Introduction to Statistics

By:
Ewa Paszek

Introduction to Statistics

By:
Ewa Paszek

Online:
< http://cnx.org/content/col10343/1.3/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: October 9, 2007
PDF generated: October 26, 2012
For copyright and attribution information for the modules contained in this collection, see p. 101.

Table of Contents

Chapter 1

Discrete Distributions

1.1 DISCRETE DISTRIBUTION[1]

1.1.1 DISCRETE DISTRIBUTION

1.1.1.1 RANDOM VARIABLE OF DISCRETE TYPE

A **SAMPLE SPACE S** may be difficult to describe if the elements of **S** are not numbers. Let discuss how one can use a rule by which each simple outcome of a random experiment, an element **s** of **S**, may be associated with a real number **x**.

Definition 1.1: DEFINITION OF RANDOM VARIABLE

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** belongs to **S** means the element **s** belongs to the set **S**.

2. It may be that the set S has elements that are themselves real numbers. In such an instance we could write $X(s) = s$ so that **X** is **the identity function** and the space of **X** is also **S**. This is illustrated in the example below.

Example 1.1

Let the random experiment be the cast of a die, observing the number of spots on the side facing up. The sample space associated with this experiment is $S = (1, 2, 3, 4, 5, 6)$. For each **s** belongs to **S**, let $X(s) = s$. The space of the random variable **X** is then {1,2,3,4,5,6}.

If we associate a probability of 1/6 with each outcome, then, for example, $P(X = 5) = 1/6, P(2 \leq X \leq 5) = 4/6$, and **s** belongs to **S** seem to be reasonable assignments, where $(2 \leq X \leq 5)$ means (**X** = 2,3,4 or 5) and $(X \leq 2)$ means (**X** = 1 or 2), in this example.

We can recognize two major difficulties:

1. In many practical situations the probabilities assigned to the event are unknown.
2. Since there are many ways of defining a function **X** on **S**, which function do we want to use?

1.1.1.1.1

Let **X** denotes a random variable with one-dimensional space **R**, a subset of the real numbers. Suppose that the space **R** contains a countable number of points; that is, **R** contains either a finite number of points or

[1]This content is available online at <http://cnx.org/content/m13114/1.5/>.

the points of **R** can be put into a one-to- one correspondence with the positive integers. Such set **R** is called **a set of discrete points** or simply **a discrete sample space**.

Furthermore, the random variable **X** is called **a random variable of the discrete type**, and **X** is said to have **a distribution of the discrete type**. For a random variable **X** of the discrete type, the probability $P(X = x)$ is frequently denoted by **f(x)**, and is called **the probability density function** and it is abbreviated **p.d.f.**.

Let **f(x)** be the p.d.f. of the random variable **X** of the discrete type, and let **R** be the space of **X**. Since, $f(x) = P(X = x)$, **x** belongs to **R**, **f(x)** must be positive for **x** belongs to **R** and we want all these probabilities to add to 1 because each $P(X = x)$ represents the fraction of times **x** can be expected to occur. Moreover, to determine the probability associated with the event $A \subset R$, one would sum the probabilities of the **x** values in **A**.

That is, we want f(x) to satisfy the properties

- $P(X = x)$,
- $\sum_{x \in R} f(x) = 1$;
- $P(X \in A) = \sum_{x \in A} f(x)$, where $A \subset R$.

Usually let $f(x) = 0$ when $x \notin R$ and thus the domain of **f(x)** is the set of real numbers. When we define the p.d.f. of **f(x)** and do not say zero elsewhere, then we tacitly mean that **f(x)** has been defined at all x's in space **R**, and it is assumed that $f(x) = 0$ elsewhere, namely, $f(x) = 0$, $x \notin R$. Since the probability $P(X = x) = f(x) > 0$ when $x \in R$ and since **R** contains all the probabilities associated with **X**, **R** is sometimes referred to as **the support of X** as well as the space of **X**.

Example 1.2

Roll a four-sided die twice and let **X** equal the larger of the two outcomes if there are different and the common value if they are the same. The sample space for this experiment is $S = [(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4]$, where each of this 16 points has probability 1/16. Then $P(X = 1) = P[(1, 1)] = 1/16$, $P(X = 2) = P[(1, 2), (2, 1), (2, 2)] = 3/16$, and similarly $P(X = 3) = 5/16$ and $P(X = 4) = 7/16$. That is, the p. d.f. of **X** can be written simply as $f(x) = P(X = x) = \frac{2x-1}{16}, x = 1, 2, 3, 4.$

We could add that $f(x) = 0$ elsewhere; but if we do not, one should take **f(x)** to equal zero when $x \notin R$.

1.1.1.1.2

A better understanding of a particular probability distribution can often be obtained with a graph that depicts the p.d.f. of **X**.

NOTE: the graph of the p.d.f. when $f(x) > 0$, would be simply the set of points $\{[x, f(x)] : x \in R$ $\}$, where **R** is the space of **X**.

Two types of graphs can be used to give a better visual appreciation of the p.d.f., namely, **a bar graph** and **a probability histogram**. A bar graph of the p.d.f. **f(x)** of the random variable **X** is a graph having a vertical line segment drawn from $(x, 0)$ to $[x, f(x)]$ at each **x** in **R**, the space of **X**. If **X** can only assume integer values, **a probability histogram** of the p.d.f. f(x) is a graphical representation that has a rectangle of height **f(x)** and a base of length 1, centered at **x**, for each $x \in R$, the space of **X**.

Definition 1.2: CUMULATIVE DISTRIBUTION FUNCTION

1. Let **X** be a random variable of the discrete type with space **R** and p.d.f. $f(x) = P(X = x)$, $x \in R$. Now take **x** to be a real number and consider the set **A** of all points in **R** that are less than or equal to **x**. That is, $A = (t : t \leq x)$ and $t \in R$.

2. Let define the function **F(x)** by

$$F(x) = P(X \leq x) = \sum_{t \in A} f(t). \tag{1.1}$$

The function **F(x)** is called **the distribution function** (sometimes **cumulative distribution function**) of the discrete-type random variable **X**.

Several properties of a distribution function **F(x)** can be listed as a consequence of the fact that probability must be a value between 0 and 1, inclusive:

- $0 \leq F(x) \leq 1$ because **F(x)** is a probability,
- **F(x)** is a nondecreasing function of **x**,
- $F(y) = 1$, where **y** is any value greater than or equal to the largest value in **R**; and $F(z) = 0$, where **z** is any value less than the smallest value in **R**;
- If **X** is a random variable of the discrete type, then **F(x)** is a step function, and the height at a step at **x**, $x \in R$, equals the probability $P(X = x)$.

NOTE: It is clear that the probability distribution associated with the random variable **X** can be described by either the distribution function **F(x)** or by the probability density function **f(x)**. The function used is a matter of convenience; in most instances, **f(x)** is easier to use than **F(x)**.

Graphical representation of the relationship between p.d.f. and c.d.f.

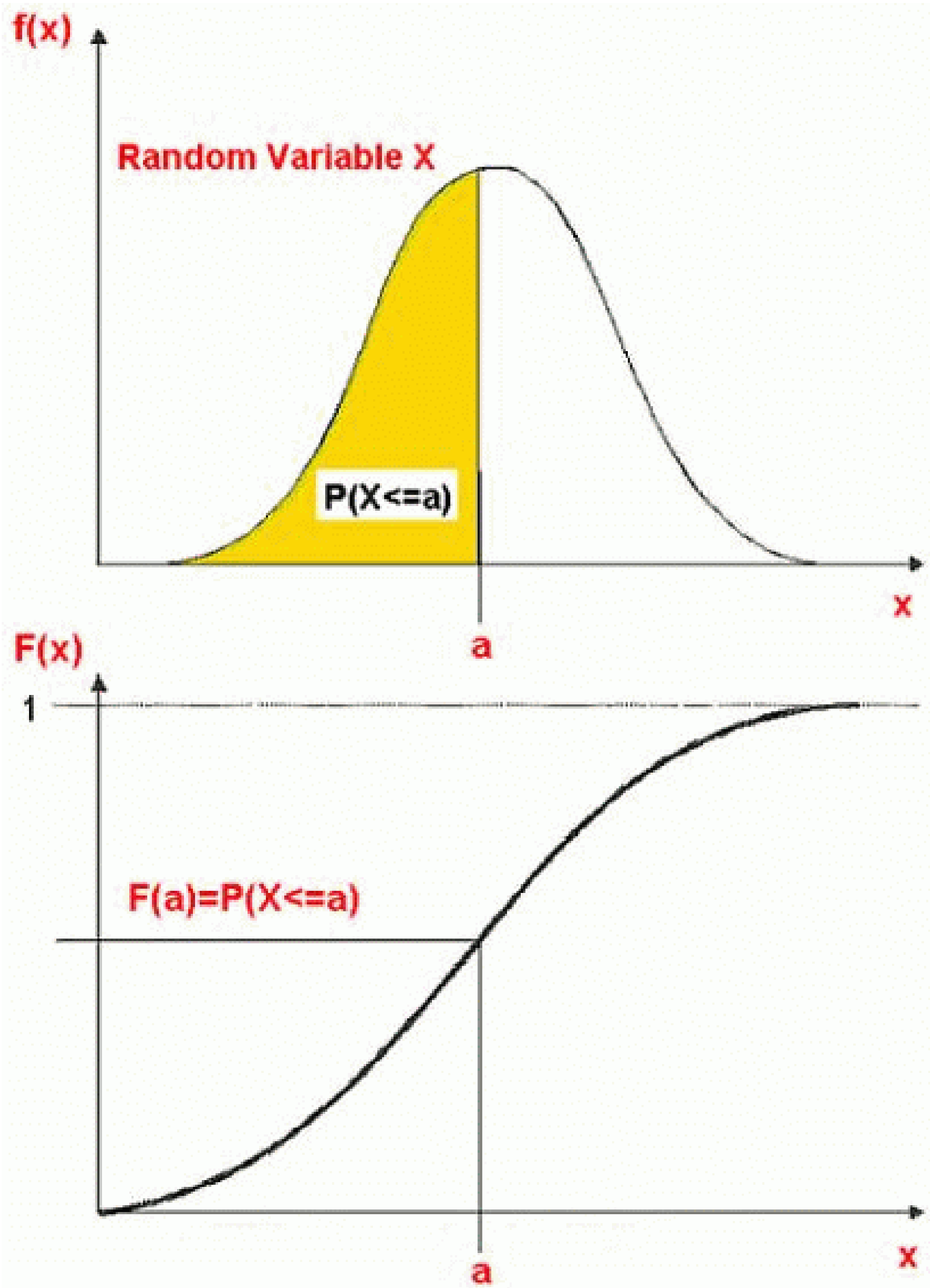

Figure 1.1: Area under p.d.f. curve to a equal to a value of c.d.f. curve at a point a.

1.1.1.1.3

Definition 1.3: MATHEMATICAL EXPECTATION
If **f(x)** is the p.d.f. of the random variable **X** of the discrete type with space **R** and if the summation

$$\sum_R u(x) f(x) = \sum_{x \in R} u(x) f(x) \tag{1.2}$$

exists, then the sum is called **the mathematical expectation** or **the expected value** of the function **u(X)**, and it is denoted by $E[u(X)]$. That is,

$$E[u(X)] = \sum_R u(x) f(x). \tag{1.3}$$

We can think of the expected value $E[u(X)]$ as a weighted mean of **u(x)**, $x \in R$, where the weights are the probabilities $f(x) = P(X = x)$.

NOTE: The usual definition of the mathematical expectation of **u(X)** requires that the sum converges absolutely; that is, $\sum_{x \in R} |u(x)| f(x)$ exists.

There is another important observation that must be made about consistency of this definition. Certainly, this function **u(X)** of the random variable **X** is itself a random variable, say **Y**. Suppose that we find the p.d.f. of **Y** to be **g(y)** on the support R_1 . Then **E(Y)** is given by the summation $\sum_{y \in R_1} y g(y)$
In general it is true that

$$\sum_R u(x) f(x) = \sum_{y \in R_1} y g(y);$$

that is, the same expectation is obtained by either method.

Example 1.3
Let **X** be the random variable defined by the outcome of the cast of the die. Thus the p.d.f. of **X** is
$f(x) = \frac{1}{6}$, $x = 1, 2, 3, 4, 5, 6.$
In terms of the observed value **x**, the function is as follows

$$u(x) = \{ \begin{array}{l} 1, x = 1, 2, 3, \\ 5, x = 4, 5, \\ 35, x = 6. \end{array}$$

The mathematical expectation is equal to

$$\sum_{x=1}^{6} u(x) f(x) = 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 35\left(\frac{1}{6}\right) = 1\left(\frac{3}{6}\right) + 5\left(\frac{2}{6}\right) + 35\left(\frac{1}{6}\right) = 8. \tag{1.4}$$

Example 1.4
Let the random variable **X** have the p.d.f. $f(x) = \frac{1}{3}$, $x \in R$, where **R** ={-1,0,1}. Let $u(X) = X^2$. Then

$$\sum_{x \in R} x^2 f(x) = (-1)^2 \left(\frac{1}{3}\right) + (0)^2 \left(\frac{1}{3}\right) + (1)^2 \left(\frac{1}{3}\right) = \frac{2}{3}. \tag{1.5}$$

However, the support of random variable $Y = X^2$ is $R_1 = (0, 1)$ and

$$P(Y=0) = P(X=0) = \frac{1}{3}$$
$$P(Y=1) = P(X=-1) + P(X=1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

That is,

$$g(y) = \{ \begin{array}{l} \frac{1}{3}, y=0, \\ \frac{2}{3}, y=1; \end{array}$$

and R_1. Hence

$\sum_{y \in R_1} y g(y) = 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right)$, which illustrates the preceding observation.

Theorem 1.1:

When it exists, mathematical expectation **E** satisfies the following properties:

1. If **c** is a constant, **E**(**c**)=**c**,
2. If **c** is a constant and **u** is a function, $E[cu(X)] = cE[u(X)]$,
3. If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$

Proof:

First, we have for the proof of (1) that

$E(c) = \sum_R cf(x) = c\sum_R f(x) = c$

because $\sum_R f(x) = 1$.

Proof:

Next, to prove (2), we see that

$E[cu(X)] = \sum_R cu(x) f(x) = c\sum_R u(x) f(x) = cE[u(X)]$.

Proof:

Finally, the proof of (3) is given by

$E[c_1 u_1(X) + c_2 u_2(X)] = \sum_R [c_1 u_1(x) + c_2 u_2(x)] f(x) = \sum_R c_1 u_1(x) f(x) + \sum_R c_2 u_2(x) f(x)$.

By applying (2), we obtain

$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(x)] + c_2 E[u_2(x)]$.

Property (3) can be extended to more than two terms by mathematical induction; That is, we have

3'. $E\left[\sum_{i=1}^{k} c_i u_i(X)\right] = \sum_{i=1}^{k} c_i E[u_i(X)]$.

Because of property (3'), mathematical expectation **E** is called a **linear** or **distributive operator**.

Example 1.5

Let **X** have the p.d.f. $f(x) = \frac{x}{10}$, **x**=1,2,3,4.

then

$E(X) = \sum_{x=1}^{4} x\left(\frac{x}{10}\right) = 1\left(\frac{1}{10}\right) + 2\left(\frac{2}{10}\right) + 3\left(\frac{3}{10}\right) + 4\left(\frac{4}{10}\right) = 3$

$E(X^2) = \sum_{x=1}^{4} x^2\left(\frac{x}{10}\right) = 1^2\left(\frac{1}{10}\right) + 2^2\left(\frac{2}{10}\right) + 3^2\left(\frac{3}{10}\right) + 4^2\left(\frac{4}{10}\right) = 10,$

and

$E[X(5-X)] = 5E(X) - E(X^2) = (5)(3) - 10 = 5.$

1.1.2

NOTE: the MEAN, VARIANCE, and STANDARD DEVIATION (Section 1.3.1: The MEAN, VARIANCE, and STANDARD DEVIATION)

1.2 MATHEMATICAL EXPECTATION[2]

1.2.1 MATHEMATICAL EXPECTIATION

Definition 1.4: MATHEMATICAL EXPECTIATION
If $f(x)$ is the p.d.f. of the random variable **X** of the discrete type with space **R** and if the summation

$$\sum_R u(x) f(x) = \sum_{x \in R} u(x) f(x). \tag{1.6}$$

exists, then the sum is called **the mathematical expectation** or **the expected value** of the function $u(X)$, and it is denoted by $E[u(x)]$. That is,

$$E[u(X)] = \sum_R u(x) f(x). \tag{1.7}$$

We can think of the expected value $E[u(x)]$ as a weighted mean of $u(x)$, $x \in R$, where the weights are the probabilities $f(x) = P(X = x)$.

NOTE: The usual definition of the mathematical expectation of $u(X)$ requires that the sum converges absolutely; that is, $\sum_{x \in R} |u(x)| f(x)$ exists.

There is another important observation that must be made about consistency of this definition. Certainly, this function $u(X)$ of the random variable **X** is itself a random variable, say **Y**. Suppose that we find the p.d.f. of **Y** to be $g(y)$ on the support R_1 . Then, $E(Y)$ is given by the summation $\sum_{y \in R_1} y g(y)$.

In general it is true that $\sum_R u(x) f(x) = \sum_{y \in R_1} y g(y)$.

This is, the same expectation is obtained by either method.

1.2.1.1

Example 1.6
Let **X** be the random variable defined by the outcome of the cast of the die. Thus the p.d.f. of **X** is

$f(x) = \frac{1}{6}$, $x = 1, 2, 3, 4, 5, 6.$

In terms of the observed value **x**, the function is as follows

$$u(x) = \begin{cases} 1, x = 1, 2, 3, \\ 5, x = 4, 5, \\ 35, x = 6. \end{cases}$$

The mathematical expectation is equal to

$\sum_{x=1}^{6} u(x) f(x) = 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 35\left(\frac{1}{6}\right) = 1\left(\frac{3}{6}\right) + 5\left(\frac{2}{6}\right) + 35\left(\frac{1}{6}\right) = 8.$

1.2.1.2

Example 1.7
Let the random variable **X** have the p.d.f.

$f(x) = \frac{1}{3}$, $x \in R$,

where, $R = (-1, 0, 1)$. Let $u(X) = X^2$. Then

$\sum_{x \in R} x^2 f(x) = (-1)^2\left(\frac{1}{3}\right) + (0)^2\left(\frac{1}{3}\right) + (1)^2\left(\frac{1}{3}\right) = \frac{2}{3}.$

However, the support of random variable $Y = X^2$ is $R_1 = (0, 1)$ and

[2] This content is available online at <http://cnx.org/content/m13530/1.2/>.

$P(Y=0) = P(X=0) = \frac{1}{3}$
$P(Y=1) = P(X=-1) + P(X=1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$

That is, $g(y) = \{ \begin{array}{l} \frac{1}{3}, y=0, \\ \frac{2}{3}, y=1; \end{array}$ and $R_1 = (0,1)$. Hence

$$\sum_{y \in R_1} y g(y) = 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right) = \frac{2}{3},$$

which illustrates the preceding observation.

1.2.1.3

Theorem 1.2:
When it exists, mathematical expectation **E** satisfies the following properties:

1. If **c** is a constant, $E(c) = c$,
2. If **c** is a constant and **u** is a function, $E[cu(X)] = cE[u(X)]$,
3. If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$.

Proof:
First, we have for the proof of (1) that

$$E(c) = \sum_R c f(x) = c \sum_R f(x) = c,$$

because $\sum_R f(x) = 1$.
Proof:
Next, to prove (2), we see that

$$E[cu(X)] = \sum_R c u(x) f(x) = c \sum_R u(x) f(x) = cE[u(X)].$$

Proof:
Finally, the proof of (3) is given by

$$E[c_1 u_1(X) + c_2 u_2(X)] = \sum_R [c_1 u_1(x) + c_2 u_2(x)] f(x) = \sum_R c_1 u_1(x) f(x) + \sum_R c_2 u_2(x) f(x).$$

By applying (2), we obtain

$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(x)] + c_2 E[u_2(x)].$$

Property (3) can be extended to more than two terms by mathematical induction; that is, we have (3')

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

Because of property (3'), mathematical expectation **E** is called **a linear** or **distributive operator**.

1.2.1.4

Example 1.8
Let **X** have the p.d.f. $f(x) = \frac{x}{10}, x = 1, 2, 3, 4$, then

$$E(X) = \sum_{x=1}^{4} x\left(\frac{x}{10}\right) = 1\left(\frac{1}{10}\right) + 2\left(\frac{2}{10}\right) + 3\left(\frac{3}{10}\right) + 4\left(\frac{4}{10}\right) = 3,$$

$$E\left(X^2\right) = \sum_{x=1}^{4} x^2\left(\frac{x}{10}\right) = 1^2\left(\frac{1}{10}\right) + 2^2\left(\frac{2}{10}\right) + 3^2\left(\frac{3}{10}\right) + 4^2\left(\frac{4}{10}\right) = 10,$$

and

$$E[X(5-X)] = 5E(X) - E\left(X^2\right) = (5)(3) - 10 = 5.$$

1.3 THE MEAN, VARIANCE, AND STANDARD DEVIATION[3]

1.3.1 The MEAN, VARIANCE, and STANDARD DEVIATION

1.3.1.1 MEAN and VARIANCE

Certain mathematical expectations are so important that they have special names. In this section we consider two of them: the mean and the variance.

1.3.1.1.1

Mean Value
If **X** is a random variable with p.d.f. $f(x)$ of the discrete type and space **R**=$(b_1, b_2, b_3, ...)$, then $E(X) = \sum_R xf(x) = b_1 f(b_1) + b_2 f(b_2) + b_3 f(b_3) + ...$ is the weighted average of the numbers belonging to **R**, where the weights are given by the p.d.f. $f(x)$.

We call $E(X)$ **the mean** of **X** (or **the mean of the distribution**) and denote it by μ. That is, $\mu = E(X)$.

NOTE: In mechanics, the weighted average of the points $b_1, b_2, b_3, ...$ in one-dimensional space is called the centroid of the system. Those without the mechanics background can think of the centroid as being the point of balance for the system in which the weights $f(b_1), f(b_2), f(b_3), ...$ are places upon the points $b_1, b_2, b_3, ...$

Example 1.9
Let **X** have the p.d.f.

$$f(x) = \{ \begin{array}{l} \frac{1}{8}, x = 0, 3, \\ \frac{3}{8}, x = 1, 2. \end{array}$$

The mean of **X** is

$$\mu = E\left[X = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{3}{2}.\right.$$

The example below shows that if the outcomes of **X** are equally likely (i.e., each of the outcomes has the same probability), then the mean of **X** is the arithmetic average of these outcomes.

[3]This content is available online at <http://cnx.org/content/m13122/1.3/>.

Example 1.10
Roll a fair die and let **X** denote the outcome. Thus **X** has the p.d.f.

$$f(x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.$$

Then,

$$E(X) = \sum_{x=1}^{6} x\left(\frac{1}{6}\right) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2},$$

which is the arithmetic average of the first six positive integers.

1.3.1.1.2

Variance
It was denoted that the mean $\mu = E(X)$ is the centroid of a system of weights of measure of the central location of the probability distribution of **X**. **A measure of the dispersion or spread of a distribution is defined as follows:**
If $u(x) = (x-\mu)^2$ and $E\left[(X-\mu)^2\right]$ exists, **the variance**, frequently denoted by σ^2 or $Var(X)$, of a random variable **X** of the discrete type (or variance of the distribution) is defined by

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_R (x-\mu)^2 f(x). \quad (1.8)$$

The positive square root of the variance is called **the standard deviation of X** and is denoted by

$$\sigma = \sqrt{Var(X)} = \sqrt{E\left[(X-\mu)^2\right]}. \quad (1.9)$$

Example 1.11
Let the p.d.f. of **X** by defined by

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

The mean of **X** is

$$\mu = E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{2}{6}\right) + 3\left(\frac{3}{6}\right) = \frac{7}{3}.$$

To find the variance and standard deviation of **X** we first find

$$E\left(X^2\right) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{2}{6}\right) + 3^2\left(\frac{3}{6}\right) = \frac{36}{6} = 6.$$

Thus the variance of **X** is

$$\sigma^2 = E\left(X^2\right) - \mu^2 = 6 - \left(\frac{7}{3}\right)^2 = \frac{5}{9},$$

and the standard deviation of **X** is

Example 1.12
Let **X** be a random variable with mean μ_x and variance σ_x^2. Of course, $Y = aX + b$, where a and b are constants, is a random variable, too. The mean of **Y** is

$$\mu_Y = E(Y) = E(aX + b) = aE(X) + b = a\mu_X + b.$$

Moreover, the variance of **Y** is

$$\sigma_Y^2 = E\left[(Y - \mu_Y)^2\right] = E\left[(aX + b - a\mu_X - b)^2\right] = E\left[a^2(X - \mu_X)^2\right] = a^2\sigma_X^2.$$

1.3.1.1.3

Moments of the distribution
Let **r** be a positive integer. If

$$E(X^r) = \sum_R x^r f(x)$$

exists, it is called **the rth moment of the distribution** about the origin. The expression moment has its origin in the study of mechanics.

In addition, the expectation

$$E[(X - b)^r] = \sum_R x^r f(x)$$

is called **the rth moment of the distribution about b**. For a given positive integer r.

$$E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)]$$

is called **the rth factorial moment**.

NOTE: The second factorial moment is equal to the difference of the second and first moments:

$$E[X(X-1)] = E(X^2) - E(X).$$

There is another formula that can be used for computing the variance that uses the second factorial moment and sometimes simplifies the calculations.

First find the values of $E(X)$ and $E[X(X-1)]$. Then

$$\sigma^2 = E[X(X-1)] + E(X) - [E(X)]^2,$$

since using the distributive property of **E**, this becomes

$$\sigma^2 = E(X^2) - E(X) + E(X) - [E(X)]^2 = E(X^2) - \mu^2.$$

Example 1.13
Let continue with example 4 (Example 1.12), it can be find that

$$E[X(X-1)] = 1(0)\left(\frac{1}{6}\right) + 2(1)\left(\frac{2}{6}\right) + 3(2)\left(\frac{3}{6}\right) = \frac{22}{6}.$$

Thus

$$\sigma^2 = E[X(X-1)] + E(X) - [E(X)]^2 = \frac{22}{6} + \frac{7}{3} - \left(\frac{7}{3}\right)^2 = \frac{5}{9}.$$