Chapter III Random Variables

3.1 Random variables

A sample space S may be difficult to describe if the elements of S are not numbers. We shall discuss how we can use a rule by which an element s of S may be associated with a number x.

Definition 3.1-1: Given a random experiment with a sample space S, a function X that assigns to each element s in S one and only one real number $X(s) = x$ is called a ***random variable***. The space of X is the set of real numbers $\{x : x = X(s), s \in S\}$, where $s \in S$ means the element s belongs to the set S. ◘

Example 3.1-1: In example 2.3-1, we had the sample space $S = \{H, T\}$. Let X be a function defined on S such that $X(H) = 1$ and $X(T) = 0$. Thus X is a real-function that has the sample space S as its domain and the space of real numbers $\{x : x = 0, 1\}$ as its range. ◘

Note that it may be that the set S has elements that are themselves real numbers. In such instance we could write $X(s) = s$ so that X is the identity function and the space of X is also S.

Example 3.1-2: In example 2.3-2, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. For each $s \in S$, let $X(s) = s$. The space of the random variable X is then $\{1, 2, 3, 4, 5, 6\}$. ◘

If we want to find the probabilities associated with events described in terms of X, we use the probabilities of those events in the original space S if they are ***known***. For instance,

$$P(a < X < b) = P\{s : s \in S \text{ and } a < X(s) < b\}.$$

The probabilities are ***induced*** on the points of the space of X by the probabilities assigned to outcomes of the sample space S through the function X. Hence, the probability $P(X = x)$ is often called an ***induced probability***.

Example 3.1-3: In example 2.3-1, we associate a probability of $1/2$ for each outcome, then, for example, $P(X = 1) = P(X = 0) = 1/2$. ◘

Example 3.1-4: Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. If we bet that at least one of the drawn balls has a number as large as or larger than 17, what is the probability that we win the bet?

Solution: Let X denote the largest number selected. Then X is a random variable taking on one of the values 3, 4,…, 20. Furthermore, if we suppose that each of the $\binom{20}{3}$ possible selections is ***equally likely to occur***, then

$$P\{X = i\} = \binom{i-1}{2} \Big/ \binom{20}{3} \qquad i = 3, 4, \ldots, 20.$$

The above equation follows because the number of the selections that result in the event $\{X = i\}$ is just the number of selections that result in ball numbered i and two of the balls numbered 1 through $i - 1$ being chosen. As there are clearly $\binom{1}{1}\binom{i-1}{2}$ such selections, we obtain the probabilities expressed in above equation. From this equation we see that

$$P\{X = 20\} = \binom{19}{2} \Big/ \binom{20}{3} = \frac{3}{20} = 0.15$$

$$P\{X = 19\} = \binom{18}{2} \Big/ \binom{20}{3} = \frac{51}{380} = 0.134$$

$$P\{X = 18\} = \binom{17}{2} \Big/ \binom{20}{3} = \frac{34}{285} = 0.119$$

$$P\{X = 17\} = \binom{16}{2} \Big/ \binom{20}{3} = \frac{2}{19} = 0.105$$

Hence, as the event $\{X \geq 17\}$ is the union of the disjoint (mutually exclusive) events $\{X = i\}$, i = 17, 18, 19, 20, it follows that the probability of our winning the bet is given by

$$P\{X \geq 17\} = 0.105 + 0.119 + 0.134 + 0.15 = 0.508. \quad ◘$$

There are two major difficulties here:

(1) In many practical situations the probabilities assigned to the events A of the sample space S are unknown.

(2) Since there are many ways of defining a function X on S, which function do we want to use?

In considering (1), we need, through repeated observations (called sampling), to estimate these probabilities or percentages. One obvious way of estimating these is by use of the relative frequency after a number of observations. If additional assumptions can be made, we will study, in this course, other ways of estimating probabilities. It is this latter aspect with which mathematical statistics is concerned. That is, if we assume certain models, we find that the theory of statistics can explain how best to draw conclusions or make predictions.

For (2), statisticians try to determine what measurement (or measurements) should be taken on an outcome; that is, how best do we "mathematize" the outcome? These measurement problems are most difficult and can only be answered by getting involved in a practical project. Nevertheless, in many instances it is clear exactly what function X the experimenter wants to define on the sample space. For example, the die game in Example 2.3-2 is concerned about the number of the spot, say X, which is up on the die.

Definition 3.1-2: A random variable X is said to be ***discrete*** if it can assume only a ***finite*** or ***countably infinite***[1] number of distinct values. ◘

---

[1] A set of elements is ***countably infinite*** if the elements in the set can be put into one-to-one correspondence with the

Definition 3.1-3: The probability that X takes on the value x, $P(X = x)$, is the defined as the ***sum of the probabilities of all sample points in S*** that are assigned the value x. ◘

For a random variable X of the ***discrete*** type, the induced probability $P(X = x)$ is frequently denoted by $f(x)$, and this function $f(x)$ is called the ***discrete probability density function*** (pdf). **Note** that some authors refer to $f(x)$ as the probability function (pf), the frequency function, or the ***probability mass function*** (pmf). We will use the terminology pmf in this course.

Example 3.1-5: A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let X denote the number of women in his selection. Find the probability distribution for X.

Solution: The supervisor can select two workers from six in $\binom{6}{2} = 15$ ways. Hence S contains 15 sample points, which we assume to be equally likely because random sampling was employed. Thus $P(E_i) = 1/15$, for $i = 1, 2, \ldots, 15$. The value for X that have nonzero probability are 0, 1, and 2. The number of ways of selecting $X = 0$ women is $\binom{3}{0}\binom{3}{2}$. Thus there are $\binom{3}{0}\binom{3}{2} = 3$ sample points in the event $X = 0$, and

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Similarly,

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5}$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Table or histogram can represent the above results, but the most concise method of representing discrete probability distributions is by means of a formula. The formula for $f(x)$ can be written as

$$f(x) = \frac{\binom{3}{x}\binom{3}{2-x}}{\binom{6}{2}}, \qquad x = 0, 1, 2.$$

Notice that the probabilities associated with all distinct value of a discrete random variable must sum to one. ◘

---

positive integers.

Theorem 3.1-1: A function $f(x)$ is a pmf if and only if it satisfies both of the following properties for at most a countably infinite set of real values $x_1, x_2, \cdots$:

1. $0 \le f(x_i) \le 1$ for all x_i.
2. $\sum_{\text{all } x_i} f(x_i) = 1$.

Proof: $0 \le f(x_i) \le 1$ follows from the fact that the value of a discrete pmf is a probability and must be nonnegative. Since $x_1, x_2, \cdots$ represent all possible values of X, the events $[X = x_1], [X = x_2], \cdots$ constitute an exhaustive partition of the sample space. Thus,

$$\sum_{\text{all } x_i} f(x_i) = \sum_{\text{all } x_i} P[X = x_i] = 1. \quad ◘$$

Definition 3.1-4: The ***cumulative distribution function*** (CDF), or simply referred to as the ***distribution function***, of a discrete random variable X is defined for any real x by

$$F(x) = P(X \le x) = \sum_{t \le x} f(t). \quad ◘$$

Example 3.1-6: The probability mass function of a random variable X is given by $f(x) = c\lambda^x / x!$, $x = 0, 1, \ldots$, where λ is some positive value. Find $P\{X = 0\}$ and $P\{X > 2\}$.

Solution: Since $\sum_{x=0}^{\infty} f(x) = 1$, we have that

$$c \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1$$

implying, because $e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$, that $ce^{\lambda} = 1$ or $c = e^{-\lambda}$. Hence,

$P\{X = 0\} = e^{-\lambda} \lambda^0 / 0! = e^{-\lambda}$.

$P\{X > 2\} = 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\} = 1 - e^{-\lambda} - \lambda e^{-\lambda} - \lambda^2 e^{-\lambda}/2$.

The cumulative distribution function $F(\cdot)$ can be expressed in terms of $f(\cdot)$ by

$$F(a) = \sum_{\text{all } x \le a} f(x), \qquad a = 0, 1, 2, \ldots. \quad ◘$$

Definition 3.1-5: A random variable X is called a continuous random variable if there is a function $f(x)$, called the probability density function (pdf) of X, such that the CDF can be represented as

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \quad ◘$$

The above defining property provides a way to derive the CDF when the pdf is given, and it follows by the Fundamental Theorem of Calculus that the pdf can be obtained from the CDF by differentiation. Specifically,

$$f(x) = \frac{d}{dx} F(x) = F'(x)$$

whenever the derivative exists.

Theorem 3.1-2: A function $f(x)$ is a pdf for some continuous random variable X if and only if it satisfies the properties

1. $f(x) \ge 0$ for all real x.
2. $\int_{-\infty}^{\infty} f(x) = 1$. ◘

Properties of a CDF $F(x)$:

(a) $0 \le F(x) \le 1$ since $F(x)$ is a probability.

(b) $F(x)$ is a non-decreasing function of x. For instance, if $a < b$, then $\{x : x \le b\} = \{x : x \le a\} \cup \{x : a < x \le b\}$ and $P(X \le b) = P(X \le a) + P(a < X \le b)$. That is, $F(b) - F(a) = P(a < X \le b) \ge 0$.

(c) From the proof of (b), it is observed that if $a < b$, then $P(a < x \le b) = F(b) - F(a)$.

(d) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$ because the set $\{x : x \le \infty\}$ is the entire one-dimensional space and the set $\{x : x \le -\infty\}$ is the null set.

(e) $F(x)$ is continuous to the right at each point x.

(f) If X is a random variable of the ***discrete*** type, then $F(x)$ is a ***step function***, and the height of a step at x, $x \in \Re$, equals the probability $P(X = x)$.

(g) If X is a continuous random variable, then $F(x)$ is a continuous function. The probability $P(a < x \le b)$ is the area bounded by the graph of $f(x)$**,** the x-axis, and the lines $x = a$ and $x = b$. Furthermore, the probability at any particular point is zero.

Example 3.1-7: Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

(a) What is the value c?

(b) Find $P\{X > 1\}$?

Solution: Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that $c\int_0^2 (4x - 2x^2)\,dx = 1$. Hence, $c\left[2x^2 - \frac{2x^3}{3}\right]\Bigg|_{x=0}^{x=2} = 1 \quad \Rightarrow c = \frac{3}{8}$.

$P\{X > 1\} = (3/8)\int_1^2 (4x - 2x^2)\,dx = 1/2$.

The cumulative distribution function F is given by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x (3/8)(4t - 2t^2)\,dt = (3/8)\left(2x^2 - \left(2x^3/3\right)\right), \quad 0 < x < 2. \quad ◘$$

Example 3.1-8: The distribution function of the random variable Y is given by

$$F(y) = \begin{cases} 0 & y < 0 \\ y/2 & 0 \le y < 1 \\ 2/3 & 1 \le y < 2 \\ 11/12 & 2 \le y < 3 \\ 1 & 3 \le y. \end{cases}$$

A graph of $F(y)$ is presented in Figure 3-1.

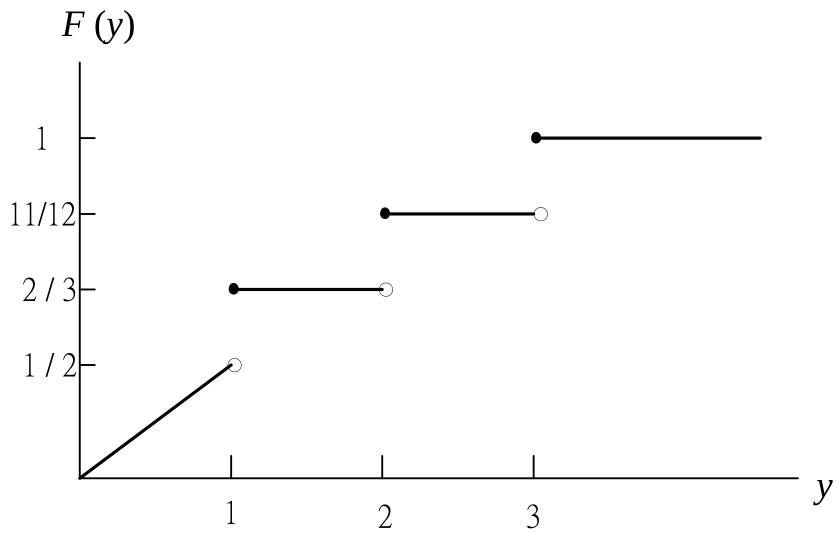

Figure 3-1 Graph of F (y)

Compute $P\{Y < 3\}$, $P\{Y = 1\}$, $P\{Y > 0.5\}$, and $P\{2 < Y \le 4\}$.

Solution: $P\{Y < 3\} = \lim_{n\to\infty} P\left\{Y \le 3 - \dfrac{1}{n}\right\} = \lim_{n\to\infty} F\left(3 - \dfrac{1}{n}\right) = \dfrac{11}{12}$.

$P\{Y = 1\} = P\{Y \le 1\} - P\{X < 1\} = F(1) - \lim_{n\to\infty} F\left(1 - \dfrac{1}{n}\right) = \dfrac{2}{3} - \dfrac{1}{2} = \dfrac{1}{6}$.

$P\{Y > 0.5\} = 1 - P\{Y \le 0.5\} = 1 - F(0.5) = 0.75$.

$P\{2 < Y \le 4\} = F(4) - F(2) = 1/12$. ◘

Example 3.1-9: An Unbounded Density Function

Let the random variable X have the distribution function $F(x)$ given by:

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x^{1/2} & \text{if } 0 < x \le 1 \\ 1 & \text{if } x > 1. \end{cases}$$

Then X has a density function $f(x)$ given by

$$f(x) = \begin{cases} \dfrac{1}{2x^{1/2}} & \text{if } 0 < x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that $f(x)$ is unbounded for x near zero. In fact, as x approaches zero by positive value, $f(x)$ tends toward infinity slowly enough so that the density function still integrates to one. An alternative example is the density of the chi-squared distribution with one degree of freedom. ◘

Definition 3.1-6: If $0 < p < 1$, a $100 \times p$th ***percentile*** of the distribution of a random variable X is the smallest value, x_p, such that $F(x_p) \ge p$. ◘

In essence, x_p is the value such that $100 \times p$% of the population values are less than or equal to x_p. We can also think in terms of a proportion p rather than a percentage $100 \times p$ of the population, and x_p is often referred to as a pth **quantile**. If X is continuous, then x_p is a solution to the equation

$$F(x_p) = p.$$

Example 3.1-10: Consider the distribution of lifetimes, X (in months), of a particular type of component. We will assume that the CDF has the form

$$F(x) = 1 - \exp\left\{-(x/3)^2\right\}, \qquad x > 0$$

and zero otherwise. The median lifetime is

$$m = 3[-\ln(1-0.5)]^{0.5} = 3\sqrt{\ln 2} = 2.498 \text{ months.}$$

It is desired to find the time t such that 10% of the components fail before t. This is the 10% percentile:

$$x_{0.10} = 3[-\ln(1-0.1)]^{0.5} = 3\sqrt{-\ln(0.9)} = 0.974 \text{ months.}$$

Thus, if the components are guaranteed for one month, slightly more than 10% will need to be replaced. ◘

3.2 Mathematical Expectation

One of the most important concepts in probability theory is that of the mathematical expectation of a random variable.

Definition 3.2-1: Let *X* be a random variable having a pdf (or pmf) $f(x)$, and let $u(X)$ be a function of *X*. Then the ***mathematical expectation*** of $u(X)$, denoted by $E[u(X)]$, is defined to be

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

if *X* is a continuous type of random variable, or

$$E[u(X)] = \sum_x u(x)\, f(x)$$

if *X* is a discrete type of random variable. ◘

Remarks: The usual definition of $E[u(X)]$ requires that the integral (or sum) converge absolutely. That is, $\int_{-\infty}^{\infty} |u(x)|\, f(x)\, dx < \infty$ (or $\sum_x |u(x)|\, f(x) < \infty$).

Theorem 3.2-1: Let *X* be a random variable having a pmf (or pdf) $f(x)$. Mathematical expectation $E(\cdot)$, if it exists, satisfies the following properties:

(a) If *c* us a constant, $E(c) = c$.

(b) If *c* is a constant and *u* is a function, $E[cu(X)] = cE[u(X)]$.

(c) If c_1 and c_2 are constants and u_1 and u_2 are functions, then

$$E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)].$$

Proof: First, we have for the proof of (a) that $E(c) = \sum_x cf(x) = c\sum_x f(x) = c$.

Next, to prove (b), we see that $E[cu(X)] = \sum_x cu(x)f(x) = c\sum_x u(x)f(x) = cE[u(X)]$.

Finally, the proof of (c) is given by

$$E[c_1u_1(X) + c_2u_2(X)] = \sum_x [c_1u_1(x) + c_2u_2(x)]f(x) = \sum_x c_1u_1(x)f(x) + \sum_x c_2u_2(x)f(x).$$

By applying (b), we obtain $E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)]$.

Property (c) can be extended to more than two terms by mathematical induction; that is, we have

(c)′ $$E\left[\sum_{i=1}^{k} c_iu_i(X)\right] = \sum_{i=1}^{k} c_iE[u_i(X)].$$

Because of property (c)′, mathematical expectation $E(\cdot)$ is called a ***linear*** or ***distributive*** operator. ◘

Certain mathematical expectations, if they exist, have special names and symbols to represent them. First, let $u(X) = X$, where *X* is a random variable of the discrete type having a pmf $f(x)$. Then

$$E[X] = \sum_x x\, f(x).$$

If the discrete points of the space of positive probabilities are $a_1, a_2, \ldots,$ then

$$E(X) = a_1 f(a_1) + a_2 f(a_2) + a_3 f(a_3) + \cdots.$$

This sum of product is seen to be a "***weighted average***" of the values $a_1, a_2, \ldots,$ the "weight" associated with each a_i being $f(a_i)$. This suggests that we call $E(X)$ the ***mean value*** of X (or the mean value of the distribution). The mean value μ of a random variable X is defined, when it exists, to be $\mu = E(X)$.

Another special mathematical expectation is obtained by taking $u(X) = (X - \mu)^2$. If X is a random variable of the discrete type having a pmf $f(x)$, then

$$E\left((X - \mu)^2\right) = \sum_x (X - \mu)^2 f(x) = (a_1 - \mu)^2 f(a_1) + (a_2 - \mu)^2 f(a_2) + \cdots,$$

where $a_1, a_2, \ldots,$ are the discrete points of the space of positive probabilities. This sum of product may be interpreted as a "***weighted average***" of the ***squares*** of the ***deviations*** of the numbers $a_1, a_2, \ldots,$ from the mean value μ of those numbers where the "weight" associated with each $(a_i - \mu)^2$ is $f(a_i)$. This mean value of the square of the deviation of X from its mean value μ is called the ***variance*** of X (or the variance of the distribution). The variance of X will be denoted by $\sigma^2 = E\left((X - \mu)^2\right)$, if it exists. The variance can be computed in another manner:

$$\sigma^2 = E\left((X - \mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right) = E\left(X^2\right) - 2\mu E(X) + \mu^2 = E\left(X^2\right) - \mu^2.$$

It frequently affords an easier way of computing the variance of X.

It is customary to call σ (the positive square root of the variance) the ***standard deviation*** of X (or the standard deviation of the distribution). The number σ is sometimes interpreted as a measure of the dispersion of the points of the space relative to the mean value μ. We note that if the space contains only one point x for which $f(x) > 0$, then $\sigma = 0$.

We next define a third special mathematical expectation, called the moment generating function of a random variable X.

Definition 3.2-2: If X is a random variable, the expected value

$$M(t) = E\left(e^{tX}\right)$$

is called the ***moment generating function*** (MGF) of X if this expected value exists for all values of t in some open interval containing 0 of the form $-h < t < h$ for some $h > 0$. ◘

It is evident that if we set $t = 0$, we have $M(0) = 1$. The moment generating function is unique and completely determines the distribution of the random variable; thus, if two random variables have the same moment generating function, they have the same distribution. If a discrete random variable X has a pmf $f(x)$ with support $\{b_1, b_2, \ldots\}$, then

$$M(t) = \sum_x e^{tx} f(x) = f(b_1)e^{tb_1} + f(b_2)e^{tb_2} + \cdots.$$

Hence, the coefficient of e^{tb_i} is $f(b_i) = P(X = b_i)$. That is, if we write a moment generating function of a discrete-type random variable *X* in the above form, the probability of any value of *X*, say b_i, is the coefficient of e^{tb_i}.

Example 3.2-1: Let the moment generating function of *X* be defined by

$$M(t) = \frac{1}{15}e^{t} + \frac{2}{15}e^{2t} + \frac{3}{15}e^{3t} + \frac{4}{15}e^{4t} + \frac{5}{15}e^{5t}.$$

Then, for example, the coefficient of e^{2t} is $2/15$. Thus $f(2) = P(X = 2) = 2/15$. In general, we see that the pmf of *X* is $f(x) = x/15$, *x* = 1, 2, 3, 4, 5. ◘

Definition 3.2-3: Let *X* be a random variable and let *r* be a positive integer. If $E\left(X^r\right)$ exists, it is called the *r*th ***moment*** of the distribution about the origin. In addition, the expectation $E\left((X - b)^r\right)$ is called the *r*th moment of the distribution about *b*. ◘

Theorem 3.2-2: If the moment generating function of *X* exists, then

$$E\left(X^r\right) = M^{(r)}(0) = \left.\frac{d^r M(t)}{dt^r}\right]_{t=0} \qquad \text{for all r =1, 2, 3,…}$$

Proof: From the theory of mathematical analysis, it can be shown that the existence of $M(t)$, for $-h < t < h$, implies that derivatives of $M(t)$ of all orders exists at *t* = 0; moreover, it is permissible to interchange of the differentiation and expectation operator. Thus,

$$\frac{dM(t)}{dt} = M'(t) = \int_{-\infty}^{\infty} xe^{tx} f(x)dx,$$

if *X* is of the continuous type, or

$$\frac{dM(t)}{dt} = M'(t) = \sum_x xe^{tx} f(x)$$

if *X* is of the discrete type. Setting *t* = 0, we have in either case

$$M'(0) = E(X) = \mu.$$

The second derivative of $M(t)$ is

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)dx \qquad \text{or} \qquad M''(t) = \sum_x x^2 e^{tx} f(x),$$

so that $M''(0) = E(X^2)$.

In general, if *r* is a positive integer, we have, by repeated differentiation with respect to *t*,

$$M^{(r)}(0) = E(X^r). \quad \blacksquare$$

Note that $\sigma^2 = M''(0) - [M'(0)]^2$. Since $M(t)$ generates the value of $E(X^r)$, it is called the moment generating function.

When the moment generating function exists, derivatives of all orders exist at $t = 0$. Thus it is possible to represent $M(t)$ as a Maclaurin's series, namely,

$$M(t) = M(0) + M'(0)\left(\frac{t}{1!}\right) + M''(0)\left(\frac{t^2}{2!}\right) + M'''(0)\left(\frac{t^3}{3!}\right) + \cdots.$$

That is, if the Maclaurin's series expansion of $M(t)$ can be found, the rth moment of X, $E(X^r)$, is the coefficient of $t^r/r!$. Or, if $M(t)$ exists and the moments are given, we can frequently sum the Maclaurin's series to obtain the closed form of $M(t)$. These points are illustrated in the next two examples.

Example 3.2-2: Suppose that the random variable X has the moment generating function

$$M(t) = \frac{1}{5}\left(e^{-2t} + e^{-t} + 1 + e^{t} + e^{2t}\right)$$

for all real t. Using the series expansion of e^u, the Maclaurin's series of $M(t)$ is easily found to be

$$M(t) = 1 + \frac{2(1+4)}{5}\left(\frac{t^2}{2!}\right) + \frac{2(1+16)}{5}\left(\frac{t^4}{4!}\right) + \cdots + \frac{2(1+2^r)}{5}\left(\frac{t^r}{r!}\right) + \cdots;$$

here r is even. Since the coefficient of $t^r/r!$ is zero when r is odd, we have

$$E(X^r) = \begin{cases} 0, & r = 1,3,5,\ldots \\ \dfrac{2(1+2^r)}{5}, & r = 2,4,6,\ldots \end{cases}$$

In particular, $\mu = E(X) = 0$ and $\sigma^2 = \dfrac{2(1+2^2)}{5} - \mu^2 = 2$. $\blacksquare$

Example 3.2-3: Let the moment of X be defined by

$$E(X^r) = 0.8, \qquad r = 1,2,3,\ldots.$$

The moment generating function of X is then

$$M(t) = M(0) + \sum_{r=1}^{\infty} 0.8\left(\frac{t^r}{r!}\right) = 1 + 0.8\sum_{r=1}^{\infty}\frac{t^r}{r!} = 0.2 + 0.8\sum_{r=0}^{\infty}\frac{t^r}{r!} = 0.2e^{0t} + 0.8e^{t}.$$

Thus, $P(X = 0) = 0.2$ and $P(X = 1) = 0.8$. $\blacksquare$

Result: Suppose that the moment generating function of X exists. Let $R(t) = \ln M(t)$ and

$R^{(k)}(0)$ denote the kth derivative of $R(t)$ evaluated for $t = 0$. Then

$$R^{(1)}(0) = E(X) = \mu \qquad \text{and} \qquad R^{(2)}(0) = E\left(X^2\right) - \mu^2 = \sigma^2 .$$

Proof: $\left.\dfrac{dR(t)}{dt}\right|_{t=0} = \left.\dfrac{M'(t)}{M(t)}\right|_{t=0} = \dfrac{M'(0)}{M(0)} = E(X)$.

$$\left.\frac{d^2R(t)}{dt^2}\right|_{t=0} = \left.\frac{M''(t)M(t) - M'(t)M'(t)}{M^2(t)}\right|_{t=0} = \frac{M''(0) - [M'(0)]^2}{M^2(0)} = E(X^2) - [E(X)]^2 = \sigma^2 . \quad \blacksquare$$

Example 3.2-4: A random variable with infinite mean

Let X have the density function

$$f(x) = \begin{cases} 1/x^2, & x \ge 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then the expected values of X is

$$E(X) = \int_1^\infty x\left(1/x^2\right) dx = \infty . \quad \blacksquare$$

Example 3.2-5: A random variable whose mean does not exist

Let the continuous random variable X have the Cauchy distribution centered at the origin with density given by

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}, \qquad -\infty < x < \infty .$$

The mean of X is then

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi\left(1 + x^2\right)} dx = \left.\frac{1}{2\pi} \log(1 + x^2)\right|_{-\infty}^{\infty}$$

and this integral does not exist since

$$\int_0^\infty x f(x) dx = \int_{-\infty}^0 - x f(x) dx = \infty . \quad \blacksquare$$

Example 3.2-6: A random variable whose first moment exists but no higher moments exist

Let the random variable X have a density given by

$$f(x) = \frac{3}{2} x^{-5/2}, \qquad 1 < x < \infty .$$

Then the expected value of X is

$$E(X) = \frac{3}{2}\int_1^\infty x^{-3/2}dx = -3x^{-1/2}\Big|_1^\infty = 3.$$

However, for integer values of $k > 1$, we find that

$$E(X^k) = \frac{3}{2}\int_1^\infty x^{k-5/2}dx = \frac{3/2}{k-3/2}x^{k-3/2}\Bigg|_1^\infty = \infty.$$

In fact, for this example the moment of order k dose exist, although it is infinite. We may modify the example slightly to achieve a case where the higher moments would be of the form $\infty - \infty$ and therefore would not exist. To do this, let the density have the same basic form but be symmetric about zero:

$$g(x) = \frac{3}{4}x^{-5/2}, \qquad 1 < |x| < \infty.$$

More generally, the Student's t-distribution with $r + 1$ degrees of freedom has moments of order 0, 1, …, r, but no higher moments exist. ◘

Example 3.2-7: A random variable whose moment generating function does not exist

Suppose the random variable X has the Cauchy distribution with density given by

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}, \qquad -\infty < x < \infty$$

The integral

$$E(e^{tX}) = \int_{-\infty}^\infty e^{tx}\frac{1}{\pi\left(1 + x^2\right)}dx$$

is infinite for any $t \neq 0$ since $e^{tx}/\pi\left(1 + x^2\right)$ is positive for $-\infty < x < \infty$ and tends to ∞ as $x \to \infty$. Thus the moment generating function does not exist in this example. ◘

Example 3.2-8: A random variable, all of whose moment exist, but whose moment generating function dose not exist

Existence (finiteness) of the moment generating function for some $t > 0$ implies that all moments exist and are finite; however, the converse is false. Consider the lognormal distribution, which is the distribution of $Y = e^X$ where X has a normal distribution. Suppose X has mean zero and variance one, so that Y has the standard lognormal distribution. The moments of Y exist for all orders $k = 1, 2, 3,\ldots$ since

$$E\left(Y^k\right) = E\left(e^{kX}\right) = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^\infty e^{kx}e^{-x^2/2}dx = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^\infty \exp\left[-\frac{(x-k)^2}{2} + \frac{k^2}{2}\right]dx = \exp\left(\frac{k^2}{2}\right).$$

However, the moment generating function does not exists since if $t > 0$,

$$E\left(e^{tY}\right) = E\left(e^{k\left(e^X\right)}\right) = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} e^{te^x - x^2/2}dx$$

$$\geq \frac{1}{(2\pi)^{1/2}}\int_0^{\infty} \exp\left[t\left(1 + x + \frac{x^2}{2} + \frac{x^3}{6}\right) - \frac{x^2}{2}\right]dx = \infty$$

since the exponential term is a third-degree polynomial in x for which the x^3 term has a positive coefficient; this exponential must then tend to ∞ as $x \to \infty$. Therefore the moment generating function of Y does not exist. ◘

Remark. In more advanced course, we would not work with the moment generating function since so many distributions do not have moment generating functions. Instead, we would let i denote the imaginary unit, t an arbitrary real, and we would define $\varphi(t) = E\left(e^{itX}\right)$. This expectation exists for every distribution and it is called the ***characteristic function*** of the distribution. To see why $\varphi(t)$ exists for all real t, we note, in the continuous case, that its absolute value

$$|\varphi(t)| = \left|\int_{-\infty}^{\infty} e^{itx} f(x)dx\right| \leq \int_{-\infty}^{\infty} \left|e^{itx} f(x)\right|dx.$$

However, $|f(x)| = f(x)$ since $f(x)$ is nonnegative and

$$\left|e^{itx}\right| = |\cos tx + i\sin tx| = \sqrt{\cos^2 tx + \sin^2 tx} = 1.$$

Thus

$$|\varphi(t)| = \left|\int_{-\infty}^{\infty} e^{itx} f(x)dx\right| \leq \int_{-\infty}^{\infty} \left|e^{itx} f(x)\right|dx \leq \int_{-\infty}^{\infty} f(x)dx = 1.$$

Every distribution has a ***unique*** characteristic function; and to each characteristic function there corresponds a ***unique*** distribution of probability. If X has a distribution with characteristic function $\varphi(t)$, then for instance, if $E(X)$ and $E(X^2)$ exist, they are given, respectively, by $iE(X) = \varphi'(0)$ and $i^2E(X^2) = \varphi''(0)$. It may write $\varphi(t) = M(it)$.

3.3 Chebyshev's Inequality

Theorem 3.3-1 (Markov's Inequality): If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof: We give a proof for the case where X is continuous with density f.

$$E(X) = \int_0^{\infty} xf(x)\,dx = \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx$$

$$\geq \int_a^\infty x f(x)dx$$

$$\geq \int_a^\infty a f(x)dx = a\int_a^\infty f(x)dx = aP(X \geq a)$$

and the result is proved. ◘

Theorem 3.3-2 (Chebyshev's Inequality): If X is a random variable with finite mean μ and variance σ^2, then for any value $b > 0$

$$P(|X - \mu| \geq b) \leq \frac{\sigma^2}{b^2}.$$

Proof: Since $(X - \mu)^2$ is nonnegative random variable, we can apply Markov's inequality (with $a = b^2$) to obtain

$$P\{(X - \mu)^2 \geq b^2\} \leq \frac{E[(X-\mu)^2]}{b^2}.$$

But since $(X - \mu)^2 \geq b^2$ if and only if $|X - \mu| \geq b$, we have

$$P(|X - \mu| \geq b) \leq \frac{E[(X-\mu)^2]}{b^2} = \frac{\sigma^2}{b^2}$$

and the proof is complete. ◘

The importance of Markov's and Chebyshev's inequalities is that they enable us to derivative bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.

Example 3.3-1: Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution: Let X be the number of items that will be produced in a week:

(a) By Markov's inequality, $P(X > 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$.

(b) By Chebyshev's inequality $P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$.

Hence, $P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$. ◘

Note that although Chebyshev's inequality is valid for all distributions of the random variable *X*, we cannot expect the bound on the probability to be very close to the actual probability in most case.