RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

1. Discrete random variables

1.1. **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.

1.2. **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). The expression p(x) is a function that assigns probabilities to each possible value x; thus it is often called the probability function for X.

1.3. **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) = 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 f(x) p(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.

1.4. **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, f(x), satisfy the conditions:

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

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

1.5. **Examples of probability mass functions.**

1.5.1. *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 table 1.
From the table we can determine the probabilities as

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

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.

$$\frac{1}{16} \; = \; \binom{4}{0}\frac{4}{16} \; = \; \binom{4}{1}\frac{6}{16} \; = \; \binom{4}{2}\frac{4}{16} \; = \; \binom{4}{3}\frac{1}{16} \; = \; \binom{4}{4}$$

We can then write the probability mass function as

Date: August 20, 2004.

TABLE 1. **Probability of a Function of the Number of Heads from Tossing a Coin Four Times.**

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

$$f(x) = \frac{\binom{4}{x}}{16} \; for \; x = 0, 1, 2, 3, 4 \tag{2}$$

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

1.5.2. *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 table 2 the outcomes are listed along with the value of the random variable associated with each outcome.

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

$$P(X=1) = \left(\frac{1}{36}\right), \; P(X=2) = \left(\frac{3}{36}\right) \; P(X=3) = \left(\frac{5}{36}\right)$$
$$P(X=4) = \left(\frac{7}{36}\right), \; P(X=5) = \left(\frac{9}{36}\right), \; P(X=6) = \left(\frac{11}{36}\right)$$

We can then write the probability mass function as

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

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

1.6. **Cumulative Distribution Functions.**

TABLE 2. Possible Outcomes of Rolling a Red Die and a Green Die – First Number in Pair is Number on Red Die

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*

1.6.1. *Definition of a Cumulative Distribution Function.* If X is a discrete random variable, the function given by

$$F(x) \;=\; P(x \;\leq\; X) \;=\; \sum_{t \,\leq\, x} f(t) \; for \;-\; \infty \;\leq\; x \;\leq\; \infty \tag{3}$$

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

1.6.2. *Properties of a Cumulative Distribution Function.* The values F(X) of the distribution function of a discrete random variable X satisfy the conditions

1: F(-∞) = 0 and F(∞) =1;
2: If a $<$ b, then F(a) $\leq$ F(b) for any real numbers a and b

1.6.3. *First example of a cumulative distribution function.* Consider tossing a coin four times. The possible outcomes are contained in table 1 and the values of f in equation 1. From this we can determine the cumulative distribution function as follows.

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

We can write this in an alternative fashion as

$$F(x) \;=\; \begin{cases} 0 & for\; x \;<\; 0 \\ \frac{1}{16} & for\; 0 \;\leq\; x \;<\; 1 \\ \frac{5}{16} & for\; 1 \;\leq\; x \;<\; 2 \\ \frac{11}{16} & for\; 2 \;\leq\; x \;<\; 3 \\ \frac{15}{16} & for\; 3 \;\leq\; x \;<\; 4 \\ 1 & for\; x \;\geq\; 4 \end{cases}$$

1.6.4. *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 $\binom{M}{x} \times \binom{N-M}{n-x}$ 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

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

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

$$\begin{aligned} F(0) &= f(0) \\ F(1) &= f(0) \;+\; f(1) \\ F(2) &= f(0) \;+\; f(1) \;+\; f(2) \\ F(3) &= f(0) \;+\; f(1) \;+\; f(2) \;+\; f(3) \\ &\vdots \\ F(n) &= f(0) \;+\; f(1) \;+\; f(2) \;+\; f(3) \;+\; \cdots \;+\; f(n) \end{aligned} \tag{5}$$

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 3. **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.

$$f(2) \;=\; P(X \,=\, 2) \;=\; \frac{\binom{3}{2}\,\binom{1}{0}}{\binom{4}{2}} = \frac{(3)(1)}{6} = \frac{1}{2} \tag{6}$$

Similarly

$$f(1) \;=\; P(X \,=\, 1) \;=\; \frac{\binom{3}{1}\,\binom{1}{1}}{\binom{4}{2}} = \frac{(3)(1)}{6} = \frac{1}{2} \tag{7}$$

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

$$\begin{aligned} F(0) \; &= f(0) \;=\; 0 \\ F(1) \; &= f(0) \;+\; f(1) \;=\frac{1}{2} \\ F(2) \; &= f(0) \;+\; f(1) \;+\; f(2) \;=\; 1 \end{aligned} \tag{8}$$

1.7. **Expected value.**

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

$$E(X) \;=\; \sum_{x} \; x\; p(x) \tag{9}$$

if it exists. The expected value exists if

$$\sum_{x} \; |\,x\,|\; p(x) \;<\; \infty \tag{10}$$

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.

1.7.2. *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

$$p\,(x) \;=\; P(X \;=\; x) \;=\; \begin{cases} \frac{1}{6} & for\; x \;=\; 1,\, 2,\, 3,\, 4,\, 5,\, 6 \\ 0 & otherwise \end{cases} \tag{11}$$

We compute the expected value as

$$\begin{aligned} E(X) \; &= \sum_{x\,\epsilon\,X} \; x\; p_X\,(x) \\ &= \sum_{i\,=\,1}^{6} i\; \left(\frac{1}{6}\right) \\ &= \frac{1 +\; 2 \;+\; 3 \;+\; 4 \;+\; 5 \;+\; 6}{6} \\ &=\frac{21}{6} \;=\; 3\;\frac{1}{2} \end{aligned} \tag{12}$$

1.7.3. *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.

$$f(x) \;=\; P(\,X \;=\; x) \;=\; \frac{\binom{2}{x}\,\binom{10}{3-x}}{\binom{12}{3}} \;\; for\; x \;=\; 0\,,\; 1\,,\; 2 \tag{13}$$

For example

$$f(x) \;=\; P(X \;=\; x) \;=\; \frac{\binom{2}{0}\,\binom{10}{3-0}}{\binom{12}{3}} \;=\; \frac{(1)\,(120)}{220} \;=\; \frac{6}{11} \tag{14}$$

We collect this information as in table 4.

TABLE 4. **Probabilities for Television Problem**

x	0	1	2
f(x)	6/11	9/22	1/22
F(x)	6/11	21/22	1

We compute the expected value as

$$\begin{aligned} E(X) \;&=\; \sum_{x\,\epsilon\,X} x\; p_X\,(x) \\ &= (0)\,\left(\frac{6}{11}\right) \,+\, (1)\,\left(\frac{9}{22}\right) \,+\, (2)\,\left(\frac{1}{22}\right) \,=\, \frac{11}{22} \,=\, \frac{1}{2} \end{aligned} \tag{15}$$

1.8. **Expected value of a function of a random variable.**

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

$$E[g(X)] \;=\; \sum_x \; g(x)\; p(x)\,. \tag{16}$$

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 $g_1, g_2, g_2, \ldots g_m$ and probability distribution

$$P[g(X) \;=\; g_i] \;=\; \sum_{\substack{\forall\, x_j such\; that \\ g(x_j) = g_i}} p(x_j) \;=\; p^*\,(g_i) \tag{17}$$

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

$$E[g(X)] \;=\; \sum_{i=1}^{m} \; g_i \; p^*(g_i). \tag{18}$$

Now substitute in to obtain

$$\begin{aligned} E[g(X)] \;&=\; \sum_{i=1}^{m} \; g_i \; p^*(g_i)\,. \\ &=\; \sum_{i=1}^{m} \; g_i \left[\sum_{\substack{\forall\, x_j \,\ni \\ g\,(\,x_j\,) = g_i}} p\,(\,x_j\,) \right] \\ &=\; \sum_{i=1}^{m} \left[\sum_{\substack{\forall\, x_j \,\ni \\ g\,(\,x_j\,) = g_i}} g_i \;\, p\,(\,x_j\,) \right] \\ &=\; \sum_{j=1}^{n} \;\; g\,(x_j\,) \;\; p(\,x_j\,). \end{aligned} \tag{19}$$

□

1.9. Properties of mathematical expectation.

1.9.1. *Constants.*

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

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

$$E[c] \;\equiv\; \sum_{x} \; c\, p(x) \;=\; c \; \sum_{x} \; p(x) \tag{20}$$

But by property 1.4b, we have

$$\sum_{x} \; p(x) \;=\; 1$$

and hence

$$E\,(c) = c \cdot (1) = c. \tag{21}$$

□

1.9.2. *Constants multiplied by functions of random variables.*

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

$$E\,[\,c\;g\,(\,X\,)\,]\;\equiv\;c\;E\;[\,(g\,(\,X\,)\,] \tag{22}$$

Proof. By theorem 1 we have

$$\begin{aligned} E[c\;g(X)] \;&\equiv \sum_x \;c\;g(x)\,p(x) \\ &= c\sum_x \;g(x)\;p(x) \\ &= c\;E[g(X)] \end{aligned} \tag{23}$$

□

1.9.3. *Sums of functions of random variables.*

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

$$E\,[g_1(X) + g_2(X) + g_3(X) + \cdots + g_k(X)] \;\equiv\; E[g_1(X)] + E[g_2(X)] + \cdots + E[g_k(X)] \tag{24}$$

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

$$\begin{aligned} E\,[g_1(X)\;+\;g_2\,(X)\;] \;&\equiv \sum_x \;[g_1\,(x)\;+\;g_2\,(x)\,]\;\;p(x) \\ &\equiv \sum_x \;g_1\,(x)\,p(x)\;+\;\sum_x \;g_2\,(x)\;\,p(x) \\ &= E\;[g_1\,(X)\,]\;+\;\;E\;[\,g_2\,(X)]\,, \end{aligned} \tag{25}$$

□

1.10. **Variance of a random variable.**

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

$$V(X) \;=\; E\;\left[(\,X\;-\;\mu\,)^2\right] \tag{26}$$

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

1.10.2. *Example 1.* Consider a random variable with the following probability distribution.

TABLE 5. Probability Distribution for X

x	p(x)
0	1/8
1	1/4
2	3/8
3	1/4

We can compute the expected value as

$$\begin{aligned} \mu \;=\; E(X) \;&=\; \sum_{x=0}^{3} x \, p_X \, (x) \\ &= (0) \left(\frac{1}{8}\right) \;+\; (1) \left(\frac{1}{4}\right) \;+\; (2) \left(\frac{3}{8}\right) \;+\; (3) \left(\frac{1}{4}\right) \;=1 \, \frac{3}{4} \end{aligned} \tag{27}$$

We compute the variance as

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

and the standard deviation as

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

1.10.3. *Alternative formula for the variance.*

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

$$V\,(X) \;\equiv\; \sigma^2 \;=\; E\left[(X \;-\; \mu\,)^2\right] \;=\; E\,\left(X^2\right) \;-\; \mu^2 \tag{28}$$

Proof. First write out the first part of equation 28 as follows

$$\begin{aligned} V\,(X) \; &\equiv \sigma^2 \;=\; E\left[(X \;-\; \mu\,)^2\right] \;=\; E\left(X^2 \;-\; 2\,\mu\,X \;+\; \mu^2\,\right) \\ &= E\left(X^2\right) \;-\; E\,(2\,\mu\,X) \;+\; E\left(\mu^2\right) \end{aligned} \tag{29}$$

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 28 to obtain

$$V\,(X) \;\equiv \sigma^2 \;=\; E\left[\,(\,X \;-\; \mu\,)^2\,\right] \;=\; E\left(X^2\right) \;-\; 2\,\mu\,E\,(X) \;+\; \mu^2 \tag{30}$$

Then making the substitution that E(X) = μ, we obtain

$$V\,(X) \;\equiv \sigma^2 \;=\; E\left(X^2\right) \;-\; \mu^2 \tag{31}$$

□

1.10.4. *Example 2.* Die toss.

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

$$p(x) \;=\; P(X \;=\; x) \;=\; \begin{cases} \frac{1}{6} & for \; x \;=\; 1,\, 2,\, 3,\, 4,\, 5,\, 6 \\ 0 & otherwise \end{cases}. \tag{32}$$

We compute the expected value as

$$
\begin{aligned}
E(X) &= \sum_{x \,\epsilon\, X} x \, p_X(x) \\
&= \sum_{i=1}^{6} i \left(\frac{1}{6}\right) \\
&= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} \\
&= \frac{21}{6} = 3\,\frac{1}{2}
\end{aligned}
\tag{33}
$$

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

$$
\begin{aligned}
E(X^2) &= \sum_{x \,\epsilon\, X} x^2 \, p_X(x) \\
&= \sum_{i=1}^{6} i^2 \left(\frac{1}{6}\right) \\
&= \frac{1 + 4 + 9 + 16 + 2 + 36}{6} \\
&= \frac{91}{6} = 15\,\frac{1}{6}
\end{aligned}
\tag{34}
$$

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 33.

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

2. The "Distribution" of Random Variables in General

2.1. 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:

$$F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0, \; F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1, \tag{36a}$$

$$F_X(a) \le F_X(b) \; for \; a < b, \; nondecreasing \; function \; of \; x, \tag{36b}$$

$$\lim_{0 < h \to 0} F_X(x+h) = F_X(x), \; continuous \; from \; the \; right, \tag{36c}$$

2.2. Example of a cumulative distribution function. Consider the following function

$$F_X(x) = \frac{1}{1 + e^{-x}} \tag{37}$$

Check condition 36a as follows.

$$\begin{aligned} \lim_{x \to -\infty} F_X(x) &= \lim_{x \to -\infty} \frac{1}{1 + e^{-x}} = \lim_{x \to \infty} \frac{1}{1 + e^{x}} = 0 \\ \lim_{x \to \infty} F_X(x) &= \lim_{x \to \infty} \frac{1}{1 + e^{-x}} = 1 \end{aligned} \tag{38}$$

To check condition 36b differentiate the cdf as follows

$$\begin{aligned} \frac{d F_X(x)}{dx} &= \frac{d\left(\frac{1}{1 + e^{-x}}\right)}{dx} \\ &= \frac{e^{-x}}{(1 + e^{-x})^2} > 0 \end{aligned} \tag{39}$$

Condition 36c is satisfied because $F_X(x)$ is a continuous function.

2.3. Discrete and continuous random variables.

2.3.1. *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.

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

2.4. Frequency (probability mass) function of a discrete random variable.

2.4.1. *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

$$p(x) = \begin{cases} P[X = x_j] & x = x_j, \;\; j = 1, 2, \ldots, n, \ldots \\ 0 & x \ne x_j \end{cases} \tag{40}$$

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

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

$$\sum_{i=1}^{\infty} P(\{x_i\}) = 1 \tag{41}$$

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 40, 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

$$P(A) = \sum_{x_i \,\epsilon\, A} p(x_i) \tag{42}$$

2.4.2. *Properties of discrete density functions.* As defined in section 1.4, a probability mass function must satisfy

$$p(x_j) > 0 \; for \; j = 1, 2, ... \tag{43a}$$

$$p(x) = 0 \; for x \neq x_j \,;\; j = 1, 2, ..., \tag{43b}$$

$$\sum_j p(x)_j = 1 \tag{43c}$$

2.4.3. *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 xth toss, we need to have a tail on the previous x-1 tosses. So the probability of the first had occurring on the xth toss is given by

$$p(x) = P(X = x) = \begin{cases} (1 - p)^{x-1} p & for \; x = 1, 2, ... \\ 0 & otherwise \end{cases} \tag{44}$$

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.

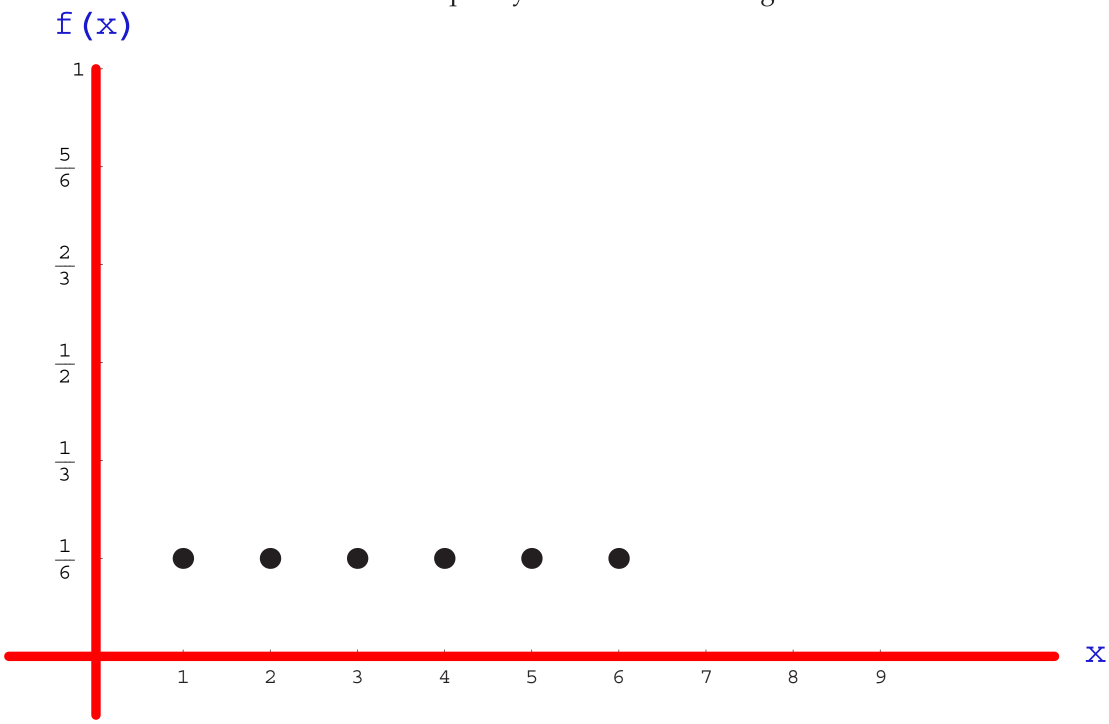
2.4.4. *Example 2 of a discrete density function.* Consider tossing a die. The sample space is {1, 2, 3, 4, 5, 6}. The elements are {1}, {2}, ... The frequency function is given by

$$p(x) = P(X = x) = \begin{cases} \frac{1}{6} & for \; x = 1, 2, 3, 4, 5, 6 \\ 0 & otherwise \end{cases}. \tag{45}$$

The density function is represented in figure 1.

2.5. **Probability density function of a continuous random variable.**

FIGURE 1. Frequency Function for Tossing a Die



2.5.1. *Alternative definition of continuous random variable.* In section 2.3.2, 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(·) such that

$$F_X(x) = \int_{-\infty}^{x} f(u)\, du \tag{46}$$

for every real number x. The integral in equation 46 is a Riemann integral evaluated from -∞ to a real number x.

2.5.2. *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

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \tag{47}$$

2.5.3. *Properties of continuous density functions.*

$$f(x) \geq 0 \; \forall x \tag{48a}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1, \tag{48b}$$

Analogous to equation 42, we can write in the continuous case

$$P(X \,\epsilon\, A) = \int_A f_X(x)\; dx \tag{49}$$

where the integral is interpreted in the sense of Lebesgue.

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

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx \tag{50}$$

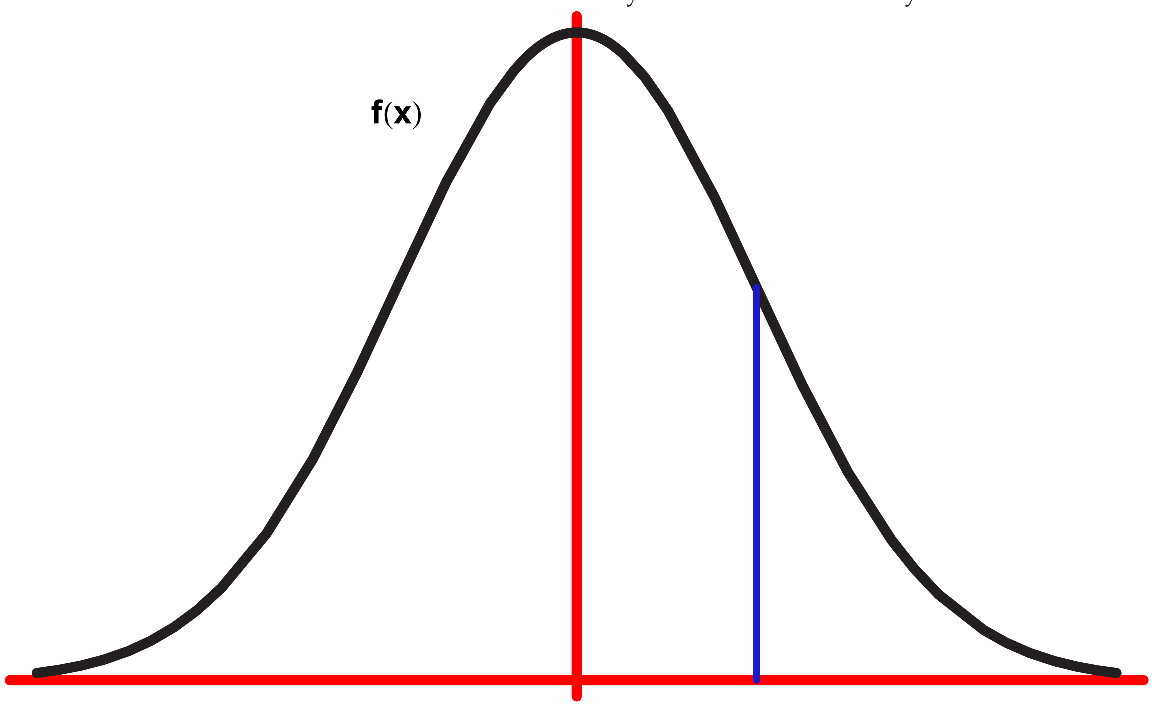
for any real constants a and b with $a \leq b$.

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

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) \tag{51}$$

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

FIGURE 2. Area under the Density Function as Probability



2.5.4. *Example 1 of a continuous density function.* Consider the following function

$$f(x) = \begin{cases} k \cdot e^{-3x} & for\ x > 0 \\ 0 & elsewhere \end{cases}. \tag{52}$$

First we must find the value of k that makes this a valid density function?
Given the condition in equation 48b we must have that

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^{\infty} k \cdot e^{-3x}\, dx = 1 \tag{53}$$

Integrate the second term to obtain

$$\int_0^{\infty} k \cdot e^{-3x}\, dx = k \cdot \lim_{t \to \infty} \frac{e^{-3x}}{-3} \Big|_0^t = \frac{k}{3} \tag{54}$$

Given that this must be equal to one we obtain

$$\begin{aligned} \frac{k}{3} &= 1 \\ \Rightarrow k &= 3 \end{aligned} \tag{55}$$

The density is then given by

$$f(x) = \begin{cases} 3 \cdot e^{-3x} & for\ x > 0 \\ 0 & elsewhere \end{cases}. \tag{56}$$

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

$$\begin{aligned} P(1 \leq X \leq 2) &= \int_1^2 3 \cdot e^{-3x}\, dx \\ &= -e^{-3x} \Big|_1^2 \\ &= -e^{-6} + e^{-3} \\ &= -0.00247875 + 0.049787 \\ &= 0.047308 \end{aligned} \tag{57}$$

2.5.5. *Example 2 of a continuous density function.* Let X have p.d.f.

$$f(x) = \begin{cases} x \cdot e^{-x} & for\ x \leq x \leq \infty \\ 0 & elsewhere \end{cases}. \tag{58}$$

This density function is shown in figure 3.

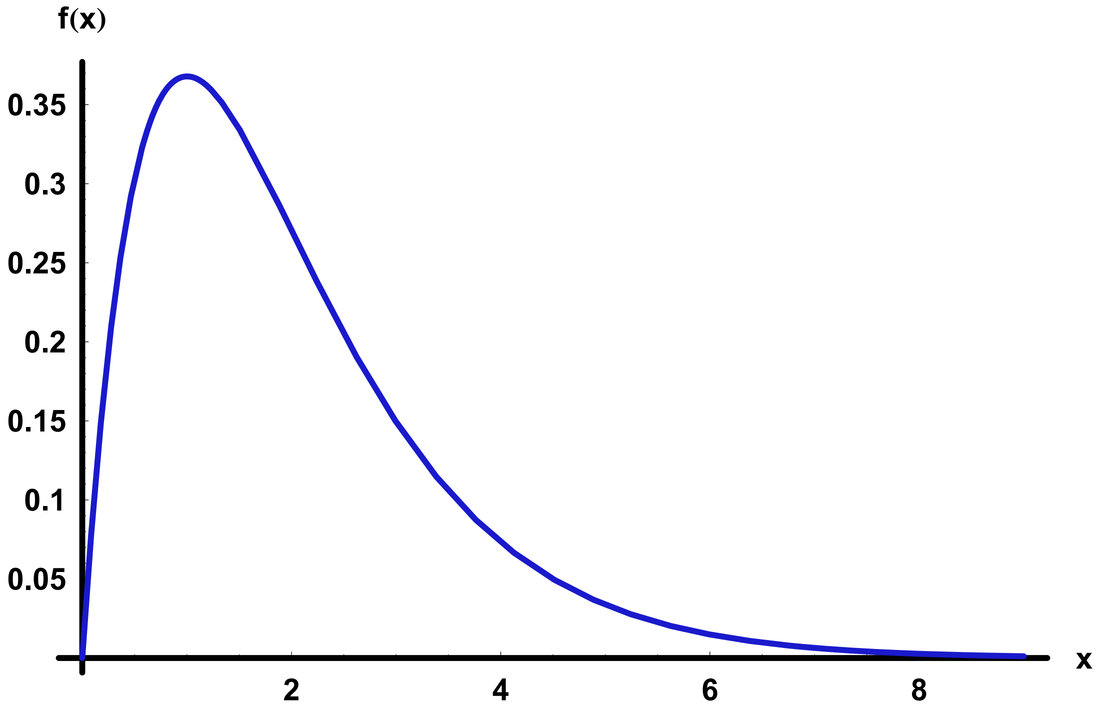
We can find the probability that ($1 \leq X \leq 2$) by integration

$$P(1 \leq X \leq 2) = \int_1^2 x \cdot e^{-x}\, dx \tag{59}$$

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

$$\begin{aligned} P(1 \leq X \leq 2) &= -x\,e^{-x} \Big|_1^2 - \int_1^2 -e^{-x}\, dx \\ &= -2\,e^{-2} + e^{-1} - \left[e^{-x} \Big|_1^2 \right] \\ &= -2\,e^{-2} + e^{-1} - e^{-2} + e^{-1} \\ &= -3\,e^{-2} + 2\,e^{-1} \\ &= \frac{-3}{e^2} + \frac{2}{e} \\ &= -0.406 + 0.73575 \\ &= 0.32975 \end{aligned} \tag{60}$$

FIGURE 3. Graph of Density Function $x\,e^{-x}$



This is represented by the area between the lines in figure 4.
We can also find the distribution function in this case.

$$F(x) \;=\; \int_0^x \; t \cdot e^{-t}\, dt \tag{61}$$

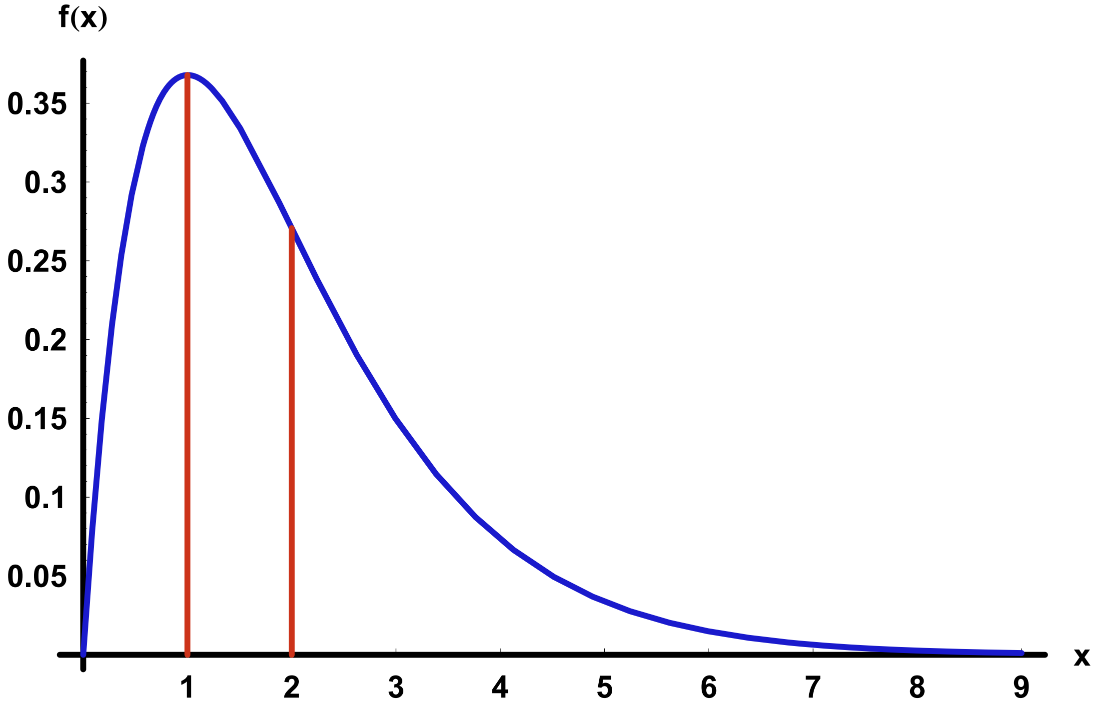
Make the u dv substitution as before to obtain

$$\begin{aligned} F(x) \;&=\; -\,t\,e^{-t}\,|_0^x \;-\; \int_0^x \; -\,e^{-t}\, dt \\ &=\; -\,t\,e^{-t}\,|_0^x \;-\; e^{-t}|_0^x \\ &= e^{-t}\,(-1 \;-\; t)|_0^x \\ &= e^{-x}\,(-1 \;-\; x) \;-\; e^{-0}\,(-1 \;-\; 0) \\ &= e^{-x}\,(-1 \;-\; x) \;+\; 1 \\ &= 1 \;-\; e^{-x}\,(1 \;+\; x) \end{aligned} \tag{62}$$

The distribution function is shown in figure 5.

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

FIGURE 4. $P(1 \leq X \leq 2)$



$$\begin{aligned} P(1 \leq X \leq 2) &= F(2) - F(1) \\ &= 1 - e^{-2}(1 + 2) - 1 + e^{-1}(1 + 1) \\ &= 2\,e^{-1} - 3\,e^{-2} \\ &= 0.73575 - 0.406 \\ &= 0.32975 \end{aligned} \tag{63}$$

We can see this as the difference in the values of F(x) at 1 and at 2 in figure 6

2.5.6. *Example 3 of a continuous density function.* Consider the normal density function given by

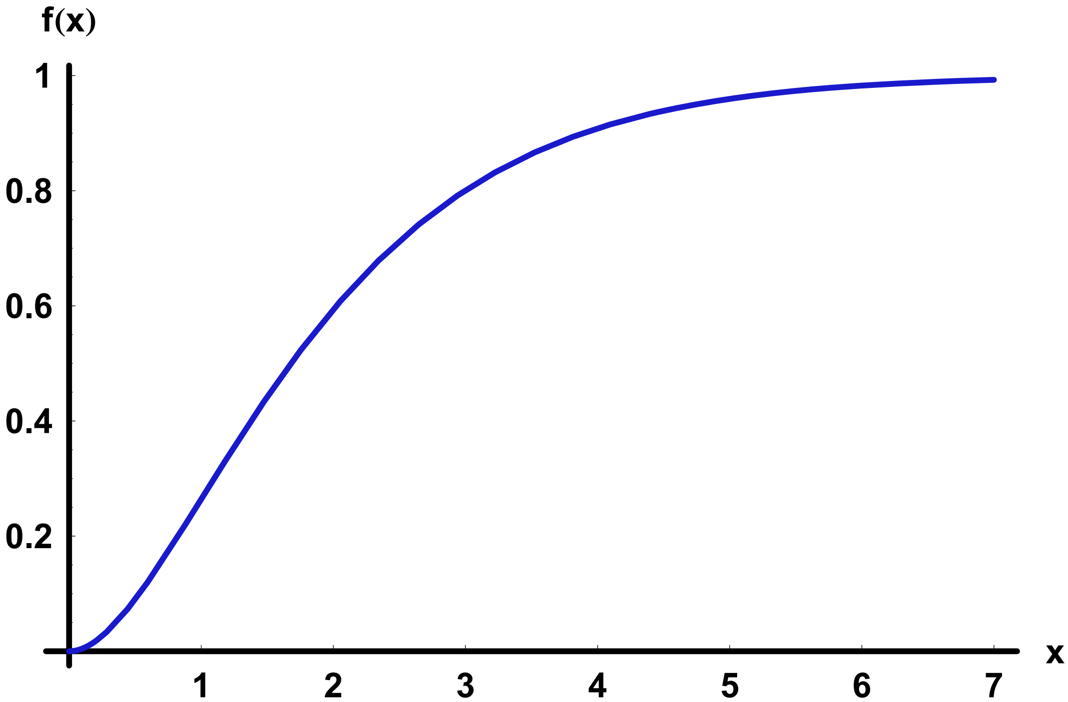
$$f(x : \mu, \sigma) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \cdot e^{\frac{-1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} \tag{64}$$

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 μ = 0, and σ = 1 and σ = 2.

2.5.7. *Example 4 of a continuous density function.* Consider a random variable with density function given by

$$f(x) = \begin{cases} (p+1)x^p & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases} \tag{65}$$

FIGURE 5. Graph of Distribution Function of Density Function $x\,e^{-x}$



where p is greater than -1. For example, if p = 0, then f(x) = 1, if p = 1, then f(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.

2.6. **Expected value.**

2.6.1. *Expectation of a single random variable.* Let X be a random variable with density f(x). The expected value of the random variable, denoted E(X), is defined to be

$$E(X) \;=\; \begin{cases} \int_{-\infty}^{\infty} x\,f(x)\,dx & if\ X\ is\ continuous \\ \sum_{x\,\epsilon\,X} x\,p_X\,(x) & if\ X\ is\ discrete \end{cases} . \tag{66}$$

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.

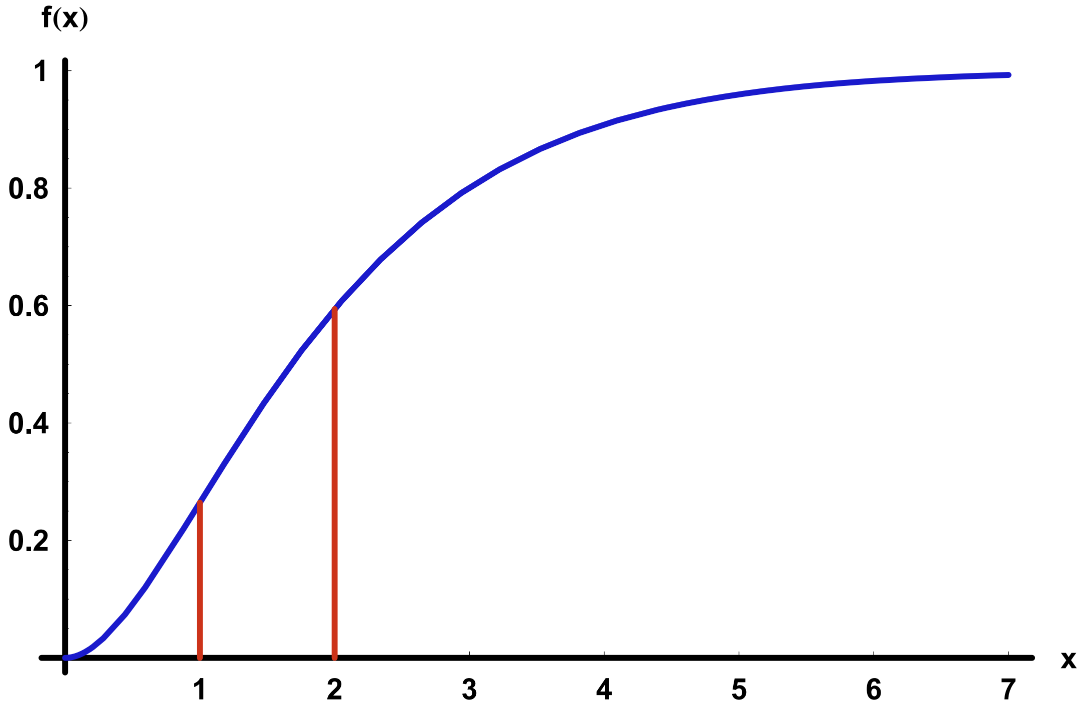
2.6.2. *Expectation of a function of a single random variable.* Let X be a random variable with density f(X). The expected value of a function g(·) of the random variable, denoted E(g(X)), is defined to be

$$E(g(X)) \;=\; \int_{-\infty}^{\infty} g(x)\,f\,(x)dx \tag{67}$$

if the integral is defined.

The expectation of a random variable can also be defined using the Riemann-Stieltjes integral where F is a monotonically increasing function of X. Specifically

FIGURE 6. $P(1 \leq X \leq 2)$ using the Distribution Function



$$E(X) = \int_{-\infty}^{\infty} x \, dF(x) = \int_{-\infty}^{\infty} x \, dF \tag{68}$$

2.7. **Properties of expectation.**

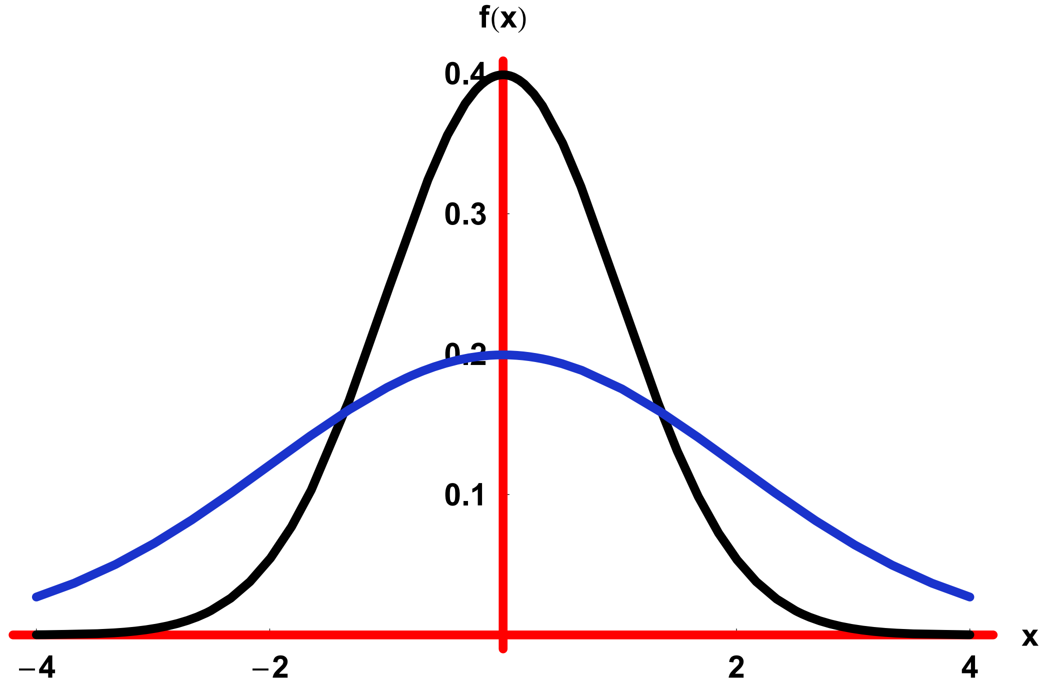
2.7.1. *Constants.*

$$\begin{aligned} E[a] &\equiv \int_{-\infty}^{\infty} a\, f(x) dx \\ &\equiv a \int_{-\infty}^{\infty} f(x) dx \\ &\equiv a \end{aligned} \tag{69}$$

2.7.2. *Constants multiplied by a random variable.*

$$\begin{aligned} E[a\, X] &\equiv \int_{-\infty}^{\infty} a\, x\, f(x) dx \\ &\equiv a \int_{-\infty}^{\infty} x\, f(x) dx \\ &\equiv a\, E[X] \end{aligned} \tag{70}$$

FIGURE 7. Normal Density Function



2.7.3. *Constants multiplied by a function of a random variable.*

$$\begin{aligned} E[a\, g(X)] &\equiv \int_{-\infty}^{\infty} a\ g(x)\ f(x) dx \\ &\equiv a \int_{-\infty}^{\infty} g(x)\, f(x) dx \\ &\equiv a\ E[g(X)] \end{aligned} \tag{71}$$

2.7.4. *Sums of expected values.* Let X be a continuous random variable with density function f(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

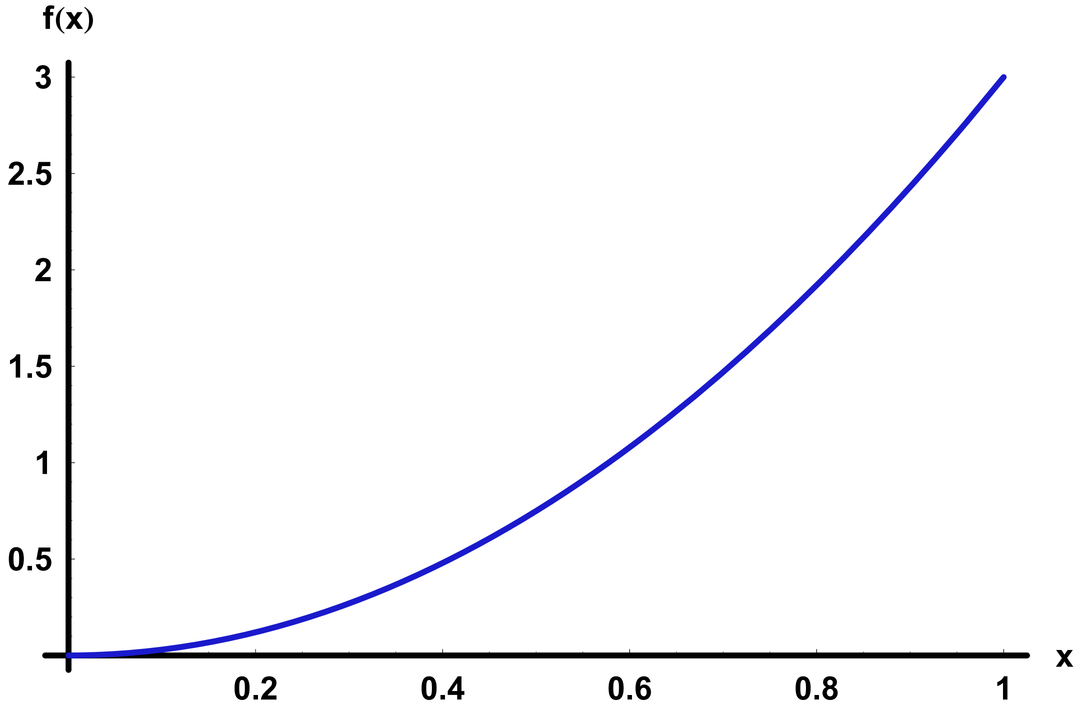
$$E\left[c_1\, g_1(X) + c_2\, g_2(X) + \cdots + c_k\, g_k(X)\ \right] \equiv\ E\left[c_1\, g_1(X)\right] + E\left[c_2\, g_2(X)\right]\ +\ \cdots\ +\ E\left[c_k\, g_k(X)\right] \tag{72}$$

2.8. **Example 1.** Consider the density function

$$f(x)\ = \begin{cases} (p+1)x^p & 0\ \le\ x\ \le\ 1 \\ 0 & otherwise \end{cases} \tag{73}$$

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

FIGURE 8. Density Function $(p+1)\,x^p$



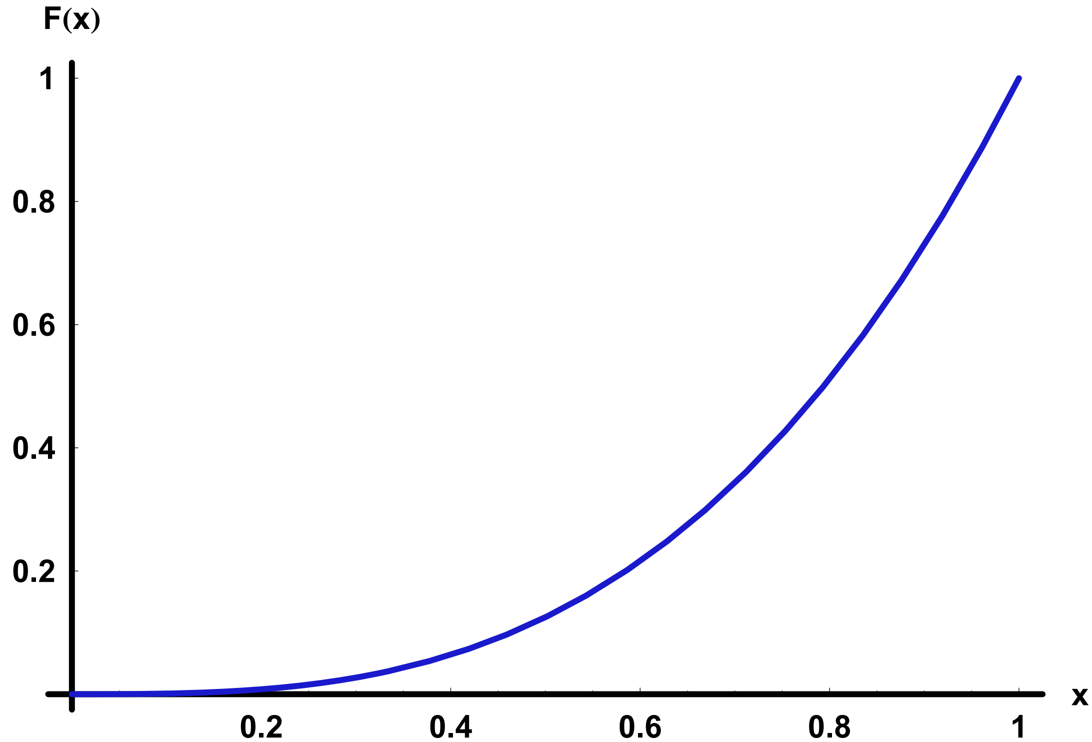
$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\,f(x)dx \\
&= \int_0^1 x(p+1)x^p dx \\
&= \int_0^1 x^{(p+1)}(p+1)dx \\
&= \frac{x^{(p+2)}(p+1)}{(p+2)} \mid {}_0^1 \\
&= \frac{p+1}{p+2}
\end{aligned} \tag{74}
$$

2.9. **Example 2.** Consider the exponential distribution which has density function

$$
f(x) \;=\; \frac{1}{\lambda}\, e^{\frac{-x}{\lambda}} \; 0 \;\leq\; x \;\leq\; \infty \,,\; \lambda \;>\; 0 \tag{75}
$$

We can compute the E(X) as follows.

FIGURE 9. Density Function $(p = 1)\, x^p$



$$
\begin{aligned}
E(X) &= \int_0^\infty x \tfrac{1}{\lambda}\, e^{\frac{-x}{\lambda}}\, dx \\
&= -\, x\ e^{\frac{-x}{\lambda}}\ |_0^\infty \ + \ \int_0^\infty\ e^{\frac{-x}{\lambda}}\, dx\ \left(u\ =\ \tfrac{x}{\lambda},\ du\ =\ \tfrac{1}{\lambda}\, dx,\ v = -\,\lambda\, e^{\frac{-x}{\lambda}},\ dv\ =\ e^{\frac{-x}{\lambda}}\, dx \right) \\
&= 0\ +\ \int_0^\infty\ e^{\frac{-x}{\lambda}}\, dx \\
&= -\,\lambda\, e^{\frac{-x}{\lambda}}\ |_0^\infty \\
&= \lambda
\end{aligned}
\tag{76}
$$

2.10. **Variance.**

2.10.1. *Definition of variance.* The variance of a single random variable X with mean μ is given by

$$
\begin{aligned}
Var(X)\ \equiv\ \sigma^2\ \equiv\ E\ &\left[(X\ -\ E\,(X))^2 \right] \\
\equiv E\ &\left[(\,X\ -\ \mu)^2 \right] \\
\equiv &\int_{-\infty}^{\infty} (x\ -\ \mu)^2\, f(x) dx
\end{aligned}
\tag{77}
$$

We can write this in a different fashion by expanding the last term in equation 77.

$$
\begin{aligned}
Var(X) &\equiv \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\
&\equiv \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx \\
&\equiv \int_{-\infty}^{\infty} x^2 f(x)\, dx - 2\mu \int_{-\infty}^{\infty} x\, f(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f(x)\, dx \\
&= E\left[X^2\right] - 2\mu E\left[X\right] + \mu^2 \\
&= E\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= E\left[X^2\right] - \mu^2 \\
&\equiv \int_{-\infty}^{\infty} x^2 f(x) dx - \left[\int_{-\infty}^{\infty} x\, f(x) dx\right]^2
\end{aligned} \tag{78}
$$

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

2.10.2. *Variance example 1.* Consider the density function

$$
f(x) = \begin{cases} (p+1)x^p & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases} \tag{79}
$$

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

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\, f(x) dx \\
&= \int_0^1 x(p+1)x^p dx \\
&= \frac{x^{(p+2)}(p+1)}{(p+2)} \Big|_0^1 \\
&= \frac{p+1}{p+2} \\
E(X^2) &= \int_0^1 x^2\, (p+1)x^p\, dx \\
&= \frac{x^{(p+3)}(p+1)}{(p+3)} \Big|_0^1 \\
&= \frac{p+1}{p+3} \\
Var(X) &= E(X^2) - E^2(X) \\
&= \frac{p+1}{p+3} - \left(\frac{p+1}{p+2}\right)^2 \\
&= \frac{p+1}{(p+2)^2\,(p+3)}
\end{aligned} \tag{80}
$$

The values of the mean and variances for various values of p are given in table 6.

TABLE 6. Mean and Variance for Distribution $f(x) = (p+1)\,x^p$ for alternative 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

2.10.3. *Variance example 2.* Consider the exponential distribution which has density function

$$f(x) = \frac{1}{\lambda}\, e^{\frac{-x}{\lambda}} \; 0 \le x \le \infty, \; \lambda > 0 \tag{81}$$

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

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\, \frac{1}{\lambda}\, e^{\frac{-x}{\lambda}}\, dx \\ &= -x^2 e^{\frac{-x}{\lambda}}\, |_0^\infty + 2 \int_0^\infty x\, e^{\frac{-x}{\lambda}}\, dx \; \left(u = \frac{x^2}{\lambda}\,, du = \frac{2\,x}{\lambda}\, dx, v = -\,\lambda\, e^{\frac{-x}{\lambda}}, dv = e^{\frac{-x}{\lambda}}\, dx \right) \\ &= 0 + 2 \int_0^\infty x\, e^{\frac{-x}{\lambda}}\, dx \\ &= -\,2\,\lambda\, x\, e^{\frac{-x}{\lambda}}\, |_0^\infty + 2 \int_0^\infty \lambda\, e^{\frac{-x}{\lambda}}\, dx \; \left(u = 2\,x\,,\; du = 2\,dx,\; v = -\,\lambda\, e^{\frac{-x}{\lambda}}, dv = e^{\frac{-x}{\lambda}}\, dx \right) \\ &= 0 \;+\; 2\,\lambda \int_0^\infty e^{\frac{-x}{\lambda}}\, dx \\ &= (2\,\lambda)\left(-\,\lambda\, e^{\frac{-x}{\lambda}}\, |_0^\infty \right) \\ &= (2\,\lambda)\,(\,\lambda\,) \\ &= 2\,\lambda^2 \end{aligned} \tag{82}$$

We can then compute the variance as

$$\begin{aligned} Var(X) &= E\,(X^2) \;-\; E^2\,(X) \\ &= 2\,\lambda^2 \;-\; \lambda^2 \\ &= \lambda^2 \end{aligned} \tag{83}$$

3. MOMENTS AND MOMENT GENERATING FUNCTIONS

3.1. **Moments.**

3.1.1. *Moments about the origin (raw moments).* The rth moment about the origin of a random variable X, denoted by μ_r', is the expected value of X^r; symbolically,

$$\begin{aligned} \mu_r' &= E(X^r) \\ &= \sum_x x^r\, f(x) \end{aligned} \tag{84}$$

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

$$\begin{aligned}\mu'_r &= E(X^r) \\ &= \int_{-\infty}^{\infty} x^r f(x)\, dx\end{aligned} \tag{85}$$

when X is continuous. The rth 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. The rth moment is sometimes written as function of θ where θ is a vector of parameters that characterize the distribution of X.

3.1.2. *Central moments.* The rth moment about the mean of a random variable X, denoted by μ_r, is the expected value of $(X - \mu_X)^r$ symbolically,

$$\begin{aligned}\mu_r &= E[(X - \mu_X)^r] \\ &= \sum_x (x - \mu_X)^r f(x)\end{aligned} \tag{86}$$

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

$$\begin{aligned}\mu_r &= E[(X - \mu_X)^r] \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^r f(x)\, dx\end{aligned} \tag{87}$$

when X is continuous. The rth moment about the mean is only defined if $E[(X - \mu_X)^r]$ exists. The rth moment about the mean of a random variable X is sometimes called the rth central moment of X. The rth central moment of X about a is defined as $E[(X-a)^r]$. If a = μ_X, we have the rth central moment of X about μ_X. Note that μ_1 = $E[(X - \mu_X)] = 0$ and μ_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.

3.1.3. *Alternative formula for the variance.*

Theorem 7.

$$\sigma_X^2 = \mu'_2 - \mu_X^2 \tag{88}$$

Proof.

$$\begin{aligned}Var(X) \equiv \sigma_X^2 &\equiv E\left[(X - E(X))^2\right] \\ &\equiv E\left[(X - \mu_X)^2\right] \\ &\equiv E\left[X^2 - 2\mu_X X + \mu_X^2\right] \\ &= E\left[X^2\right] - 2\mu_X E\left[X\right] + \mu^2 \\ &= E\left[X^2\right] - 2\mu_X^2 + \mu_X^2 \\ &= E\left[X^2\right] - \mu_X^2 \\ &= \mu'_2 - \mu_X^2\end{aligned} \tag{89}$$

□

3.2. **Moment generating functions.**

3.2.1. *Definition of a moment generating function.* The moment generating function of a random variable X is given by

$$M_X(t) = E\, e^{t\,X} \tag{90}$$

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E\, e^{tX}$ exists. We can write $M_X(t)$ as

$$M_X(\,t\,) = \begin{cases} \int_{-\infty}^{\infty} e^{t\,x}\, f_X\,(x)\; dx & if\ X\ is\ continuous \\ \sum_x e^{t\,x}\, P\,(X = x) & if\ X\ is\ discrete \end{cases} . \tag{91}$$

To understand why we call this a moment generating function consider first the discrete case. We can write e^{tx} in an alternative way using a Maclaurin series expansion. The Maclaurin series of a function f(t) is given by

$$\begin{aligned} f(t) &= \sum_{n=0}^{\infty} \frac{f^{(n)}\,(0)}{n\,!}\, t^n = \sum_{n=0}^{\infty} f^{(n)}\,(0)\, \frac{t^n}{n\,!} \\ &= f\,(\,0\,) + \frac{f^{(1)}\,(0)}{1\,!}\, t + \frac{f^{(\,2\,)}\,(0)}{2\,!}\, t^2 + \frac{f^{(3)}\,(0)}{3\,!}\, t^3 + \cdots + \\ &= f\,(\,0\,) + f^{(1)}(0)\, \frac{t}{1\,!} + f^{(\,2\,)}(0)\, \frac{t^2}{2\,!} + f^{(3)}(0)\, \frac{t^3}{3\,!} + \cdots + \end{aligned} \tag{92}$$

where $f^{(n)}$ is the nth derivative of the function with respect to t and $f^{(n)}(0)$ is the nth derivative of f with respect to t evaluated at t = 0. For the function e^{tx}, the requisite derivatives are

$$\begin{aligned} \frac{d\,e^{t\,x}}{d\,t} &= x\,e^{t\,x}, & \left.\frac{d\,e^{t\,x}}{d\,t}\right]_{t\,=\,0} &= x \\ \frac{d^2\,e^{t\,x}}{d\,t^2} &= x^2\,e^{t\,x}, & \left.\frac{d^2\,e^{t\,x}}{d\,t^2}\right]_{t\,=\,0} &= x^2 \\ \frac{d^3\,e^{t\,x}}{d\,t^3} &= x^3\,e^{t\,x}, & \left.\frac{d^3\,e^{t\,x}}{d\,t^3}\right]_{t\,=\,0} &= x^3 \\ & & \vdots & \\ \frac{d^j\,e^{t\,x}}{d\,t^j} &= x^j\,e^{t\,x}, & \left.\frac{d^j\,e^{t\,x}}{d\,t^j}\right]_{t\,=\,0} &= x^j \end{aligned} \tag{93}$$

We can then write the Maclaurin series as

$$\begin{aligned} e^{t\,x} &= \sum_{n=0}^{\infty} \frac{d^n e^{t\,x}}{d\,t^n}\,(0)\, \frac{t^n}{n\,!} \\ &= \sum_{n=0}^{\infty} x^n\, \frac{t^n}{n\,!} \\ &= 1 + t\,x + \frac{t^2\,x^2}{2\,!} + \frac{t^3\,x^3}{3\,!} + \cdots + \frac{t^r\,x^r}{r\,!} + \cdots \end{aligned} \tag{94}$$

We can then compute $E(e^{tx}) = M_X\,(t)$ as

$$\begin{aligned} E\left[e^{t\,x}\right] &= M_X(t) \;=\; \sum_x e^{t\,x}\; f(x) \\ &= \sum_x \left[1 + t\,x + \frac{t^2\,x^2}{2\,!} + \frac{t^3\,x^3}{3!} \;+\; \cdots + \frac{t^r\,x^r}{r!} + \cdots\right] f(x) \\ &= \sum_x f(x) + t\sum_x x\,f(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \cdots + \frac{t^r}{r!}\sum_x x^r f(x) + \cdots \\ &= 1 + \mu t + \mu_2'\,\frac{t^2}{2!} + \mu_3'\,\frac{t^3}{3!} + \cdots + \mu_r'\frac{t^r}{r!} \;+\; \cdots \end{aligned} \tag{95}$$

In the expansion, the coefficient of $\frac{t^r}{t!}$ is μ_r', the rth moment about the origin of the random variable X.

3.2.2. *Example derivation of a moment generating function.* Find the moment-generating function of the random variable whose probability density is given by

$$f(x) \;=\begin{cases} e^{-x} & for\ x\ >\ 0 \\ 0 & elsewhere \end{cases} \tag{96}$$

and use it to find an expression for μ_r'. By definition

$$\begin{aligned} M_X(t) \;=\; E\left(e^{tX}\right) &= \int_{-\infty}^{\infty} e^{t\,x}\cdot e^{-x}\,dx \\ &= \int_{o}^{\infty} e^{-x\,(1\,-\,t)}dx \\ &= \frac{-\,1}{t\;-\;1}\;e^{-x\,(1\,-\,t)}\,\Big|_0^{\infty} \\ &= 0\;-\;\left[\frac{-1}{1\;-\;t}\right] \\ &= \frac{1}{1\;-\;t}\;for\;t\;<\;1 \end{aligned} \tag{97}$$

As is well known, when $|t| < 1$ the Maclaurin's series for $\frac{1}{1\,-\,t}$ is given by

$$\begin{aligned} M_x(t) \;=\; \frac{1}{1\;-\;t} &= 1\;+\;t\;+\;t^2\;+\;t^3\;+\;\cdots\;+\;t^r + \cdots \\ &= 1\;+\;1!\;\cdot\frac{t}{1!}\;+\;2!\;\cdot\;\frac{t^2}{2!}\;+\;3!\;\cdot\;\frac{t^3}{3}!\;+\;\cdots\;+\;r!\;\cdot\;\frac{t^r}{r!}\;+\;\cdots \end{aligned} \tag{98}$$

or we can derive it directly using equation 92. To derive it directly utilizing the Maclaurin series we need the all derivatives of the function $\frac{1}{1\,-\,t}$ evaluated at 0. The derivatives are as follows

$$\begin{aligned} f(t) \; = \frac{1}{1 \; - \; t} &= (1 \; - \; t)^{-\,1} \\ f^{(1)} &= (1 \; - \; t)^{-\,2} \\ f^{(2)} &= 2 \; (1 \; - \; t)^{-3} \\ f^{(3)} &= 6 \; (1 \; - \; t)^{-4} \\ f^{(4)} &= 24 \; (1 \; - \; t)^{-5} \\ f^{(5)} &= 120 \; (1 \; - \; t)^{-6} \\ &\vdots \\ f^{(n)} &= n\,! \; (1 \; - \; t)^{(n \, + \, 1)} \\ &\vdots \end{aligned} \tag{99}$$

Evaluating them at zero gives

$$\begin{aligned} f(0) \; = \frac{1}{1 \; - \; 0} &= (1 \; - \; 0)^{-\,1} \; = \; 1 \\ f^{(1)} &= (1 \; - \; 0)^{-2} \; = \; 1 \; = \; 1! \\ f^{(2)} &= 2 \; (1 \; - \; 0)^{-3} \; = \; 2 \; = \; 2! \\ f^{(3)} &= 6 \; (1 \; - \; 0)^{-4} \; = \; 6 \; = \; 3! \\ f^{(4)} &= 24 \; (1 \; - \; 0)^{-5} \; = \; 24 \; = \; 4! \\ f^{(5)} &= 120 \; (1 \; - \; 0)^{-6} \; = \; 120 \; = \; 5! \\ &\vdots \\ f^{(n)} &= n! \; (1 \; - \; 0)^{-\,(n \, + \, 1)} \; = \; n! \\ &\vdots \end{aligned} \tag{100}$$

Now substituting in appropriate values for the derivatives of the function $f(t) \; = \frac{1}{1 \, - \, t}$ we obtain

$$\begin{aligned} f(t) \; &= \sum_{n \, = \, 0}^{\infty} \frac{f^{(n)} \, (0)}{n!} \; t^n \\ &= f \, (0) \; + \; \frac{f^{(1)} \, (0)}{1!} \; t \; + \; \frac{f^{(2)} \, (0)}{2!} \; t^2 \; + \; \frac{f^{(3)} \, (0)}{3!} \; t^3 \; + \; \cdots \; + \\ &= 1 \; + \; \frac{1!}{1!} \; t \; + \; \frac{2!}{2!} \; t^2 \; + \; \frac{3!}{3!} \; t^3 \; + \; \cdots \; + \\ &= 1 \; + \; t \; + \; t^2 \; + \; t^3 \; + \; \cdots \; + \end{aligned} \tag{101}$$

A further issue is to determine the radius of convergence for this particular function. Consider an arbitrary series where the nth term is denoted by a_n. The ratio test says that

$$If \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1, \text{ then the series is absolutely convergent} \tag{102a}$$

$$\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = L > 1 \text{ or } \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = \infty, \text{ then the series is divergent} \tag{102b}$$

Now consider the nth term and the (n+1)th term of the Maclaurin series expansion of $\frac{1}{1-t}$.

$$\begin{aligned} a_n &= t^n \\ \lim_{n \to \infty} \left| \frac{t^{n+1}}{t^n} \right| &= \lim_{n \to \infty} |t| = L \end{aligned} \tag{103}$$

The only way for this to be less than one in absolute value is for the absolute value of t to be less than one, i.e., $|t| < 1$. Now writing out the Maclaurin series as in equation 98 and remembering that in the expansion, the coefficient of $\frac{t^r}{r!}$ is μ_r', the rth moment about the origin of the random variable X

$$\begin{aligned} M_x(t) = \frac{1}{1-t} &= 1 + t + t^2 + t^3 + \cdots + t^r + \cdots \\ &= 1 + 1! \cdot \frac{t}{1!} + 2! \cdot \frac{t^2}{2!} + 3! \cdot \frac{t^3}{3}! + \cdots + r! \cdot \frac{t^r}{r!} + \cdots \end{aligned} \tag{104}$$

it is clear that μ_r' = r! for r = 0, 1, 2, ... For this density function E[X] = 1 because the coefficient of $\frac{t^1}{1!}$ is 1. We can verify this by finding E[X] directly by integrating.

$$E(X) = \int_0^\infty x \cdot e^{-x}\, dx \tag{105}$$

To do so we need to integrate by parts with u = x and dv = $e^{-x} dx$. Then du = dx and v = $-e^{-x}\, dx$. We then have

$$\begin{aligned} E(X) &= \int_0^\infty x \cdot e^{-x}\, dx,\ u = x,\ du = dx,\ v = -e^{-x},\ dv = e^{-x} dx \\ &= -x e^{-x} |_0^\infty - \int_0^\infty -e^{-x}\, dx \\ &= [0 - 0] - \left[e^{-x} |_0^\infty \right] \\ &= 0 - [0 - 1] = 1 \end{aligned} \tag{106}$$

3.2.3. *Moment property of the moment generating functions for discrete random variables.*

Theorem 8. *If $M_X(t)$ exists, then for any positive integer k,*

$$\left. \frac{d^k M_X(t))}{dt^k} \right]_{t=0} = M_X^{(k)}(0) = \mu_k' . \tag{107}$$

In other words, if you find the kth derivative of $M_X(t)$ with respect to t and then set t = 0, the result will be μ_k' .

Proof. $\frac{d^k M_X(t)}{dt^k}$, or $M_X^{(k)}(t)$, is the kth derivative of $M_X(t)$ with respect to t. From equation 95 we know that

$$M_X(t) = E\left(e^{tX}\right) = 1 + t\,\mu_1' + \frac{t^2}{2!}\,\mu_2' + \frac{t^3}{3!}\,\mu_3' + \cdots \tag{108}$$

It then follows that

$$M_X^{(1)}(t) = \mu_1' + \frac{2\,t}{2!}\,\mu_2' + \frac{3\,t^2}{3!}\,\mu_3' + \cdots \tag{109a}$$

$$M_X^{(2)}(t) = \mu_2' + \frac{2\,t}{2!}\,\mu_3' + \frac{3\,t^2}{3!}\,\mu_4' + \cdots \tag{109b}$$

where we note that $\frac{n}{n!} = \frac{1}{(n-1)!}$. In general we find that

$$M_X^{(k)}(t) = \mu_k' + \frac{2\,t}{2\,!}\,\mu_{k+1}' + \frac{3\,t^2}{3\,!}\,\mu_{k+2}' + \cdots . \tag{110}$$

Setting t = 0 in each of the above derivatives, we obtain

$$M_X^{(1)}(0) = \mu_1' \tag{111a}$$

$$M_X^{(2)}(0) = \mu_2' \tag{111b}$$

and, in general,

$$M_X^{(k)}(0) = \mu_k' \tag{112}$$

□

These operations involve interchanging derivatives and infinite sums, which can be justified if $M_X(t)$ exists.

3.2.4. *Moment property of the moment generating functions for continuous random variables.*

Theorem 9. *If X has mgf $M_X(t)$, then*

$$E\,X^n = M_X^{(n)}(0), \tag{113}$$

where we define

$$M_X^{(n)}(0) = \frac{d^n}{dt^n}\,M_X(t)\,|_{t=0} \tag{114}$$

The nth moment of the distribution is equal to the nth derivative of $M_X(t)$ evaluated at t = 0.

Proof. We will assume that we can differentiate under the integral sign and differentiate equation 91.

$$\begin{aligned}
\frac{d}{dt}\,M_X(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx}\,f_X(x)\,dx \\
&= \int_{-\infty}^{\infty}\left(\frac{d}{dt}\,e^{tx}\right) f_X(x)\,dx \\
&= \int_{-\infty}^{\infty}\left(x\,e^{tx}\right) f_X(x)\,dx \\
&= E\left(X\,e^{tX}\right)
\end{aligned} \tag{115}$$

Now evaluate equation 115 at t = 0.

$$\frac{d}{dt}\, M_X\,(t)\,|_{t=0} \;=\; E\,\left(X\,e^{\,t\,X}\right)\,|t\;=\;0\;=\;E\;X \tag{116}$$

We can proceed in a similar fashion for other derivatives. We illustrate for n = 2.

$$\begin{aligned}
\frac{d^2}{dt^2}\,M_X\,(t)\; &= \frac{d^2}{dt^2}\,\int_{-\infty}^{\infty}\,e^{t\,x}\,f_X\,(x\,)\,dx \\
&= \int_{-\infty}^{\infty}\,\left(\frac{d^2}{dt^2}\;e^{t\,x}\right)\,f_X\,(x)\,dx \\
&= \int_{-\infty}^{\infty}\,\left(\frac{d}{dt}\;x\,e^{t\,x}\right)\,f_X\,(x)\,dx \\
&= \int_{-\infty}^{\infty}\,\left(x^2\,e^{t\,x}\right)\,f_X\,(x)\,dx \\
&= E\,\left(\,X^2\,e^{\,t\,X}\,\right)
\end{aligned} \tag{117}$$

Now evaluate equation 117 at t = 0.

$$\frac{d^2}{dt^2}\,M_X\,(t)\,|_{t=0}\;=\;E\left(X^2\,e^{\,t\,X}\right)\,|t\;=\;0\;=\;E\;X^2 \tag{118}$$

□

3.3. **Some properties of moment generating functions.** If a and b are constants, then

$$M_{X+a}\,(t)\;=E\,\left(e^{\,(X\,+\,a\,)\,t}\right)\;=\;e^{\,a\,t}\,\cdot\,M_X\,(t) \tag{119a}$$

$$M_{bX}\,(t)\;=E\,\left(e^{\,b\,X\,t}\right)\;=\;M_X(\,b\,t\,) \tag{119b}$$

$$M_{\frac{X\,+\,a}{b}}\,(t)\;=E\,\left(e^{\left(\frac{X+a}{b}\right)\,t}\right)\;=\;e^{\frac{a}{b}\,t}\,\cdot\,M_X\,\left(\frac{t}{b}\right) \tag{119c}$$

3.4. **Examples of moment generating functions.**

3.4.1. *Example 1.* Consider a random variable with two possible values, 0 and 1, and corresponding probabilities f(1) = p, f(0) = 1-p. For this distribution

$$\begin{aligned}
M_X\,(t)\;&=E\,\left(e^{t\,X}\,\right) \\
&=e^{t\,\cdot\,1}\,f\,(1)\,+\,e^{t\,\cdot\,0}\,f\,(0) \\
&=e^t\,p\,+\,e^0\,(1\,-\,p) \\
&=e^0\,(1\,-\,p)\,+\,e^t\,p \\
&=1\,-\,p\,+\,e^t\,p \\
&=1\,+\,p\,\left(e^t\,-\,1\right)
\end{aligned} \tag{120}$$

The derivatives are

$$\begin{aligned} M_X^{(1)}(t) &= p\, e^t \\ M_X^{(2)}(t) &= p\, e^t \\ M_X^{(3)}(t) &= p\, e^t \\ &\vdots \\ M_X^{(k)}(t) &= p\, e^t \\ &\vdots \end{aligned} \tag{121}$$

Thus

$$E\left[X^k\right] \;=\; M_X^{(k)}(0) \;=\; p\, e^0 \;=\; p \tag{122}$$

We can also find this by expanding $M_X(t)$ using the Maclaurin series for the moment generating function for this problem

$$\begin{aligned} M_X(t) &= E\left(e^{t\,X}\right) \\ &= 1 \,+\, p\,\left(e^t \,-\, 1\right) \end{aligned} \tag{123}$$

To obtain this we first need the series expansion of e^t. All derivatives of e^t are equal to e^t. The expansion is then given by

$$\begin{aligned} e^t \;&=\; \sum_{n=0}^{\infty} \frac{d^n\, e^t}{d\,t^n}(0)\,\frac{t^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \\ &= 1 \,+\, t \,+\, \frac{t^2}{2!} \,+\, \frac{t^3}{3!} \,+\, \cdots \,+\, \frac{t^r}{r!} \,+\, \cdots \end{aligned} \tag{124}$$

Substituting equation 124 into equation 123 we obtain

$$\begin{aligned} M_X(t) &= 1 \,+\, p\, e^t \,-\, p \\ &= 1 \,+\, p\,\left[1 \,+\, t \,+\, \frac{t^2}{2!} \,+\, \frac{t^3}{3!} \,+\, \cdots \,+\, \frac{t^r}{r\,!} \,+\, \cdots\right] \,-\, p \\ &= 1 \,+\, p \,+\, p\,t \,+\, p\,\frac{t^2}{2!} \,+\, p\,\frac{t^3}{3!} \,+\, \cdots \,+\, p\,\frac{t^r}{r!} \,+\, \cdots \,-\, p \\ &= 1 \,+\, p\,t \,+\, p\,\frac{t^2}{2!} \,+\, p\,\frac{t^3}{3!} \,+\, \cdots \,+\, p\,\frac{t^r}{r!} \,+\, \cdots \end{aligned} \tag{125}$$

We can then see that all moments are equal to p. This is also clear by direct computation

$$
\begin{aligned}
E\left(X\right) &= (1)\,p\,+\,(0)\,(1\,-\,p)\,=\,p \\
E\left(X^2\right) &= (1^2)\,p\,+\,(0^2)\,(1\,-\,p)\,=\,p \\
E\left(X^3\right) &= (1^3)\,p\,+\,(0^3)\,(1\,-\,p)\,=\,p \\
&\vdots \\
E\left(X^k\right) &= (1^k)\,p\,+\,(0^k)\,(1\,-\,p)\,=\,p \\
&\vdots
\end{aligned}
\tag{126}
$$

3.4.2. *Example 2.* Consider the exponential distribution which has a density function given by

$$
f(x)\;=\;\frac{1}{\lambda}\,e^{\frac{-x}{\lambda}}\;0\;\leq\;x\;\leq\;\infty\,,\;\lambda\;>\;0 \tag{127}
$$

For $\lambda\,t < 1$, we have

$$
\begin{aligned}
M_X(t)\;&=\;\int_0^{\infty}\;e^{t\,x}\;\frac{1}{\lambda}\;e^{\frac{-x}{\lambda}}\;dx \\
&=\frac{1}{\lambda}\;\int_0^{\infty}\;e^{-\left(\frac{1}{\lambda}\,-\,t\right)\,x}\;dx \\
&=\frac{1}{\lambda}\;\int_0^{\infty}\;e^{-\left(\frac{1\,-\,\lambda\,t}{\lambda}\right)\,x}\;dx \\
&=\frac{1}{\lambda}\;\left[\frac{-\,\lambda}{1\,-\,\lambda\,t}\right]\;e^{-\left(\frac{1\,-\,\lambda\,t}{\lambda}\right)\,x}\,\Big|_0^{\infty} \\
&=\;\left[\frac{-\,1}{1\,-\,\lambda\,t}\right]\;e^{-\left(\frac{1\,-\,\lambda\,t}{\lambda}\right)\,x}\,\Big|_0^{\infty} \\
&=0\;-\;\left[\frac{-\,1}{1\,-\,\lambda\,t}\right]\;e^{0} \\
&=\frac{1}{1\,-\,\lambda\,t}
\end{aligned}
\tag{128}
$$

We can then find the moments by differentiation. The first moment is

$$
\begin{aligned}
E(X)\;&=\frac{d}{dt}\;(1\;-\;\lambda\,t\,)^{-1}\,|_{t\,=\,0} \\
&=\lambda\,(1\;-\;\lambda\,t)^{-2}\,|_{t\,=\,0} \\
&=\lambda
\end{aligned}
\tag{129}
$$

The second moment is

$$
\begin{aligned}
E(X^2) &= \frac{d^2}{dt^2}\ (1\ -\ \lambda\, t)^{\,-1}\ |_{t\,=\,0} \\
&= \frac{d}{dt}\ \left(\lambda\ (1\ -\ \lambda\, t)^{-2}\right)\ |_{t\,=\,0} \\
&= 2\,\lambda^2\,(1\ -\ \lambda\, t)^{\,-3}\,|_{t\,=\,0} \\
&= 2\,\lambda^2
\end{aligned} \tag{130}
$$

3.4.3. *Example 3.* Consider the normal distribution which has a density function given by

$$
f(x\,;\ \mu,\,\sigma^2)\ =\ \frac{1}{\sqrt{2\pi\sigma^2}}\ \cdot\ e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^{\,2}} \tag{131}
$$

Let g(x) = X - μ, where X is a normally distributed random variable with mean μ and variance σ^2. Find the moment-generating function for (X - μ). This is the moment generating function for central moments of the normal distribution.

$$
M_X(t)\ =\ E[e^{t\,(X\ -\ \mu)}]\ =\ \frac{1}{\sqrt{2\pi\sigma^2}}\ \int_{-\infty}^{\infty}\ e^{t\ (x\,-\,\mu)}\ e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^{\,2}}dx \tag{132}
$$

To integrate, let u = x - μ. Then du = dx and

$$
\begin{aligned}
M_X(t)\ &= \frac{1}{\sigma\sqrt{2\pi}}\ \int_{-\infty}^{\infty}\ e^{tu}e^{\frac{-\,u^2}{2\sigma^2}}\ du \\
&= \frac{1}{\sigma\sqrt{2\pi}}\ \int_{-\infty}^{\infty}\ e^{\left[t\,u\ -\ \frac{u^2}{2\,\sigma^2}\right]}\ du \\
&= \frac{1}{\sigma\sqrt{2\pi}}\ \int_{-\infty}^{\infty}\ e^{\left[\frac{1}{2\,\sigma^2}\ \left(2\,\sigma^2\,t\,u\ -\ u^2\right)\right]}\ du \\
&= \frac{1}{\sigma\sqrt{2\pi}}\ \int_{-\infty}^{\infty}\ \exp\ \left[\left(\frac{-1}{2\,\sigma^2}\right)\ (u^2\ -\ 2\sigma^2\,t\,u\,)\right]\ du
\end{aligned} \tag{133}
$$

To simplify the integral, complete the square in the exponent of e. That is, write the second term in brackets as

$$
\left(u^2\ -\ 2\sigma^2\,t\,u\right)\ =\ \left(u^2\ -\ 2\sigma^2\,t\,u\ +\ \sigma^4\,t^2\ -\ \sigma^4\,t^2\right) \tag{134}
$$

This then will give

$$
\begin{aligned}
\exp\left[\left(\frac{-1}{2\sigma^2}\right)(u^2-2\sigma^2tu)\right] &= \exp\left[\left(\frac{-1}{2\sigma^2}\right)(u^2-2\sigma^2tu+\sigma^4t^2-\sigma^4t^2)\right] \\
&= \exp\left[\left(\frac{-1}{2\,\sigma^2}\right)\ (u^2\ -\ 2\sigma^2\,t\,u+\sigma^4\,t^2)\right]\cdot\exp\ \left[\left(\frac{-1}{2\sigma^2}\right)(-\,\sigma^4t^2)\right] \\
&= \exp\ \left[\left(\frac{-1}{2\sigma^2}\right)(u^2-2\sigma^2tu+\sigma^4\,t^2)\right]\cdot\exp\left[\frac{\sigma^2t^2}{2}\right]
\end{aligned} \tag{135}
$$

Now substitute equation 135 into equation 133 and simplify. We begin by making the substitution and factoring out the term $\exp\left[\frac{\sigma^2t^2}{2}\right]$.

$$
\begin{aligned}
M_X(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[\left(\frac{-1}{2\,\sigma^2}\right)\,(u^2 - 2\sigma^2\,t\,u)\right]\,du \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[\left(\frac{-1}{2\,\sigma^2}\right)\,(u^2 \;-\; 2\sigma^2\,t\,u + \;\sigma^4\,t^2)\right]\cdot\exp\left[\frac{\sigma^2\,t^2}{2}\right]\,du \\
&= \exp\left[\frac{\sigma^2\,t^2}{2}\right]\;\left[\frac{1}{\sigma\sqrt{2\pi}}\right]\int_{-\infty}^{\infty} \exp\left[\left(\frac{-1}{2\,\sigma^2}\right)(u^2 \;-\; 2\sigma^2\,t\,u + \;\sigma^4\,t^2)\right]\,du
\end{aligned} \tag{136}
$$

Now move $\left[\frac{1}{\sigma\sqrt{2\pi}}\right]$ inside the integral sign, take the square root of $(u^2 \;-\; 2\sigma^2\,t\,u + \;\sigma^4\,t^2)$ and simplify

$$
\begin{aligned}
M_X(t) &= \exp\left[\frac{\sigma^2\,t^2}{2}\right]\int_{-\infty}^{\infty} \frac{\exp\left[\left(\frac{-1}{2\,\sigma^2}\right)\,(u^2 - 2\sigma^2\,t\,u + \sigma^4\,t^2)\right]}{\sigma\sqrt{2\pi}}\,du \\
&= \exp\left[\frac{\sigma^2\,t^2}{2}\right]\int_{-\infty}^{\infty} \frac{\exp\left[\left(\frac{-1}{2\,\sigma^2}\right)\,(u - \sigma^2\,t\,)^2\right]}{\sigma\sqrt{2\pi}}\,du \\
&= e^{\frac{t^2\,\sigma^2}{2}}\int_{-\infty}^{\infty} \frac{e^{\frac{-1}{2}\left[\frac{u-\sigma^2 t}{\sigma}\right]^2}}{\sigma\sqrt{2\pi}}\,du
\end{aligned} \tag{137}
$$

The function inside the integral is a normal density function with mean and variance equal to $\sigma^2 t$ and σ^2, respectively. Hence the integral is equal to 1. Then

$$
M_X(t) \;=\; e^{\frac{t^2\,\sigma^2}{2}}\,. \tag{138}
$$

The moments of u = x - μ can be obtained from M_X(t) by differentiating. For example the first central moment is

$$
\begin{aligned}
E(X \;-\; \mu\,) &= \frac{d}{dt}\,\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\;|_{t\,=\,0} \\
&= t\;\sigma^2\;\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\;|_{t\,=\,0} \\
&= 0
\end{aligned} \tag{139}
$$

The second central moment is

$$
\begin{aligned}
E(X \;-\; \mu\,)^2 &= \frac{d^2}{dt^2}\;\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\;|_{t\,=\,0} \\
&= \frac{d}{dt}\;\left(t\;\sigma^2\;\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\right)\;|_{t\,=\,0} \\
&= \left(t^2\,\sigma^4\;\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\;+\;\sigma^2\;\left(e^{\frac{t^2\,\sigma^2}{2}}\right)\right)\;|_{t\,=\,0} \\
&= \sigma^2
\end{aligned} \tag{140}
$$

The third central moment is

$$\begin{aligned} E(X - \mu)^3 &= \frac{d^3}{dt^3} \left(e^{\frac{t^2 \sigma^2}{2}} \right) |_{t=0} \\ &= \frac{d}{dt} \left(t^2 \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + \sigma^2 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= \left(t^3 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 2\, t\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + t\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= \left(t^3 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, t\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= 0 \end{aligned} \tag{141}$$

The fourth central moment is

$$\begin{aligned} E(X - \mu)^4 &= \frac{d^4}{dt^4} \left(e^{\frac{t^2 \sigma^2}{2}} \right) |_{t=0} \\ &= \frac{d}{dt} \left(t^3 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, t\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= \left(t^4 \sigma^8 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, t^2 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, t^2 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= \left(t^4 \sigma^8 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 6\, t^2 \sigma^6 \left(e^{\frac{t^2 \sigma^2}{2}} \right) + 3\, \sigma^4 \left(e^{\frac{t^2 \sigma^2}{2}} \right) \right) |_{t=0} \\ &= 3\, \sigma^4 \end{aligned} \tag{142}$$

3.4.4. *Example 4.* Now consider the raw moments of the normal distribution. The density function is given by

$$f(x\,;\,\mu,\,\sigma^2\,) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \tag{143}$$

To find the moment-generating function for X we integrate the following function.

$$M_X(t) = E[e^{tX}] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx}\, e^{\frac{-1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} dx \tag{144}$$

First rewrite the integral as follows by putting the exponents over a common denominator.

$$\begin{aligned} M_X(t) = E[e^{tX}\,] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx}\, e^{\frac{-1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2} (x-\mu)^2 + tx} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2} (x-\mu)^2 + \frac{2\sigma^2 t x}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2} \left[(x-\mu)^2 - 2\sigma^2 t x \right]} dx \end{aligned} \tag{145}$$

Now square the term in the exponent and simplify

$$
\begin{aligned}
M_X(t) \;=\; E[e^{tX}] \;&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \; e^{\frac{-1}{2\,\sigma^2}\left[x^2\,-\,2\,\mu\,x\,+\,\mu^2\,-\,2\,\sigma^2\,t\,x\right]}\,dx \\
&=\frac{1}{\sqrt{2\pi\sigma^2}}\;\int_{-\infty}^{\infty} \; e^{\frac{-1}{2\,\sigma^2}\left[\;x^2\,-\,2\,x\left(\,\mu\,+\,\sigma^2\,t\,\right)\,+\,\mu^2\,\right]}\,dx
\end{aligned} \tag{146}
$$

Now consider the exponent of e and complete the square for the portion in brackets as follows.

$$
\begin{aligned}
x^2 - 2x\,\left(\mu + \sigma^2\,t\right) + \mu^2 =& x^2 - 2\,x\,\left(\,\mu + \sigma^2\,t\right) + \mu^2 + \mathbf{2}\mu\sigma^{\mathbf{2}}\mathbf{t} + \sigma^{\mathbf{4}}\mathbf{t}^{\mathbf{2}} - 2\mu\sigma^2 t - \sigma^4 t^2 \\
=& \left(x^2 - (\mu + \sigma^2\,t)\right)^2 - 2\,\mu\,\sigma^2\,t - \sigma^4\,t^2
\end{aligned} \tag{147}
$$

To simplify the integral, complete the square in the exponent of e by multiplying and dividing by

$$
\left[e^{\frac{2\,\mu\,\sigma^2\,t+\sigma^4\,t^2}{2\,\sigma^2}}\right]\,\left[e^{\frac{-2\,\mu\,\sigma^2\,t\,-\,\sigma^4\,t^2}{2\,\sigma^2}}\right] \;=\; 1 \tag{148}
$$

in the following manner

$$
\begin{aligned}
M_X\,(t) \;&=\; \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{-1}{2\,\sigma^2}\left[x^2-2x\left(\mu+\sigma^2 t\right)\,+\,\mu^2\right]}\,dx \\
&= \left[e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}}\right]\,\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\left[x^2-2x\left(\mu+\sigma^2 t\right)+\mu^2\,\right]}\,\left[e^{\frac{-\,2\mu\sigma^2 t-\sigma^4 t^2}{2\,\sigma^2}}\right]\,dx \\
&= \left[\,e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}}\right]\,\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\left[x^2-2x\left(\,\mu+\sigma^2 t\right)+\mu^2+2\mu\sigma^2 t+\sigma^4 t^2\right]}\,dx
\end{aligned} \tag{149}
$$

Now find the square root of

$$
x^2 - 2x\left(\mu + \sigma^2 t\right) + \mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2 \tag{150}
$$

Given we would like to have $(x - \mathit{something})^2$, try squaring $x - (\mu + \sigma^2 t)$ as follows

$$
\begin{aligned}
\left[x - (\mu + \sigma^2 t)\right] \;&= x^2 \;-\; 2\left(x(\mu + \sigma^2 t)\right) \;+\; \left(\mu + \sigma^2 t\right)^2 \\
&= x^2 \;-\; 2x\left(\mu - \sigma^2 t\right) \;+\; \mu^2 \;+\; 2\mu\sigma^2 t \;+\; \sigma^4 t^2
\end{aligned} \tag{151}
$$

So $\left[x - (\mu + \sigma^2 t)\right]$ is the square root of $x^2 \;-\; 2x\left(\mu - \sigma^2 t\right) \;+\; \mu^2 \;+\; 2\mu\sigma^2 t \;+\; \sigma^4 t^2$. Making the substitution in equation 149 we obtain

$$
\begin{aligned}
M_X\,(t) \;&=\; \left[e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}}\right]\,\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\left[x^2-2x\left(\mu+\sigma^2 t\right)+\mu^2+2\mu\sigma^2 t+\sigma^4 t^2\right]}\,dx \\
&= \left[e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}}\right]\,\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}\left(\left[x-(\mu+\sigma^2 t)\right]\right)}
\end{aligned} \tag{152}
$$

The expression to the right of $e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\,\sigma^2}}$ is a normal density function with mean and variance equal to $\mu + \sigma^2 t$ and σ^2, respectively. Hence the integral is equal to 1. Then

$$
\begin{aligned}
M_X(t) &= \left[e^{\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}} \right]. \\
&= e^{\mu t + \frac{t^2 \sigma^2}{2}}
\end{aligned} \tag{153}
$$

The moments of X can be obtained from $M_X(t)$ by differentiating with respect to t. For example the first raw moment is

$$
\begin{aligned}
E(X) &= \frac{d}{dt} \left(e^{\mu\, t \,+\, \frac{t^2 \sigma^2}{2}} \right) |_{t\,=\,0} \\
&= (\mu + t\,\sigma^2) \left(e^{\mu t + \frac{t^2\,\sigma^2}{2}} \right) |_{t\,=\,0} \\
&= \mu
\end{aligned} \tag{154}
$$

The second raw moment is

$$
\begin{aligned}
E(x^2) &= \frac{d^2}{dt^2} \left(e^{\mu\, t \,+\, \frac{t^2 \sigma^2}{2}} \right) |_{t=0} \\
&= \frac{d}{dt} \left((\mu + t\,\sigma^2) \left(e^{\mu\, t + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t=0} \\
&= \left((\mu \;+\; t\;\sigma^2)^2 \left(e^{\mu\, t + \frac{t^2\,\sigma^2}{2}} \right) + \sigma^2 \left(e^{\mu t + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t\,=\,0} \\
&= \mu^2 \;+\; \sigma^2
\end{aligned} \tag{155}
$$

The third raw moment is

$$
\begin{aligned}
E(X^3) &= \frac{d^3}{dt^3} \left(e^{\mu t + \frac{t^2 \sigma^2}{2}} \right) |_{t=0} \\
&= \frac{d}{dt} \left(\left(\mu + t\sigma^2\right)^2 \left(e^{\mu t + \frac{t^2\,\sigma^2}{2}} \right) + \sigma^2 \left(e^{\mu t + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t\,=\,0} \\
&= \left[\left(\mu + t\sigma^2\right)^3 \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) + 2\,\sigma^2 \left(\mu + t\sigma^2\right) \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) + \sigma^2 \left(\mu + t\sigma^2\right) \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) \right] |_{t=0} \\
&= \left(\left(\mu + t\,\sigma^2\right)^3 \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) + 3\,\sigma^2 \left(\mu + t\sigma^2\right) \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) \right) |_{t\,=\,0} \\
&= \mu^3 \;+\; 3\sigma^2 \mu
\end{aligned} \tag{156}
$$

The fourth raw moment is

$$
\begin{aligned}
E(X^4) &= \frac{d^4}{dt^4} \left(e^{\mu + \frac{t^2 \sigma^2}{2}} \right) |_{t=0} \\
&= \frac{d}{dt} \left(\left(\mu + t\,\sigma^2\right)^3 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) + 3\,\sigma^2 \left(\mu + t\,\sigma^2\,\right) \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t=0} \\
&= \left(\left(\mu + t\,\sigma^2\right)^4 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) + 3\sigma^2 \left(\mu + t\,\sigma^2\right)^2 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t=0} \\
&+ \left(3\sigma^2 \left(\mu + t\,\sigma^2\,\right)^2 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) + 3\,\sigma^4 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t=0} \\
&= \left(\left(\mu + t\,\sigma^2\right)^4 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) + 6\,\sigma^2 \left(\mu + t\,\sigma^2\,\right)^2 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) + 3\,\sigma^4 \left(e^{\mu + \frac{t^2\,\sigma^2}{2}} \right) \right) |_{t=0} \\
&= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4
\end{aligned} \tag{157}
$$

4. Chebyshev's inequality

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

4.1. **A Theorem of Chebyshev.**

Theorem 10. *Let X be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,*

$$P(|X - \mu| < k\,\sigma) \geq 1 - \frac{1}{k^2} \qquad or \qquad P(|X - \mu| \geq k\,\sigma) \leq \frac{1}{k^2}. \tag{158}$$

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.

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, Chcbyshev's inequality provides meaningful bounds for probabilities of interest.

Proof. Let f(x) denote the density function of X. Then

$$\begin{aligned} V(X) = \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 \, f(x)\, dx \\ &= \int_{-\infty}^{\mu - k\,\sigma} (x - \mu)^2 \, f(x)\, dx \\ &+ \int_{\mu - k\,\sigma}^{\mu + k\,\sigma} (x - \mu)^2 \, f(x)\, dx \\ &+ \int_{\mu + k\,\sigma}^{\infty} (x - \mu)^2 \, f(x)\, dx. \end{aligned} \tag{159}$$

The second integral is always greater than or equal to zero.

Now consider relationship between $(x - \mu)^2$ and $k\sigma^2$.

$$\begin{aligned} x &\leq \mu - k\,\sigma \\ \Rightarrow -x &\geq k\,\sigma - \mu \\ \Rightarrow \mu - x &\geq k\,\sigma \\ \Rightarrow (\mu - x)^2 &\geq k^2\,\sigma^2 \\ \Rightarrow (x - \mu)^2 &\geq k^2\,\sigma^2 \end{aligned} \tag{160}$$

And similarly,

$$\begin{aligned} x\, &\geq \mu\, +\, k\,\sigma \\ \Rightarrow\, x\, -\, \mu\, &\geq k\,\sigma \\ \Rightarrow\, (\,x\, -\, \mu\,)^2\, &\geq k^2\,\sigma^2 \end{aligned} \tag{161}$$

Now replace $(x - \mu)^2$ with $k\sigma^2$ in the first and third integrals of equation 159 to obtain the inequality

$$V(X)\; =\; \sigma^2\; \geq\; \int_{-\infty}^{\mu\,-\,k\,\sigma}\; k^2\,\sigma^2\, f(x)\,dx\; +\; \int_{\mu\,+\,k\,\sigma}^{\infty}\; k^2\,\sigma^2\; f(x)\,dx\; . \tag{162}$$

Then

$$\sigma^2\; \geq\; k^2\,\sigma^2\; \left[\int_{-\infty}^{\mu\,-\,k\,\sigma}\; f(x)\,dx\; +\; \int_{\mu\,+\,k\,\sigma}^{+\infty}\; f(x)\,dx\right] \tag{163}$$

W can write this in the following useful manner

$$\begin{aligned} \sigma^2\; &\geq k^2\,\sigma^2\{P\,(\,X\; \leq\; \mu\; -\; k\,\sigma)\; +\; P\,(\,X +\; \geq\; \mu\; +\; k\,\sigma)\} \\ &= k^2\,\sigma^2\,P\,(\,|\,X\; -\; \mu\,|\; \geq\; k\,\sigma). \end{aligned} \tag{164}$$

Dividing by $k^2\sigma^2$, we obtain

$$P\;(\,|\,X\; -\; \mu\,|\; \geq\; k\,\sigma\;)\; \leq \frac{1}{k^2}, \tag{165}$$

or, equivalently,

$$P\;(\,|\,X\; -\; \mu\,|\; <\; k\,\sigma\;)\; \geq\; 1\; -\frac{1}{k^2}. \tag{166}$$

□

4.2. **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 158 in the following useful manner.

$$P\;[(\mu\; -\; k\,\sigma\,)\; <\; X\; <\; (\,\mu\; +\; k\,\mu)]\; \geq\; 1\; -\; \frac{1}{k^2} \tag{167}$$

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

$$\begin{aligned} \mu - k\,\mu &= 12 - (k)\,(2) = 8 \\ \Rightarrow\ 2\,k &= 4 \\ \Rightarrow\ k &= 2 \\ & and \\ 12 + (\,k\,)\,(\,2\,) &= 16 \\ \Rightarrow\ 2\,k - 4 & \\ \Rightarrow\ k &= 2 \end{aligned} \tag{168}$$

We then obtain

$$P\,[(8) < X < (16)] \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} \tag{169}$$

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

4.3. **Alternative statement of Chebyshev's inequality.**

Theorem 11. *Let X be a random variable and let g(x) be a non-negative function. Then for* $r > 0$*,*

$$P\,[g(X) \geq r] \leq \frac{E\,g\,(X)}{r} \tag{170}$$

Proof.

$$\begin{aligned} E\,g\,(X) &= \int_{-\infty}^{\infty} g\,(x)\ f_X\,(x)\ dx \\ &\geq \int_{[x\,:\ g\,(x)\ \geq\ r]} g\ (x)\ f_X\,(x)\ dx \qquad (g\ is\ nonnegative) \\ &\geq r \int_{[\,x\,:\ g(x)\ \geq\ r\,]} f_X\,(x)\ dx \qquad (g\,(x) \geq r) \\ &= r\,P\ [\,g\,(X) \geq r\,] \\ \Rightarrow\ P\ [g\,(X) \geq r] &\leq \frac{E\,g\,(X)}{r} \end{aligned} \tag{171}$$

□

4.4. **Another version of Chebyshev's inequality as special case of general version.**

Corollary 1. Let X be a random variable with mean μ and variance σ^2. Then for any $k > 0$ or any $\varepsilon > 0$

$$P\ [\,|X - \mu| \geq k\,\sigma\,] \leq \frac{1}{k^2} \tag{172a}$$

$$P\ [\,|X - \mu| \geq \varepsilon\,] \leq \frac{\sigma^2}{\varepsilon^2} \tag{172b}$$

Proof. Let g(x) = $\frac{(x-\mu)^2}{\sigma^2}$, where μ = E(X) and σ^2 = Var (X). Then let r = k^2. Then

$$\begin{aligned} P\left[\frac{(X-\mu)^2}{\sigma^2} \geq k^2\right] &\leq \frac{1}{k^2} E\left(\frac{(X-\mu)^2}{\sigma^2}\right) \\ &= \frac{1}{k^2}\,\frac{E(X-\mu)^2}{\sigma^2} = \frac{1}{k^2} \end{aligned} \tag{173}$$

because $E(X-\mu)^2 = \sigma^2$. We can then rewrite equation 173 as follows

$$\begin{aligned} P\left[\frac{(X-\mu)^2}{\sigma^2} \geq k^2\right] &\leq \frac{1}{k^2} \\ \Rightarrow P\left[(X-\mu)^2 \geq k^2\sigma^2\right] &\leq \frac{1}{k^2} \\ \Rightarrow P\left[|X-\mu| \geq k\sigma\right] &\leq \frac{1}{k^2} \end{aligned} \tag{174}$$

□

REFERENCES

[1] Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.
[2] Bickel P.J., and K.A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Vol 1).* 2^{nd} Edition. Upper Saddle River, NJ: Prentice Hall, 2001.
[3] Billingsley, P. *Probability and Measure.* 3rd edition. New York: Wiley, 1995.
[4] Casella, G. And R.L. Berger. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2002.
[5] Cramer, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.
[6] Goldberger, A.S. *Econometric Theory.* New York: Wiley, 1964.
[7] Lindgren, B.W. *Statistical Theory* 4^{th} edition. Boca Raton, FL: Chapman & Hall/CRC, 1993.
[8] Rao, C.R. *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley, 1973.
[9] Theil, H. *Principles of Econometrics*. New York: Wiley, 1971.