SOME THEORY AND PRACTICE OF STATISTICS
by Howard G. Tucker

CHAPTER 3. EXPECTATION

3.1 Definition of the Expectation of a Random Variable. As we have seen, the distribution function F_X of a random variable X spreads a unit of mass over the real line by assigning to every interval $[a, b]$ of real numbers the mass $F_X(b) - F_X(a)$. The expectation of this random variable will be defined as the first moment of the spread of this mass about the origin, i.e., the expectation will be defined as the center of gravity of this spread of mass. This is not an artificial notion; it will play an important role with respect to the notion of relative frequency, a notion that is a central concept of statistics. So, after exhibiting what appears at first to be a rather strange definition of expectation, we shall show how it is computed in the two special cases of absolutely continuous and discrete distributions. Finally, we develop basic properties of expectation *when it exists.*

As we pointed out above, a random variable X determines a distribution function F_X which spreads one unit of mass along the real line according to the formula $P([a < X \leq b]) = F_X(b) - F_X(a)$. The expectation of this random variable or of its distribution function is nothing other than the centroid or center of gravity of this spread of unit mass. The problem is in defining it. If, say, the random variable were discrete, taking values a_1, a_2, a_3, a_4 with corresponding probabilities p_1, p_2, p_3, p_4, (where $\sum_{i=1}^{4} p_i = 1$), then the centroid, or center of gravity, or first moment of this spread of mass is $\sum_{i=1}^{4} a_i p_i$. Let us suppose that $a_1 < a_2 < 0 < a_3 < a_4$. Then the distribution function F_X would be as graphed in figure 1.

SPACE FOR FIGURE 1

Now one notices that the center of gravity given by the above formula is nothing other than the sum of the areas in rectangles denoted by 3 and 4 minus the sum of the areas in the rectangles marked off as 1 and 2. In other words, the center of gravity is the area between the graph of $y = F_X(x)$, the line $y = 1$ and the y-axis minus the area below the graph of $y = F_X(x)$, above the x-axis and to the left of the y-axis. For this and any discrete distribution function, this is always the case. Since any distribution function can be approximated quite closely by a discrete distribution function, it would seem

almost natural to define the centroid in the general case by such a difference in areas. Accordingly, we arrive at the following definition.

Definition. Let X be a random variable with distribution function $F_X(x)$. The expectation $E(X)$ or EX of X is defined by

$$\begin{aligned} E(X) &= \int_0^\infty (1 - F_X(x))dx - \int_{-\infty}^0 F_X(x)dx \\ &= \int_0^\infty P([X > x])dx - \int_{-\infty}^0 P([X \leq x])dx \ , \end{aligned}$$

and is said to exist if the two integrals are finite. In cases where at least one of the integrals is infinite, we shall say that the expectation of X does not exist.

In other words, we say that the expectation of X exists if and only if the two integrals in the above definition are finite. For technical reasons later on we shall need the following proposition.

Lemma 1. *If X is a random variable that satisfies*

$$\int_0^\infty x^k P([X > x])dx < \infty \text{ and } \int_{-\infty}^0 \mid x \mid^k P([X \leq x])dx < \infty$$

for some integer $k \geq 0$, then

$$\int_0^\infty x^k P([X > x])dx = \int_0^\infty x^k P([X \geq x])dx$$

and

$$\int_{-\infty}^0 x^k P([X < x])dx = \int_{-\infty}^0 x^k P([X \leq x])dx \ .$$

Proof: We shall only prove the second equality. We first prove this in the case where $k = 0$. Let $\epsilon > 0$ be arbitrary. Since $[X < x] \subset [X \leq x] \subset [X < x + \epsilon]$, we have

$$\int_{-\infty}^0 P([X < x])dx \leq \int_{-\infty}^0 P([X \leq x])dx \leq \int_{-\infty}^0 P([X < x + \epsilon])dx.$$

Let us make the change of variable $y = x + \epsilon$ in this last integral. Then this last integral becomes

$$\int_{-\infty}^{\epsilon} P([X < y])dy = \int_{-\infty}^{0} P([X < x])dx + \int_{0}^{\epsilon} P([X < x])dx \leq \int_{-\infty}^{0} P([X < x])dx + \epsilon \ .$$

These last two displays imply

$$\int_{-\infty}^{0} P([X < x])dx \leq \int_{-\infty}^{0} P([X \leq x])dx \leq \int_{-\infty}^{0} P([X < x])dx + \epsilon \ .$$

Letting $\epsilon \to 0$, we obtain the conclusion in the case of $k = 0$. We now prove the second equation in the conclusion of the theorem in the case where $k \geq 1$. Letting $\epsilon > 0$ be arbitrary, but keeping $\epsilon < 1$, we have

$$\begin{aligned} \int_{-\infty}^{0} \mid x \mid^k P([X < x])dx &\leq \int_{-\infty}^{0} \mid x \mid^k P([X \leq x])dx \\ &\leq \int_{-\infty}^{0} \mid x \mid^k P([X < x + \epsilon])dx \ . \end{aligned}$$

Letting $y = x + \epsilon$, this last integral is

$$\begin{aligned} &= \int_{-\infty}^{\epsilon} \mid y - \epsilon \mid^k P([X < y])dy \\ &= \int_{-\infty}^{0} \mid y - \epsilon \mid^k P([X < y])dy + \int_{0}^{\epsilon} \mid y - \epsilon \mid^k P([X < y])dy \ . \end{aligned}$$

Let us bound the second integral from above. If $0 \leq y \leq \epsilon$, then $\mid y - \epsilon \mid < \epsilon$. Also, $P([X < y]) \leq 1$, so

$$\int_{0}^{\epsilon} \mid y - \epsilon \mid^k P([X < y])dy \leq \epsilon^{k+1}.$$

Note that for $y \leq 0$ the inequality $\mid y - \epsilon \mid^k - \mid y \mid^k \geq 0$, so the first integral is equal to

$$\int_{-\infty}^{0} \mid y \mid^k P([X < y])dy + \int_{-\infty}^{0} \mid\mid y - \epsilon \mid^k - \mid y \mid^k \mid P([X < y])dy \ .$$

We shall bound the second integral in this last expression. By the triangle inequality, $\mid y-\epsilon \mid \leq \mid y \mid +\epsilon$ and so by the binomial theorem,

$$\mid y-\epsilon \mid^k \leq (\mid y \mid +\epsilon)^k = \sum_{j=0}^{k} \binom{k}{j} \mid y \mid^j \epsilon^{k-j}.$$

Thus,

$$\mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid \leq \sum_{j=0}^{k-1} \binom{k}{j} \mid y \mid^j \epsilon^{k-j}.$$

Now we may write

$$\begin{aligned} \int_{-\infty}^{0} \mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid P([X<y])dy &= \int_{-\infty}^{-1} \mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid P([X<y])dy \\ &+ \int_{-1}^{0} \mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid P([X<y])dy \end{aligned} .$$

Over the interval $(-\infty,-1]$ we have $\mid y \mid \geq 1$, and over $[-1,0]$ we have $\mid y \mid \leq 1$; keeping in mind that $0<\epsilon<1$, we have

$$\begin{aligned} \int_{-\infty}^{-1} \mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid P([X<y])dy &\leq \sum_{j=0}^{k-1} \tbinom{k}{j} \epsilon^{k-j} \int_{-\infty}^{-1} \mid y \mid^j P([X<y])dy \\ &\leq \epsilon(2^k-1) \int_{-\infty}^{-1} \mid y \mid^k P([X<y])dy \end{aligned}$$

Also, as above, but keeping in mind that $\mid y \mid \leq 1$ over $[-1,0]$, we have

$$\begin{aligned} \int_{-1}^{0} \mid\mid y-\epsilon \mid^k - \mid y \mid^k \mid P([X<y])dy &\leq \epsilon(2^k-1) \int_{-1}^{0} P([X<y])dy \\ &\leq \epsilon(2^k-1). \end{aligned}$$

Combining the above, we obtain

$$\begin{aligned} \int_{-\infty}^{0} \mid x \mid^k P([X<x])dx &\leq \int_{-\infty}^{0} \mid x \mid^k P([X \leq x])dx \\ &\leq \int_{-\infty}^{0} \mid x \mid^k P([X<x])dx + \epsilon(2^k-1) \ . \end{aligned}$$

Since by hypothesis $\int_{-\infty}^{0} \mid x \mid^k P([X \leq x])dx < \infty$, we obtain the conclusion by taking the limit as $\epsilon \to 0$.

This proposition is of great use to us. It allows us to replace $\leq$ by $<$ and $\geq$ by $>$.

Lemma 2. *If $\{a_n\}$ is a denumerable sequence of nonnegative real numbers, and if $Q_n = \sum_{j=n}^{\infty} a_j$, then*

$$\sum_{n=1}^{\infty} Q_n = \sum_{n=1}^{\infty} n a_n \ .$$

Proof: Since infinite series of nonnegative terms can be arranged without altering convergence properties, we have

$$\begin{aligned} \textstyle\sum_{n=1}^{\infty} Q_n &= a_1 + a_2 + a_3 + a_4 + \cdots \\ &\qquad + a_2 + a_3 + a_4 + \cdots \\ &\qquad\qquad + a_3 + a_4 + \cdots \\ &\qquad\qquad\qquad + a_4 + \cdots \\ &= a_1 + 2a_2 + 3a_3 + 4a_4 + \cdots \ , \end{aligned}$$

which completes the proof.

Theorem 1. *Let X be a random variable with a discrete distribution, i.e., there exists a finite or denumerable sequence of distinct real numbers $\{x_n\}$ and corresponding probabilities $\{p_n\}$ such that $P([X = x_n]) = p_n$ for all values of n and $\sum_{n \geq 1} p_n = 1$. Then the expectation of X exists and*

$$E(X) = \sum_{n \geq 1} x_n p_n$$

if and only if this series is absolutely convergent.

Proof: Let

$$I_1 = \int_{0}^{\infty} P([X > x])dx \text{ and } I_2 = \int_{-\infty}^{0} P([X \leq x])dx \ .$$

We shall use the fact that $P([X > x])$ is nonincreasing in x, i.e., if $x_1 < x_2$, then $P([X > x_1]) \geq P([X \geq x_2])$. Because of the definition of the improper

Riemann integral, we may write for every positive integer n,

$$I_1 = \sum_{k\geq 1} \int_{(k-1)/n}^{k/n} P([X > x])dx \ .$$

Because of the nonincreasing property of $P([X > x])$ in x that was mentioned above, we have the following inequality:

$$\frac{1}{n}P([X > \frac{k}{n}]) \leq \int_{(k-1)/n}^{k/n} P([X > x])dx \leq \frac{1}{n}P([X > \frac{k-1}{n}]) \ .$$

Thus, if we denote

$$L_n = \sum_{k\geq 1} \frac{1}{n}P([X > \frac{k}{n}])$$

and

$$U_n = \sum_{k\geq 1} \frac{1}{n}P([X > \frac{k-1}{n}]) \ ,$$

we easily see because of the above inequality that for every positive integer n,

$$L_n \leq I_1 \leq U_n \ .$$

However, we may write

$$L_n = \frac{1}{n}\sum_{k\geq 1}\sum_{j\geq k} P([\frac{j}{n} < X \leq \frac{j+1}{n}]) \ .$$

Now if we let $a_j = P([\frac{j}{n} < X \leq \frac{j+1}{n}])$, we may apply lemma 2 to obtain

$$\begin{aligned} L_n &= \sum_{k=1}^{\infty} \tfrac{k-1}{n} P\left(\left[\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right]\right) \\ &= \sum_{k=1}^{\infty} \tfrac{k}{n} P\left(\left[\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right]\right) - \sum_{k=1}^{\infty} \tfrac{1}{n} P\left(\left[\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right]\right) \ . \end{aligned}$$

However

$$\begin{aligned} \tfrac{k}{n} P\left(\left[\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right]\right) &= \tfrac{k}{n} \sum \{p_j : (k-1)/n < x_j \leq k/n\} \\ &\geq \sum \{x_j p_j : (k-1)/n < x_j \leq k/n\} \ . \end{aligned}$$

Hence

$$\begin{aligned} L_n &\geq \textstyle\sum_{k\geq 1}\sum\{x_j p_j : (k-1)/n < x_j \leq k/n\} - \frac{1}{n}P([X>0]) \\ &\geq \textstyle\sum\{x_m p_m : x_m > 0\} - \frac{1}{n} \ . \end{aligned}$$

In a similar fashion one can prove that

$$U_n \leq \sum\{x_m p_m : x_m > 0\} + \frac{1}{n} \ .$$

Thus, for every positive integer n,

$$\sum\{x_m p_m : x_m > 0\} - \frac{1}{n} \leq I_1 \leq \sum\{x_m p_m : x_m > 0\} + \frac{1}{n} \ ,$$

from which there follows, by taking the limit as $n \to \infty$, that

$$I_1 = \sum\{x_m p_m : x_m > 0\} \ .$$

In an entirely similar manner, one can prove that

$$I_2 = -\sum\{x_m p_m : x_m \leq 0\} \ .$$

Thus absolute convergence of the series $\sum_{n\geq 1} x_m p_m$ is necessary and sufficient for the integrals I_1 and I_2 to exist and be finite, and we obtain

$$E(X) = I_1 - I_2 = \sum_{m\geq 1} x_m p_m \ ,$$

which proves the theorem.

Theorem 2. *Let X be a random variable with an absolutely continuous distribution function and whose density $f_X(x)$ is piecewise continuous and is bounded over every bounded interval. Then the expectation of X exists and*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

if and only if this integral exists as an improper Riemann integral.

Proof: By the definition of expectation given above, the expectation, $E(X)$, exists and equals $\int_0^\infty P([X>x])dx - \int_{-\infty}^0 P([X \leq x])dx$, provided that

both of these improper Riemann integrals exist and are finite. For $A > 0$, and making use of integration by parts, the hypothesis that $f_X(x)$ is bounded over every bounded interval and is continuous everywhere except at a countable set of points, and the fundamental theorem of calculus, we obtain

$$\begin{aligned}
\int_0^A P([X > x])dx &= \int_0^A \left(\int_x^\infty f_X(t)dt \right) dx \\
&= -\int_0^A \left(\int_\infty^x f_X(t)dt \right) dx \\
&= -x \int_\infty^x f_X(t)dt \,|_0^A + \int_0^A x f_X(x)dx \\
&= A \int_A^\infty f_X(t)dt + \int_0^A x f_X(x)dx.
\end{aligned}$$

Both quantities on the right hand side are nonnegative. First let us suppose that $E(X)$ exists and is finite. Then as $A \to \infty$, we have $\int_0^A P([X > x])dx \to \int_0^\infty P([X > x])dx < \infty$. Since $0 \leq \int_0^A x f_X(x)dx \leq \int_0^A P([X > x])dx < \infty$ we have, taking $A \to \infty$, $\int_0^\infty x f_X(x)dx \leq \int_0^\infty P([X > x])dx < \infty$, i.e., $\int_0^\infty x f_X(x)dx < \infty$. Since $A \int_A^\infty f_X(t)dt \leq \int_A^\infty x f_X(x)dx < \infty$, it follows that $A \int_A^\infty f_X(t)dt \to 0$. Thus by the displayed equation above, $\int_0^\infty x f_X(x)dx = \int_0^\infty P([X > x])dx < \infty$. Now let us assume that $\int_0^\infty x f_X(x)dx < \infty$. Then because $A \int_A^\infty f_X(t)dt \leq \int_A^\infty x f_X(x)dx < \infty$, we have $A \int_A^\infty f_X(t)dt \to 0$ as $A \to \infty$. By the displayed equation above, it follows that $\int_0^\infty P([X > x])dx < \infty$ and $\int_0^\infty P([X > x])dx = \int_0^\infty x f_X(x)dx$. Similarly, $\int_{-\infty}^0 P([X \leq x])dx < \infty$ if and only if $-\int_{-\infty}^0 x f_X(x)dx < \infty$, in which case the two integrals are equal, which proves the theorem

We now give some examples of expectations of random variables.

Example 1. The binomial distribution. Suppose X is a random variable with the binomial distribution, $Bin(n,p)$, i.e.,

$$P([X=x]) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0,1,2,\cdots,n \\ 0 & \text{otherwise .} \end{cases}$$

In this case we have a discrete distribution, so

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP([X=k]) \\ &= \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= np(p+(1-p))^{n-1} \\ &= np \ . \end{aligned}$$

Example 2. The geometric distribution. Let X be a random variable whose distribution is $geom(p)$. Then $P([X=n]) = (1-p)^{n-1}p$ for $n = 1,2,\cdots$. Recalling that the derivative of a power series is equal to the series of the derivatives inside the interval of absolute convergence, and recalling that the geometric series $\sum_{n\geq 0} x^n$ has radius of convergence 1, then for $0 < p < 1$,

$$E(X) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = -p\frac{d}{dp}\sum_{n=0}^{\infty}(1-p)^n = -p\frac{d}{dp}\left(\frac{1}{1-(1-p)}\right) = \frac{1}{p} \ .$$

Example 3. The Poisson distribution. Let X be a random variable whose distribution is $\mathcal{P}(\lambda)$ for some $\lambda > 0$, i.e., $P([X=n]) = e^{-\lambda}\lambda^n/n!$ for $n = 0,1,2,\cdots$. Then

$$E(X) = \sum_{n=0}^{\infty} ne^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda \ .$$

We next consider expectation of random variables with discrete distributions that arise from sampling without replacement. We shall first find the expectation of the general sample survey sum distribution and then draw off as special cases the expectations of the Wilcoxon and hypergeometric distributions. But first we need a combinatorial lemma.

Lemma 3. *If* $x_1, x_2, \cdots, x_N$ are any numbers, and if $1 \leq n \leq N$, then

$$\sum_{1 \leq i_1 < i_2 < \cdots i_n \leq N} \sum_{k=1}^{n} x_{i_k} = \binom{N-1}{n-1} \sum_{j=1}^{N} x_j \ .$$

Proof: For every j, $1 \leq j \leq N$, there are obviously $\binom{N-1}{n-1}$ sums of the form $\sum\limits_{k=1}^{n} x_{i_k}$, where $1 \leq i_1 < i_2 < \cdots i_n \leq N$, that include x_j. This proves the lemma.

Example 4. The sample survey sum distribution. Let X denote the sum of n numbers selected at random without replacement from the (not necessarily distinct) set of N numbers $x_1, x_2, \cdots, x_N$. Each sum of the form $\sum\limits_{k=1}^{n} x_{i_k}$ is selected with probability $1/\binom{N}{n}$. Therefore, by lemma 3,

$$E(X) = \sum_{1 \leq i_1 < i_2 < \cdots i_n \leq N} \frac{1}{\binom{N}{n}} \sum_{k=1}^{n} x_{i_k} = \frac{n}{N} \sum_{j=1}^{N} x_j = n\overline{x},$$

where $\overline{x} = \frac{1}{N} \sum\limits_{j=1}^{N} x_j$.

Example 5. The Wilcoxon Distribution. In this case, a random variable W is the sum of n numbers selected at random without replacement from $1, 2, \cdots, N$. We recall that

$$\sum_{j=1}^{N} j = \frac{N(N+1)}{2},$$

and so, applying the result of example 4, if W has the $Wilcoxon(n, N)$ distribution, then $E(W) = \frac{n(N+1)}{2}$.

Example 6. The hypergeometric distribution. Recall that X is said to have the hypergeometric distribution if it has the same distribution as the number of red balls in a sample of size n taken at random without replacement from an urn that contains r red balls and b black balls, where $1 \leq n \leq r + b$. In this case, we assign the number 1 to each of the r red balls and the number 0 to each of the black balls. Now the number X of red balls in the sample of size n is equal to the sum of the numbers in the sample. Thus, taking $N = r + b$, we have

$$E(X) = n\frac{r}{r+b} \ .$$

Example 7. The normal distribution, $\mathcal{N}(\mu, \sigma^2)$. This absolutely continuous distribution has a density given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad , -\infty < x < \infty,$$

where $\mu \in (-\infty, \infty)$ and $\sigma^2 > 0$ are constants. Using theorem 2, we obtain

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} x e^{-(x-\mu)^2/2\sigma^2} dx = \mu.$$

Example 8. The exponential distribution. This absolutely continuous distribution has a density given by

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} \text{ if } \ x \geq 0 \\ 0 \text{ if } \ x < 0, \end{cases}$$

where $\alpha > 0$ is some constant. Using theorem 2, we obtain

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \alpha \int_{-\infty}^{\infty} x e^{-\alpha x} dx = \frac{1}{\alpha} \ .$$

Expectations do not always exist. As an example of a random variable with an absolutely continuous distribution for which the expectation does not exist, let us consider the Cauchy distribution given in the previous section. Recall that its density is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

for all real values of x. Now remembering the care we must take with an improper integral, we note that

$$\int_{0}^{\infty} x f_X(x) dx = \int_{0}^{\infty} x \frac{1}{\pi(1+x^2)} dx = \infty \ ,$$

and $\int_{-\infty}^{0} x \frac{1}{\pi(1+x^2)} dx = -\infty$. Thus the condition given by theorem 2 is violated. An example of a random variable with a discrete distribution for

which the expectation does not exist, consider a random variable Y with discrete density given by $P([Y = 2^n]) = \frac{1}{2^n}$ for $n = 1, 2, \cdots$. In this case, $\sum_{n=1}^{\infty} 2^n P([Y = 2^n]) = \infty$, thus violating the condition of theorem 1.

EXERCISES

1. Let X be a random variable with distribution function defined by

$$F_X(x) = \begin{cases} 0 \text{ if } x < -3.5 \\ .2 \text{ if } -3.5 \leq x < -1 \\ .35 \text{ if } -1 \leq x < 0 \\ .45 \text{ if } 0 \leq x < 2 \\ .65 \text{ if } 2 \leq x < 3.5 \\ 1.00 \text{ if } x \geq 3.5 \ . \end{cases}$$

(i) Compute $E(X)$ by using theorem 1.
(ii) Compute $E(X)$ by using the definition of expectation.

2. Let X be a random variable with distribution function defined by

$$F_X(x) = \begin{cases} \frac{1}{2}e^x \text{ if } x < 0 \\ \frac{1}{2} + \frac{1}{2}\sqrt{x} \text{ if } 0 \leq x < 1 \\ 1 \text{ if } x \geq 1 \ . \end{cases}$$

(i) Find the density of $F_X(x)$.
(ii) Compute $E(X)$ using theorem 2.
(iii) Compute $E(X)$ using the definition of expectation.

3. Let X be a random variable with absolutely continuous distribution with density given by

$$f_X(x) = \begin{cases} x^3 \text{ if } 0 \leq x \leq \sqrt{2} \\ 0 \text{ otherwise.} \end{cases}$$

Determine $F_X(x)$ and $E(X)$,

4. Let X be a random variable with absolutely continuous distribution with density given by

$$f_X(x) = \frac{1}{2}e^{-|x-2|} \text{ for all real } x \ .$$

(i) Find the distribution function of X.
(ii) Compute $E(X)$.

3.2 Basic Properties of Expectation. We now develop some linearity and order properties of expectation that will be used repeatedly from here on.

Theorem 1. *If X is a random variable that is bounded, i.e., if there exists a constant $K > 0$ such that $P([| X |< K]) = 1$, then the expectation of X exists, and $-K < E(X) < K$.*

Proof: In this case it is easy to see that $P([X > K]) = 0$ and $P([X < -K]) = 0$. Thus

$$\int_0^\infty P([X > x])dx = \int_0^K P([X > x])dx \leq K$$

and

$$\int_{-\infty}^0 P([X \leq x])dx = \int_{-K}^0 P([X \leq x])dx \leq K \quad ,$$

which proves the finiteness of both integrals in the definition of expectation. The second conclusion is easily obtained from the above inequalities.

Theorem 2. *If X is a random variable whose expectation exists, and if c is a constant, then the expectation of cX exists and $E(cX) = cE(X)$.*

Proof: We shall prove this only for the case $c < 0$; the proof for $c \geq 0$ is the very same and is left as an exercise. We observe that $\int_0^\infty P([cX > x])dx = \int_0^\infty P([X < x/c])dx$. Making the change of variable $y = x/c$, we obtain, using proposition 1 in section 3.1,

$$\int_0^\infty P([X < x/c])dx = -c \int_{-\infty}^0 P([X < y])dy = -c \int_{-\infty}^0 P([X \leq y])dy$$

which is finite by hypothesis and the definition of expectation. Similarly, $\int_{-\infty}^0 P([cX \leq x])dx = c \int_0^\infty P([X > y])dy$, which yields our conclusion.

Theorem 3. *If X, Y and Z are random variables, if expectations exist for X and Z, and if $X \leq Y \leq Z$, (i.e., if*

$$P([X \leq Y \leq Z]) = 1,$$

then the expectation of Y exists, and $E(X) \leq E(Y) \leq E(Z)$.

Proof: The hypothesis $X \leq Y \leq Z$ is easily seen to imply

$$[X > x] \subset [Y > x] \subset [Z > x] \text{ and } [Z \leq x] \subset [Y \leq x] \subset [X \leq x] \ .$$

for every real number x. Thus

$$P([X > x]) \leq P([Y > x]) \leq P([Z > x])$$

and

$$P([Z \leq x]) \leq P([Y \leq x]) \leq P([X \leq x]).$$

The hypothesis that X and Z have expectations, the definition of expectation and the above inequalities easily imply the conclusions.

Corollary. *If X and Z are random variables whose expectations exist, and if $X \leq Z$, then $E(X) \leq E(Z)$.*

Proof: This follows from theorem 3 by taking $Y = X$.

Lemma 1. *If X and Y are random variables, then*

$$\lim_{K \to \infty} \int_0^\infty P([X > x] \cap [\mid Y \mid < K])dx = \int_0^\infty P([X > x])dx$$

and

$$\lim_{K \to \infty} \int_{-\infty}^0 P([X \leq x] \cap [\mid Y \mid < K])dx = \int_{-\infty}^0 P([X \leq x])dx \ ,$$

no matter whether in each case the right hand side is finite or infinite.

Proof: We shall only prove the first conclusion; a similar proof holds for the second conclusion. We first observe that $[X > x] \cap [\mid Y \mid \leq K] \subset [X > x]$, so that $P([X > x] \cap [\mid Y \mid \leq K]) \leq P([X > x])$. From this we obtain that

$$\int_0^\infty P([X > x] \cap [\mid Y \mid \leq K])dx \leq \int_0^\infty P([X > x])dx$$

for all $K > 0$. Since the right side does not depend on K, and since the left side is nondecreasing in K (prove this!), it follows that the following limit exists and

$$lim_{K\to\infty} \int_0^\infty P([X > x] \cap [\mid Y \mid \leq K])dx \leq \int_0^\infty P([X > x])dx.$$

We now prove the reverse inequality. Note that

$$P([X > x]) = P([X > x] \cap [\mid Y \mid \leq K]) + P([X > x] \cap [\mid Y \mid > K]) \ .$$

Now $P([X > x] \cap [\mid Y \mid > K]) \leq P([\mid Y \mid > K])$, so

$$P([X > x] \cap [\mid Y \mid \leq K]) \geq P([X > x]) - P([\mid Y \mid > K]) \ .$$

Let $A > 0$ and $\epsilon > 0$ be fixed (with A as large as one wishes, and $\epsilon > 0$ being as small as one wishes). Then, since $P([\mid Y \mid \leq K]) \to 1$ as $K \to \infty$, it follows that $P([\mid Y \mid > K]) \to 0$ as $K \to \infty$. Hence there exists a $K_0 > 0$ such that for all $K \geq K_0$, the inequality $P([\mid Y \mid > K]) < \frac{\epsilon}{A}$ is true. Hence for all $K \geq K_0$,

$$\begin{aligned} \int_0^\infty P([X > x] \cap [\mid Y \mid \leq K])dx &\geq \int_0^A (P([X > x]) - P([\mid Y \mid > K]))dx \\ &\geq \int_0^A P([X > x])dx - \epsilon \ . \end{aligned}$$

Since the first expression in this last display is nondecreasing as K increases, we have

$$lim_{K\to\infty} \int_0^\infty P([X > x] \cap [\mid Y \mid \leq K])dx \geq \int_0^A P([X > x])dx - \epsilon \ .$$

Since this holds for all A, we can take the limit of both sides as $A \to \infty$ to obtain

$$lim_{K\to\infty} \int_0^\infty P([X > x] \cap [\mid Y \mid \leq K])dx \geq \int_0^\infty P([X > x])dx - \epsilon \ .$$

Next, take the limit of both sides as $\epsilon \to 0$, and the reverse inequality is established, which concludes the proof of the first conclusion.

Corollary. *If X is a random variable whose expectation exists, then*

$$E(X) = lim_{K\to\infty} E(XI_{[|X|<K]}) \ .$$

Proof: We first observe the following easily proved identities: for $x > 0, K > 0$,

$$[X > x] \cap [|\ X\ | < K] = [XI_{[|X|<K]} > x]$$
$$\text{and}$$
$$[X \leq -x] \cap [|\ X\ | < K] = [XI_{[|X|<K]} \leq -x].$$

Using these identities, lemma 1 and the definition of expectation, we obtain

$$\begin{aligned} E(X) &= \int_0^\infty P([X > x])dx - \int_{-\infty}^0 P([X \leq x])dx \\ &= lim_{K\to\infty} \int_0^\infty P([X > x] \cap [|\ X\ | < K])dx \\ &- lim_{K\to\infty} \int_{-\infty}^0 P([X \leq x] \cap [|\ X\ | < K])dx \\ &= lim_{K\to\infty} \left\{ \int_0^\infty P([XI_{[|X|<K]} > x])dx - \int_{-\infty}^0 P([XI_{[|X|<K]} \leq x])dx \right\} \\ &= lim_{K\to\infty} E(XI_{[|X|<K]}), \end{aligned}$$

which proves the corollary.

Theorem 4. *If X and Y are random variables whose expectations exist, then the expectation of $X + Y$ exists, and $E(X + Y) = E(X) + E(Y)$.*

Proof: We first prove that the expectation of $X + Y$ exists. In order to do this, we need only prove that

$$\int_0^\infty P([X + Y > x])dx < \infty \text{ and } \int_{-\infty}^0 P([X + Y \leq x])dx < \infty.$$

We shall prove only the second inequality. We observe that for every $x \leq 0$,

$$[X + Y \leq x] \subset [X \leq \frac{x}{2}] \cup [Y \leq \frac{x}{2}].$$

Thus

$$P([X + Y \leq x]) \leq P([X \leq \frac{x}{2}]) + P([Y \leq \frac{x}{2}]),$$

and, by properties of integration,

$$\int_{-\infty}^{0} P([X+Y \le x])dx \le \int_{-\infty}^{0} P([X \le \frac{x}{2}])dx + \int_{-\infty}^{0} P([Y \le \frac{x}{2}])dx.$$

By the change of variable $y = x/2$, this last inequality becomes

$$\int_{-\infty}^{0} P([X+Y \le x])dx \le 2\int_{-\infty}^{0} P([X \le y])dy + 2\int_{-\infty}^{0} P([Y \le y])dy,$$

which is finite by hypothesis. We now prove the theorem in special cases.

Case (i) Let X be as in the hypothesis, and let $Y = cI_A$, where c is a constant and A is an event. We next observe that

$$\begin{aligned}
E(X+cI_A) &= \textstyle\int_0^\infty P([X+cI_A > x])dx - \int_{-\infty}^0 P([X+cI_A \le x])dx \\
&= \textstyle\int_0^\infty P([X+cI_A > x] \cap A)dx + \int_0^\infty P([X > x] \cap A^c)dx \\
&\quad \textstyle- \int_{-\infty}^0 P([X+cI_A \le x] \cap A)dx - \int_{-\infty}^0 P([X \le x] \cap A^c)dx \\
&= \textstyle\int_0^\infty P([X+c > x] \cap A)dx + \int_0^\infty P([X > x] \cap A^c)dx \\
&\quad \textstyle- \int_{-\infty}^0 P([X+c \le x] \cap A)dx - \int_{-\infty}^0 P([X \le x] \cap A^c)dx \\
&= \textstyle\int_{-c}^\infty P([X > y] \cap A)dy + \int_{-c}^\infty P([X > y] \cap A^c)dy \\
&\quad \textstyle- \int_{-\infty}^{-c} P([X \le y] \cap A)dy - \int_{-\infty}^{-c} P([X \le y] \cap A^c)dy \\
&= \textstyle\int_{-c}^0 P([X > y] \cap A)dy + \int_0^\infty P([X > y])dy \\
&\quad \textstyle- \int_0^{-c} P([X \le y] \cap A)dy - \int_{-\infty}^0 P([X \le y])dy \\
&= \textstyle\int_{-c}^0 P([X > y] \cap A)dy + \int_{-c}^0 P([X \le y] \cap A)dy + E(X) \\
&= \textstyle\int_{-c}^0 P(A)dy + E(X) = E(X) + cP(A) = E(X) + E(cI_A)\ .
\end{aligned}$$

Thus we have proved the theorem in case(i).

Case(ii) Again, let X be as in the hypothesis, and let Y be a discrete random variable of the form

$$Y = \sum_{j=1}^{n} y_j I_{A_j}\ ,$$

where n is finite, $A_1, \cdots, A_n$ are events, and $y_1, \cdots, y_n$ are constants. The random variable Y is bounded, so its expectation exists. Applying case (i) again and again, we have $E(X + y_1 I_{A_1}) = E(X) + y_1 P(A_1)$, $E(X + y_1 I_{A_1} + y_2 I_{A_2}) = E(X + y_1 I_{A_1}) + y_2 P(A_2) = E(X) + y_1 P(A_1) + y_2 P(A_2)$, and after $n-2$ more applications of case 1 we obtain

$$E(X+Y) = E(X + \sum_{j=1}^{n} y_j I_{A_j}) = E(X) + \sum_{j=1}^{n} y_j P(A_j) = E(X) + E(Y)\ ,$$

which proves the theorem in case (ii).

Case (iii) Now suppose X is as in the hypothesis, and suppose Y is a bounded random variable, i.e., suppose there is an integer $N > 0$ such that $\mid Y(\omega) \mid < N$ for all $\omega \in \Omega$. The expectation of Y exists since it is bounded. In this case we define two random variables, Y_n^+ and Y_n^- for every positive integer n by:

$$Y_n^+ = \sum_{j=-N2^n+1}^{N2^n} \frac{j}{2^n} I_{[(j-1)/2^n < Y \leq j/2^n]}$$

and

$$Y_n^- = \sum_{j=-N2^n+1}^{N2^n} \frac{j-1}{2^n} I_{[(j-1)/2^n < Y \leq j/2^n]} \ .$$

It is easy to verify that $Y_n^-(\omega) \leq Y(\omega) \leq Y_n^+(\omega)$ over Ω, and $0 \leq Y_n^+(\omega) - Y_n^-(\omega) \leq 1/2^n$ for all n and for all $\omega \in \Omega$. Thus by theorem 3 and case (ii) we have

$$\begin{aligned} &E(Y_n^+ - Y_n^-) \leq 1/2^n \ , \\ &E(Y_n^+ - Y_n^-) = E(Y_n^+) - E(Y_n^-) \text{ and} \\ &E(Y_n^-) \leq E(Y) \leq E(Y_n^+) \ , \end{aligned}$$

which imply

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

Since $X + Y_n^- \leq X + Y \leq X + Y_n^+$, we have by case (ii) and theorem 3 that $X + Y$ has finite expectation, and hence by the inequalities above and theorem 5, one obtains

$$\begin{aligned} E(X+Y) &\leq E(X + Y_n^+) \\ &= E(X) + E(Y_n^+) \\ &\leq E(X) + E(Y) + 1/2^n \end{aligned}$$

for every positive integer n. Similarly $E(X+Y) \geq E(X) + E(Y) - 1/2^n$, and thus for every positive integer n,

$$-1/2^n \leq E(X+Y) - E(X) - E(Y) \leq 1/2^n \ .$$

Letting $n \to \infty$, we obtain $E(X+Y) = E(X) + E(Y)$ for case (iii).

Case (iv) We now prove the theorem in the general case. By lemma 1, we obtain

$$\begin{aligned} \int_0^\infty P([X+Y > x])dx &= \lim_{K\to\infty} \int_0^\infty P([X+Y > x] \cap [\mid Y \mid < K])dx \\ &= \lim_{K\to\infty} \int_0^\infty P([X + Y I_{[\mid Y \mid < K]} > x] \cap [\mid Y \mid < K])dx \ , \end{aligned}$$

no matter whether the left side is finite or infinite. Similarly, finite or infinite, we have by lemma 1,

$$\int_{-\infty}^{0} P([X+Y \leq x])dx = \lim_{K\to\infty} \int_{-\infty}^{0} P([X+YI_{[|Y|<K]} \leq x] \cap [\mid Y \mid < K])dx.$$

Now let us denote $D_K = U_K - L_K$ where

$$U_K = \int_{0}^{\infty} P([X+YI_{[|Y|<K]} > x] \cap [\mid Y \mid < K])dx$$

and

$$L_K = \int_{-\infty}^{0} P([X+YI_{[|Y|<K]} \leq x] \cap [\mid Y \mid < K])dx \ .$$

We first notice that this difference is finite. This follows from case (iii). We next observe that

$$\begin{aligned} D_K &= \textstyle\int_0^\infty P([X+YI_{[|Y|<K]} > x])dx \\ &- \textstyle\int_0^\infty P([X+YI_{[|Y|<K]} > x] \cap [\mid Y \mid \geq K])dx \\ &- \textstyle\int_{-\infty}^0 P([X+YI_{[|Y|<K]} \leq x])dx \\ &+ \textstyle\int_{-\infty}^0 P([X+YI_{[|Y|<K]} \leq x] \cap [\mid Y \mid \geq K])dx \\ &= E(X+YI_{[|Y|<K]}) \\ &- \textstyle\int_0^\infty P([X > x])dx + \int_{-\infty}^0 P([X \leq x])dx \\ &+ \textstyle\int_0^\infty P([X > x] \cap [\mid Y \mid < K])dx - \int_0^\infty P([X \leq x] \cap [\mid Y \mid < K])dx, \end{aligned}$$

which, by case (iii),

$$\begin{aligned} &= E(X) + E(YI_{[|Y|<K]}) - E(X) + \textstyle\int_0^\infty P([X > x] \cap [\mid Y \mid < K])dx \\ &- \textstyle\int_{-\infty}^0 P([X \leq x] \cap [\mid Y \mid < K])dx \ . \end{aligned}$$

Now, applying lemma 1 and its corollary, we obtain

$$\begin{aligned} \lim_{K\to\infty} D_K &= E(Y) + \textstyle\int_0^\infty P([X > x])dx - \int_{-\infty}^0 P([X \leq x])dx \\ &= E(X) + E(Y) \ . \end{aligned}$$

Thus $E(X+Y) = E(X) + E(Y)$, which proves case (iv) and the theorem.

Corollary. *If $X_1, \cdots X_n$ are random variables whose expectations exist, and if $c_1, \cdots, c_n$ are constants, then the expectation of $\sum_{k=1}^n c_k X_k$ exists, and*

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

Proof: This is an easy consequence of theorems 2 and 4.

EXERCISES

1. Use the definition of expectation to prove: if $0 < a < b$ are constants, if X is a random variable, and if $P([a \leq X \leq b]) = 1$, then the expectation of X exists and $a \leq E(X) \leq b$.

2. Prove: if X is a random variable whose expectation exists, if b is a constant, and if $P([X \leq b]) = 1$, then $E(X) \leq b$.

3. A random variable X is said to be symmetric or symmetrically distributed about the number c if, for every $x > 0$, $P([c - x \leq X \leq c]) = P([c \leq X \leq c + x])$. Prove: if X is a random variable whose expectation exists, and if X is symmetric about c, then $E(X) = c$.

4. Prove: if X is a random variable whose expectation exists, and if c is a constant, then $X + c$ is a random variable whose expectation exists, and $E(X + c) = E(X) + c$.

5. Suppose X is a random variable with absolutely continuous distribution and whose density is

$$f_X(x) = \begin{cases} \frac{1}{2} sin(x - 2.3) \text{ if } 2.3 \leq x \leq 2.3 + \pi \\ 0 \text{ otherwise.} \end{cases}$$

Evaluate $E(X)$.

6. Prove: if K and x are positive constants, and if X is a random variable, then

$$[X > x] \cap [| X |< K] = [X I_{[|X|<K]} > x] \ .$$

7. Verify the following: if c and x are real numbers, if A is an event, and if X is a random variable, then

(i) $[X + cI_A > x] \cap A = [X + c > x] \cap A$,

(ii) $[X + cI_A > x] \cap A^c = [X > x] \cap A^c$ and

(iii) $\int_{-c}^{0} P([X \leq y] \cap A)dy + \int_{-c}^{0} P([X > y] \cap A)dy = cP(A)$.

8. Prove: if Y is a random variable that satisfies $0 \leq Y(\omega) \leq n$ for all $\omega \in \Omega$, where n is a positive integer, then

$$\sum_{j=1}^{n} (j-1) I_{[j-1<Y\leq j]}(\omega) \leq Y(\omega) \leq \sum_{j=1}^{n} j I_{[j-1<Y\leq j]}(\omega)$$

for all $\omega \in \Omega$.

9. In problem 5 above, prove that the random variable X is symmetrically distributed about $2.3 + \frac{\pi}{2}$.

3.3 Moments and Central Moments. The purpose of this section is to define the various kinds of moments used in statistics and to derive some of their properties. We note first that powers of random variables are also random variables; we obtained this in theorem 5 in section 2.1.

Definition. If r is a positive integer, and if X is a random variable, we define the rth moment of X to be $m_r = E(X^r)$, provided the expectation exists. We define the rth central moment to be $\mu_r = E((X - E(X))^r)$, provided $E(X)$ and $E((X-E(X))^r)$ both exist. The second central moment, μ_2, of X is called the variance of X and is denoted by $VarX$ or $Var(X)$.

Theorem 1. *If X is a random variable, if $r < s$ are positive integers, and if $E(X^s)$ exists, then $E(X^r)$ exists.*

Proof: Since by hypothesis $E(X^s)$ exists, then by definition,

$$E(X^s) = \int_0^\infty P([X^s > x])dx - \int_{-\infty}^0 P([X^s \leq x])dx \ ,$$

and both integrals are finite. We now prove that $\int_0^\infty P([X^r > x])dx < \infty$. Indeed, by the same argument used to prove proposition 1 in section 2.3, we have

$$0 \leq \int_0^1 P([X^r > x])dx = \int_0^1 P([X^r \geq x])dx \leq 1$$

and

$$0 \leq \int_{-1}^0 P([X^r < x])dx = \int_{-1}^0 P([X^r \leq x])dx \leq 1 \ .$$

Now for $1 \leq x < \infty$, $x^{1/r} \geq x^{1/s}$, and thus $P([X > x^{1/r}]) \leq P([X > x^{1/s}])$ and $P([X < -x^{1/r}]) \leq P([X < -x^{1/s}])$. Hence if r is odd,

$$\begin{aligned} \textstyle\int_0^\infty P([X^r > x])dx &\leq 1 + \textstyle\int_1^\infty P([X > x^{1/r}])dx \\ &\leq 1 + \textstyle\int_1^\infty P([X > x^{1/s}])dx \\ &\leq 1 + \textstyle\int_0^\infty P([X^s > x])dx < \infty \ . \end{aligned}$$

But if r is even, then

$$\begin{aligned} \textstyle\int_0^\infty P([X^r > x])dx &\leq 1 + \textstyle\int_1^\infty (P([X < -x^{1/r}]) + P([X > x^{1/r}]))dx \\ &\leq 1 + \textstyle\int_1^\infty (P([X < -x^{1/s}]) + P([X > x^{1/s}]))dx \ . \end{aligned}$$

If s is odd, then by lemma 1 in section 3.1, we have

$$\int_1^\infty P([X < -x^{1/s}])dx = \int_1^\infty P([X^s \leq -x])dx = \int_{-\infty}^{-1} P([X^s \leq x])dx < \infty \ ,$$

and if s is even, then

$$\int_1^\infty P([X < -x^{1/s}])dx = \int_1^\infty P([X > x^{1/s}])dx < \infty.$$

No matter whether s is odd or even, we have

$$\int_1^\infty P([X > x^{1/s}])dx \leq \int_1^\infty P([X^s > x])dx < \infty \ ,$$

and thus we have shown that $\int_0^\infty P([X^r > x])dx < \infty$. In a similar manner, one can show that $\int_{-\infty}^0 P([X^r \leq x])dx < \infty$, and the theorem is proved.

Corollary to Theorem 1. *If the rth moment of a random variable exists, then its rth central moment exists.*

Proof: This follows easily by theorem 1 and the binomial theorem.

Theorem 2. *If X is a discrete random variable that takes values $x_1, x_2, \cdots$ with probabilities $p_1, p_2, \cdots$ respectively, i.e., $P([X = x_n]) = p_n$ for all n and $\sum_{n\geq 1} p_n = 1$, then the rth moment of X exists if and only if the series $\sum_{n\geq 1} \mid x_n \mid^r p_n < \infty$, in which case $E(X^r) = \sum_{n\geq 1} x_n^r p_n$.*

Proof: By the definition of a discrete random variable, $\Omega = \cup_{n\geq 1}[X = x_n]$. Now

$$[X^r = x] = \bigcup_{\{n: x_n^r = x\}} [X = x_n].$$

Hence if $\{y_1, y_2, \cdots\}$ are the distinct values of the set $\{x_1^r, x_2^r, \cdots\}$, and if we denote $q_n = \sum\{p_j : x_j^r = y_n\}$, we have

$$P([X^r = y_n]) = P\left(\bigcup_{\{j: x_j^r = y_n\}} [X = x_j]\right) = \sum_{\{j: x_j^r = y_n\}} p_j = q_n.$$

Hence by theorem 1 in section 3.2, $E(X^r)$ exists if and only if the series $\sum\limits_n \mid y_n \mid q_n < \infty$. But

$$\sum_n |y_n| q_n = \sum_n |y_n| \left\{ \sum_{\{j: x_j^r = y_n\}} p_j \right\} = \sum_n \left\{ |y_n| \sum_{\{j: x_j^r = y_n\}} p_j \right\} = \sum_j |x_j^r| p_j \ ,$$

from which we obtain the theorem.

Lemma 1. *If Z is a random variable with an absolutely continuous distribution, and if $h : [a, b] \to \mathbf{R}^1$ is a nondecreasing function over the interval $[a, b]$, then*

$$h(a)P([a \leq Z < b]) \leq \int_a^b h(x)f_Z(x)dx \leq h(b)P([a \leq Z < b]) \ .$$

Proof: This easily follows from the following:

$$h(a)P([a \leq Z < b]) = h(a)\int_a^b f_Z(x)dx = \int_a^b h(a)f_Z(x)dx \leq \int_a^b h(x)f_Z(x)dx \ ,$$

the last inequality following from the fact that the density is nonnegative and h is nondecreasing. The second inequality follows by similar reasoning.

Lemma 2. *If Z is a random variable, then*

$$\sum_{k=1}^{\infty} \frac{1}{n} P\left(\left[Z \geq \frac{k}{n}\right]\right) \leq \int_0^{\infty} P([Z > x])dx.$$

Proof: Let N be a positive integer, and define

$$g(x) = P\left(\left[Z \geq \frac{k}{n}\right]\right) \text{ whenever } \frac{k-1}{n} \leq x < \frac{k}{n} \text{ for all } x \in [0, \infty).$$

It follows that $g(x) \leq P([Z \geq x])$ for all $x \geq 0$. Thus

$$\int_0^N g(x)dx = \sum_{k=1}^{N} \frac{1}{n} P\left(\left[Z \geq \frac{k}{n}\right]\right).$$

But $\int_0^N g(x)dx \leq \int_0^N P([Z \geq x])\, dx$, and by the definition of the improper integral, the conclusion follows.

Theorem 3. *If X is a random variable with absolutely continuous distribution function, and if r is a positive integer, then $E(X^r)$ exists if and only if $\int_{-\infty}^{\infty} \mid x \mid^r f_X(x)dx < \infty$, in which case*

$$E(X^r) = \int_{-\infty}^{\infty} x^r f_X(x)dx \ .$$

Proof: Case (i) r is even. In this case, $X^r(\omega) \geq 0$ for every $\omega \in \Omega$, and $h(x)$ defined by $h(x) = x^r$ is increasing over $[0, \infty)$ and is decreasing over $(-\infty, 0]$. Thus, if we denote $I = \int_0^\infty P([X^r > x])dx$, it follows from the definition of expectation that $E(X^r)$ exists if and only if $I < \infty$, in which case $E(X^r) = I$. Let us denote

$$\pi(x) = \begin{cases} (-\infty, -x^{1/r}) \cup (x^{1/r}, \infty) \text{ if } x \geq 0 \\ (-\infty, \infty) \text{ if } \ x < 0 \ . \end{cases}$$

Then, if $0 \leq x' < x''$, it follows that $\pi(x'') \subset \pi(x')$. Also $P([X^r > x]) = P([X \in \pi(x)])$. Let us define $g(x) = P([X^r > x])$; then $g(x) = \int_{\pi(x)} f_X(t)dt$ is a nonincreasing function of x, $g(x) = 1$ if $x = 0$, and $g(x) \to 0$ as $x \to \infty$. Now denote

$$L_n = \sum_{k=1}^{\infty} \frac{1}{n} \int_{\pi(\frac{k}{n})} f_X(t)dt$$

and

$$U_n = \sum_{k=1}^{\infty} \frac{1}{n} \int_{\pi(\frac{k-1}{n})} f_X(t)dt \ .$$

By the above properties of g, it follows that $L_n \leq I \leq U_n$ for $n = 1, 2, \cdots$. (Proof: $L_n \leq I$ follows from the fact that by lemma 2 and the fact that r is even,

$$\begin{aligned} \sum_{k=1}^{\infty} \tfrac{1}{n} \textstyle\int_{\pi(\frac{k}{n})} f_X(t)dt &= \sum_{k=1}^{\infty} \tfrac{1}{n} P([X \in \pi(\tfrac{k}{n})]) \\ &= \sum_{k=1}^{\infty} \tfrac{1}{n} P([X^r > \tfrac{k}{n}]) \\ &\leq \int_0^{\infty} P([X^r > x])dx.) \end{aligned}$$

Now let $a_k = \frac{1}{n} \int_{\pi(\frac{k-1}{n}) \setminus \pi(\frac{k}{n})} f_X(t)dt$, and applying lemma 1 above and lemma 2 in section 3.1, we obtain

$$\begin{aligned} L_n &= \sum_{k=1}^{\infty} \tfrac{k-1}{n} \textstyle\int_{\pi(\frac{k-1}{n}) \setminus \pi(\frac{k}{n})} f_X(t)dt \\ &= \sum_{k=1}^{\infty} \tfrac{k}{n} \textstyle\int_{\pi(\frac{k-1}{n}) \setminus \pi(\frac{k}{n})} f_X(t)dt - \tfrac{1}{n} \\ &\geq \sum_{k=1}^{\infty} \textstyle\int_{\pi(\frac{k-1}{n}) \setminus \pi(\frac{k}{n})} t^r f_X(t)dt - \tfrac{1}{n} \\ &= \textstyle\int_{-\infty}^{\infty} t^r f_X(t)dt - \tfrac{1}{n} \ . \end{aligned}$$

In a similar manner, $U_n \leq \int_{-\infty}^{\infty} t^r f_X(t)dt + \frac{1}{n}$. Thus, for every positive integer n,

$$\int_{-\infty}^{\infty} t^r f_X(t)dt - \frac{1}{n} \leq I \leq \int_{-\infty}^{\infty} t^r f_X(t)dt + \frac{1}{n} \ ,$$

which yields the conclusions of the theorem in case r is even. Case (ii) r is odd. In this case, define

$$I_1 = \int_0^{\infty} P([X > x])dx \text{ and } I_2 = \int_{-\infty}^{0} P([X \leq x])dx \ .$$

Then by definition of expectation, $E(X^r)$ exists if and only if both $I_1 < \infty$ and $I_2 < \infty$.For every $x > 0$, define $\pi_1(x) = (x^{1/r}, \infty)$ and for $x < 0$ define $\pi_2(x) = (-\infty, x^{1/r})$. Using $\pi_1(x)$ and $\pi_2(x)$ as we did in the proving the theorem in case (i) when r was even, we obtain

$$I_1 = \int_0^{\infty} x^r f_X(x)dx \text{ and } \ I_2 = -\int_{-\infty}^{0} x^r f_X(x)dx \ ,$$

which yields the conclusion of the theorem in this case, thus concluding the proof of the theorem.

EXERCISES

1.Prove: if U is a random variable whose nth moment exists, then the nth central moment exists and

$$E((U - E(U))^n) = \sum_{j=0}^{n} \binom{n}{j} E(U^j)(-E(U))^{n-j} \ .$$

2. Suppose Y is a random variable that has the *uniform distribution* over $[0, 1)$, i.e., it has an absolutely continuous distribution with density given by

$$f_Y(y) = \begin{cases} 1 \text{ if } 0 \leq y < 1 \\ 0 \text{ otherwise.} \end{cases}$$

Compute $E(Y^n)$ for any positive integer n.

3. In problem 2, compute the expectation and the variance of Y.

4. Suppose Z is a random variable that is uniformly distributed over the set of numbers $\{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n}{n}\}$, i.e., $P([Z = \frac{i}{n}]) = \frac{1}{n}$ for $1 \leq i \leq n$, where n is a positive integer. Derive the formula for $E(Z)$.

5. Prove: if Y and Z_n are random variables as in problems 2 and 4, prove that $0 \leq F_Y(y) - F_{Z_n}(y) < \frac{1}{n}$ for all real values of y.

6. Prove: if Y and Z_n are random variables as in problems 2 and 4, then $F_{Z_n}(y) \to F_Y(y)$ uniformly in y as $n \to \infty$, i.e., prove that for every $\epsilon > 0$ there exists a positive integer N such that $\mid F_Y(y) - F_{Z_n}(y) \mid < \epsilon$ for all values of y and for all $n > N$.

7. **Definition.** If X is a nonnegative integer-valued random variable, its probability generating function $g_X(u)$ is defined by

$$g_X(u) = \sum_{n=0}^{\infty} P([X = n])u^n = E(u^X)$$

for all values of $u \in [-1, 1]$. Prove that

(ii) $g_X(1) = 1$,

(ii) if the expectation of X exists, then $E(X) = g'_X(1)$.

3.4 Variances of Some Special Distributions. In section 3.3 we introduced the notions of moments and central moments. A special case of central moment was singled out, namely that which is usually referred to as the variance of a random variable. This is an important concept in probability and statistics. It is a natural measure of the dispersion of probability mass along the real line, and in mechanics it is called the moment of inertia. Our purpose in this section is to establish a few basic properties of variance and to derive the formulas for the variances of six important discrete distributions. As a reminder, let us recall that the variance of a random variable X, $Var(X)$, is defined by $Var(X) = E((X - E(X))^2)$

Theorem 1. *If X is a random variable with finite second moment, then* $Var(X) = E(X^2) - (E(X))^2$.

Proof: By the corollary to theorem 1 in section 3.3, the existence of $E(X^2)$ implies the same for $Var(X)$. Using results already established for expectation we have

$$\begin{aligned} Var(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2E(X)X + (E(X))^2) \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2) \\ &= E(X^2) - (E(X))^2, \end{aligned}$$

which yields or formula.

Theorem 2. *If X is a random variable with finite second moment, and if C is any constant, then $Var(CX) = C^2 Var(X)$.*

Proof: By already established properties of expectation, we see that $E((CX)^2) = E(C^2X^2) = C^2E(X^2)$ and $E(CX) = CE(X)$. By these results and by theorem 4,

$$\begin{aligned} Var(CX) &= E((CX)^2) - (E(CX))^2 \\ &= C^2E(X^2) - C^2(E(X))^2 \\ &= C^2Var(X) \ . \end{aligned}$$

Theorem 3. *If X is a random variable with finite second moment, and if C is any constant, then $X + C$ has a finite second moment, and $Var(X + C) = Var(X)$.*

Proof: If X has finite second moment, then $E(X^2) - 2CE(X) + C^2$ is a finite well-defined number. But $E(X^2) - 2CE(X) + C^2 = E(X^2 + 2CX + C^2) = E((X - E(X))^2) = Var(X)$. The second conclusion follows directly from the definition of variance and the fact that $E(X + C) = E(X) + C$.

Now let us derive formulas for variances of the six important discrete distributions.

Example 1. The binomial distribution. Suppose X is a random variable with the binomial distribution, i.e.,

$$P([X = k]) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0, 1, 2, \cdots, n.$$

We shall compute $Var(X)$. In section 3.1 we found that $E(X) = np$. Now

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{n} k^2 \tbinom{n}{k} p^k (1-p)^{n-k} \\ &= p \sum_{k=1}^{n} (k(k-1) + k) \tbinom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k-1=0}^{n-1} (k-1) \tbinom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &\quad + np \sum_{k-1=0}^{n-1} \tbinom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \ . \end{aligned}$$

Now the first of the two sums in this last expression is the expectation of a random variable whose distribution is $Bin(n-1, p)$ whose expectation is $(n-1)p$. The second sum is just $(p + 1 - p))^{n-1} = 1$. Thus

$$E(X^2) = np(n-1)p + np = (np)^2 - np^2 + np \ .$$

Using theorem 1 we obtain $Var(X) = (np)^2 - np^2 + np - (np)^2 = np(1-p)$.

Example 2. The geometric distribution. We now find the variance of a random variable, X, whose distribution is $geom(p)$, i.e., $P([X = n]) = (1-p)^{n-1}p$ for $n = 1, 2, \cdots$, where $0 < p < 1$. By theorem 2,

$$\begin{aligned} E(X^2) &= \sum_{n=1}^{\infty} n^2(1-p)^{n-1}p = \sum_{n=1}^{\infty} n((n-1)+1)(1-p)^{n-1}p \\ &= (1-p)p\sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2} + p\sum_{n=1}^{\infty} n(1-p)^{n-1}. \end{aligned}$$

Since $0 < 1-p < 1$, the two series in this last expression can be expressed as term by term second and first derivatives respectively, and thus

$$\begin{aligned} E(X^2) &= (1-p)p\tfrac{d^2}{dp^2}\left(\sum_{n=0}^{\infty}(1-p)^n\right) - p\tfrac{d}{dp}\left(\sum_{n=0}^{\infty}(1-p)^n\right) \\ &= (1-p)p\tfrac{d^2}{dp^2}\left(\tfrac{1}{1-(1-p)}\right) - p\tfrac{d}{dp}\left(\tfrac{1}{1-(1-p)}\right) \\ &= (1-p)\tfrac{2}{p^2} + p\tfrac{1}{p^2} \\ &= \tfrac{2-p}{p^2}. \end{aligned}$$

By example 2 in section 3.1, $E(X) = \frac{1}{p}$. Thus, by theorem 1,

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= \tfrac{2-p}{p^2} - \tfrac{1}{p^2}\ , \text{ or} \\ Var(X) &= \tfrac{1-p}{p^2}\ . \end{aligned}$$

Example 3. The Poisson distribution. The discrete density of the Poisson distribution is given by

$$P([X = n]) = e^{-\lambda}\frac{\lambda^n}{n!} \text{ for } n = 0, 1, 2, \cdots\ .$$

In example 3 in section 3.1, we found that $E(X) = \lambda$. In order to find the formula for $Var(X)$, we shall find $E(X^2)$ and then apply theorem 1. By theorem 2, by the same differentiating techniques as in the previous example, and remembering the expansion $e^z = \sum_{n=0}^{\infty} z^n/n!$, we have

$$\begin{aligned} E(X^2) &= \sum_{n=0}^{\infty} n^2 e^{-\lambda}\tfrac{\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=0}^{\infty} n((n-1)+1)\tfrac{\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=2}^{\infty} n(n-1)\tfrac{\lambda^n}{n!} + e^{-\lambda}\sum_{n=1}^{\infty} n\tfrac{\lambda^n}{n!} \\ &= e^{-\lambda}\lambda^2\tfrac{d^2}{d\lambda^2}e^{\lambda} + e^{-\lambda}\lambda\tfrac{d}{d\lambda}e^{\lambda} = \lambda^2 + \lambda\ . \end{aligned}$$

Thus, by theorem 1, we have $Var(X) = \lambda$.

We next wish to derive formulas for the variances of distributions of the sample survey sum distribution and its special cases. For this we need the following lemma.

Lemma 1. *If* $x_1, x_2, \cdots, x_N$ *are a finite sequence of not necessarily distinct numbers, if* $2 \leq n \leq N$, *and if* $f(x, y)$ *is any function of two variables, then*

$$\sum_{1 \leq i_1 < \cdots < i_n \leq N} \left\{ \sum_{1 \leq u < v \leq n} f(x_{i_u}, x_{i_v}) \right\} = \binom{N-2}{n-2} \sum_{1 \leq u < v \leq N} f(x_u, x_v) \ .$$

Proof: Let u, v be integers that satisfy $1 \leq u < v \leq N$. Then $f(x_u, x_v)$ can appear in exactly $\binom{N-2}{n-2}$ sums of the form

$$\sum_{1 \leq u < v \leq n} f(x_{i_u}, x_{i_v}) \ ,$$

which yields the above result.

Example 4. The sample survey sum distribution. Let X denote the sum of a simple random sample of size n taken without replacement from the N not necessarily distinct numbers $x_1, x_2, \cdots, x_N$. In section 3.1 we obtained $E(X) = n\overline{x}$, where $\overline{x} = \frac{1}{N} \sum_{j=1}^{N} x_j$. By the lemma from section 3.1 and by the lemma above, we have

$$\begin{aligned}
E(X^2) &= \sum_{1 \leq i_1 < \cdots < i_n \leq N} \left(\sum_{j=1}^{n} x_{i_j} \right)^2 \frac{1}{\binom{N}{n}} \\
&= \sum_{1 \leq i_1 < \cdots < i_n \leq N} \sum_{j=1}^{n} x_{i_j}^2 \frac{1}{\binom{N}{n}} \\
&\quad +2 \sum_{1 \leq i_1 < \cdots < i_n \leq N} \sum_{1 \leq u < v \leq n} x_{i_u} x_{i_v} \frac{1}{\binom{N}{n}} \\
&= \frac{\binom{N-1}{n-1}}{\binom{N}{n}} \sum_{j=1}^{N} x_j^2 + 2 \frac{\binom{N-2}{n-2}}{\binom{N}{n}} \sum_{1 \leq u < v \leq N} x_u x_v \\
&= \frac{n}{N} \sum_{j=1}^{N} x_j^2 + 2 \frac{n(n-1)}{N(N-1)} \sum_{1 \leq u < v \leq N} x_u x_v \ .
\end{aligned}$$

Now

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
&= \tfrac{n}{N}\sum_{j=1}^{N} x_j^2 + \tfrac{n(n-1)}{N(N-1)}\left\{\left(\sum_{j=1}^{N} x_j\right)^2 - \sum_{j=1}^{N} x_j^2\right\} - n^2\overline{x}^2 \\
&= \tfrac{n(N-n)}{N(N-1)}\sum_{j=1}^{N} x_j^2 - \overline{x}^2\tfrac{n(N-n)}{N-1} \\
&= \tfrac{n(N-n)}{N}\tfrac{1}{N-1}\left\{\sum_{j=1}^{N} x_j^2 - N\overline{x}^2\right\}.
\end{aligned}$$

Thus

$$Var(X) = n\frac{(N-n)}{N}\frac{1}{N-1}\left\{\sum_{j=1}^{N} x_j^2 - N\overline{x}^2\right\} .$$

Recalling that

$$\sum_{j=1}^{N} x_j^2 - N\overline{x}^2 = \sum_{i=1}^{N}(x_i - \overline{x})^2 ,$$

we also have

$$Var(X) = n\left(1 - \frac{n}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 .$$

Example 5. The Wilcoxon distribution. If W is a random variable with the $Wilcoxon(n, N)$ distribution, then it is the sum of a simple random sample of size n taken from the integers $1, 2, \cdots, N$ without replacement. We obtained in section 3.1 that $E(W) = n(N+1)/2$. Recall from freshman calculus courses that

$$\sum_{i=1}^{N} i = \frac{N(N+1)}{2} \text{ and } \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6} .$$

We now apply the formula for variance obtained in example 4, where we take $x_i = i$ for $1 \le i \le N$, to obtain

$$Var(W) = n\frac{N-n}{N}\frac{1}{N-1}\left\{\frac{N(N+1)(2N+1)}{6} - N\frac{(N+1)^2}{4}\right\} .$$

After a little algebra, we arrive at

$$Var(W) = \frac{1}{12} n(N-n)(N+1)\,.$$

Example 6. The hypergeometric distribution. Let H denote the number of red balls in a simple random sample of size n selected without replacement from an urn that contains N balls, of which R are red. The problem is to find the formula for $Var(H)$. We determined in section 3.1 that $E(H) = n\frac{R}{N}$. We now apply the result of example 4, where we let R of the x_i's equal 1 and the remaining $N-R$ of them equal 0. Then $\sum_{j=1}^{N} x_j = \sum_{j=1}^{N} x_j^2 = R$ and

$$Var(H) = n\frac{N-n}{N}\frac{1}{N-1}\left\{R - N\frac{R^2}{N^2}\right\}.$$

Again, after a little algebra, we obtain the formula

$$Var(H) = n\frac{N-n}{N(N-R)}R(N-R)\ .$$

Example 7. The uniform$[0,1]$ distribution. Let X be a random variable that is uniformly distributed over $[0,1]$, i.e.,

$$f_X(x) = \begin{cases} 1 \text{ if } 0 \le x \le 1 \\ 0 \text{ if } x < 0 \text{ or } x > 1. \end{cases}$$

Then

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x dx = \frac{1}{2}.$$

Also,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^1 x^2 dx = \frac{1}{3}.$$

Thus

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{12}.$$

Example 8. The exponential distribution. This absolutely continuous distribution function has a density given by

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} \text{ if } x \ge 0 \\ 0 \text{ if } x < 0, \end{cases}$$

where $\alpha > 0$ is a positive constant. By example 8 in section 3.1, $E(X) = 1/\alpha$. Now

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^{\infty} \alpha x^2 e^{-\alpha x} dx \\ &= \tfrac{1}{\alpha^2} \int_0^{\infty} y^2 e^{-y} dy = \tfrac{2}{\alpha^2}. \end{aligned}$$

Thus,

$$Var(X) = \frac{2}{\alpha^2} - (\frac{1}{\alpha})^2 = \frac{1}{\alpha^2}.$$

EXERCISES

1. Let X be a random variable with discrete density $f_X(k) = \frac{1}{3}$ for $k = 1, 2$ and 3. Define $h(x) = E((X - x)^2)$.

(i) Express h in the form $h(x) = ax^2 + bx + c$, and solve for a, b, c.

(ii) Compute $E(X)$ and $Var(X)$.

(iii) Find the minimum value of $h(x)$, and find the value of x at which this minimum value is achieved.

2. Prove: if X is a random variable with finite second moment, then the smallest value that the function $h(x) = E((X - x)^2)$ can achieve is $h(E(X)) = Var(X)$.

3. Prove: if X is a random variable with finite variance, then $-\sqrt{E(X^2)} \leq E(X) \leq \sqrt{E(X^2)}$.

4. Compute the variance of a random variable Z that has the uniform distribution over $[0, 1]$.

5. Compute the variance of a random variable W with absolutely continuous distribution with density

$$f_W(w) = \begin{cases} \frac{1}{6} \text{ if } 14 \leq w \leq 20 \\ 0 \text{ otherwise.} \end{cases}$$

6. Let U and V be random variables with absolutely continuous distributions, and suppose that

$$f_U(t) = f_V(-t) \text{ for all real } t.$$

Prove that $E(U) = -E(V)$ and $Var(U) = Var(V)$.

7. Find the expectation and the variance of a random variable Z that has an absolutely continuous distribution with density given by

$$f_Z(z) = \begin{cases} \frac{1}{2} z^2 e^{-z} \text{ if } z \geq 0 \\ 0 \text{ if } z < 0. \end{cases} .$$

8. Find the expectation and variance of a random variable W, the sum of the upturned faces resulting from tossing a pair of fair dice.

9. Prove that if X is a non-constant random variable whose expectation exists, then $P([X < E(X)]) > 0$.