that models and reflects the values that are obtained from the sample. We will develop later in the book the commonly used probability distributions and the modeling techniques that determine which one to use.

Please note that the central limit theorem as it is nearly always stated requires that the events (the random variables) are independent. We know little about confidence intervals and hypothesis testing if the variables are not independent. To simplify matters, we usually assume that the random variables are independent. If X and Y are independent events,

$$P(X \cap Y) = P(X) \cdot P(Y)$$

This notation means that whenever X and Y are independent, the probability of X and Y occurring equals the product of the probability of X and the probability of Y.

Discrete Case

Let X and Y represent the result of a coin toss on the first (X) and second (Y) flip of a fair coin. X and Y are independent events. Therefore, P(X=head, Y=tail) = P(X=head) • P(Y=tail) = 1/2 • 1/2 = 1/4.

Consider a graduate modeling and simulation class with ten women and six men. Suppose we select two students at random and determine the probability that both are women. The two events are only independent if we allow the possibility that the same person can be picked twice. Otherwise, P(2 consecutive women) = 10/16 • 9/15 $\neq$ 10/16 • 10/16. We conclude that the two events are not independent.

The ***binomial probability density***

$$P(X) = \binom{n}{x} p^n (1 - p)^{n-x}, \quad x = 0,1,\ldots n$$

is used to calculate the probability of x successes in n trials. It can only be used if each trial is independent and p, the probability of success, is the same for each trial.

Example 1: To calculate the probability of obtaining seven heads in ten coin flips:

$$P(X=7) = \binom{10}{7} (1/2)^7 (1\text{-}1/2)^3 = .1172$$

(One would use the table for the binomial distribution to calculate this probability with n=10, X=7, p=.05)

Example 2: Consider the joint probability density table below. If x and y are independent, then $f(x,y) = f_x(x) \cdot f_y(y)$. To check independence, consider the following:

$f(3,3) = 1/16$

$f_x(3) = 5/32$

$f_y(3) = 13/64$

$f(x,y) = f(3,3) = 1/16 \neq 5/32 \cdot 13/64$

Therefore, x and y are not independent.

x \ y	1	2	3	4	Total
1	1/32	1/16	1/16	1/32	3/16
2	1/8	1/16	1/16	1/16	5/16
3	1/64	3/64	1/16	1/32	5/32
4	1/16	1/64	1/64	1/4	11/32
Totals	15/64	3/16	13/64	3/8	1

Continuous Case

If x and y are continous random variables, we can extend the formula from the discrete case as follows. For independent random variables, $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$.

This leads to the marginal distribution functions $F_x(X)$, $F_y(Y)$ and the equivalent formula for independent continuous random variables:

$$F(X,Y) = F_x(X) \cdot F_y(Y)$$

The converse of this result holds - namely if $F_x(X) \bullet F_y(Y) = F(X,Y)$ X,Y, then X and Y are independent events. If $F_x(X) \bullet F_y(Y) \neq F(X,Y)$ for any value of (X,Y), then X and Y are said to be **<u>dependent</u>** events.

Consider the joint probability function that is represented by the table below:

		x			
		0	1	2	
	0	1/16	0	1/16	2/16
y	1	0	1/2	1/16	9/16
	2	1/16	1/4	0	5/16
		2/16	3/4	2/16	

Let us consider the value of the joint probability function, $f(x,y) = f(1,1)$. The previous result - if $F_x(X) \bullet F_y(Y) = F(X,Y)$, then X and Y are independent events - extends to the marginal distributions of x and y. This could be written as follows: if $f(x_1, x_2, \ldots x_n)$ represents the joint distribution of n random variables $x_1, x_2, \ldots x_n$ and $f_i(x_i)$ represents the marginal distribution of the random variable x_i, then the n random variables are independent if and only if $f(x_1, x_2, \ldots x_n)$ $= \prod_{i=1}^{n} f_i(x_i)$. The use of the phrase if any only if means that both the original statement and its converse are true. The notation $\prod$ means the product $f_1(x_1) \bullet f_2(x_2) \bullet f_3(x_3) \bullet \ldots f_n(x_n)$.

In the example above, we have two variables. Therefore, we should let $x_1 = x$, $x_2 = y$.

We want to check if $f(1,1) = \underset{x=1}{f(x)} \bullet \underset{y=1}{f(y)}$. We see $\underset{x=1}{f(x)} = 0 + 1/2 + 1/4 = 3/4$. It follows that

$\underset{y=1}{f(y)} = 9/16$. $f(1,1) = 1/2 \neq 3/4 \bullet 9/16 = \underset{x=1}{f(1)} \bullet \underset{y=1}{f(1)}$. Therefore, we conclude that x and y are

not independent.

Example 3: Suppose x and y are continuous random variables with joint density

function: $f(x,y) = cxy, \; 0 < x < 2, \; 0 < y < 1.$

$= 0$ elsewhere

To find c,

$$\int_0^2 \int_0^1 cxy \, dy \, dx = 1 \rightarrow \int_0^2 \left(cx \left.\frac{y^2}{2}\right|_0^1\right) dx$$

$$\rightarrow c \int_0^2 x\,[1/2]\, dx$$

$$\rightarrow \frac{c}{2} \left.\frac{x^2}{2}\right|_0^2 = 1 \qquad \rightarrow c = 1$$

Therefore, $f(x,y) = xy, \quad 0 < x < 2, 0 < y < 1$

$= 0$ elsewhere

$$F_x(x) = \int_0^x \int_0^1 xy \, dy \, dx$$

$$= \int_0^x x \left.\frac{y^2}{2}\right|_0^1 dx$$

$$\rightarrow F_x(x) = \int_0^x \frac{x}{2} \, dx = \frac{x^2}{4}$$

$$F_y(y) = \int_0^y \int_0^2 xy \, dx \, dy$$

$$= \int_0^y \frac{x^2}{2} \; y \Big|_0^2 \; dy$$

$$F_y(y) = \int_0^y 2y \, dy = y^2$$

Now we have:

$$(1) \qquad F_x(x) = x^2 / 4, \;\; F_y(y) = y^2$$

To check for independence, consider

$$F(1,1/2) \qquad = \qquad \int_0^1 \int_0^{½} xy \, dy \, dx = \int_0^1 \frac{xy^2}{2} \Big|_0^{½} \; dx$$

$$= \int_0^1 1/8 \; x \, dx \qquad = \qquad \int_0^1 x/8 \, dx = 1/16$$

We can substitute in (1) and find for (1,1/2), $F_x(1) = 1/4$, $F_y(1/2) = 1/4$. $F_x(1) \cdot F_y(1/2) = 1/4 \cdot 1/4$ $= 1/16 = F(1,1/2)$

From this example, we can conclude little. But if $F_x(1) \cdot F_x(1/2) \neq F(1,1/2)$, we can conclude that x and y were not independent random variables.

Consider the general result that if x and y are independent, $F(x,y) = F_x(x) \cdot F_y(y)$.

$$F(x,y) = \qquad \int_0^x \int_0^y xy \; dy \; dx$$

$$= \qquad \int_0^x \frac{xy^2}{2} \Big|_0^y dx \;\; = \;\; \int_0^x \frac{xy^2}{2} \; dx$$

$$F(x,y) = \qquad \frac{x^2y^2}{4}$$

We know that $F_x(x) \cdot F_y(y) = \frac{x^2}{4} \cdot y^2$

From this calculus exercise, the result proves the independence of random variables x and y for this example.

To glimpse the complexity of independence, inspect problem 6 of Exercises 2.6 later in this chapter. This illustrates the problems with the intuitively appealing but false statement if x and y have correlation 0, then x and y are independent.

Dependence is a nightmare for the theory of statistics. The major results in statistics rely upon the central limit theorem and the central limit theorem is nearly always stated with independence assumptions. In mathematical modeling we typically assume that the random variables that we are studying are independent. This simplifies matters greatly and allows us to proceed. However, in using real-world samples, pure independence is rare.

To learn about recent attempts in mathematical modeling to address the consequences of dependence in one sample, please read "Dependent Random Variables and the Central Limit Theorem," written by this author and Tim Sheehan (a graduate student from Iona College). This article has been reprinted later in this book together with commentary. The commentary and suggestions should hopefully lead advanced undergraduates to participate in current research. We thank Pergamon Press and the *International Journal for Mathematical and Computer Modeling* for generously granting permission for us to include this article in this book.

EXERCISES 2.4

1. If you play blackjack poorly and give the house a 60%-40% edge each hand, what are your chances of winning 5 hands out of 10? (You have a probability of 40% of winning each hand.)

2. Let x and y be discrete random variables with joint probability function:

   $f(x,y) = cxy, x = 1,2,3; \quad y = 1,2,3$

   a) Find c

   b) Find $F_x(x)$

   c) Find $F_y(y)$

   d) Does $F(x,y) = F_x(x) \bullet F_y(y)$? Are x and y independent?

3. Let x and y be continuous random variables with joint density function:

   $$f(x,y) = \begin{cases} cxy & 0 < x < 5, \; 1 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

   a) Find c

   b) Find $F_x(x)$

   c) Find $F_y(y)$

   d) Does $F(x,y) = F_x(x) \bullet F_y(y)$? Are x and y independent?

4. For exercise 3 above, find $P(x+y < 2)$.

5. Let x be a continuous random variable with probability density function

   $f(x) = x^2, \; 0 \leq x \leq b$

   a) Find b

   b) Plot f(x)

c) Calculate $P(0 \leq x \leq 1)$

2.5 Mathematical Expectation

Mathematical expectation, also called ***expected value***, is a central concept in statistics. To illustrate the idea of expected value, suppose a coin is flipped. If the result is a head, you win \$10; if the result is a tail, you lose \$10. A gambling casino would go broke with this game because the expected value of the coin flip is 0, meaning that the casino would expect a profit of zero.

We define ***expected value*** as follows:

$$\text{Expected value} = E(x) = \sum_{i=1}^{n} x_i \bullet F(x_i)$$

$$= x_1 \bullet P(x_1) + x_2 \bullet P(x_2) + \ldots + x_n \bullet P(x_n)$$

In our example there are two possibilities: $x_1 = 10$ if there is a head and $x_2 = (-10)$ if there is a tail. $P(x_1) = 1/2$ and $P(x_2) = 1/2$. Therefore, $E(x) = \sum_{i=1}^{2} x_i \bullet P(x_i) = 10 \bullet 1/2 + (-10) \bullet 1/2 = 0$. This is the key to a fair game, an expectation of zero. But most gambling ventures have high negative expectation. That is why most people lose.

If each event $x_1, x_2, \ldots x_n$ is equally probable, $P(x_1) = 1/n$. This leads to the result:

$E(x) = \dfrac{x_1 + x_2 + \ldots x_n}{n}$ or the arithmetic mean of the values x_i.

The expectation of x is also written μ_x or μ and represents a weighted average of the value of x.

Consider the example of "over, under and seven" - a common game at fundraisers. Two dice are rolled, and you can bet over (8,9,10,11,12), under (2,3,4,5,6) or seven. If you win on over or under, you win the amount you bet. You win four times your money if you play and win on 7. If one plays \$1 on over, the expected value is computed as follows:

$$P(\text{seven}) = 6/36 = 1/6$$

$$P(\text{over}) + P(\text{under}) = 1 - 1/6 = 5/6$$

$$P(\text{over}) = P(\text{under}) = 5/12$$

$$E(x) = (1/6)(-1) + (5/12)(1) + (5/12)(-1)$$

$$E(x) = -1/6 \approx -.17$$

If you wager \$1 on seven, the expected value is:

$$E(x) = (1/6)(4) + (5/12)(-1) + (5/12)(-1)$$

$$E(x) = -1/6 \approx -.17$$

Either way, one can expect to lose 17% of the total money one wagers. Of course, you could win and walk away with your gains. But mathematical expectation enables one to determine the mean (or what one could expect) if probabilities of events and the outcomes associated with such events can be estimated.

Continuous Case

For a continuous random variable x with f(x) the associated probability density function,

$$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

This is the natural extension of the finite case:

$$E(x) = \sum_{i=1}^{n} x_i\, P(x_i)$$

For example, consider the uniform density function,

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

$$= \quad 0 \qquad \text{elsewhere}$$

This is called the ***uniform density*** and is very useful. For our example,

$$E(x) = \int_{-\infty}^{a} x\ f(x)\ dx + \int_{a}^{b} x\ f(x)\ dx + \int_{b}^{\infty} x\ f(x)\ dx$$

$$= \int_{-\infty}^{a} x \bullet 0\ dx + \int_{a}^{b} x \bullet f(x)\ dx + \int_{b}^{\infty} x \bullet 0\ dx$$

$$= 0 + \int_{a}^{b} x\ f(x)\, dx + 0$$

$$= \int_{a}^{b} x\ \frac{1}{b-a}\ dx$$

$$\frac{1}{b-a} \bullet \frac{x^2}{2} \Big|_{a}^{b} = \frac{a+b}{2}$$

Therefore, $E(x) = \dfrac{b+a}{2}$

For an application of the uniform density, consider a bank that opens at 8 A.M. and closes at 4 P.M. Let 8 A.M. = a = 0. Let 4 P.M. = b = 8 [8 hours after opening]. Let x represent the time of customer entry to the bank. For example, if a customer entered at 10 A.M., x = 2.

$$E(x) = \frac{b+a}{2} = \frac{8+0}{2} = 4$$

This means that if 100 customers entered the bank uniformly from 8 A.M. until closing at 4 P.M., and you assigned a real number from 0 - 8 to the time each customer entered, the best estimate of the average time of customer entry is 4 or 12 noon. Of course, we could as easily have let a = 8 and b = 16.

Properties of Expected Value

Some important results on expectation include:

1) If c is a real number, $E(cx) = c\,E(x)$

2) If x and y are random variables, then $E(x+y) = E(x) + E(y)$

3) If x and y are independent random variables, $E(xy) = E(x) \bullet E(y)$

4) If a and b are constants, $E(ax + b) = a\,E(x) + b$.

Generalizations to functions of two or more variables can be made. For example, if x, y are two continuous random variables with joint density function f(x,y), then the expectation of h(x,y) is given by:

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

If x,y are two discrete random variables with joint density function f(x,y), then the expectation of g(x,y) is :

$$E(g(x,y)) = \sum_{y} \sum_{x} g(x,y)\, f(x,y)$$

Countably Infinite Sample Space

Though finite and continuous sample spaces are common in modeling, a third possibility exists - the countably infinite sample space. ***Countable infinity*** means a one-to-one correspondence can be made with the natural numbers. Therefore, the set of all odd numbers, the set of all even numbers, and even the set of rational numbers can be placed in a 1-1 correspondence (or pairing) with the natural numbers and form a countably infinite sample space. To illustrate how the concept of expectation could be applied to a countably infinite sample space, consider a game where you flip a coin and continue until you obtain a head. If you get a head on the first flip, you win \$1. Otherwise if the first flip is a tail, you start counting the trials after the first tail. You pay \$x when the first head is tossed on the xth trial ($x \geq 2$). The mathematical expectation for this game could be expressed as follows:

$$E(x) = \sum (x) \cdot P(x)$$

$$= 1 \cdot 1/2 + (-2)\, 1/4 + (-3)\, 1/8 + (-4)\, 1/16 + (-5)\, 1/32 + \ldots$$

$$E(x) = -3/8 - 4/16 - 5/32 - 6/64 + \ldots$$

$$= -1\,(3/8 + 4/16 + 5/32 + 6/64 + \ldots)$$

$$= -1 \left(\sum_{n=3}^{\infty} \frac{n}{2^n} \right)$$

Ratio Test

Let us examine whether you have set up the possibility of infinite negative expectation. For determining whether an infinite series converges, the ***ratio test*** is very useful and nicely suits this example. Consider $\sum_{n=3}^{\infty} \frac{n}{2^n}$

The ratio test considers the limit:

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

For our example, $a_n = \frac{n}{2^n}$, $a_{n+1} = \frac{n+1}{2^{n+1}}$

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n\to\infty} \left| \frac{\frac{n+1}{2^{n+1}}}{\frac{n}{2^n}} \right|$$

$$= \lim_{n\to\infty} \frac{n+1}{n} \cdot 1/2 = 1/2$$

The series converges since $\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = 1/2 < 1.$

Therefore, E(x) = -1 (L), where L is some positive constant to be determined.

The exact limit was determined by Dr. Henry Ricardo in the following elegant demonstration:

To find $\sum_{n=3}^{\infty} \frac{n}{2^n}$

Step 1 For $|x| < 1, \quad \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ (1)

by properties of infinite geometric progressions.

Step 2 Differentiate both sides of equation (1):

$$\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n \; x^{n-1} \qquad (2)$$

This is justified by the uniform convergence of the series of derivatives in any interval $-1 < -r \leq x \leq r < 1$

Step 3 Multiply both sides of equation (2) by x:

$$\frac{x}{(1-x)^2} = x \sum_{n=1}^{\infty} n \; x^{n-1} = \sum_{n=1}^{\infty} n \; x^n \text{ for } |x| < 1.$$

Step 4 Let $x = 1/2$.

$$\frac{1/2}{(1-1/2)^2} = \sum_{n=1}^{\infty} \frac{n}{2^n} \quad \rightarrow \quad 2 = \sum_{n=1}^{\infty} \frac{n}{2^n}$$

Step 5 $S = \sum_{n=3}^{\infty} \frac{n}{2^n} = 2 - 1/2 - 2/4 = 1$

This game would not be a good idea for you to play, unless you are the person getting paid the \$x if the first head comes on the xth trial. However, at least no one will be in the extremely disadvantageous position of infinite loss expectation and an expectation of (-1) isn't that bad.

EXERCISES 2.5

1.

H	T
20	(-12)

Find the expected value for the game with above payoff if Dr. Persi Diaconis (Harvard) is flipping the coin and can obtain heads 52% of the time.

For this game, if a head is tossed, Dr. Diaconis wins \$20. He pays \$12 if a tail is tossed. This poor expectation is typical of many common bets such as the lottery or if one plays blackjack poorly at the casino.

2. Find the payoffs that should be normally associated with the rolls of a pair of dice if the game were fair - i.e., $E(x) = 0$. For example, $P(\text{seven}) = 1/6$. Therefore, \$6 should be paid if one plays and wins with seven.

3. Calculate $E(x)$ for the density function:

$$f(x) = 3x^2 \qquad 0 < x < 1$$
$$= 0 \qquad \text{elsewhere}$$

4. Prove the three properties of expected value.

5. Determine the mathematical expectation for a roll of dice where the house loses \$1 with over 7 (8,9,10,11,12) and wins #x with a roll of 7.

6. Prove properties 1-3 for mathematical expectation (continuous case) using calculus.

2.6 Variance, Standard Deviation, Covariance and Correlation Coefficient

We have introduced sample variance and standard deviation in the first section of the text. For the discrete case we define ***sample variance*** s^2:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - x)^2}{n - 1}$$

Remember that $\overline{x} = \sum \frac{x}{n}$ and is called the arithmetic mean.

The sample ***standard deviation*** s is simply the square root of the variance:

$$s = \sqrt{\frac{\sum_{i=1} (x_i - x)^2}{n - 1}}$$

The variance or standard deviation tells you how spread out the scores are around the mean, $\overline{x}$. We could also write the definition of variance in terms of our recently developed concept of expectation.

$$\mathrm{Var}(x) = E\,[(x - \mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2\ f(x_i)$$

In the special case where each x_i is equiprobable, we obtain the definition of population variance:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Continuous Case

We can extend the definition of variance from discrete to continuous variables quite naturally. We use the same definition:

$$\sigma^2 = E\,[\,(x - \mu)^2\,]$$

But in the continuous case, we use the integral:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \; f(x) \; dx$$

For example, consider the probability density function:

$$f(x) = 3 \; x^2 \quad 0 < x < 1$$
$$= 0 \quad \text{elsewhere}$$

$$\mu = E(x) = \int_{-\infty}^{\infty} 3 \; x^2 \; (x) \; dx = \int_{0}^{1} 3 \; x^3 \; dx$$

$$= 3 \; \frac{x^4}{4} \Big|_0^1$$

$$\mu = 3/4$$

$$\sigma^2 = \int_0^1 (x - 3/4)^2 \; (3x^2) \; dx$$

$$\sigma^2 = \int_0^1 (3x^4 - 9/2 \; x^3 + 27/16 \; x^2) \; dx$$

$$\text{variance} = \sigma^2 = \frac{3x^5}{5} - \frac{9x^4}{8} + \frac{9x^3}{16} \Big|_0^1 = 3/80$$

$$\text{standard deviation} = \sigma = \sqrt{3/80}$$

Certain Properties of Variance

Six properties of variance that may be useful are:

1) $\sigma^2 = E(x^2) - \mu^2 = E(x^2) - [E(x)]^2$

Note: $E(x) = \mu$

If x and y are independent variables, properties (2) and (3) hold:

2) $Var(x + y) = Var(x) + Var(y)$

3) $Var(x - y) = Var(x) + Var(y)$

Note that property (3) is different from what our intuition would guess. We present an outline of a proof for property (3) below:

First we must prove property (2). For this we assume that x and y are independent random variables. Let μ_x = mean for x; μ_y = mean for y.

$\text{Var}(x + y) = E(([x+y] - [\mu_x + \mu_y])^2)$

Group the terms and we obtain:

$\text{Var}(x + y) = E(((x - \mu_x) + (y - \mu_y))^2)$

Multiply and we derive:

$= E(((x - \mu_x)^2 + 2(x - \mu_x)(y - \mu_y) + (y - \mu_y)^2))$

Using the properties of expected value:

$= E((x - \mu_x)^2) + 2E((x - \mu_x)(y - \mu_y)) + E((y - \mu_y)^2)$

We know that $E((x - \mu_x)(y - \mu_y))$

$= E((x - \mu_x)) \bullet E((y - \mu_y))$

We also know that:

$E(x - \mu_x) = 0$ and $E(y - \mu_y) = 0$

Therefore $\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y)$

To complete our proof that $\text{Var}(x - y) = \text{Var}(x + y)$, use property (5) that:

$\text{Var}(cx) = c^2 \text{Var}(x)$, letting $c = -1$ and substituting as follows:

$\text{Var}(x - y) = \text{Var}(x + (-1y))$

$= \text{Var}(x) + (-1)^2 \text{Var } y = \text{Var}(x) + \text{Var}(y)$

4) We can extend (2) and (3) to n variables:

$\text{Var}(x_1 \pm x_2 \pm \ldots \pm x_n) = \text{Var}(x_1) + \text{Var}(x_2) + \ldots + \text{Var}(x_n)$

5) $\text{Var}(cx) = c^2 (\text{Var } x)$

6) $E[(x - c)^2]$ is a minimum if $c = \mu$

To illustrate property (1), consider our previous example.

$$f(x) = 3x^2, \qquad 0 < x < 1$$
$$= 0 \qquad \text{elsewhere}$$

We computed $\sigma^2 = 3/80$ by the formula $\sigma^2 = E(x - \mu)^2$

Compare the procedure using (1).

$$\sigma^2 = E(x^2) - \mu^2$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 \; f(x) \; dx$$

$$= \int_0^1 x^2 \; (3x^2) \; dx = \int_0^1 3x^4 \; dx = 3x^5 \Big|_0^1 = 3/5$$

$$\sigma^2 = E(x^2) - \mu^2$$

$$\sigma^2 = 3/5 - (3/4)^2 = 3/5 - 9/16 = 48/80 - 45/80 = 3/80$$

Covariance

We can extend these ideas to two or more variables. Let us consider the case of two continuous random variables x and y having joint density function $f(x,y)$. The means of x and y are:

$$\mu_x = E(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \; f(x,y) \; dxdy$$

$$\mu_y = E(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \; f(x,y) \; dxdy$$

The variances are written using the standard definition:

$\infty \qquad \infty$

$$\sigma_x^2 = E[(x - \mu_x)^2] = \int_{-\infty} \int_{-\infty} (x - \mu_x)^2 \; f(x,y) \; dxdy$$

$$\sigma y^2 = E[(y - \mu_y)^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_y)^2 \; f(x,y) \; dxdy$$

An important quantity that allows us to extend our earlier concept from Chapter One of correlation coefficient (discrete) to continuous variables is the covariance of x and y, σ_{xy}.

$$\text{cov}(x,y) = \sigma_{xy} = E[(x - \mu_x)(y - \mu_y)]$$

In the continuous case,

$$\sigma_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) \; f(x,y) \; dxdy$$

In the discrete case,

$$\mu_x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \; f(x_i,y_j)$$

$$\mu_y = \sum_{i=1}^{n} \sum_{j=1}^{n} y_j \; f(x_i,y_j)$$

and covariance is defined:

$$\sigma_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - \mu_x)(y_j - \mu_y) \; f(x_i,y_j)$$

Let us illustrate covariance with an example.

$$\text{Let } f(x,y) = a\,x\,y \qquad 0 < x < 1$$
$$0 < y < 1$$
$$= 0 \qquad \text{elsewhere}$$

First, to find a,

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

$$\int_0^1 \int_0^1 axy \; dxdy = a \int_0^1 x^2/2 \; y \Big|_0^1 dy$$

$$= a \int_0^1 1/2 \, y \, dy = a/2 \;\; y^2/2 \Big|_0^1 = 1.$$

so that $a/4 = 1$ and $a = 4$.

Thus, $f(x,y) = 4 \; xy \qquad 0 < x < 1$

$\qquad\qquad\qquad\qquad 0 < y < 1$

$\qquad\quad = 0 \qquad\qquad$ elsewhere

and $\mu_x = E(x) = \int_0^1 \int_0^1 x \; 4xy \; dxdy$

$$= \int_0^1 4x^3/3 \Big|_0^1 y \; dy = \int_0^1 4/3 \; y \; dy$$

$$= 4/3 \;\; y^2/2 \Big|_0^1 = 2/3$$

From observation (or calculation) we obtain $\mu_y = 2/3$, so that

$$\sigma_{xy} = \int_0^1 \int_0^1 (x - 2/3)\,(y - 2/3) \;\; 4 \;\; xy \;\; dxdy$$

$$= \int_0^1 \int_0^1 (4x^2y^2 - 8/3\,xy^2 - 8/3\;x^2y + 16/9\,xy) \;\; dxdy$$

$$= \int_0^1 4x^3/3 \;\; y^2 \Big|_0^1 dy - 8/3 \int_0^1 x^2/2 \;\; y^2 \Big|_0^1 dy \; -$$

$$8/3 \int_0^1 x^3/3 \, y \Big|_0^1 dy + 16/9 \int_0^1 x^2/2 \;\; y \Big|_0^1 dy$$

$$= 0.$$

Let us consider an example of covariance for a discrete random variable. The table below gives the joint probability function of two random variables x and y:

		x		
		0	1	2
y	0	1/6	1/3	1/6
	1	0	0	0
	2	1/6	0	1/6

$$\mu_x = E(x) = \underset{x=0}{0}\;\underset{f(x=0)}{[1/6 + 0 + 1/6]} + \underset{x=1}{1}\;\underset{f(x=1)}{[1/3 + 0 + 0]} + \underset{x=2}{2}\;\underset{f(x=2)}{[1/6 + 0 + 1/6]}$$

$$= 0 + 1/3 + 2/3 = 1$$

$\mu_x = \Sigma\, x\, f_x$ where f_x is the marginal probability function for x

$\mu_y = \Sigma\, y\, f_y$ where f_y is the marginal probability function for y.

$\mu_y = E(y) = 0\ [1/6 + 1/3 + 1/6] + 1\ [0 = 0 = 0] + 2\ [1/6 + 0 + 1/6] = 2/3$

$$E(xy) = \sum_x \sum_y xy\ f(x,y)$$

$$= \underset{x=0\ \ y=0\ \ f(0,0)}{0 \cdot 0 \cdot 1/6} + \underset{x=0\ \ y=1\ \ f(0,1)}{0 \cdot 1 \cdot 0} + \underset{x=0\ \ y=2\ \ f(0,2)}{0 \cdot 2 \cdot 1/6} + \underset{x=1\ \ y=0\ \ f(1,0)}{1 \cdot 0 \cdot 1/3}$$

$$+ \underset{x=1\ \ y=1\ \ f(1,1)}{1 \cdot 1 \cdot 0} + \underset{x=1\ \ y=2\ \ f(1,2)}{1 \cdot 2 \cdot 0} + \underset{x=2\ \ y=0\ \ f(2,0)}{2 \cdot 0 \cdot 1/6} + \underset{x=2\ \ y=1\ \ f(2,1)}{2 \cdot 1 \cdot 0}$$

$$+ \underset{x=2\ \ y=2\ \ f(2,2)}{2 \cdot 2 \cdot 1/6} = 2/3$$

Covariance of x and y $= \text{cov}(x,y) = E(xy) - E(x)\,(E(y)) = 2/3 - 1\,(2/3) = 0.$

Now correlation between x and y $\quad \text{cor}(x,y) = \dfrac{\text{cov}(xy)}{\sigma_x\ \sigma_y} = 0$

Note the covariance between x and y = 0, but x and y are not independent.

To show $f(x,y) \neq f_x \cdot g_y$

Let $x = 2, y = 2$

$f(2,2) = 1/6$

$f(x{=}2) = f_x = 1/6 + 0 + 1/6 = 1/3$

$f(y{=}2) = g_y = 1/6 + 0 + 1/6 = 1/3$

$1/6 \neq 1/3 \cdot 1/3$ x, y are not independent.

To find $\sigma_x^2 = E(x^2) - (E(x))^2$

$$E(x^2) = \sum_x \sum_y x^2 \ f(x,y)$$

$$= \underset{0^2}{x{=}0\ y{=}0} + \underset{0^2}{x{=}0\ y{=}1} + \underset{0^2}{x{=}0\ y{=}2} + \underset{1^2 \cdot f(1,0)}{x{=}1\ y{=}0} + \underset{1^2 \cdot f(1,1)}{x{=}1\ y{=}1}$$

$$+ \underset{1^2 f(1,2)}{x{=}1\ y{=}2} + \underset{2^2 f(2,0)}{x{=}2\ y{=}0} + \underset{2^2 f(2,1)}{x{=}2\ y{=}1} + \underset{2^2 f(2,2)}{x{=}2\ y{=}2}$$

$$= 1/3 + 4/6 + 4/6 = 10/6 = 5/3$$

Similarly

$$E(y^2) = \sum_x \sum_y y^2 \ f(x,y)$$

$$= \underset{0^2 \cdot 1/6}{x{=}0\ y{=}0} + \underset{1^2 \cdot 0}{x{=}0\ y{=}1} + \underset{2^2 \cdot 1/6}{x{=}0\ y{=}2} + \underset{0^2 \cdot 1/3}{x{=}1\ y{=}0} + \underset{1^2 \cdot 0}{x{=}1\ y{=}1}$$

$$+ \underset{2^2 \cdot 0}{x{=}1\ y{=}2} + \underset{0^2 \cdot 1/6}{x{=}2\ y{=}0} + \underset{1^2 \cdot 0}{x{=}2\ y{=}1} + \underset{2^2 \cdot 1/6}{x{=}2\ y{=}2} = 4/3 = E(y^2$$

$\sigma_x^2 = E(x^2) - (E(x))^2 = 5/3 - 1^2 = 2/3$

$\sigma_y^2 = E(y^2) - (E(y))^2 = 4/3 - (2/3)^2 = 12/9 - 4/9 = 8/9$

$$\text{Cor}(x,y) = \frac{\text{Cov}(x,y)}{\sigma_x \ \sigma_y} = \frac{0}{\sqrt{2/3}\ \sqrt{8/9}}$$

$\text{Cor}(x,y) = 0$

$\text{Cov}(x,y) = 0$

However, x and y are not independent. The converse holds. If x and y are independent, cov(x,y) = 0.

Properties of Covariance

1. $\sigma_{xy} = E(x,y) - E(x)\ E(y)$

2. If x and y are independent random variables, cov(x,y) = 0.

3. If cov(x,y) = 0, then x and y are not necessarily independent.

4. $\text{var}\,(x \pm y) = \text{var}\,x + \text{var}\,y \pm 2\,\text{cov}\,(x,y)$

For example, to show property #1, calculate:

$$
\begin{aligned}
\text{cov}(x,y) &= E[\,(x - \mu_x)\,(y - \mu_y)] \\
&= E[xy - \mu_x y - \mu_y x + \mu_x \mu_y] \\
&= E(xy) - E(\mu_x\, y) - E(\mu_y\, x) + E\,(\mu_x \mu_y) \\
&= E(xy) - \mu_x\, E(y) - \mu_y\, E(x) + \mu_x\, \mu_y \\
&= E(xy) - E(x)\, E(y) - \mu_y \mu_x + \mu_x \mu_y
\end{aligned}
$$

$$\therefore\ \text{cov}(x,y) = E(xy) - E(x)\, E(y)$$

Let us prove the following important theorem:

Th. Let x, y be independent random variables. Then cov(x,y) = 0.

Pf. Since x,y are independent, $E(xy) = E(x)\ E(y)$

$$
\begin{aligned}
\text{cov}(x,y) &= E(xy) - E(x)\, E(y) \\
&= E(x)\, E(y) - E(x)\, E(y) = 0
\end{aligned}
$$

Note from our two examples the converse does not hold.

If cov(x,y) = 0, we cannot conclude that x,y are independent.

According to the great statistician Kolmogorov, independence is the central problem in statistics. The Central Limit Theorem rises or falls based upon the independence or dependence of the random variables.

Correlation - Continuous Case

Correlation for discrete random variables was treated in Chapter One. To measure correlation for continuous random variables, we use the covariance, which is zero if x and y are independent.

We define the ***correlation*** between x and y:

$$\text{cor}(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

For our earlier example,

$$f(x,y) = \begin{cases} 4xy & 0 < x < 1 \\ & 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

we found $\text{cov}(x,y) = 0, \quad \mu_x = \mu_y = 2/3$

$$\sigma_x^2 = E[(x - \mu_x)]^2$$

$$\sigma_x^2 = E[(x - 2/3)]^2$$

$$\sigma_x^2 = E(x^2 - 4/3\,x + 4/9)$$

Though we know $\text{cor}(x,y) = \dfrac{0}{\sigma_x \sigma_y} = 0$, let us calculate σ_x, σ_y

$$\sigma_x^2 = \int_0^1 \int_0^1 (x^2 - 4/3\,x + 4/9)\,(4xy)\,dxdy$$

$$= \int_0^1 \int_0^1 (4x^2y - 16/3\,x^2y + 16/9\,xy)\,dxdy$$

$$\sigma_x^2 = \int_0^1 x^4y - 16/9\ x^3y + 8/9\ x^2y \Big|_0^1 dy$$

$$\sigma_x^2 = \int_0^1 (y - 16/9\ y + 8/9\ y)\ dy = 1/9 \int_0^1 y\ dy$$

$$\sigma_x^2 = 1/9\ \frac{y^2}{2} \Big|_0^1 = 1/18$$

$$\sigma_x^2 = 1/18 = \sigma_y^2$$

$$\sigma_x = \sqrt{1/18} = \frac{\sqrt{2}}{6} = \sigma_y$$

$$\text{cor}\ (x,y) = \frac{\text{cov}\ (x,y)}{\sigma_x\ \sigma_y} = \frac{0}{\frac{\sqrt{2}}{6}\ \frac{\sqrt{2}}{6}} = 0$$

Dependence

Independence is a necessary assumption for the all important Central Limit Theorem, that is a foundation for statistics. My article, "Dependent Random Variables and the Central Limit Theorem" (Fall 1989, International Journal of Mathematical and Computer Modeling), created a computer simulation to link levels of dependence (autocorrelation) with corresponding adjusted Type I error - alpha levels. I am grateful for Dr. Lehman (U of California Berkeley) for his sage advice as to the problems with generalizability of such results. Dr. Lehman envisioned a future where statisticians and computer scientists adjust statistical results based upon dependence. We are far from the realization of his vision.

EXERCISES 2.6

1. Calculate a, μ, σ^2 for the probability density function:

$$f(x) = a\,e^{2x} \quad 0 < x < 1$$
$$= 0 \quad \text{elsewhere}$$

2. Show that var (cx) = c^2 var x

3. Calculate a, μ_x, μ_y, σ_x^2, σ_y^2, cov (x,y) and cor (x,y) for the probability density function:

$$f(x,y) = a\,x^2\,y^3 \quad 0 \le x \le 1$$
$$0 \le y \le 1$$
$$= 0 \quad \text{elsewhere}$$

4. Using the information from example 3, compute f_x (x) and f_y (y). Are x and y independent?

5. Calculate a, E(x), var (x), E(y), var (y), cov (x,y) and cor (x,y) for the following probability density function:

$$f(x,y) = a\,(x + y) \quad x = 0, 1, 2$$
$$y = 0, 1, 2$$
$$= 0 \quad \text{elsewhere}$$

6. Suppose x is a discrete random variable with two values x = -4 and x = 4. Suppose the probability of each equals .5. Let y be a discrete random variable with $y = x^2$. Show that cov (x,y) = E (xy) - E (x) E (y). [This demonstrates the third property of covariance. cov (x,y) = 0, but x and y are clearly not independent. This single counterexample disproves the intuitively appealing (but wrong) converse of property two of covariance.]