Chapter 5: Discrete Probability Distributions

Random Variables

A random variable, X, is a variable that has a numerical value, determined by chance, for each outcome of an experiment or observation.

There are two types of random variables: discrete and continuous.

Discrete random variable—a random variable that can take on only a finite number of values or a countable number of values. Discrete random variables are typically counts.

Continuous random variable—a random variable that can take on any of the infinite number of values in an interval. Continuous random variables are typically measurements.

Examples of random variables:

- The number of cups of coffee that a fast-food restaurant serves each day *(discrete)*
- The number of nursing majors at Massasoit *(discrete)*
- The blood pressures of all patients admitted to a hospital on a specific day *(continuous)*
- The time it takes to complete an exercise session *(continuous)*
- The speed of a race car *(continuous)*

Probability Distributions

A probability distribution is an assignment of probabilities to the specific values of a random variable or to a range of values of the random variable.

Example: Let X = the number showing when a die is rolled. Create a probability distribution for X.

Solution: The sample space contains six equally likely outcomes: {1, 2, 3, 4, 5, 6}.

X	$P(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Features of the probability distribution of a discrete random variable

1. The probability distribution has a probability assigned to each value of the random variable.
2. The sum of all the assigned probabilities must be 1.
3. Individual probabilities must be between 0 and 1 inclusive.

Example: Let X = the number of heads in two tosses of a coin. Create a probability distribution for X.

Solution: The sample space contains four equally likely outcomes: {HH, HT, TH, TT}.

X	$P(X)$	
0	0.25	*(or 1/4)*
1	0.50	*(or 2/4)*
2	0.25	*{or 1/4)*

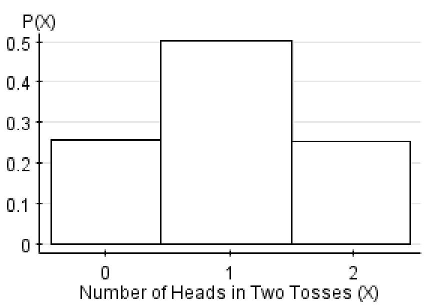
Using the probability distribution above, what is the probability that you toss either zero heads or one head in two tosses of a coin?

P(0 or 1) = P(0) + P(1) = 0.25 + 0.50 = 0.75

Probability Histograms

Example: Using the probability distribution above , create a probability histogram for X.

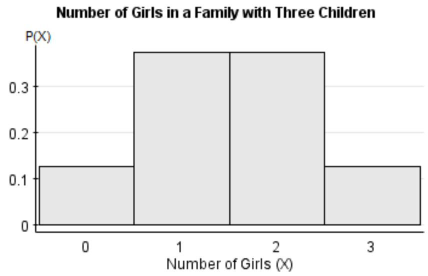
Solution:



Example: A couple plans to have three children. Let X represent the number of girls out of three children. Create a probability distribution and a probability histogram for X.

Solution: There are eight possible outcomes: {BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG}. The probability distribution is below.

X	$P(X)$	
0	0.125	*(or 1/8)*
1	0.375	*(or 3/8)*
2	0.375	*(or 3/8)*
3	0.125	*(or 1/8)*



Mean, Variance and Standard Deviation of a Discrete Probability Distribution

Mean: $\mu = \sum X \cdot P(X)$

Variance: $\sigma^2 = \sum (X - \mu)^2 P(X)$

Standard deviation: $\sigma = \sqrt{\sum (X - \mu)^2 P(X)}$

Note: μ is also called the expected value of X.

Example: Let X represent the number of heads in two tosses of a coin. What are the mean, variance, and standard deviation of X?

Solution:

To find the mean and the standard deviation of X, create the following table. Note that the mean is the sum of the third column and the variance is the sum of the last column.

X	$P(X)$	$X \cdot P(X)$	$X - \mu$	$(X - \mu)^2$	$(X - \mu)^2\ P(X)$
0	0.25	0	-1	1	0.25
1	0.50	0.50	0	0	0
2	0.25	0.50	1	1	0.25

$\mu = \sum X \cdot P(X) = 1$ $\sigma^2 = \sum (X - \mu)^2 P(X) = 0.5$

Mean = $\mu = \sum X \cdot P(X) = 1$

Variance = $\sigma^2 = \sum (X - \mu)^2 P(X) = 0.5$

Standard Deviation = $\sigma = \sqrt{\sigma^2} = \sqrt{0.5} = 0.7$

<u>Example:</u> A couple plans to have three children. Let X represent the number of girls out of three children. Compute the mean and standard deviation of X.

<u>Solution:</u> There are eight possible outcomes: {BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG}. The probability distribution is below.

X	$P(X)$
0	0.125
1	0.375
2	0.375
3	0.125

To find the mean and the standard deviation of X, create the following table. Note that the mean is the sum of the third column and the variance is the sum of the last column.

X	$P(X)$	$X \cdot P(X)$	$X - \mu$	$(X-\mu)^2$	$(X-\mu)^2\ P(X)$
0	0.125	0	-1.5	2.25	0.28125
1	0.375	0.375	-0.5	0.25	0.09375
2	0.375	0.75	0.5	0.25	0.09375
3	0.125	0.375	1.5	2.25	0.28125

$\mu = \sum X \cdot P(X) = 1.5$ $\qquad$ $\sigma^2 = \sum (X-\mu)^2 P(X) = 0.75$

Mean = $\mu = \sum X \cdot P(X) = 1.5$

Variance = $\sigma^2 = \sum (X-\mu)^2 P(X) = 0.75$ (0.8 rounded)

Standard Deviation = $\sigma = \sqrt{\sigma^2} = \sqrt{0.75} = 0.9$

Expected Value

The mean of a probability distribution is also referred to as the expected value. It is the "long-term average" if a statistical experiment were carried out many times.

Example: Suppose someone asks you to play the following game. Roll a die. If the die shows a 1, 2, 3, or 4, you pay $10. If the die shows a 5, you win $15. If the die shows a 6, you win $30. What is your expected value?

Solution: Let X represent winnings. Arrange these results in a table.

X	$P(X)$
-10	4/6
15	1/6
30	1/6

Expected value = μ = (-10)(4/6) + (15)*(1/6) + 30 * (1/6) = 0.83. On average, you would win $0.83 per game.

Example: A lottery offers three prizes: one $1000 prize, one $500 prize, and five $100 prizes. One thousand tickets are sold at $3 each. What is your expected profit if you buy one ticket?

Solution: If you spend $3 on a ticket and win the $1000 prize, your net winnings are $997. If you spend $3 on a ticket and win the $500 prize, your net winnings are $497. If you spend $3 on a ticket and win the $100 prize, your net winnings are $97. If you do not win, your net winnings are -$3 since you spent three dollars on the ticket.

Let X represent winnings. Arrange these results in a table.

X	$P(X)$
997	1/1000
497	1/1000
97	5/1000
-3	993/1000

Expected value = μ = (-997)(1/1000) + (497)*(1/1000) + (-97)(5/1000) + (-3)(993/1000) = -1

Example: A 28-year-old man pays $500 for a one-year life insurance policy with coverage of $100,000. If the probability that he will live through the year is 0.9992, what is the expected value for the insurance policy?

Solution: There are two possibilities: the man lives through the year or he doesn't. Let X represent the insurance payout.

Outcome	X	$P(X)$
Lives	-500	0.9992
Dies	99,500	0.0008

Expected value = μ = (-500)*(0.9992) + (999,500)*(0.008) = **-**$420

Binomial Probabilities

<u>Binomial probability experiment</u>—An experiment that deal with situations that result in two possible outcomes (e.g. boy/girl, for/against, heads/tails, right/wrong)

Examples of questions involving binomial probabilities.

1. A couple plans to have four children. What is the probability that they have exactly 3 boys?

2. A biased coin is tossed 10 times. Suppose that for a single toss, the probability of heads is 70% and the probability of tails is 30%. What is the probability that exactly 2 tails are tossed?

3. A student guesses on a twelve-question multiple-choice exam where there are four possible answers for each question. What is the probability that he or she gets all 12 correct?

4. According to a recent study, about 40 percent of college students have a credit card. Suppose that four college students are selected at random. What is the probability that exactly one has a credit card?

5. According to a recent poll, approximately seventy percent of U.S. adults drink alcohol. Suppose 5 U.S. adults are randomly selected. What is the probability that exactly two adults in the sample drink alcohol?

Characteristics of a binomial experiment

1. There are a fixed number of observations, called trials. The number of trials is denoted by the letter n.
2. The n trials must be independent.
3. Each trial must have only two possible outcomes labeled success or failure.
4. The probability of success and failure must remain constant for each trial. The probability of success is denoted by p and the probability of failure is denoted by q. Note: $q = 1 - p$.
5. The central problem of a binomial experiment is to find the probability of x successes in n trials.

Formula for the Binomial Probability Distribution

$$P(X) = \frac{n!}{x!\,(n-x)!}p^x q^{n-x}$$

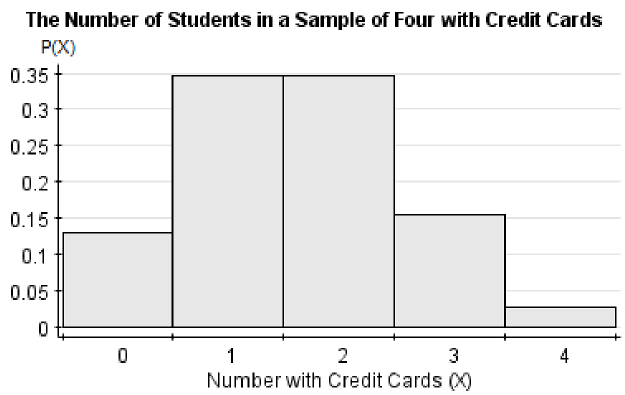
The *factorial* of a positive integer n, denoted by n!, is the product of all positive integers less than or equal to n. For example,

4! = 4 * 3 * 2 * 1 = 24
5! = 5 * 4 * 3 * 2 * 1 = 120

Example: According to a recent study, about 40 percent of college students have a credit card. Suppose that four college students are selected at random. Let X represent the number in the sample that have a credit card. Create a probability distribution and a probability histogram for X.

Solution: $n = 4, p = 0.40, q = 1 - 0.40 = 0.60$. Use the binomial probability formula, the binomial probability table, or the calculator (binompdf) to find the probabilities.

X	$P(X)$
0	0.1296
1	0.3456
2	0.3456
3	0.1536
4	0.0256



Example: Refer to the previous problem. If four college students are selected at random, find the following probabilities. Use either the binomial probability formula or the binomial probability table.

a. That exactly one has a credit card.

P(1) = 0.3456

b. That none of them have a credit card.

P(0) = 0.1296

c. That two or fewer have a credit card.

$P(X \leq 2) = P(0) + P(1) + P(2) = 0.1296 + 0.3456 + 0.3456 = 0.8208$

d. at least one has a credit card?

$P(at\ least\ 1) = 1 - P(none) = 1 - 0.1296 = 0.8704$

e. That more than two have a credit card?

$P(three) = P(3) + P(4) = 0.1536 + 0.0256 = 0.1792$

Example: According to a recent poll, approximately seventy percent of U.S. adults drink alcohol. Suppose 5 U.S. adults are randomly selected. Let X represent the number of adults in the sample who drink alcohol. Use the binomial probability formula, the binomial probability table, or your calculator to find the following probabilities.

a. That exactly 2 adults in the sample drink alcohol.

$P(2)$ = 0.1323

b. That at least three adults in the sample drink alcohol.

P(at least 3) = P(3) + P(4) + P(5) = 0.3087 + 0.36015 + 0.16807 = 0.83692

Alternatively, you can use binomcdf on the calculator:

P(at least 3) = 1 – P(2 or fewer) = 1 – binomcdf(5, 0.70, 2) = 0.83692

c. That everyone in the sample drinks alcohol.

$P(5)$ = 0.16807

Mean, Variance, and Standard Deviation of the Binomial Distribution

For the binomial probability distribution, the mean, variance, and standard deviation are given by the formulas below.

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard Deviation: $\sigma = \sqrt{npq}$

- X is a random variable representing the number of successes in a binomial distribution.
- n is the number of trials.
- p is the probability of success on a single trial.
- $q = 1 - p$ is the probability of failure on a single trial.

Example: According to a recent study, about 40 percent of college students have a credit card. Suppose that four college students are selected at random. Let X represent the number in the sample that have a credit card. Find the mean, variance, and standard deviation of X.

Solution:
Mean: μ = (4)(0.40) = 1.6 students

Variance: $\sigma^2 = npq$ = (4)(0.40)(0.60) = 0.96 *(round to 1.0)*

Standard Deviation: $\sigma = \sqrt{npq} = \sqrt{(4)(0.40)(0.60)}$ = 0.979 *(round to 1.0)*

Question: In a sample of four college students, would it be "unusual" for all four to have a credit card?

Recall that an unusual value is one that is more than two standard deviations from the mean.

Because the value 4 is more than two standard deviations above the mean, it would be considered unusual.

Example: According to a recent poll, approximately seventy percent of U.S. adults drink alcohol. Suppose 5 U.S. adults are randomly selected. Let X represent the number of adults in the sample who drink alcohol. Find the mean, variance, and standard deviation of X.

Solution:
Mean: μ = (5)(0.70) = 3.5 students

Variance: $\sigma^2 = npq$ = (5)(0.70)(0.30) = 1.05 *(round to 1.1)*

Standard Deviation: $\sigma = \sqrt{npq} = \sqrt{(5)(0.70)(0.30)}$ = *1.0247* *(round to 1.0)*

Question: In a sample of 5 U.S. adults, would it be "unusual" for none of them to drink alcohol?

Recall that an unusual value is one that is more than two standard deviations from the mean.

Because the value 0 is more than two standard deviations below the mean, it would be considered unusual.