Random Variables and Probability Distributions

August 26, 2006

1 Random Variables

1.1 Definition of a Random Variable

Consider an experiment with a sample space Ω. Then consider a function X that assigns to each possible outcome in an experiment ($\omega \in \Omega$) one and only one real number, $X(\omega) = x$. The space of X is the set of real numbers $\{x : X(\omega) = x,\, \omega \in \Omega\}$, where $\omega \in \Omega$ means that the element ω belongs to the set Ω. If X may assume any value in some given interval I (the interval may be bounded or unbounded), it is called a continuous random variable. If it can assume only a number of separated values, it is called a discrete random variable.

An even simpler (but less precise) definition of a random variable is as follows. A random variable is a real-valued function for which the domain is a sample space.

Alternatively one can think of a random variable as a set function, it assigns a real number to a set. Consider the following example.

Roll a red die and a green die. Let the random variable assigned to each outcome be the sum of the numbers on the dice. There are 36 points in the sample space. In table 1 the outcomes are listed along with the value of the random variable associated with each outcome.

We then write $X = 8$ for the subset of Ω given by

$$\{(6,\,2)(5,\,3)(4,\,4)(3,\,5)(2,\,6)\}$$

Table 1: Possible outcomes of rolling a red die and a green die. First number in pair is number on red die.

Green (A) / Red (D)	1	2	3	4	5	6
1	1 1 *2*	1 2 *3*	1 3 *4*	1 4 *5*	1 5 *6*	1 6 *7*
2	2 1 *3*	2 2 *4*	2 3 *5*	2 4 *6*	2 5 *7*	2 6 *8*
3	3 1 *4*	3 2 *5*	3 3 *6*	3 4 *7*	3 5 *8*	3 6 *9*
4	4 1 *5*	4 2 *6*	4 3 *7*	4 4 *8*	4 5 *9*	4 6 *10*
5	5 1 *6*	5 2 *7*	5 3 *8*	5 4 *9*	5 5 *10*	5 6 *11*
6	6 1 *7*	6 2 *8*	6 3 *9*	6 4 *10*	6 5 *11*	6 6 *12*

Thus, $X = 8$ is to be interpreted as the set of elements of Ω for which the total is 8.

Or consider another example where a student has 5 pairs of brown socks and 6 pairs of black socks or 22 total socks. He does not fold his socks but throws them all in a drawer. Each morning he draws two socks from his drawer. What is the probability he will choose two black socks? We will first develop the sample space where we use the letter B to denote black and the letter R (for russet) to denote brown. There are four possible outcomes to this experiment: BB, BR, RB and RR. Let the random variable defined on this sample space be the number of black socks drawn.

Table 2: Sample space, values of random variable and probabilities for sock experiment

Element of Sample Space	**Value of Random Variable**	**Probability of Random Variable**
BB	2	$\left(\frac{12}{22}\right)\left(\frac{11}{21}\right) = \frac{22}{77}$
BR	1	$\left(\frac{12}{22}\right)\left(\frac{10}{21}\right) = \frac{20}{77}$
RB	1	$\left(\frac{10}{22}\right)\left(\frac{12}{21}\right) = \frac{20}{77}$
RR	0	$\left(\frac{10}{22}\right)\left(\frac{9}{21}\right) = \frac{15}{77}$

We can compute the probability of two black socks by multiplying 12/22 by 11/21 to obtain 22/77. We can compute the other probabilities in a similar manner.

1.2 Notation

We will normally use *uppercase letters*, such as X to denote *random variables*, and *lowercase letters*, such as x, to denote particular values that a random variable may assume. For example, let X denote any one of the six possible values that can result from tossing a die. After the die is tossed, the number actually observed will be denoted by the symbol x. Note that X is a random variable, but the specific observed value, s, is not random. We sometimes write $P(X = x)$ for the probability that the random variable X takes on the value x.

1.3 Random Vector

A random vector $X = (X_1, X_2, \ldots, X_k)'$ is a k-tuple of random variables. For $k = 1$, random vectors are just random variables. The event $X^{-1}(B)$ (meaning the event that leads to the function X mapping it into the set B in R^k) is usually written $[X \epsilon B]$. This means that when we write $[X \epsilon B]$ we mean that the event that occurred lead to an outcome that mapped into the set B in R^k. And when we write $P[X \epsilon B]$, we mean the probability of the event that will map into the set B in R^k.

1.4 Probability Distribution of a Random Vector

The probability distribution of a random vector S is defined as the probability measure in the probability model $(R^k,\, \mathcal{F}^k,\, P_X)$. This is given by

$$P_X(B) = P([X \,\epsilon\, B]) \tag{1}$$

2 Random Sampling

2.1 Simple Random Sampling

Simple random sampling is the basic sampling technique where we select a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. Every possible sample of a given size has the same chance of selection, i.e. each member of the population is equally likely to be chosen at any stage in the sampling process.

If, for example, numbered pieces of cardboard are drawn from a hat, it is important that they be thoroughly mixed, that they be identical in every respect except for the number printed on them and that the person selecting them be well blindfolded.

Second in order to meet the equal opportunity requirement, it is important that the sampling be done with replacement. That is, each time an item is selected, the relevant measure is taken and recorded. Then the item must be replaced in the population and be thoroughly mixed with the other items before the next item is drawn. If the items are not replaced in the population, each time an item is withdrawn, the probability of being selected, for each of the remaining items, will have been increased. For example, if in the black sock brown sock problem, we drew two black socks from the drawer and put them on, and then repeated the experiment, the probability of two black socks would now be 10/20 multiplied by 9/19 rather than $12/22 \times 11/21$. Of course, this kind of change in probability becomes trivial if the population is very large.

More formally consider a population with N elements and a sample of size n. If the sampling is conducted in such a way that each of the $\binom{N}{n}$

samples has an equal probability of bring selected, the sampling is said to be random and the result is said to be a simple random sample.

2.2 Random Sampling

Random sampling is a sampling technique where we select a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, **but possibly non-equal**, chance of being included in the sample.

2.3 Definition of a Discrete Random Variable

A random variable X is said to be *discrete*if it can assume only a finite or countable infinite number of distinct values. A discrete random variable can be defined on both a countable or uncountable sample space.

2.4 Probability for a discrete random variable

The probability that X takes on the value x, $P(X = x)$, is defined as the sum of the probabilities of all sample points in Ω that are assigned the value x. We may denote $P(X = x)$ by $p(x)$ or $p_X(x)$. The expression $p_X(x)$ is a function that assigns probabilities to each possible value x; thus it is often called the probability function for the random variable X.

2.5 Probability distribution for a discrete random variable

The probability distribution for a discrete random variable X can be represented by a formula, a table, or a graph, which provides $p_X(x) = P(X = x)$ for all x. The probability distribution for a discrete random variable assigns nonzero probabilities to only a countable number of distinct x values. Any value x not explicitly assigned a positive probability is understood to be such that $P(X = x) = 0$.

The function $p_X(x) = P(X = x)$ for each x within the range of X is called the *probability distribution*of X. It is often called the probability mass function for the discrete random variable X.

2.6 Properties of the probability distribution for a discrete random variable

A function can serve as the probability distribution for a discrete random variable X if and only if it's values, $p_X(x)$, satisfy the conditions:

a $p_X(x) \geq 0$ for each value within its domain

b $\sum_x p_X(x) = 1$, where the summation extends over all the values within its domain

2.7 Examples of probability mass functions

2.7.1 Example 1

Find a formula for the probability distribution of the total number of heads obtained in four tosses of a balanced coin.

The sample space, probabilities and the value of the random variable are given in the table below:

TABLE 2

Table R.1
Tossing a Coin Four Times

Element of sample space	Probability	Value of random variable X (x)
HHHH	1/16	4
HHHT	1/16	3
HHTH	1/16	3
HTHH	1/16	3
THHH	1/16	3
HHTT	1/16	2
HTHT	1/16	2
HTTH	1/16	2
THHT	1/16	2
THTH	1/16	2
TTHH	1/16	2
HTTT	1/16	1
THTT	1/16	1
TTHT	1/16	1
TTTH	1/16	1
TTTT	1/16	0

From the table we can determine the probabilities as

$$P(X=0) = \frac{1}{16},\ P(X=1) = \frac{4}{16},\ P(X=2) = \frac{6}{16},\ P(X=3) = \frac{4}{16},\ P(X=4) = \frac{1}{16} \quad (2)$$

Notice that the denominators of the five fractions are the same and the numerators of the five fractions are $1, 4, 6, 4, 1$. The numbers in the numerators is a set of binomial coefficients.

$$\begin{aligned} \frac{1}{16} &= \binom{4}{0} \cdot \frac{1}{16}, & \frac{4}{16} &= \binom{4}{1} \cdot \frac{1}{16}, & \frac{6}{16} &= \binom{4}{2} \cdot \frac{1}{16}, \\ \frac{4}{16} &= \binom{4}{3} \cdot \frac{1}{16}, & \frac{1}{16} &= \binom{4}{4} \cdot \frac{1}{16} \end{aligned}$$

We can then write the probability mass function as

$$p_X(x) = \frac{\binom{4}{x}}{16} \ for\ x = 0,\ 1,\ 2,\ 3,\ 4 \quad (3)$$

Note that all the probabilities are positive and that they sum to one.

2.7.2 Example 2

Roll a red die and a green die. Let the random variable be the larger of the two numbers if they are different and the common value if they are the same. There are 36 points in the sample space. In the table below the outcomes are listed along with the value of the random variable associated with each outcome.

Green (A)	1	2	3	4	5	6
Red (D)						
1	1 1 *1*	1 2 *2*	1 3 *3*	1 4 *4*	1 5 *5*	1 6 *6*
2	2 1 *2*	2 2 *2*	2 3 *3*	2 4 *4*	2 5 *5*	2 6 *6*
3	3 1 *3*	3 2 *3*	3 3 *3*	3 4 *4*	3 5 *5*	3 6 6
4	4 1 *4*	4 2 *4*	4 3 *4*	4 4 *4*	4 5 *5*	4 6 *6*
5	5 1 *5*	5 2 *5*	5 3 *5*	5 4 *5*	5 5 *5*	5 6 *6*
6	6 1 *6*	6 2 *6*	6 3 *6*	6 4 *6*	6 5 *6*	6 6 *6*

The probability that $X = 1$, $P(X = 1) = P[(1,1)] = 1/36$. The probability that $X = 2$, $P(X = 2) = P[(1,2),(2,1),(2,2)] = 3/36$. Continuing we obtain

$$\begin{aligned} P(X=1) &= \frac{1}{36}, \quad P(X=2) = \frac{3}{36}, \quad P(X=3) = \frac{5}{36}, \\ P(X=4) &= \frac{7}{36}, \quad P(X=5) = \frac{9}{36}, \quad P(X=6) = \frac{11}{36} \end{aligned}$$

We can then write the probability mass function as

$$p_X(x) = P(X = x) = \frac{2x - 1}{36} \quad for\ x = 1, 2, 3, 4, 5, 6$$

Note that all the probabilities are positive and that they sum to one.

2.8 Cumulative Distribution Functions

2.8.1 Definition of a Cumulative Distribution Function

If X is a discrete random variable, the function given by

$$F_X(x) = P(x \leq X) = \sum_{t \leq x} p(t) \quad for - \infty \leq x \leq \infty$$

where p(t) is the value of the probability distribution of X at t, is called the *cumulative distribution function* of X. The function F_X(x) is also called the *distribution function* of X.

2.8.2 Properties of a Cumulative Distribution Function

The values $F_X(X)$ of the distribution function of a discrete random variable X satisfy the conditions

1 $F(-\infty) = 0$ and $F(\infty) = 1$;

2 If $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b

2.8.3 First example of a cumulative distribution function

Consider tossing a coin four times. The possible outcomes are contained in Table 2 and the values of $p(\cdot)$ in equation 3. From this we can determine the cumulative distribution function as follows.

$$
\begin{aligned}
F(0) &= p(0) = \frac{1}{16} \\
F(1) &= p(0) + p(1) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16} \\
F(2) &= p(0) + p(1) + p(2) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16} \\
F(3) &= p(0) + p(1) + p(2) + p(3) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{6} = \frac{15}{16} \\
F(4) &= p(0) + p(1) + p(2) + p(3) + p(4) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{6} + \frac{1}{16} = \frac{16}{16}
\end{aligned}
$$

We can write this in an alternative fashion as

$$
F_X(x) = \begin{cases} 0 & for\ x < 0 \\ \frac{1}{16} & for\ 0 \le x < 1 \\ \frac{5}{16} & for\ 1 \le x < 2 \\ \frac{11}{16} & for\ 2 \le x < 3 \\ \frac{15}{16} & for\ 3 \le x < 4 \\ 1 & for\ x \ge 4 \end{cases}
$$

2.8.4 Second example of a cumulative distribution function

Consider a group of N individuals, M of whom are female. Then $N - M$ are male. Now pick n individuals from this population without replacement. Let x be the number of females chosen. There are $\binom{M}{x}$ ways of choosing x females from the M in the population and $\binom{N-M}{n-x}$ ways of choosing $n - x$ of the $N - M$ males. Therefore, there are $\binom{M}{x} \cdot \binom{N-M}{n-x}$ ways of choosing x females and $n - x$ males. Because there are $\binom{N}{n}$ ways of choosing n of the N elements in the set, and because we will assume that they all are equally likely, the probability of x females in a sample of size n is given by

$$p_X(x) \;=\; P(X = x) \;=\; \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} for\ x \;=\; 0\,,\ 1\,,\ 2\,,\ 3\,,\cdots\,,\ n$$
$$and\ x \;\leq\; M,\ and\ n \;-\; x \;\leq\; N \;-\; M. \tag{4}$$

For this discrete distribution we compute the cumulative density by adding up the appropriate terms of the probability mass function.

$$\begin{aligned}
F(0) &= \quad p(0) \\
F(1) &= \quad p(0) \;+\; p(1) \\
F(2) &= \quad p(0) \;+\; p(1) \;+\; p(2) \\
F(3) &= \quad p(0) \;+\; p(1) \;+\; p(2) \;+\; px(3) \\
&\vdots \\
F(n) &= \quad p(0) \;+\; p(1) \;+\; p(2) \;+\; p(3) \;+\; \cdots \;+\; p(n)
\end{aligned}$$

Consider a population with four individuals, three of whom are female, denoted respectively by A, B, C, D where A is a male and the others are females. Then consider drawing two from this population. Based on equation 4 there should be $\binom{4}{2} = 6$ elements in the sample space. The sample space is given by

Table 2: **Drawing Two Individuals from a Population of Four where Order Does Not Matter (no replacement)**

Element of sample space	Probability	Value of random variable X
AB	1/6	1
AC	1/6	1
AD	1/6	1
BC	1/6	2
BD	1/6	2
CD	1/6	2

We can see that the probability of 2 females is $\frac{1}{2}$. We can also obtain this using the formula as follows.

$$p(2) = P(X = 2) = \frac{\binom{3}{2}\binom{1}{0}}{\binom{4}{2}} = \frac{(3)(1)}{6} = \frac{1}{2}$$

Similarly

$$p(1) = P(X = 1) = \frac{\binom{3}{1}\binom{1}{1}}{\binom{4}{2}} = \frac{(3)(1)}{6} = \frac{1}{2}$$

We cannot use the formula to compute $P(0)$ because $(2-0) \not\leq (4-3)$. $P(0)$ is then equal to 0. We can then compute the cumulative distribution function as

$$\begin{aligned} F(0) &= p(0) = 0 \\ F(1) &= p(0) + p(1) = \frac{1}{2} \\ F(2) &= p(0) + p(1) + p(2) = 1 \end{aligned}$$

2.9 Expected value

2.9.1 Definition of expected value

Let X be a discrete random variable with probability function $p_X(x)$. Then the *expected value* of X, E(X), is defined to be

$$E(X) = \sum_x x\, p_X(x) \tag{5}$$

if it exists. The expected value exists if

$$\sum_x |\, x\,|\; p_X(x) < \infty$$

The expected value is kind of a weighted average. It is also sometimes referred to as the population mean of the random variable and denoted μ_X.

2.9.2 First example computing an expected value

Toss a die that has six sides. Observe the number that comes up. The probability mass or frequency function is given by

$$p_X(x) \; = \; P(X \; = \; x) \; = \begin{cases} \frac{1}{6} & for \quad x \; = \; 1, 2, 3, 4, 5, 6 \\ 0 & otherwise \end{cases}$$

We compute the expected value as

$$\begin{aligned} E(X) \; &= \; \sum_{x \; \epsilon \; X} x \; p_X\,(x) \\ &= \; \sum_{i \; = \; 1}^{6} i \; \left(\frac{1}{6}\right) \\ &= \; \frac{1 + \; 2 \; + \; 3 \; + \; 4 \; + \; 5 \; + \; 6}{6} \\ &= \; \frac{21}{6} \; = \; 3 \; \frac{1}{2} \end{aligned}$$

2.9.3 Second example computing an expected value

Consider a group of 12 television sets, two of which have white cords and ten which have black cords. Suppose three of them are chosen at random and shipped to a care center. What are the probabilities that zero, one, or two of the sets with white cords are shipped? What is the expected number with white cords that will be shipped?

It is clear that x of the two sets with white cords and $3 - x$ of the ten sets with black cords can be chosen in $\binom{2}{x} \times \binom{10}{3-x}$ ways. The three sets can be chosen in $\binom{12}{3}$ ways. So we have a probability mass function as follows.

$$p_X(x) \; = \; P(X \; = \; x) \; = \; \frac{\binom{2}{x} \; \binom{10}{3-x}}{\binom{12}{3}} \quad for \;\; x \; = \; 0\,, \; 1\,, \; 2$$

For example

$$p(0) \; = \; P(X \; = \; 0) \; = \; \frac{\binom{2}{0} \; \binom{10}{3-0}}{\binom{12}{3}} \; = \; \frac{(1)\,(120)}{220} \; = \; \frac{6}{11}$$

We collect this information as in table 3.
We compute the expected value as

Table 3: **Probabilities for Television Problem**

x	0	1	2
$p_X(x)$	6/11	9/22	1/22
$F_X(x)$	6/11	21/22	1

$$\begin{aligned} E(X) &= \sum_{x \,\epsilon\, X} x \; p_X\,(x) \\ &= (0)\left(\frac{6}{11}\right) + (1)\left(\frac{9}{22}\right) + (2)\left(\frac{1}{22}\right) = \frac{11}{22} = \frac{1}{2} \end{aligned}$$

Note that the expected value is not in the sample space.

2.10 Expected value of a function of a random variable

Theorem 1 *Let X be a discrete random variable with probability mass function $p_X(x)$ and $g(X)$ be a real-valued function of X. Then the expected value of $g(X)$ is given by*

$$E[g(X)] = \sum_x g(x)\; p_X(x)\,.$$

Proof for case of finite values of X. Consider the case where the random variable X takes on a finite number of values $x_1, x_2, x_3, \cdots x_n$. The function $g(x)$ may not be one-to-one (the different values of x_i may yield the same value of $g(x_i)$. Suppose that $g(X)$ takes on m different values ($m \leq n$). It follows that $g(X)$ is also a random variable with possible values g_1, g_2, g_3, . . . g_m and probability distribution

$$P[g(X) = g_i] = \sum_{\substack{\forall j \; such\;\; that \\ g(x_j) = g_i}} p(x_j) = p^*\,(g_i)$$

for all $i = 1, 2,$ . . . m. Here $p^*(g_i)$ is the probability that the experiment results in a value for the function f of the initial random variable of g_i. Using the definition of expected value in equation we obtain

$$E[g(X)] \;=\; \sum_{i=1}^{m} \; g_i \; p^*(g_i).$$

Now substitute in to obtain

$$\begin{aligned} E[g(X)] \quad &= \quad \sum_{i=1}^{m} (\, g_i \sum_{\substack{\forall j \; such \;\; that \\ g(x_j) = g_i}} p(x_j)) \\ &= \quad \sum_{i=1}^{m} \sum_{\substack{\forall j \; such \;\; that \\ g(x_j) = g_i}} g_i \; p(x_j) \\ &= \quad \sum_{j=1}^{n} g(x_j) \; p(x_j) \end{aligned}$$

■

2.10.1 Example demonstrating that $E[g(X)] \neq g\left(E\left[X\right]\right)$

Consider the following example of a discrete random variable X:

$Values$	4	6	8	10	12
Probabilities	0.2	0.2	0.2	0.2	0.2

The expected value can be calculated as:

$$E\left[X\right] = 0.2 \cdot 4 + 0.2 \cdot 6 + 0.2 \cdot 8 + 0.2 \cdot 10 + 0.2 \cdot 12 = 8$$

Now consider $g(X) = 1/X$

$Values$	1/4	1/6	1/8	1/10	1/12
Probabilities	0.2	0.2	0.2	0.2	0.2

The expected value can be calculated as:

$$E\left[X\right] = 0.2 \cdot \frac{1}{4} + 0.2 \cdot \frac{1}{6} + 0.2 \cdot \frac{1}{8} + 0.2 \cdot \frac{1}{10} + 0.2 \cdot \frac{1}{12} = 0.658\bar{3}$$

Note that

$$0.658\bar{3} = E\left[g\left(x\right)\right] \neq g\left(E\left(x\right)\right) = \frac{1}{8}$$

2.11 Properties of mathematical expectation

2.11.1 Constants

Theorem 2 *Let X be a discrete random variable with probability function $p_X(x)$ and c be a constant. Then $E(c) = c$.*

Proof. Consider the function $g(X) = c$. Then by Theorem 1

$$E[c] \equiv \sum_x c\, p_X(x) = c \sum_x p_X(x)$$

But by property 2.6b, we have

$$\sum_x p_X(x) = 1$$

and hence

$$E(c) = c \cdot (1) = c.$$

■

2.11.2 Constants multiplied by functions of random variables

Theorem 3 *Let X be a discrete random variable with probability function $p_X(x)$, $g(X)$ be a function of X, and let c be a constant. Then*

$$E[c\, g(X)] \equiv c\, E[(g(X)]$$

Proof.
By Theorem 1 we have

$$\begin{aligned} E[c\, g(X)] &\equiv \sum_x c\, g(x)\, p_X(x) \\ &= c \sum_x g(x)\, p_X(x) \\ &= c\, E[g(X)] \end{aligned}$$

■

2.11.3 Sums of functions of random variables

Theorem 4 *Let X be a discrete random variable with probability function $p_X(x)$, $g_1(X), g_2(X), g_3(X), \cdots, g_k(X)$ be k functions of X. Then*

$$E\left[g_1(X) + g_2(X) + g_3(X) + \cdots + g_k(X)\right] \equiv E[g_1(X)] + E[g_2(X)] + \cdots + E[g_k(X)]$$

Proof for the case of k = 2.
By Theorem 1 we have we have

$$\begin{aligned} E\left[g_1(X) + g_2(X)\right] &\equiv \sum_x \left[g_1(x) + g_2(x)\right] p_X(x) \\ &\equiv \sum_x g_1(x)\, p_X(x) + \sum_x g_2(x)\, p_X(x) \\ &= E\left[g_1(X)\right] + E\left[g_2(X)\right], \end{aligned}$$

■

2.12 Variance of a random variable

2.12.1 Definition of variance

The variance of a random variable X is defined to be the expected value of $(X - \mu)^2$. That is

$$V(X) = E\left[(X - \mu)^2\right]$$

The standard deviation of X is the square root of $V(X)$.

2.12.2 Example 1

Consider a random variable with the following probability distribution.

x	$p_X(x)$
0	1/8
1	1/4
2	3/8
3	1/4

We can compute the expected value as

$$\begin{aligned}\mu = E(X) = & \sum_{x=0}^{3} x\, p_X(x) \\ = & (0)\left(\frac{1}{8}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{3}{8}\right) + (3)\left(\frac{1}{4}\right) \\ = & 1\frac{3}{4}\end{aligned}$$

We compute the variance as

$$\begin{aligned}\sigma^2 &= E[X - \mu)^2] = \Sigma_{x=0}^{3} (x - \mu)^2 p_X(x) \\ &= (0 - 1.75)^2 \left(\frac{1}{8}\right) + (1 - 1.75)^2 \left(\frac{1}{4}\right) \\ &\quad + (2 - 1.75)^2 \left(\frac{3}{8}\right) + (3 - 1.75)^2 \left(\frac{1}{4}\right) \\ &= .9375\end{aligned}$$

and the standard deviation as

$$\begin{aligned}\sigma^2 &= 0.9375 \\ \sigma &= +\sqrt{\sigma^2} = \sqrt{.9375} = 0.97.\end{aligned}$$

2.12.3 Alternative formula for the variance

Theorem 5 *Let X be a discrete random variable with probability function $p_X(x)$; then*

$$V(X) \equiv \sigma^2 = E\left[(X - \mu)^2\right] = E\left(X^2\right) - \mu^2 \tag{6}$$

Proof.

First write out the first part of equation 6 as follows

$$\begin{aligned}V(X) \equiv & \sigma^2 = E\left[(X - \mu)^2\right] = E\left(X^2 - 2\mu X + \mu^2\right) \\ = & E\left(X^2\right) - E(2\mu X) + E\left(\mu^2\right)\end{aligned}$$

where the last step follows from theorem 4. Note that μ is a constant, then apply theorems 3 and 2 to the second and third terms in equation 6 to obtain

$$V(X) \equiv \sigma^2 = E\left[(X - \mu)^2\right] = E\left(X^2\right) - 2\mu E(X) + \mu^2$$

Then making the substitution that $E(X) = \mu$, we obtain

$$V(X) \equiv \sigma^2 = E\left(X^2\right) - \mu^2$$

■

2.12.4 Example 2

Toss a die that has six sides. Observe the number that comes up. The probability mass or frequency function is given by

$$p_X(x) = P(X = x) = \begin{cases} \frac{1}{6} & for \quad x = 1, 2, 3, 4, 5, 6 \\ 0 & otherwise \end{cases} .$$

We compute the expected value as

We compute the variance by then computing the $E(X^2)$ as follows

$$\begin{aligned} E(X^2) &= \sum_{x \,\epsilon\, X} x^2 \, p_X(x) \\ &= \sum_{i=1}^{6} i^2 \left(\frac{1}{6}\right) \\ &= \frac{1 + 4 + 9 + 16 + 25 + 36}{6} \\ &= \frac{91}{6} = 15\,\frac{1}{6} \end{aligned}$$

The expected value as

$$\begin{aligned} E(X) &= \sum_{x \,\epsilon\, X} x \, p_X \, (x) \\ &= \sum_{i\,=\,1}^{6} i \, \left(\frac{1}{6}\right) \\ &= \frac{1 + \ 2 \ + \ 3 \ + \ 4 \ + \ 5 \ + \ 6}{6} \\ &= \frac{21}{6} \ = \ 3\,\frac{1}{2} \end{aligned} \tag{7}$$

We can then compute the variance using the formula $Var(X) = E(X^{\,2}) - E^{\,2}(X)$ and the fact the $E(X) = 21/6$ from equation 7.

$$\begin{aligned} Var(X) &= E\,(X^2) \ - \ E^2(X) \\ &= \frac{91}{6} \ - \ \left(\frac{21}{6}\right)^2 \\ &= \frac{91}{6} \ - \ \left(\frac{441}{36}\right) \\ &= \frac{546}{36} \ - \frac{441}{36} \\ &= \frac{105}{36} \ = \frac{35}{12} \ = \ 2.916\overline{6} \end{aligned}$$

3 The "Distribution" of Random Variables in General

3.1 Cumulative distribution function

The cumulative distribution function (cdf) of a random variable X, denoted by $F_X(\cdot)$, is defined to be the function with domain the real line and range the interval $[0, 1]$, which satisfies $F_X(x) = P_X[X \le x] = P[\{\omega : X(\omega) \le x\}]$ for every real number x. F has the following properties:

$$F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0, \; F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1, \tag{8a}$$

$$F_X(a) \le F_X(b) \; for \; a < b, \;\; nondecreasing \; function \; of \; x, \tag{8b}$$

$$\lim_{0<h \to 0} F_X(x+h) = F_X(x), \; continuous \; from \; the \; right, \tag{8c}$$

3.2 Example of a cumulative distribution function

Consider the following function

$$F_X(x) = \frac{1}{1 + e^{-x}}$$

Check condition 8a as follows.

$$\begin{aligned} \lim_{x \to -\infty} F_X(x) &= \lim_{x \to -\infty} \frac{1}{1 + e^{-x}} = \lim_{x \to \infty} \frac{1}{1 + e^{x}} = 0 \\ \lim_{x \to \infty} F_X(x) &= \lim_{x \to \infty} \frac{1}{1 + e^{-x}} = 1 \end{aligned}$$

To check condition 8b differentiate the cdf as follows

$$\begin{aligned} \frac{d F_X(x)}{dx} &= \frac{d\left(\frac{1}{1 + e^{-x}}\right)}{dx} \\ &= \frac{e^{-x}}{(1 + e^{-x})^2} > 0 \end{aligned}$$

Condition 8c is satisfied because $F_X(x)$ is a continuous function.

3.3 Discrete and continuous random variables

3.3.1 Discrete random variable

A random variable X will be said to be discrete if the range of X is countable, that is if it can assume only a finite or countably infinite number of values. Alternatively, a random variable is discrete if $F_X(x)$ is a step function of x.

3.3.2 Continuous random variable

A random variable X is continuous if $F_X(x)$ is a continuous function of x.

3.4 Frequency (probability mass) function of a discrete random variable

3.4.1 Definition of a frequency (discrete density) function

If X is a discrete random variable with the distinct values, $x_1, x_2, \cdots, x_n, \cdots$, then the function denoted by $p(\cdot)$ and defined by

$$p_X(x) \;=\; \begin{cases} P[X = x_j] & if\ x = x_j, \quad j = 1,\, 2\,, \ldots,\, n, \ldots \\ 0 & if\ x \neq x_j \end{cases} \tag{9}$$

is defined to be the frequency, discrete density, or probability mass function of X. We will often write $f_X(x)$ for $p_X(x)$ to denote frequency as compared to probability.

A discrete probability distribution on R^k is a probability measure P such that

$$\sum_{i=1}^{\infty} P\left(\{x_i\}\right) \;=\; 1$$

for some sequence of points in R^k, i.e. the sequence of points that occur as an outcome of the experiment. Given the definition of the frequency function in equation 9, we can also say that any non-negative function p on R^k that vanishes except on a sequence $x_1, x_2, \cdots, x_n, \cdots$ of vectors and that satisfies

$$\sum_{i=1}^{\infty} p(x_i) \;=\; 1$$

defines a unique probability distribution by the relation

$$P(A) = \sum_{x_i \,\epsilon\, A} p(x_i) \tag{10}$$

3.4.2 Properties of discrete density functions

As defined in section 2.6, a probability mass function must satisfy

$$p_X(x_j) > 0, \text{ for } j = 1, 2, \ldots \tag{11a}$$

$$p_X(x) = 0, \text{ for } x \neq x_j\,;\; j = 1, 2, \ldots, \tag{11b}$$

$$\sum_j p_X(x)_j = 1 \tag{11c}$$

3.4.3 Example 1 of a discrete density function

Consider a probability model where there are two possible outcomes to a single action (say heads and tails) and consider repeating this action several times until one of the outcomes occurs. Let the random variable be the number of actions required to obtain a particular outcome (say heads). Let p be the probability that outcome is a head and $(1-p)$ the probability of a tail. Then to obtain the first head on the x^{th} toss, we need to have a tail on the previous $x-1$ tosses. So the probability of the first head occurring on the x^{th} toss is given by

$$p_X(x) = P(X = x) = \begin{cases} (1 - p)^{x-1}\, p & for\ x = 1, 2, \ldots \\ 0 & otherwise \end{cases}$$

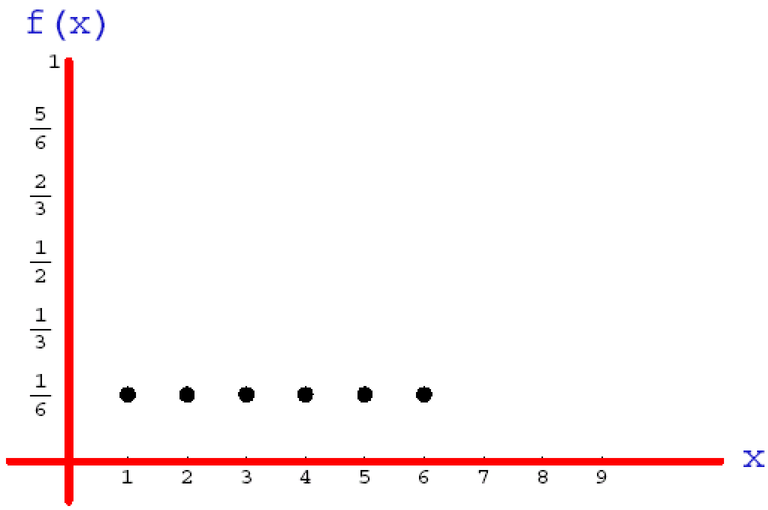
For example the probability that it takes 4 tosses to get a head is 1/16 while the probability it takes 2 tosses is 1/4.

3.4.4 Example 2 of a discrete density function

Consider tossing a die. The sample space is {1, 2, 3, 4, 5, 6}. The elements are {1}, {2}, ... The frequency function is given by

$$p(x) = P(X = x) = \begin{cases} \frac{1}{6} & for \quad x = 1, 2, 3, 4, 5, 6 \\ 0 & otherwise \end{cases}.$$

The density function is represented in Figure 1 below.



3.5 Probability density function of a continuous random variable

3.5.1 Alternative definition of continuous random variable

In Section 3.3.2, we defined a random variable to be continuous if $F_X(x)$ is a continuous function of x. We also say that a random variable X is continuous if there exists a function $f(\cdot)$ such that

$$F_X(x) = \int_{-\infty}^{x} f(u)\, du \tag{12}$$

for every real number x. The integral in equation 12 is a Riemann integral evaluated from $-\infty$ to a real number x.

3.5.2 Definition of a probability density frequency function (pdf)

The probability density function, $f_X(x)$, of a continuous random variable X is the function $f(\cdot)$ that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

3.5.3 Properties of continuous density functions

$$f_X(x) \geq 0 \; \forall x$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1, \tag{13a}$$

Analogous to equation 10, we can write in the continuous case

$$P(X \epsilon A) = \int_A f_X(x)\; dx$$

Theorem 6 *For a density function $f_X(x)$ defined over the set of all real numbers the following holds*

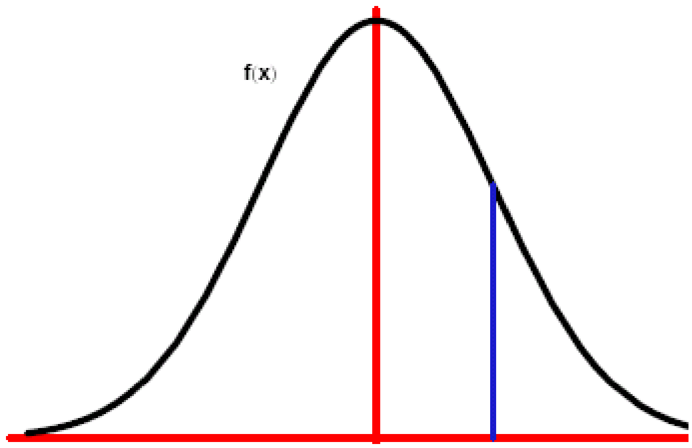
$$P(a \leq X \leq b) = \int_a^b f_X(x)\; dx$$

for any real constants a and b with $a \leq b$.

Also note that for a continuous random variable X the following are equivalent

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

Note that we can obtain the various probabilities by integrating the area under the density function as seen in Figure 2.



3.5.4 Example 1 of a continuous density function

Consider the following function

$$f_X(x) \;=\; \begin{cases} k \cdot e^{-3x} & for\ x > 0 \\ 0 & elsewhere \end{cases}.$$

First we must find the value of k that makes this a valid density function. Given the condition in equation 13a we must have that

$$\int_{-\infty}^{\infty} f_X(x)\,dx \;=\; \int_0^{\infty} k \cdot e^{-3x}\,dx \;=\; 1$$

Integrate the second term to obtain

$$\int_0^{\infty} k \cdot e^{-3x}\,dx \;=\; k \cdot \lim_{t\to\infty} \frac{e^{-3x}}{-3} \,\Big|_0^t \;=\; \frac{k}{3}$$

Given that this must be equal to one we obtain

$$\begin{aligned} \frac{k}{3} &= 1 \\ \Rightarrow k &= 3 \end{aligned}$$

The density is then given by

$$f_X(x) \;=\; \begin{cases} 3 \cdot e^{-3x} & for\ x > 0 \\ 0 & elsewhere \end{cases}.$$

Now find the probability that $1 \leq X \leq 2$.

$$
\begin{aligned}
P(1 \;\leq\; X \;\leq\; 2) \;&=\; \int_{1}^{2} \; 3 \cdot e^{-3x} \, dx \\
&=\; - \, e^{-3x} \, |_{1}^{2} \\
&=\; - \, e^{-6} \; + \; e^{-3} \\
&=\; - \; 0.00247875 \; + \; 0.049787 \\
&=\; 0.047308
\end{aligned}
$$

3.5.5 Example 2 of a continuous density function

Let X have pdf:

$$
f_X(x) \;=\; \begin{cases} x \cdot e^{-x} & for\ 0 \;\leq x \leq \infty \\ 0 & elsewhere \end{cases} .
$$

This density function is shown in Figure 3 below.

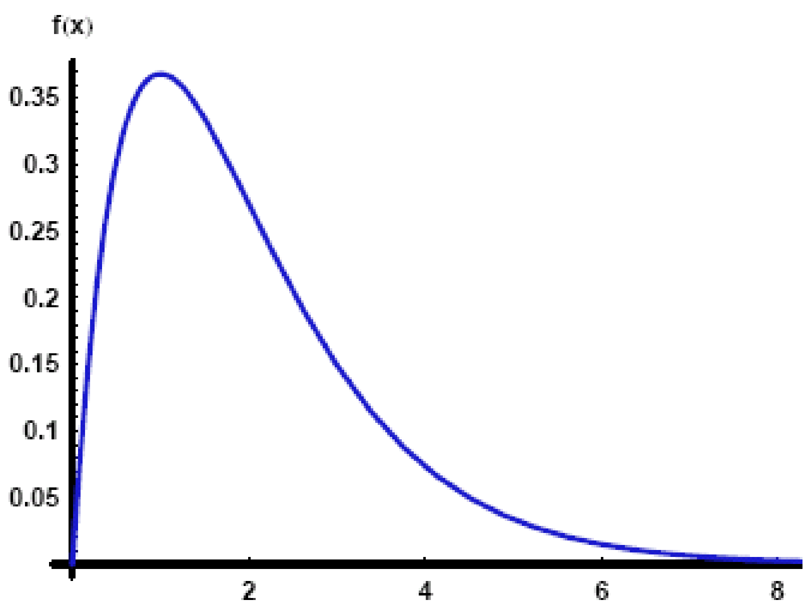


Figure 3

We can find the probability that $1 \leq X \leq 2$ by integration

$$P(1 \ \leq \ X \ \leq \ 2) \ = \ \int_1^2 \ x \cdot e^{-x} \, dx$$

First integrate the expression on the right by parts letting $u = x$ and $dv = e^{-x}\, dx$. Then $du = dx$ and $v = -e^{-x}\, dx$. We then have

$$\begin{aligned} P(1 \ \leq \ X \ \leq \ 2) &= -x\,e^{-x}\,|_1^2 - \int_1^2 \left(-e^{-x}\right)\, dx \\ &= -2\,e^{-2} + e^{-1} - \left[\, e^{-x}\,|_1^2 \,\right] \\ &= -2\,e^{-2} + e^{-1} - e^{-2} + e^{-1} \\ &= -3\,e^{-2} + 2\,e^{-1} \\ &= \frac{-3}{e^2} + \frac{2}{e} \\ &= -0.406 + 0.73575 \\ &= 0.32975 \end{aligned}$$

This is represented by the area between the lines in Figure 4.

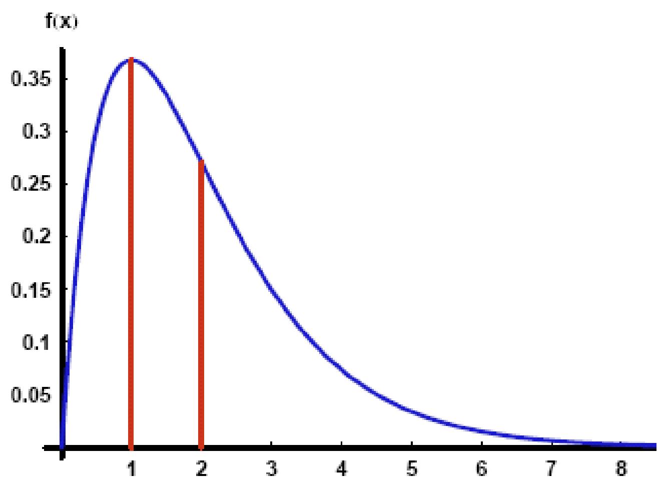


Figure 4

We can also find the distribution function in this case.

$$F_X(x) \;=\; \int_0^x \; t \cdot e^{-t}\, dt$$

Make the $u\ dv$ substitution as before to obtain

$$\begin{aligned}
F_X(x) \;&=\; -\,t\,e^{-t}\,|_0^x \;-\; \int_0^x \left(-\,e^{-t}\right)\; dt \\
&=\; -\,t\,e^{-t}\,|_0^x \;-\; e^{-t}|_0^x \\
&=\; -\,e^{-t}\,(1 \;+\; t)|_0^x \\
&=\; -\,e^{-x}\,(1 \;+\; x) \;+\; e^{-0}\,(\,1 \;+\; 0) \\
&=\; -\,e^{-x}\,(1 \;+\; x) \;+\; 1 \\
&=\; 1 \;-\; e^{-x}\,(1 \;+\; x)
\end{aligned}$$

The distribution function is shown in Figure 5.

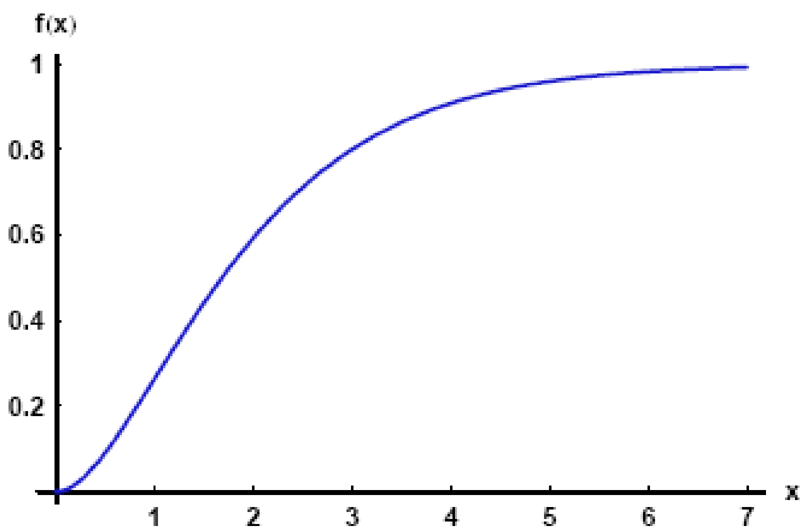


Figure 5

Now consider the probability that $1 \leq X \leq 2$

$$\begin{aligned} P(1 \leq X \leq 2) &= F(2) - F(1) \\ &= \left[1 - e^{-2}(1 + 2)\right] - \left[1 - e^{-1}(1 + 1)\right] \\ &= 2e^{-1} - 3e^{-2} \\ &= 0.73575 - 0.406 \\ &= 0.32975 \end{aligned}$$

We can see this as the difference in the values of $F_X(x)$ at 1 and at 2 in Figure 6

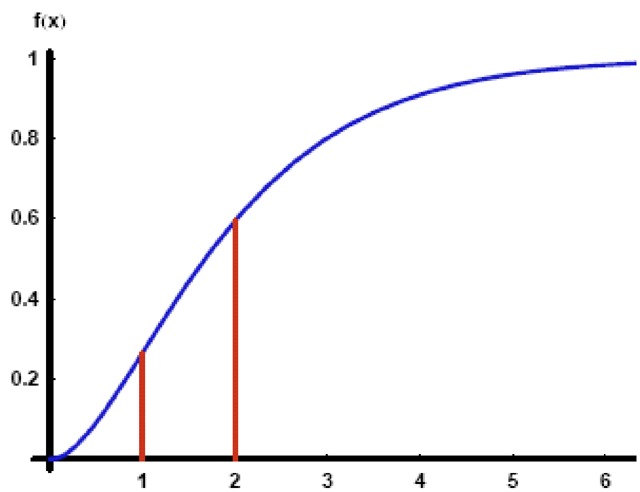


Figure 6

3.5.6 Example 3 of a continuous density function

Consider the normal density function given by

$$f(x : \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2}$$

where μ and σ are parameters of the function. The shape and location of the density function depends on the parameters μ and σ. In Figure 7 the diagram the density is drawn for $\mu = 0$ and $\sigma = 1$ as well as $\mu = 0$ and $\sigma = 2$.

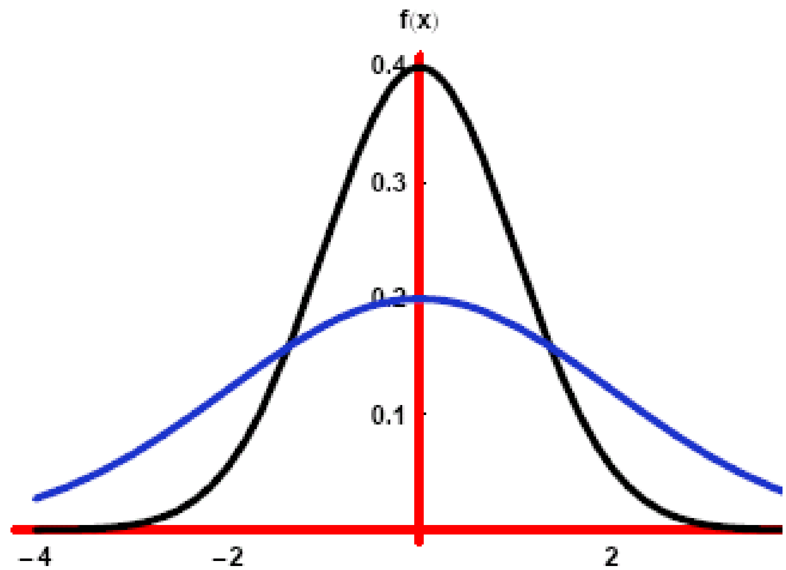


Figure 7

3.5.7 Example 4 of a continuous density function

Consider a random variable with density function given by

$$f_X(x) \;=\; \begin{cases} (p+1)x^p & 0 \;\leq\; x \;\leq\; 1 \\ 0 & otherwise \end{cases}$$

where p is greater than -1. For example, if $p = 0$, then $f_X(x) = 1$, if $p = 1$, then $f_X(x) = 2x$ and so on. The density function with $p = 2$ is shown in Figure 8. The distribution function with $p = 2$ is shown in Figure 9.

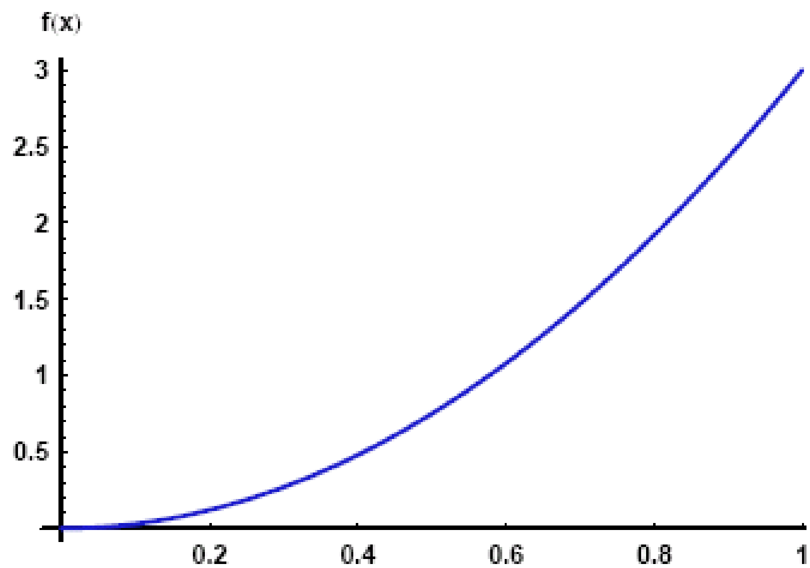


Figure 8, pdf

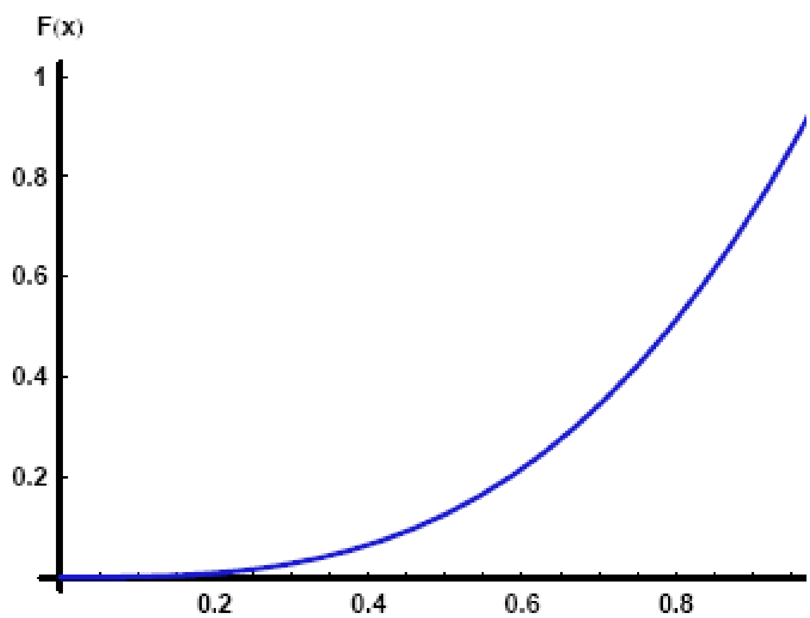


Figure 9, Cdf

3.6 Expected value

3.6.1 Expectation of a single random variable

Let X be a random variable with density $f_X(x)$. The expected value of the random variable, denoted $E(X)$, is defined to be

$$E(X) \;=\; \begin{cases} \int_{-\infty}^{\infty} x\, f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum\limits_{x\,\epsilon\,X} x\, p_X(x) & \text{if } X \text{ is discrete} \end{cases} .$$

provided the sum or integral is defined. The expected value is kind of a weighted average. It is also sometimes referred to as the population mean of the random variable and denoted μ_X.

3.6.2 Expectation of a function of a single random variable

Let X be a random variable with density $f_X(X)$. The expected value of a function $g(\cdot)$ of the random variable, denoted $E(g(X))$, is defined to be

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

if the integral is defined.

3.7 Properties of expectation

3.7.1 Constants

$$\begin{aligned} E[a] \;&\equiv\; \int_{-\infty}^{\infty} a\, f_X(x) dx \\ &=\; a \int_{-\infty}^{\infty} f_X(x) dx \\ &=\; a \end{aligned}$$

3.7.2 Constants multiplied by a random variable

$$
\begin{aligned}
E[a\,X] &\equiv \int_{-\infty}^{\infty} a\,x\,f_X(x)dx \\
&\equiv a\int_{-\infty}^{\infty} x\,f_X(x)dx \\
&\equiv a\,E[X]
\end{aligned}
$$

3.7.3 Constants multiplied by a function of a random variable

$$
\begin{aligned}
E[a\,g(X)] &\equiv \int_{-\infty}^{\infty} a\;g(x)\;f_X(x)dx \\
&= a\int_{-\infty}^{\infty} g(x)\,f_X(x)dx \\
&= a\;E[g(X)]
\end{aligned}
$$

3.7.4 Sums of expected values

Let X be a continuous random variable with density function $f_X(x)$ and let $g_1(X), g_2(X), g_3(X), \cdots, g_k(X)$ be k functions of X. Also let $c_1,\ c_2,\ c_3,\ \cdots c_k$ be k constants. Then

$$E\left[c_1\,g_1(X) + c_2\,g_2(X) + \cdots + c_k\,g_k(X)\;\right] \equiv E\left[c_1\,g_1(X)\right] + E\left[c_2\,g_2(X)\right] + \cdots + E\left[c_k\,g_k(X)\right]$$

3.8 Example 1

Consider the density function

$$f_X(x) = \begin{cases} (p+1)x^p & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

where p is greater than -1. We can compute the $E(X)$ as follows.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\, f_X(x)dx = \int_0^1 x(p+1)x^p dx \\
&= \int_0^1 x^{(p+1)}(p+1)dx = \frac{x^{(p+2)}(p+1)}{(p+2)} \Big|\, {}_0^1 \\
&= \frac{p+1}{p+2}
\end{aligned}
$$

3.9 Example 2

Consider the exponential distribution which has density function

$$
f_X(x) = \frac{1}{\lambda}\, e^{-\frac{x}{\lambda}} \quad 0 \le x \le \infty\,, \quad \lambda > 0
$$

We can compute $E(X)$ as follows.

$$
\begin{aligned}
E(X) &= \int_0^{\infty} x \frac{1}{\lambda}\, e^{\frac{-x}{\lambda}}\, dx \\
&= -x\, e^{\frac{-x}{\lambda}}\, |_0^{\infty} + \int_0^{\infty} e^{\frac{-x}{\lambda}}\, dx \quad \left(u = \frac{x}{\lambda},\ du = \frac{1}{\lambda}\, dx,\ v = -\lambda\, e^{\frac{-x}{\lambda}},\ dv = e^{\frac{-x}{\lambda}}\, dx \right) \\
&= 0 + \left(-\lambda\, e^{\frac{-x}{\lambda}}\, |_0^{\infty} \right) \\
&= \lambda
\end{aligned}
$$

3.10 Variance

3.10.1 Definition of variance

The variance of a single random variable X with mean μ is given by

$$
\begin{aligned}
Var(X) &\equiv \sigma^2 \equiv E\left[(X - E(X))^2 \right] \qquad (14) \\
&\equiv E\left[(X - \mu)^2 \right] \\
&\equiv \int_{-\infty}^{\infty} (x - \mu)^2\, f_X(x)dx
\end{aligned}
$$

We can write this in a different fashion by expanding the last term in equation 14.

$$
\begin{aligned}
Var(X) &\equiv \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f_X(x) dx \\
&= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - 2\mu \int_{-\infty}^{\infty} x\, f_X(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f_X(x)\, dx \\
&= E\left[X^2\right] - 2\mu E\left[X\right] + \mu^2 \\
&= E\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= E\left[X^2\right] - \mu^2 \\
&\equiv \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left[\int_{-\infty}^{\infty} x\, f_X(x) dx\right]^2
\end{aligned}
$$

so

$$Var(X) = E\left[X^2\right] - E\left[X\right]^2$$

The variance is a measure of the dispersion of the random variable about the mean.

3.10.2 Variance example 1

Consider the density function

$$f_X(x) = \begin{cases} (p+1)x^p & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

where p is greater than -1. We can compute $Var(X)$ as follows. From before recall that

$$E(X) = \frac{p+1}{p+2}$$

Now,

$$
\begin{aligned}
E(X^2) &= \int_0^1 x^2\,(p\,+\,1)x^p\;dx = \int_0^1 x^2\,(p\,+\,1)x^{p+2}\;dx \\
&= \frac{x^{(p+3)}(p\,+\,1)}{(p\,+\,3)}\Big|\,_0^1 = \frac{p+1}{p+3} \\
Var\,(\,X\,) &= E\,(X^2\,) - E(\,X\,)^2 \\
&= \frac{p+1}{p+3} - \left(\frac{p+1}{p+2}\right)^2 = (p+1)\,\frac{(p+2)^2 - (p+1)\,(p+3)}{(p+2\,)^2\;(p+3\,)} \\
&= \frac{(p+1)\,(p^2+4p+4-p^2-3-4p)}{(p+2\,)^2\;(p+3\,)} \\
&= \frac{p+1}{(p+2\,)^2\;(p+3\,)}
\end{aligned}
$$

The values of the mean and variances for various values of p are given in the table below:

Table 4: Mean and Variance for Distribution $f_X(x) = (p+1)\,x^p$ for various values of p

p	-.5	0	1	2	∞
E(x)	0.333	0.5	0.66667	0.75	1
Var(x)	0.08888	0.833333	0.277778	0.00047	0

3.10.3 Variance example 2

Consider the exponential distribution which has density function

$$
f_X(x) \;=\; \frac{1}{\lambda}\,e^{\frac{-x}{\lambda}}\quad 0\;\le\;x\;\le\;\infty\,,\quad \lambda\;>\;0
$$

We can compute the $E(X^2)$ as follows

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \, \frac{1}{\lambda} \, e^{\frac{-x}{\lambda}} \, dx \\
&= -x^2 e^{\frac{-x}{\lambda}} \, |_0^\infty + 2 \int_0^\infty x \, e^{\frac{-x}{\lambda}} \, dx \;\; \left(u = \frac{x^2}{\lambda} \, , du = \frac{2\,x}{\lambda} \, dx, v = -\,\lambda \, e \right. \\
&= 0 + 2 \int_0^\infty x \, e^{\frac{-x}{\lambda}} \, dx \\
&= -\,2\,\lambda\,x\,e^{\frac{-x}{\lambda}} \, |_0^\infty + 2 \int_0^\infty \lambda \, e^{\frac{-x}{\lambda}} \, dx \;\; \left(u \; = 2\,x \, , \;\; du = 2\,dx, \;\; v = - \right. \\
&= 0 \; + \; 2\,\lambda \int_0^\infty e^{\frac{-x}{\lambda}} dx = \;\; (2\,\lambda) \left(-\,\lambda\,e^{\frac{-x}{\lambda}} \, |_0^\infty \right) \\
&= (2\,\lambda)\,(\,\lambda\,) = \;\; 2\,\lambda^2
\end{aligned}
$$

We can then compute the variance as

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

4 Moments

4.1 Moments

4.1.1 Moments about the origin (raw moments)

The r^{th} moment about the origin of a random variable X, denoted by μ'_r, is the expected value of X^r; symbolically,

$$\mu'_r = E(X^r) = \sum_x x^r f_X(x)$$

for $r = 0, 1, 2, ...$ when X is discrete and

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f_X(x)\, dx$$

when X is continuous. The r^{th} moment about the origin is only defined if $E[X^r]$ exists. A moment about the origin is sometimes called a raw moment. Note that $\mu'_1 = E(X) = \mu_X$, the mean of the distribution of X, or simply the mean of X. Similarly, $E(X^2)$, which we sometimes use to calculate the variance of a stochastic variable, is the second raw moment of X. The r^{th} moment is sometimes written as a function of θ where θ is a vector of parameters that characterize the distribution of X.

4.1.2 Central moments

The r^{th} moment about the mean of a random variable X, denoted by μ_r, is the expected value of $(X - \mu_X)^r$ symbolically,

$$\mu_r = E[(X - \mu_X)^r] = \sum_x (x - \mu_X)^r f_X(x)$$

for $r = 0, 1, 2, ...$ when X is discrete and

$$\mu_r = E[(X - \mu_X)^r] = \int_{-\infty}^{\infty} (x - \mu_X)^r f_X(x)\, dx$$

when X is continuous. The r^{th} moment about the mean is only defined if $E[(X - \mu_X)^r]$ exists. The r^{th} moment about the mean of a random variable X is sometimes called the r^{th} central moment of X. Note that $\mu_1 = E[(X - \mu_X)] = 0$ and $\mu_2 = E[(X - \mu_X)^2] = Var[X]$. Also note that all odd moments of X around its mean are zero for symmetrical distributions, provided such moments exist.

4.1.3 Alternative formula for the variance

Theorem 7

$$\sigma_X^2 \;=\; \mu_2' \;-\; \mu_X^2$$

Proof.

$$\begin{aligned}
Var(X) &\equiv \sigma_X^2 \equiv E\left[(X - E(X))^2\right] \\
&\equiv E\left[(X - \mu_X)^2\right] \\
&= E\left[X^2 - 2\,\mu_X X + \mu_X^2\right] \\
&= E\left[X^2\right] - 2\,\mu_X E\left[X\right] + \mu_X^2 \\
&= E\left[X^2\right] - 2\,\mu_X^2 + \mu_X^2 \\
&= E\left[X^2\right] - \mu_X^2 \\
&= \mu_2' - \mu_X^2
\end{aligned}$$

■

5 Markov's inequality

Markov's inequality applies equally well to discrete and continuous random variables. We state it here as a theorem.

Theorem 8 *Let X be a random variable and let $g(x)$ be a non-negative function. Then for $r > 0$,*

$$P\left[g(X) \;\geq\; r\right] \;\leq\; \frac{E\,g\,(X)}{r} \tag{15}$$

Proof.

$$\begin{aligned}
E\,g\,(X) &= \int_{-\infty}^{\infty} g\,(x)\; f_X\,(x)\; dx \\
&\geq \int_{[x\,:\,g\,(x)\,\geq\,r]} g\,(x)\; f_X\,(x)\; dx && (g \text{ is nonnegative}) \\
&\geq r \int_{[\,x\,:\,g(x)\,\geq\,r\,]} f_X\,(x)\; dx && (g\,(x) \;\geq\; r) \\
&= r\,P\,[\,g\,(X) \;\geq\; r\,]
\end{aligned}$$

So, we have

$$
\begin{aligned}
E\,g\,(X) \;\; &\geq \;\; r\,P\;[\,g\,(X) \;\geq\; r\,] \Leftrightarrow \\
P\;[g\,(X) \;\geq\; r\,] \;\; &\leq \;\; \frac{E\,g\,(X)}{r}
\end{aligned}
$$

and the proof is complete. ■

5.1 Chebyshev's inequality

Corollary 9 *Let X be a random variable with mean μ and variance σ^2. Then for any $k > 0$ or any $\varepsilon > 0$*

$$
\begin{aligned}
P\;[\,|X \;-\; \mu| \;\geq\; k\,\sigma\,] \;\; &\leq \;\; \frac{1}{k^2} \\
P(|X \;-\; \mu| \;<\; k\,\sigma) \;\; &\geq \;\; 1 \;-\frac{1}{k^2} \qquad (16) \\
P\;[\,|X \;-\; \mu| \;\geq\; \varepsilon\,] \;\; &\leq \;\; \frac{\sigma^2}{\varepsilon^2}
\end{aligned}
$$

Proof. Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$, where $\mu = E(X)$ and $\sigma^2 = Var(X)$. Note that $g(x) \geq 0$. Then let $r = k^2$. Then, by Markov's inequality

$$
\begin{aligned}
P\left[\frac{(\,X \;-\; \mu\,)^2}{\sigma^2} \;\;\geq\;\; k^2\right] \;\; &\leq \;\; \frac{1}{k^2}\,E\left(\frac{(X \;-\; \mu\,)^2}{\sigma^2}\right) \qquad (17) \\
&= \;\; \frac{1}{k^2}\,\frac{E\,(X \;-\; \mu\,)^2}{\sigma^2} \;\; = \;\; \frac{1}{k^2}
\end{aligned}
$$

because $E(X - \mu)^2 = \sigma^2$. We can then rewrite equation 17 as follows

$$
\begin{aligned}
P\left[\frac{(X \;-\; \mu)^2}{\sigma^2} \;\;\geq\;\; k^2\right] \;\; &\leq \;\; \frac{1}{k^2} \\
\Rightarrow\; P\;\left[(\,X \;-\; \mu\,)^2 \;\;\geq\;\; k^2\,\sigma^2\right] \;\; &\leq \;\; \frac{1}{k^2} \\
\Rightarrow\; P\;[|\,X \;-\; \mu\,| \;\;\geq\;\; k\,\sigma] \;\; &\leq \;\; \frac{1}{k^2}
\end{aligned}
$$

■

The result applies for any probability distribution, whether the probability histogram is bell-shaped or not. The results of the theorem are very conservative in the sense that the actual probability that X is in the interval $\mu \pm k\sigma$ usually exceeds the lower bound for the probability, $1 - 1/k^2$, by a considerable amount.

Markov and Chebyshev's theorem enables us to find bounds for probabilities that ordinarily would have to be obtained by tedious mathematical manipulations (integration or summation). We often can obtain estimates of the means and variances of random variables without specifying the distribution of the variable. In situations like these, Chebyshev's inequality provides meaningful bounds for probabilities of interest.

5.2 Example

The number of accidents that occur during a given month at a particular intersection, X, tabulated by a group of Boy Scouts over a long time period is found to have a mean of 12 and a standard deviation of 2. The underlying distribution is not known. What is the probability that, next month, X will be greater than eight but less than sixteen. We thus want $P[8 < X < 16]$. We can write equation 16 in the following useful manner.

$$P\,[(\mu \;-\; k\,\sigma\,) \;<\; X \;<\; (\,\mu \;+\; k\,\mu)] \;\geq\; 1 \;-\; \frac{1}{k^2}$$

For this problem $\mu = 12$ and $\sigma = 2$ so $\mu - k\sigma = 12 - 2k$. We can solve this equation for the k that gives us the desired bounds on the probability.

$$\begin{aligned} \mu \;-\; k\,\mu \quad &= \quad 12 - (k)\,(2) \;=\; 8 \\ &\Rightarrow \quad 2\,k \;=\; 4 \Rightarrow \; k \;=\; 2 \end{aligned}$$

and

$$\begin{gathered} 12 \;+\; (\,k\,)\,(\,2\,) \;=\; 16 \\ \Rightarrow\; 2\,k \;-\; 4 \Rightarrow\; k \;=\; 2 \end{gathered}$$

We then obtain

$$P\,[(8) \;<\; X \;<\; (16)\,] \;\geq\; 1 \;-\; \frac{1}{2^2} \;=\; 1 \;-\; \frac{1}{4} \;=\; \frac{3}{4}$$

Therefore the probability that X is between 8 and 16 is at least 3/4.