Arkansas Tech University
MATH 3513: Applied Statistics I
Dr. Marcel B. Finan

2.4 Random Variables

By definition, a **random variable** X is a function with domain the sample space and range a subset of the real numbers. For example, in rolling two dice X might represent the sum of the points on the two dice. Similarly, in taking samples of college students X might represent the number of hours per week a student studies, a student's GPA, or a student's height.
The notation $X(s) = x$ means that x is the value associated with the outcome s by the random variable X.
We consider two types of random variables: discrete random variables and continuous random variables. A **discrete** random variable is a random variable whose range has the property that between any two values in the range there is a gap. A **continuous** random variable is a random variable whose range is an interval or union of intervals in $\mathbb{R}$.

Example 2.4.1
State whether the random variables are discrete or continuous.
(a) A coin is tossed ten times. The random variable X is the number of tails that are noted.
(b) A light bulb is burned until it burns out. The random variable Y is its lifetime in hours.

Solution.
(a) X can only take the values 0, 1, ..., 10, so X is a discrete random variable.
(b) Y can take any positive real value, so Y is a continuous random variable ■

Example 2.4.2
The sample space of the experiment of tossing a coin 3 times is given by

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let X = # of Heads in 3 tosses. Find the range of X.

Solution.
We have

$$\begin{array}{cccccccccccc} X(HHH) & = & 3 & X(HHT) & = & 2 & X(HTH) & = & 2 & X(HTT) & = & 1 \\ X(THH) & = & 2 & X(THT) & = & 1 & X(TTH) & = & 1 & X(TTT) & = & 0 \end{array}$$

Thus, the range of X consists of $\{0, 1, 2, 3\}$ so that X is a discrete random variable ■

We use upper-case letters X, Y, Z, etc. to represent random variables. We use small letters x, y, z, etc to represent possible values that the corresponding random variables X, Y, Z, etc. can take. The statement $X = x$ defines an event consisting of all outcomes with $X-$measurement equal to x which is the set $\{s \in S : X(s) = x\}$. For instance, considering the random variable of the previous example, the statement "X = 2" is the event $\{HHT, HTH, THH\}$. Because the value of a random variable is determined by the outcomes of the experiment, we may assign probabilities to the possible values of the random variable. For example, $P(X = 2) = \frac{3}{8}$.

Example 2.4.3
Consider the experiment consisting of 2 rolls of a fair 4-sided die. Let X be a random variable, equal to the maximum of the 2 rolls. Complete the following table

x	1	2	3	4
P(X=x)				

Solution.
The sample space of this experiment is

$$S = \{(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)\}.$$

Thus,

x	1	2	3	4
P(X=x)	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

<u>Probability Mass Function</u>
For a discrete random variable X, we define the **probability distribution** or the **probability mass function**(abbreviated pmf)by the equation

$$p(x) = P(X = x).$$

That is, a probability mass function gives the probability that a discrete random variable is exactly equal to some value.
The pmf can be described by an equation, a table, or a graph that shows how probability is assigned to possible values of the random variable.

Example 2.4.4
Suppose a variable X can take the values 1, 2, 3, or 4. The probabilities associated with each outcome are described by the following table:

x	1	2	3	4
$p(x)$	0.1	0.3	0.4	0.2

Draw the probability histogram.

Solution.
The probability histogram is shown in Figure 2.4.1 ■

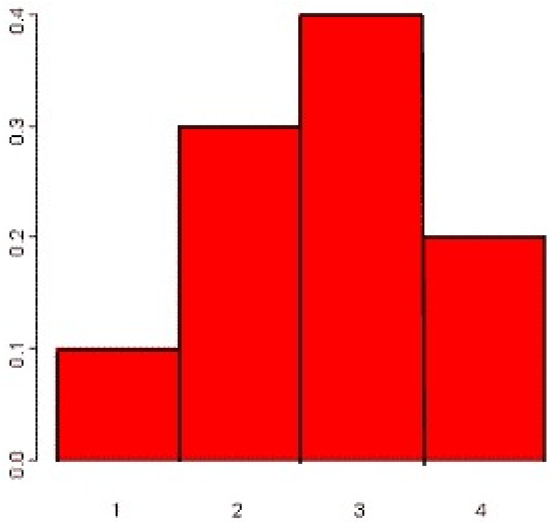

Figure 2.4.1

Example 2.4.5
A committee of 4 is to be selected from a group consisting of 5 men and 5 women. Let X be the random variable that represents the number of women in the committee. Create the probability mass distribution.

Solution.
For $x = 0, 1, 2, 3, 4$ we have

$$p(x) = \frac{\binom{5}{x}\binom{5}{4-x}}{\binom{10}{4}}.$$

The probability mass function can be described by the table

x	0	1	2	3	4
$p(x)$	$\frac{5}{210}$	$\frac{50}{210}$	$\frac{100}{210}$	$\frac{50}{210}$	$\frac{5}{210}$

Cumulative Distribution Function

All random variables (discrete or continuous) have a **distribution function** or a **cumulative distribution function**, abbreviated cdf. It is a function giving the probability that the random variable X is less than or equal to x, for every value x. For a discrete random variable, the cumulative distribution function is found by summing up the probabilities. That is,

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t).$$

Note that $F(x)$ is defined for any number x, not necessarily in the range of X.

Example 2.4.6

Given the following pmf

$$p(x) = \begin{cases} 1, & \text{if } x = a \\ 0, & \text{otherwise.} \end{cases}$$

Find a formula for $F(x)$ and sketch its graph.

Solution.

If $x < a$ then $F(x) = P(X \leq x) = P(X \neq a) = 0$. If $x \geq a$ then $F(x) = P(X \leq x) = 1 - P(X > x) = 1 - 0 = 1$. The graph of $F(x)$ is given in Figure 2.4.2 ■

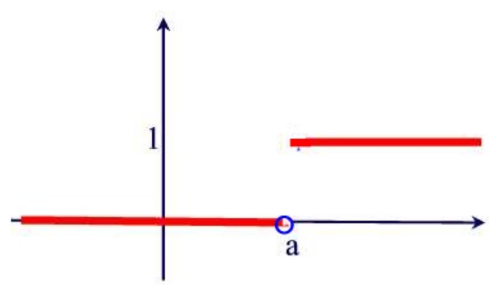

Figure 2.4.2

For discrete random variables the cumulative distribution function will always be a step function with jumps at each value of x that has probability greater than 0. Note the value of $F(x)$ is assigned to the top of the jump.

Example 2.4.7
Consider the following probability mass function

x	1	2	3	4
$p(x)$	0.25	0.5	0.125	0.125

Find a formula for $F(x)$ and sketch its graph.

Solution.
The cdf is given by

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.25 & 1 \leq x < 2 \\ 0.75 & 2 \leq x < 3 \\ 0.875 & 3 \leq x < 4 \\ 1 & 4 \leq x. \end{cases}$$

Its graph is given in Figure 2.4.3 ■

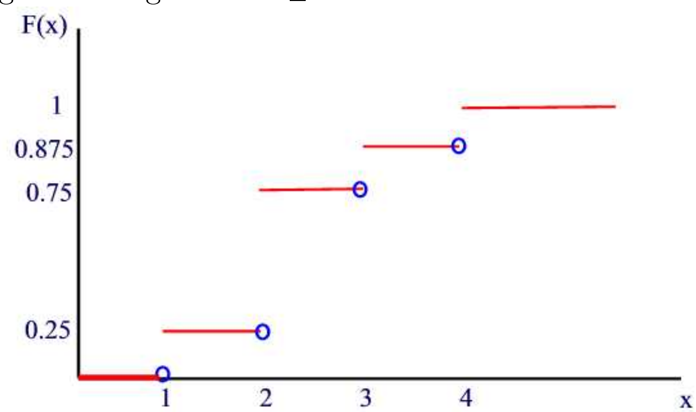

Figure 2.4.3

The Mean for Discrete Random Variables
A cube has three red faces, two green faces, and one blue face. A game consists of rolling the cube twice. You pay \$2 to play. If both faces are the same color, you are paid \$5(that is you win \$3). If not, you lose the \$2 it costs to play. Will you win money in the long run? Let W denote the event that you win. Then $W = \{RR, GG, BB\}$ and

$$P(W) = P(RR) + P(GG) + P(BB) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{1}{6} = \frac{7}{18} \approx 39\%.$$

Thus, $P(L) = \frac{11}{18} = 61\%$. Hence, if you play the game 18 times you expect to win 7 times and lose 11 times on average. So your winnings in dollars will be $3 \times 7 - 2 \times 11 = -1$. That is, you can expect to lose \$1 if you play the game 18 times. On the average, you will lose \$ $\frac{1}{18}$ per game (about 6 cents). This can be found also using the equation

$$3 \times \frac{7}{18} - 2 \times \frac{11}{18} = -\frac{1}{18}$$

If we let X denote the winnings of this game then the range of X consists of the two numbers 3 and -2 which occur with respective probability 0.39 and 0.61. Thus, we can write

$$E(X) = 3 \times \frac{7}{18} - 2 \times \frac{11}{18} = -\frac{1}{18}.$$

We call this number the expected value of X. More formally, let the range of a discrete random variable X be a sequence of numbers $x_1, x_2, \cdots, x_k$, and let $P(x)$ be the corresponding probability mass function. Then the **expected value** of X is

$$E(X) = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_k p(x_k).$$

The following is a justification of the above formula. Suppose that X has k possible values $x_1, x_2, \cdots, x_k$ and that

$$p_i = P(X = x_i) = p(x_i), i = 1, 2, \cdots, k.$$

Suppose that in n repetitions of the experiment, the number of times that X takes the value x_i is n_i. Then the sum of the values of X over the n repetitions is

$$n_1 x_1 + n_2 x_2 + \cdots + n_k x_k$$

and the average value of X is

$$\frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n} = \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \cdots + \frac{n_k}{n} x_k.$$

But $P(X = x_i) = \lim_{n \to \infty} \frac{n_i}{n}$. Thus, the average value of X approaches

$$E(X) = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_k p(x_k).$$

The expected value of X is also known as the **mean** value.

Example 2.4.8
Suppose that an insurance company has broken down yearly automobile claims for drivers from age 16 through 21 as shown in the following table.

Amount of claim	Probability
\$ 0	0.80
\$ 2000	0.10
\$ 4000	0.05
\$ 6000	0.03
\$ 8000	0.01
\$ 10000	0.01

How much should the company charge as its average premium in order to break even on costs for claims?

Solution.
Let X be the random variable of the amount of claim. Finding the expected value of X we have

$$E(X) = 0(.80)+2000(.10)+4000(.05)+6000(.03)+8000(.01)+10000(.01) = 760$$

Since the average claim value is \$760, the average automobile insurance premium should be set at \$760 per year for the insurance company to break even ■

The mean of any function of X is provided by the following result.

Theorem 2.4.1
If X is a discrete random variable with range D and pmf $p(x)$. For any function $g(x)$, $g(X)$ is a discrete random variable with mean given by

$$E(g(X)) = \sum_{x \in D} g(x)p(x).$$

Example 2.4.9
Let X be a discrete random variable. Show that (a) $E(aX + b) = aE(X) + b$ and (b) $E(aX^2 + bX + c) = aE(X^2) + bE(X) + c$.

Solution.
Let D denote the range of X. Then (a)

$$\begin{aligned}E(aX+b)=&\sum_{x\in D}(ax+b)p(x)\\=&a\sum_{x\in D}xp(x)+b\sum_{x\in D}p(x)\\=&a\sum_{x\in D}xp(x)+b.\end{aligned}$$

(b) We have

$$\begin{aligned}E(aX^2+bX+c)=&\sum_{x\in D}(ax^2+bx+c)p(x)\\=&\sum_{x\in D}ax^2p(x)+\sum_{x\in D}bxp(x)+\sum_{x\in D}cp(x)\\=&a\sum_{x\in D}x^2p(x)+b\sum_{x\in D}xp(x)+c\sum_{x\in D}p(x)\\=&aE(X^2)+bE(X)+c\ \blacksquare\end{aligned}$$

Remark 2.4.1
The expected value (or mean) is related to the physical property of center of mass. If we have a weightless rod in which weights of mass $p(x)$ located at a distance x from the left endpoint of the rod then the point at which the rod is balanced is called the **center of mass.** If α is the center of mass then we must have $\sum_x(x-\alpha)p(x)=0$. This equation implies that $\alpha=\sum_x xp(x)=E(X)$. Thus, the expected value tells us something about the center of the probability mass function.

<u>The Variance and Standard Deviation for Discrete Random Variables</u>
Recall Section 1.2 that the variance of a population is the expected value of the squared deviations. By analogy, the expected squared distance between the random variable and its mean is called the **variance** of the random variable. The positive square root of the variance is called the **standard deviation** of the random variable. If σ_X denotes the standard deviation then the variance is given by the formula

$$\text{Var(X)}=\sigma_X^2=E\left[(X-E(X))^2\right]=\sum_x(x-E(X))^2p(x).$$

The variance of a random variable is typically calculated using the following formula

$$\begin{aligned}\mathrm{Var}(X) =& E[(X - E(X))^2] \\ =& E[X^2 - 2XE(X) + (E(X))^2] \\ =& E(X^2) - 2E(X)E(X) + (E(X))^2 \\ =& E(X^2) - (E(X))^2\end{aligned}$$

where we used Example 2.4.9.

Example 2.4.10
Find the variance of the random variable X with probability distribution $P(X = 1) = P(X = -1) = \frac{1}{2}$.

Solution.
Since $E(X) = 1 \times \frac{1}{2} - 1 \times \frac{1}{2} = 0$ and $E(X^2) = 1^2\frac{1}{2} + (-1)^2 \times \frac{1}{2} = 1$ we find $\mathrm{Var}(X) = 1 - 0 = 1$ ■

A useful identity is given in the following result

Theorem 2.4.2
If X is a discrete random variable then for any constants a and b we have

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Proof.
Since $E(aX + b) = aE(X) + b$, we have

$$\begin{aligned}\mathrm{Var}(aX + b) =& E\left[(aX + b - E(aX + b))^2\right] \\ =& E[a^2(X - E(X))^2] \\ =& a^2E((X - E(X))^2) \\ =& a^2\mathrm{Var}(X) \ ■\end{aligned}$$

Remark 2.4.2
Note that the units of Var(X) is the square of the units of X. This motivates the definition of the standard deviation $\sigma_X = \sqrt{\mathrm{Var}(X)}$ which is measured in the same units as X.

Example 2.4.11
In a recent study, it was found that tickets cost to the Dallas Cowboys football games averages \$80 with a variance of 105 square dollar. What will be the variance of the cost of tickets if 3% tax is charged on all tickets?

Solution.
Let X be the current ticket price and Y be the new ticket price. Then $Y = 1.03X$. Hence,

$$\text{Var}(Y) = \text{Var}(1.03X) = 1.03^2\text{Var}(X) = (1.03)^2(105) = 111.3945 \blacksquare$$

Example 2.4.12
In the experiment of rolling one die, let X be the number on the face that comes up. Find the variance and standard deviation of X.

Solution.
We have

$$E(X) = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

and

$$E(X^2) = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) \cdot \frac{1}{6} = \frac{91}{6}.$$

Thus,

$$\text{Var}(X) = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

The standard deviation is

$$\sigma_X = \sqrt{\frac{35}{12}} \approx 1.7078 \blacksquare$$

<u>Continuous Random Variables</u>
Continuous random variables are random quantities that are measured on a continuous scale. They can usually take on any value over some interval, which distinguishes them from discrete random variables, which can take on only a sequence of values, usually integers.
We say that a random variable is **continuous** if there exists a nonnegative function f (not necessarily continuous) defined for all real numbers and having the property that for any set B of real numbers we have

$$P(X \in B) = \int_B f(x)dx.$$

We call the function f the **probability density function** (abbreviated pdf) of the random variable X.
If we let $B = (-\infty, \infty)$ then

$$\int_{-\infty}^{\infty} f(x)dx = P[X \in (-\infty, \infty)] = 1.$$

Now, if we let $B = [a, b]$ then

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

That is, areas under the probability density function represent probabilities as illustrated in Figure 2.4.4.

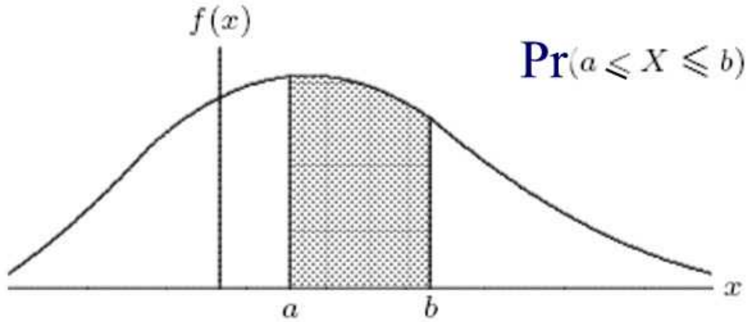

Figure 2.4.4

Now, if we let $a = b$ in the previous formula we find

$$P(X = a) = \int_a^a f(x)dx = 0.$$

It follows from this result that

$$P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b)$$

and

$$P(X \leq a) = P(X < a) = \int_{-\infty}^{a} f(x)dx \quad \text{and} \quad P(X \geq a) = P(X > a) = \int_a^{\infty} f(x)dx.$$

Example 2.4.13
Suppose that the function $f(t)$ defined below is the density function of some random variable X.

$$f(t) = \begin{cases} e^{-t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

Compute $P(-10 \leq X \leq 10)$.

Solution.

$$\begin{aligned} P(-10 \leq X \leq 10) &= \int_{-10}^{10} f(t)dt \\ &= \int_{-10}^{0} f(t)dt + \int_{0}^{10} f(t)dt \\ &= \int_{0}^{10} e^{-t}dt \\ &= -e^{-t}\Big|_0^{10} = 1 - e^{-10} \blacksquare \end{aligned}$$

<u>**Cumulative Distribution Function**</u>
The **cumulative distribution function** or simply the **distribution function** (abbreviated cdf) $F(t)$ of the random variable X is defined as follows

$$F(t) = P(X \leq t)$$

i.e., $F(t)$ is equal to the probability that the variable X assumes values, which are less than or equal to t. From this definition we can write

$$F(t) = \int_{-\infty}^{t} f(y)dy.$$

Geometrically, $F(t)$ is the area under the graph of f to the left of t.

Example 2.4.14
Find the distribution functions corresponding to the following density functions:
(a) $f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$
(b) $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty$
(c) $f(x) = \frac{a-1}{(1+x)^a}, \quad 0 < x < \infty$
(d) $f(x) = k\alpha x^{\alpha-1}e^{-kx^\alpha}, \quad 0 < x < \infty, k > 0, \alpha > 0.$

Solution.
(a)

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{1}{\pi(1+y^2)} dy \\
&= \left[\frac{1}{\pi}\arctan y\right]_{-\infty}^{x} \\
&= \frac{1}{\pi}\arctan x - \frac{1}{\pi}\cdot\frac{-\pi}{2} \\
&= \frac{1}{\pi}\arctan x + \frac{1}{2}
\end{aligned}$$

(b)

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{e^{-y}}{(1+e^{-y})^2} dy \\
&= \left[\frac{1}{1+e^{-y}}\right]_{-\infty}^{x} \\
&= \frac{1}{1+e^{-x}}
\end{aligned}$$

(c) For $x \geq 0$

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{a-1}{(1+y)^a} dy \\
&= \left[-\frac{1}{(1+y)^{a-1}}\right]_{0}^{x} \\
&= 1 - \frac{1}{(1+x)^{a-1}}
\end{aligned}$$

For $x < 0$ it is obvious that $F(x) = 0$, so we could write the result in full as

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{(1+x)^{a-1}} & x \geq 0. \end{cases}$$

(d) For $x \geq 0$

$$\begin{aligned}
F(x) &= \int_{0}^{x} k\alpha y^{\alpha-1} e^{-ky^{\alpha}} dy \\
&= \left[-e^{-ky^{\alpha}}\right]_{0}^{x} \\
&= 1 - e^{-kx^{\alpha}}
\end{aligned}$$

For $x < 0$ we have $F(x) = 0$ so that

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - ke^{-kx^{\alpha}} & x \geq 0 \end{cases} \blacksquare$$

Example 2.4.15
(a) Determine the value of c so that the following function is a pdf.

$$f(x) = \begin{cases} \frac{15}{64} + \frac{x}{64} & -2 \leq x \leq 0 \\ \frac{3}{8} + cx & 0 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(b) Determine $P(-1 \leq X \leq 1)$.
(c) Find $F(x)$.

Solution.
(a) Observe that f is discontinuous at the points -2 and 0, and is potentially also discontinuous at the point 3. We first find the value of c that makes f a pdf.

$$\begin{aligned} 1 =& \int_{-2}^{0} \left(\frac{15}{64} + \frac{x}{64} \right) dx + \int_{0}^{3} \left(\frac{3}{8} + cx \right) dx \\ =& \left[\frac{15}{64}x + \frac{x^2}{128} \right]_{-2}^{0} + \left[\frac{3}{8}x + \frac{cx^2}{2} \right]_{0}^{3} \\ =& \frac{30}{64} - \frac{2}{64} + \frac{9}{8} + \frac{9c}{2} \\ =& \frac{100}{64} + \frac{9c}{2} \end{aligned}$$

Solving for c we find $c = -\frac{1}{8}$.
(b) The probability $P(-1 \leq X \leq 1)$ is calculated as follows.

$$P(-1 \leq X \leq 1) = \int_{-1}^{0} \left(\frac{15}{64} + \frac{x}{64} \right) dx + \int_{0}^{1} \left(\frac{3}{8} - \frac{x}{8} \right) dx = \frac{69}{128}$$

(c) For $-2 \leq x \leq 0$ we have

$$F(x) = \int_{-2}^{x} \left(\frac{15}{64} + \frac{t}{64} \right) dt = \frac{x^2}{128} + \frac{15}{64}x + \frac{7}{16}$$

and for $0 < x \leq 3$

$$F(x) = \int_{-2}^{0} \left(\frac{15}{64} + \frac{x}{64} \right) dx + \int_{0}^{x} \left(\frac{3}{8} - \frac{t}{8} \right) dt = \frac{7}{16} + \frac{3}{8}x - \frac{1}{16}x^2.$$

Hence the full cdf is

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{x^2}{128} + \frac{15}{64}x + \frac{7}{16} & -2 \leq x \leq 0 \\ \frac{7}{16} + \frac{3}{8}x - \frac{1}{16}x^2 & 0 < x \leq 3 \\ 1 & x > 3 \end{cases}$$

Observe that at all points of discontinuity of the pdf, the cdf is continuous. That is, even when the pdf is discontinuous, the cdf is continuous ■

Mean and Variance for Continuous Random Variable

As with discrete random variables, the expected value of a continuous random variable is a measure of location. It defines the balancing point of the distribution.

Suppose that a continuous random variable X has a density function $f(x)$ defined in $[a, b]$. Let's try to estimate $E(X)$ by cutting $[a, b]$ into n equal subintervals, each of width Δx, so $\Delta x = \frac{(b-a)}{n}$. Let $x_i = a + i\Delta x, i = 0, 1, ..., n,$ be the partition points between the subintervals. Then, the probability of X assuming a value in $[x_i, x_{i+1}]$ is

$$P(x_i \leq X \leq x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x)dx \approx \Delta x f\left(\frac{x_i + x_{i+1}}{2}\right)$$

where we used the midpoint rule to estimate the integral. An estimate of the desired expectation is approximately

$$E(X) \approx \sum_{i=0}^{n-1} \left(\frac{x_i + x_{i+1}}{2}\right) \Delta x f\left(\frac{x_i + x_{i+1}}{2}\right).$$

A better estimate is obtained by letting $n \to \infty$. Thus, we obtain

$$E(X) = \int_a^b xf(x)dx.$$

The above argument applies if either a or b are infinite. In this case, one has to make sure that all improper integrals in question converge.

Since the domain of f consists of all real numbers, we define the **expected value** of X by the improper integral

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

provided that the improper integral converges.

Example 2.4.16
Find $E(X)$ when the density function of X is

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Solution.
Using the formula for $E(X)$ we find

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

Example 2.4.17
A continuous random variable has the pdf

$$f(x) = \begin{cases} \frac{600}{x^2}, & 100 < x < 120 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine the mean and variance of X.
(c) Find $P(X > 110)$.

Solution.
(a) We have

$$E(X) = \int_{100}^{120} x \cdot 600x^{-2} dx = 600 \ln x|_{100}^{120} \approx 109.39$$

and

$$\sigma_X^2 = E(X^2) - (E(X))^2 = \int_{100}^{120} x^2 \cdot 600x^{-2} dx - 109.39^2 \approx 33.19.$$

(b) The desired probability is

$$P(X > 110) = \int_{110}^{120} 600x^{-2} dx = \frac{5}{11} \blacksquare$$

<u>Median and Percentiles for Continuous Random Variables</u>
In addition to the information provided by the mean and variance of a distribution, some other metrics such as the median, the mode, the percentile, and the quantile provide useful information. For a continuous random variable X, the **median** is the number M such that $P(X \leq M) = P(X \geq M) = 0.5$. Generally, M is found by solving the equation $F(M) = 0.5$ where F is the cdf of X.

Example 2.4.18
Let X be a continuous random variable with pdf $f(x) = \frac{1}{b-a}$ for $a < x < b$ and 0 otherwise. Find the median of X.

Solution.
We must find a number M such that $\int_a^M \frac{dx}{b-a} = 0.5$. This leads to the equation $\frac{M-a}{b-a} = 0.5$. Solving this equation we find $M = \frac{a+b}{2}$ ■

In statistics, a percentile is the value of a variable below which a certain percent of observations fall. For example, if a score is in the 85$^{\text{th}}$ percentile, it is higher than 85% of the other scores. For a random variable X and $0 < p < 1$, the 100p$^{\text{th}}$ **percentile** (or the p^{th} **quantile**) is the number x such

$$P(X < x) \leq p \leq P(X \leq x).$$

For a continuous random variable, this is the solution to the equation $F(x) = p$. The 25$^{\text{th}}$ percentile is also known as the **first quartile**, the 50$^{\text{th}}$ percentile as the median or second quartile, and the 75$^{\text{th}}$ percentile as the third quartile.

Example 2.4.19
A loss random variable X has the density function

$$f(x) = \begin{cases} \frac{2.5(200)^{2.5}}{x^{3.5}} & x > 200 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate the difference between the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of X.

Solution.
First, the cdf is given by

$$F(x) = \int_{200}^{x} \frac{2.5(200)^{2.5}}{t^{3.5}} dt.$$

If Q_1 is the 25^{th} percentile then it satisfies the equation

$$F(Q_1) = \frac{1}{4}$$

or equivalently

$$1 - F(Q_1) = \frac{3}{4}.$$

This leads to

$$\frac{3}{4} = \int_{Q_1}^{\infty} \frac{2.5(200)^{2.5}}{t^{3.5}} dt = -\left(\frac{200}{t}\right)^{2.5}\Bigg|_{Q_1}^{\infty} = \left(\frac{200}{Q_1}\right)^{2.5}.$$

Solving for Q_1 we find $Q_1 = 200(4/3)^{0.4} \approx 224.4$. Similarly, the third quartile (i.e. 75^{th} percentile) is given by $Q_3 = 348.2$, The **interquartile range** (i.e., the difference between the 25^{th} and 75^{th} percentiles) is $Q_3 - Q_1 = 348.2 - 224.4 = 123.8$ ∎

Example 2.4.20
What percentile is 0.63 quantile?

Solution.
0.63 quantile is 63^{rd} percentile ∎