

Econ 205

Discrete Random Variables and Probability Distributions – 1

Course Outline

- **Probability**

- Probability and Inference
- The Discrete Case
- Calculating the Probability of an Event
- Conditional Probability and the Independence of Events
- Laws of Probability
- The Law of Total Probability and Bayes' Rule
- Introduction to Random Variables and their distributions

- **Discrete Random Variables and Their Probability Distributions**

- Discrete Random Variable
- The Probability Distribution for Discrete Random Variable
- The Expected Value of Random Variable
- Moments and Moment-Generating Functions
- The Bernoulli Probability Distribution
- The Binomial Probability Distribution
- The Geometric Probability Distribution
- The Negative Binomial Probability Distribution
- The Hypergeometric Probability Distribution
- The Poisson Probability Distribution
- Probability-Generating Functions

- **Continuous Random Variables and Their Probability Densities**

- Continuous Random Variable
- The Probability Distribution for Continuous Random Variable
- The Expected Value for Continuous Random Variable
- Moments and Moment-Generating Functions
- The Uniform Probability Distribution
- The Normal Probability Distribution
- Exponential Probability Distribution
- Chi-Square Probability Distribution
- Tchebysheff's Theorem

Course Outline (cont'd)

- **Multivariate Probability Distributions**

- Bivariate and Multivariate Probability Distributions
- Independent Random Variables
- Marginal and Conditional Probability Distributions
- The Expected Value of a Function of Random Variables
- Special Theorems
- The Covariance of Two Random Variables
- The Expected Value and Variance of Linear Functions of Random Variables
- The Multinomial Probability Distribution
- The Bivariate Normal Distribution
- Conditional Expectations

- **Functions of Random Variables**

- Finding the Probability Distribution of a Function of Random Variables
- The Method of Distribution Functions
- The Method of Transformations
- The Method of Moment-Generating Functions
- Multivariable Transformations Using Jacobians

Table of Contents of Devore & Berk

2	Probability	50
	Introduction	50
2.1	Sample Spaces and Events	51
2.2	Axioms, Interpretations, and Properties of Probability	56
2.3	Counting Techniques	66
2.4	Conditional Probability	74
2.5	Independence	84
3	Discrete Random Variables and Probability Distributions	96
	Introduction	96
3.1	Random Variables	97
3.2	Probability Distributions for Discrete Random Variables	101
3.3	Expected Values of Discrete Random Variables	112
3.4	Moments and Moment Generating Functions	121
3.5	The Binomial Probability Distribution	128
3.6	Hypergeometric and Negative Binomial Distributions	138
3.7	The Poisson Probability Distribution	146
4	Continuous Random Variables and Probability Distributions	158
	Introduction	158
4.1	Probability Density Functions and Cumulative Distribution Functions	159
4.2	Expected Values and Moment Generating Functions	171
4.3	The Normal Distribution	179
4.4	The Gamma Distribution and Its Relatives	194
4.5	Other Continuous Distributions	202
4.6	Probability Plots	210
4.7	Transformations of a Random Variable	220
5	Joint Probability Distributions	232
	Introduction	232
5.1	Jointly Distributed Random Variables	233
5.2	Expected Values, Covariance, and Correlation	245
5.3	Conditional Distributions	253
5.4	Transformations of Random Variables	265
5.5	Order Statistics	271

Content of this presentation

- Discrete Random Variables and Probability Distributions
 - Random Variables
 - Discrete and Continuous Random Variables
 - Probability Distributions for Discrete Random Variables
 - The Cumulative Distribution Function for Discrete Random Variables
 - Expected Values of Discrete Random Variables
 - The Variance of a Discrete Random Variable
 - Moments and Moment-Generating Functions
- Examples

Random Variables

Definition: A **random variable** is a real valued function Defined over the sample space.

We use uppercase letters to denote random variables and lowercase letters to denote particular values of random variables.

Definition: Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random** variable.

TWO TYPES OF RANDOM VARIABLES:

DEFINITION

A **discrete** random variable is an rv whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on.

A random variable is **continuous** if *both* of the following apply:

1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).
2. No possible value of the variable has positive probability, that is, $P(X = c) = 0$ for any possible value c .

Definitions from [WMS]:

A random variable is **discrete** if it can assume only a finite or countably infinite number of distinct values.

A random variable is **continuous** if its distribution function is continuous.

Probability Distributions for Discrete Random Variables

Definition:

The **probability distribution** or **probability mass function** (pmf) of a discrete rv is defined for every number x by $p(x) = P(X = x) = P(\text{all } s \in \mathcal{S}: X(s) = x)$.¹

Properties:

- 1) $p(x) \geq 0$ for all x
- 2) $\sum_x p(x) = 1$

Definition from [WMS]:

The **probability distribution** of a random variable is the collection of probabilities for each value of a random variable.

It can be represented by a formula, table or a graph that provides $p(y) = P(Y=y)$ for all y .

DEFINITION

Suppose $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

The Cumulative Distribution Function for Discrete Random Variables

Definition:

The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y)$$

For any number x , $F(x)$ is the probability that the observed value of X will be at most x .

Property:

For X a discrete rv, the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a **step function**.

The Cdf can be derived from pmf. Also, this process can be reversed to obtain the pmf from the cdf whenever cdf is available. For example, if the possible values of X are 0, 1, 2, 3 and 4,

$$\begin{aligned} p(2) &= P(X=2) = [p(0) + p(1) + p(2)] - [p(0) + p(1)] \\ &= P(X \leq 2) - P(X \leq 1) \\ &= F(2) - F(1) \end{aligned}$$

Proposition:

For any two numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

where $F(a-)$ represents the maximum of $F(x)$ values to the left of a . Equivalently, if a is the limit of values of x approaching from the left, then $F(a-)$ is the limiting value of $F(x)$. In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a \text{ or } a + 1 \text{ or } \dots \text{ or } b) \\ &= F(b) - F(a - 1) \end{aligned}$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a - 1)$ in this case.

Expected Values of Discrete Random Variables:

Definition: Expected Value of X

Let X be a discrete rv with set of possible values D and pmf $p(x)$.

The **expected value** or **mean value** of X , denoted by $E(X)$ or μ_X , is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

This expected value will exist provided that $\sum_{x \in D} |x| \cdot p(x) < \infty$

Proposition: Expected Value of a function of X

If the rv X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

assuming that $\sum_D |h(x)| \cdot p(x)$ is finite.

Proposition:

$$E(aX + b) = a \cdot E(X) + b$$

(Or, using alternative notation, $\mu_{aX+b} = a \cdot \mu_X + b$.)

Proof

$$\begin{aligned} E(aX + b) &= \sum_D (ax + b) \cdot p(x) = a \sum_D x \cdot p(x) + b \sum_D p(x) \\ &= aE(X) + b \end{aligned}$$


Variance of a Discrete Random Variable:

Definition:

Let X have pmf $p(x)$ and expected value μ . Then the **variance** of X , denoted by $V(X)$ or σ_X^2 , or just σ^2 , is

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E(X - \mu)^2$$

The **standard deviation** (SD) of X is

$$\sigma_X = \sqrt{\sigma_X^2}$$

Proposition: Shortcut Formula

$$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x) \right] - \mu^2 = E(X^2) - [E(X)]^2$$

Proof of the Shortcut Formula Expand $(x - \mu)^2$ in the definition of σ^2 to obtain $x^2 - 2\mu x + \mu^2$, and then carry Σ through to each of the three terms:

$$\begin{aligned} \sigma^2 &= \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2 \end{aligned}$$

■

The Variance of a function of X:

$$V[h(X)] = \sigma_{h(X)}^2 = \sum_D \{h(x) - E[h(x)]\}^2 \cdot p(x)$$

Proposition:

$$V(aX + b) = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2 \text{ and } \sigma_{aX+b} = |a| \cdot \sigma_X$$

Moments

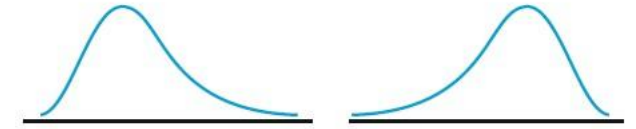
Moments about the mean: Expected values of powers of $X - \mu$

Moments about 0: Expected values of powers of X

- 1) First moment about 0 is the **mean**.
- 2) Second moment about the mean is the **variance**.
- 3) A measure of departure from symmetry called the **skewness** is obtained from the third moment about the mean by dividing it by σ^3 : (Since the third moment about the mean depends on the scale of measurement, we divide by σ^3 .)

$$\frac{E[(X - \mu)^3]}{\sigma^3} = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$$

When the skewness is negative, we say the distribution is negatively skewed or that it is skewed to the left.



Positively skewed

Negatively skewed

Moment Generating Functions - 1

Moment generating functions help in calculation of the moments and understanding of statistical distributions.

Definition:

The moment generating function (mgf) of a discrete random variable X is defined to be

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} p(x)$$

where D is the set of possible X values. We will say that the moment generating function exists if $M_X(t)$ is defined for an interval of numbers that includes zero as well as positive and negative values of t (an interval including 0 in its interior).

Property:

When $t = 0$, for any random variable X

$$M_X(0) = E(e^{0X}) = \sum_{x \in D} e^{0x} p(x) = \sum_{x \in D} 1p(x) = 1$$

Proposition:

If the mgf exists and is the same for two distributions, then the two distributions are the same. That is, the moment generating function uniquely specifies the probability distribution; there is a one-to-one correspondence between distributions and mgf's.

Moment Generating Functions - 2

Theorem If the mgf exists,

$$E(X^r) = M_X^{(r)}(0)$$

Proof We show that the theorem is true for $r = 1$ and $r = 2$. A proof by mathematical induction can be used for general r . Differentiate

$$\frac{d}{dt}M_X(t) = \frac{d}{dt} \sum_{x \in D} e^{xt} p(x) = \sum_{x \in D} \frac{d}{dt} e^{xt} p(x) = \sum_{x \in D} x e^{xt} p(x)$$

where we have interchanged the order of summation and differentiation. This is justified inside the interval of convergence, which includes 0 in its interior. Next we set $t = 0$ and get the first moment

$$M_X'(0) = M_X^{(1)}(0) = \sum_{x \in D} x p(x) = E(X)$$

Differentiate again:

$$\frac{d^2}{dt^2}M_X(t) = \frac{d}{dt} \sum_{x \in D} x e^{xt} p(x) = \sum_{x \in D} x \frac{d}{dt} e^{xt} p(x) = \sum_{x \in D} x^2 e^{xt} p(x)$$

Set $t = 0$ to get the second moment

$$M_X''(0) = M_X^{(2)}(0) = \sum_{x \in D} x^2 p(x) = E(X^2)$$

■

Sometimes, when doing the differentiation, defining $R_X(t) = \ln[M_X(t)]$ makes the computation easier.

We can show that if the mgf exists, then

$$\mu = E(X) = R_X'(0)$$

$$\sigma^2 = V(X) = R_X''(0)$$

Proposition:

Let X have the mgf $M_X(t)$ and let $Y = aX + b$. Then $M_Y(t) = e^{bt} M_X(at)$.

EXAMPLES

Examples 1-2: Random Variables

Example 1: A discrete random variable

Consider the experiment in which batteries are examined until a good one (S) is obtained. The sample space is

$$\mathcal{S} = \{S, FS, FFS, \dots\}$$

Define an rv X by

X = the number of batteries examined before the experiment terminates

Then $X(S) = 1, X(FS) = 2, X(FFS) = 3, \dots, X(FFFFFFFS) = 7$, and so on. Any positive integer is a possible value of X , so the set of possible values is infinite. ■

Example 2: A continuous random variable

Consider the daily rainfall in a specified location. Theoretically, with measuring equipment with perfect accuracy, the random variable Y measuring the amount of rainfall can take on any value between 0 and 5 inches. Each of the uncountably infinite number of points in the interval $(0,5)$ represents a distinct possible value of the amount of rainfall in a day. Y is a continuous random variable.

Example 3: Probability Distributions for Discrete Random Variables : PMF

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

Solution:

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

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

Similarly,

$$p(1) = P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5},$$

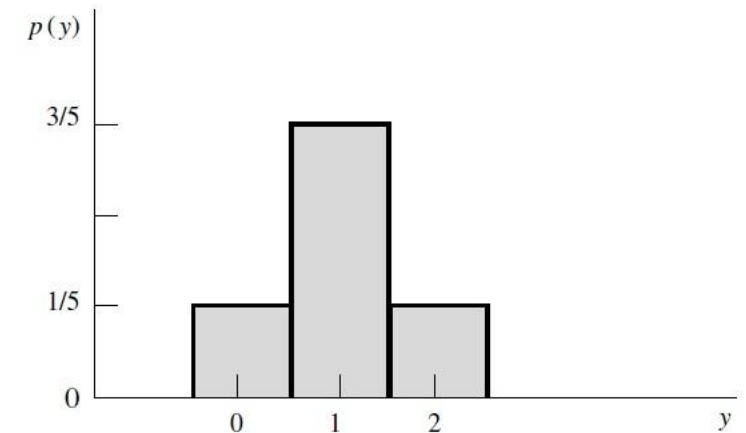
$$p(2) = P(Y = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Notice that $(Y = 1)$ is by far the most likely outcome. This should seem reasonable since the number of women equals the number of men in the original group. ■

Probability distribution

y	$p(y)$
0	$1/5$
1	$3/5$
2	$1/5$

Probability histogram



Examples 4-5: Probability Distributions for Discrete Random Variables

Example 4: PMF (Probability Distribution)

Consider a group of five potential blood donors—A, B, C, D, and E—of whom only A and B have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified. Let the rv Y = the number of typings necessary to identify an O+ individual. Then the pmf of Y is

$$\begin{aligned} p(1) &= P(Y=1) = P(\text{A or B typed first}) = \frac{2}{5} = .4 \\ p(2) &= P(Y=2) = P(\text{C, D, or E first, and then A or B}) \\ &= P(\text{C, D, or E first}) \cdot P(\text{A or B next} | \text{C, D, or E first}) = \frac{3}{5} \cdot \frac{2}{4} = .3 \\ p(3) &= P(Y=3) = P(\text{C, D, or E first and second, and then A or B}) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = .2 \\ p(4) &= P(Y=4) = P(\text{C, D, and E all done first}) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = .1 \\ p(y) &= 0 \text{ for } y \neq 1, 2, 3, 4. \end{aligned}$$

The pmf can be presented compactly in tabular form:

y	1	2	3	4
p(y)	.4	.3	.2	.1

where any y value not listed receives zero probability.

Example 5: Parameter of a probability distribution

Starting at a fixed time, we observe the gender of each newborn child at a certain hospital until a boy (B) is born. Let $p = P(B)$, assume that successive births are independent, and define the rv X by X = number of births observed. Then

$$\begin{aligned} p(1) &= P(X=1) = P(B) = p \\ p(2) &= P(X=2) = P(GB) = P(G) \cdot P(B) = (1-p)p \end{aligned}$$

and

$$p(3) = P(X=3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1-p)^2p$$

Continuing in this way, a general formula emerges:

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

The quantity p in Expression (3.2) represents a number between 0 and 1 and is a parameter of the probability distribution. In the gender example, $p = .51$ might be appropriate, but if we were looking for the first child with Rh-positive blood, then we might have $p = .85$. ■

Example 6: The Cumulative Distribution Function

A store carries flash drives with either 1, 2, 4, 8, or 16 GB of memory. The accompanying table gives the distribution of Y = the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \leq y < 2 \\ .15 & 2 \leq y < 4 \\ .50 & 4 \leq y < 8 \\ .90 & 8 \leq y < 16 \\ 1 & 16 \leq y \end{cases}$$

Let's first determine $F(y)$ for each of the five possible values of Y :

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = .05$$

$$F(2) = P(Y \leq 2) = P(Y = 1 \text{ or } 2) = p(1) + p(2) = .15$$

$$F(4) = P(Y \leq 4) = P(Y = 1 \text{ or } 2 \text{ or } 4) = p(1) + p(2) + p(4) = .50$$

$$F(8) = P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = .90$$

$$F(16) = P(Y \leq 16) = 1$$

Now for any other number y , $F(y)$ will equal the value of F at the closest possible value of Y to the left of y . For example,

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2) = F(2) = .15$$

$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4) = F(4) = .50$$

If y is < 1 , $F(y) = 0$ [e.g. $F(.58) = 0$], and if y is at least 16, $F(y) = 1$ [e.g. $F(25) = 1$]. The cdf is thus

A graph of this cdf is shown in Figure 3.5.

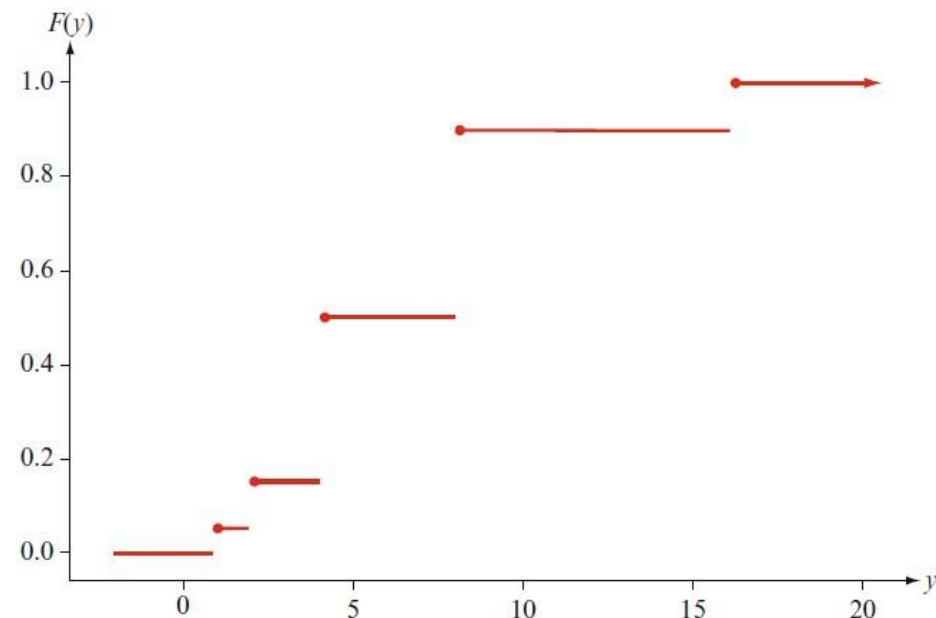


Figure 3.5 A graph of the cdf

Example 7: Expected Values of Discrete Random Variables and Variance

EXAMPLE The probability distribution for a random variable Y is given in Table 3.3. Find the mean, variance, and standard deviation of Y .

Table 3.3 Probability distribution for Y

y	$p(y)$
0	1/8
1	1/4
2	3/8
3	1/4

Solution By Definitions

$$\mu = E(Y) = \sum_{y=0}^3 yp(y) = (0)(1/8) + (1)(1/4) + (2)(3/8) + (3)(1/4) = 1.75,$$

$$\begin{aligned}\sigma^2 &= E[(Y - \mu)^2] = \sum_{y=0}^3 (y - \mu)^2 p(y) \\ &= (0 - 1.75)^2(1/8) + (1 - 1.75)^2(1/4) + (2 - 1.75)^2(3/8) + (3 - 1.75)^2(1/4) \\ &= .9375, \\ \sigma &= +\sqrt{\sigma^2} = \sqrt{.9375} = .97.\end{aligned}$$

Alternatively, the variance can be found in the following way:

$$E(Y^2) = \sum_y y^2 p(y) = (0)^2(1/8) + (1)^2(1/4) + (2)^2(3/8) + (3)^2(1/4) = 4,$$

$$\sigma^2 = E(Y^2) - \mu^2 = 4 - (1.75)^2 = .9375. \quad \blacksquare$$

Examples 8-9-10: Expected Value of a Function of a Random Variable

Example 8: Prove the following theorem:

Let Y be a discrete random variable with probability function $p(y)$ and c be a constant. Then $E(c) = c$.

Proof:

Consider the function $g(Y) \equiv c$.

$$E(c) = \sum_y cp(y) = c \sum_y p(y).$$

But $\sum_y p(y) = 1$ and, hence, $E(c) = c(1) = c$.

Example 9: Prove the following theorem:

Let Y be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of Y , and c be a constant. Then

$$E[cg(Y)] = cE[g(Y)].$$

Proof:

$$E[cg(Y)] = \sum_y cg(y)p(y) = c \sum_y g(y)p(y) = cE[g(Y)].$$

Example 10: Prove the following theorem:

Let Y be a discrete random variable with probability function $p(y)$ and $g_1(Y), g_2(Y), \dots, g_k(Y)$ be k functions of Y . Then

$$E[g_1(Y) + g_2(Y) + \dots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \dots + E[g_k(Y)].$$

Proof:

We will demonstrate the proof only for the case $k = 2$, but analogous steps will hold for any finite k .

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)]. \end{aligned}$$

Example 11: Expected Value of a Function of a Random Variable

The manager of an industrial plant is planning to buy a new machine of either type A or type B . If t denotes the number of hours of daily operation, the number of daily repairs Y_1 required to maintain a machine of type A is a random variable with mean and variance both equal to $.10t$. The number of daily repairs Y_2 for a machine of type B is a random variable with mean and variance both equal to $.12t$. The daily cost of operating A is $C_A(t) = 10t + 30Y_1^2$; for B it is $C_B(t) = 8t + 30Y_2^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consists of (a) 10 hours and (b) 20 hours?

Solution:

The expected daily cost for A is

$$\begin{aligned} E[C_A(t)] &= E[10t + 30Y_1^2] = 10t + 30E(Y_1^2) \\ &= 10t + 30\{V(Y_1) + [E(Y_1)]^2\} = 10t + 30[.10t + (.10t)^2] \\ &= 13t + .3t^2. \end{aligned}$$

In this calculation, we used the known values for $V(Y_1)$ and $E(Y_1)$ and the fact that $V(Y_1) = E(Y_1^2) - [E(Y_1)]^2$ to obtain that $E(Y_1^2) = V(Y_1) + [E(Y_1)]^2 = .10t + (.10t)^2$. Similarly,

$$\begin{aligned} E[C_B(t)] &= E[8t + 30Y_2^2] = 8t + 30E(Y_2^2) \\ &= 8t + 30\{V(Y_2) + [E(Y_2)]^2\} = 8t + 30[.12t + (.12t)^2] \\ &= 11.6t + .432t^2. \end{aligned}$$

Thus, for scenario (a) where $t = 10$,

$$E[C_A(10)] = 160 \quad \text{and} \quad E[C_B(10)] = 159.2,$$

which results in the choice of machine B .

For scenario (b), $t = 20$ and

$$E[C_A(20)] = 380 \quad \text{and} \quad E[C_B(20)] = 404.8,$$

resulting in the choice of machine A .

In conclusion, machines of type B are more economical for short time periods because of their smaller hourly operating cost. For long time periods, however, machines of type A are more economical because they tend to be repaired less frequently. ■

Examples 12-13-14: Moment Generating Functions

Example 12: Finding the mgf

Let X be a Bernoulli random variable with $p(0) = \frac{1}{3}$ and $p(1) = \frac{2}{3}$. Then

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} p(x) = e^{t \cdot 0} \frac{1}{3} + e^{t \cdot 1} \frac{2}{3} = \frac{1}{3} + e^t \frac{2}{3}$$

It should be clear that a Bernoulli random variable will always have an mgf of the form $p(0) + p(1)e^t$. This mgf exists because it is defined for all t . ■

Example 13: Finding the mgf of a linear function of X

Recall: Let X have the mgf $M_X(t)$ and let $Y = aX + b$. Then $M_Y(t) = e^{bt} M_X(at)$.

Let X be a Bernoulli random variable with $p(0) = \frac{20}{38}$ and $p(1) = \frac{18}{38}$. Think of X as the number of wins, 0 or 1, in a single play of roulette. If you play roulette at an American casino and you bet red, then your chances of winning are $\frac{18}{38}$ because 18 of the 38 possible outcomes are red. Then $M_X(t) = \frac{20}{38} + e^t \frac{18}{38}$. Let your bet be \$5 and let Y be your winnings. If $X = 0$ then $Y = -5$, and if $X = 1$ then $Y = +5$. The linear equation $Y = 10X - 5$ gives the appropriate relationship.

The equation is of the form $Y = aX + b$ with $a = 10$ and $b = -5$, so by the proposition

$$\begin{aligned} M_Y(t) &= e^{bt} M_X(at) = e^{-5t} M_X(10t) \\ &= e^{-5t} \left[\frac{20}{38} + e^{10t} \frac{18}{38} \right] = e^{-5t} \frac{20}{38} + e^{5t} \frac{18}{38} \end{aligned}$$

From this we can read off the probabilities for Y : $p(-5) = \frac{20}{38}$ and $p(5) = \frac{18}{38}$. ■

Example 14: Finding the mean and the variance using mgf

X represents the number of claims in

a year with pmf and mgf

x	0	1	2
$p(x)$	.7	.2	.1

$$M_X(t) = .7 + .2e^t + .1e^{2t}$$

First, find the derivatives

$$\begin{aligned} M'_X(t) &= .2e^t + .1(2)e^{2t} \\ M''_X(t) &= .2e^t + .1(2)(2)e^{2t} \end{aligned}$$

Setting t to 0 in the first derivative gives the first moment

$$E(X) = M'_X(0) = M_X^{(1)}(0) = .2e^0 + .1(2)e^{2(0)} = .2 + .1(2) = .4$$

Setting t to 0 in the second derivative gives the second moment

$$E(X^2) = M''_X(0) = M_X^{(2)}(0) = .2e^0 + .1(2)(2)e^{2(0)} = .2 + .1(2)(2) = .6$$

To get the variance recall the shortcut formula from the previous section:

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = .6 - .4^2 = .6 - .16 = .44$$

References

- Devore, J.L. and Berk, K.N. (2012) Modern Mathematical Statistics with Applications, 2nd Edition, New York: Springer.
- Wackerly, D.D., Mendenhall, W. and Scheaffer, R.L., (2008) Mathematical Statistics With Applications, 7th Edition, Pacific Grove: Duxbury.