## 5.1 Random Variables and Probability Distributions

### Statistical Experiment

A **statistical experiment** is any process by which an observation or a measurement is made.

### Example A Statistical Experiment

a. Measure the daily rainfall in inches.
b. Count the number of eggs in a nest.
c. Measure the weight in kg of bear cubs.
d. Count the number of defective light bulbs in a case of bulbs.

### Random Variable

A **random variable**, $x$, represents a quantity being measured.

### Example B Random Variables

a. $x$ = amount of rain each day in inches.
b. $x$ = the number of eggs in a nest.
c. $x$ = the weight in kg of bear cubs.
d. $x$ = the number of defective light bulbs in a case.

### Discrete and Continuous Random Variables

i. When a random variable $x$ can take on only countable values (such as 0, 1, 2, 3, . . .), then $x$ is said to be a **Discrete Random Variable**.

ii. When a random variable $x$ can take on any value in an interval, then $x$ is said to be a **Continuous Random Variable**.

### Example C Discrete vs Continuous Random Variables

Which of the random variables in Example A, parts a-d, are discrete and which are continuous?

### Example D  Discrete vs Continuous Random Variables

Which of the following random variables are discrete and which are continuous?

a. The number of students in a section of a statistics course.
b. The air pressure in an automobile tire.
c. The number of osprey chicks living in a nest.
d. The height of students at Palomar.
e. The mpg of randomly selected vehicles on a highway.
f. The time it takes a student to register for spring semester.

### Probability Distribution

A **probability distribution** is an assignment of probabilities to specific values of a random variable (discrete) or to a range of values of a random variable (continuous). A probability distribution is basically a relative frequency distribution organized in a table. Recall: The sum of all probabilities must be one.

| Roll a Die *x* | *P(x)* |
|---|---|
| 1 | 1/6 |
| 2 | 1/6 |
| 3 | 1/6 |
| 4 | 1/6 |
| 5 | 1/6 |
| 6 | 1/6 |

### Mean and Standard Deviation of a Discrete Probability Distribution

For a discrete random variable $x$ and probability of that variable, $P(x)$:

$$\text{mean} = \textbf{expected value} = \mu = \sum x \cdot P(x)$$

$$\text{standard deviation} = \sigma = \sqrt{\sum (x-\mu)^2 \cdot P(x)}$$

When the random variable is given as ranges of numbers, set $x$ equal to the midpoint of each range. See Table.

| Age Range | Midpoint of the Range, *x* | *P(x)* |
|---|---|---|
| 18-24 years | 21 years | .26 |
| 25-34 years | 29.5 years | .34 |

## Example 1

Dr. Fidget developed a test to measure boredom tolerance. He administered it to a group of 20,000 adults. The possible scores were 0, 1, 2, 3, 4, 5, and 6, with 6 indicating the highest tolerance for boredom. The results are shown.

| **Score** $x$ | **Number (frequency, *f*)** | **Probability** $P(x) = f / 20000$ |
|---|---|---|
| 0 | 1400 | |
| 1 | 2600 | |
| 2 | 3600 | |
| 3 | 6000 | |
| 4 | 4400 | |
| 5 | 1600 | |
| 6 | 400 | |

a. Find the probability (relative frequency) of each score and construct a probability distribution in the table above. Let $L_1$ be the scores, $x$, $L_2$ be the frequency, and $L_3$ be the probabilities, P(x). On the TI-83: $\text{L}_2 / 20{,}000 \to \text{L}_3$.

b. Graph the probability distribution as a histogram of probability versus test score. What is the total area of the bars?

c. Compute the expected value of the test scores and the standard deviation. TI-83: 1-Var Stats $\text{L}_\text{x}, \text{L}_{\text{P(x)}}$. Use $\bar{x} = \mu$, and $\sigma_\text{x} = \sigma$.

d. Topnotch Clothing Co. wants to hire someone with a score of 5 or 6 to operate is machinery. What is the probability that a randomly selected person has a score of 5 or 6?

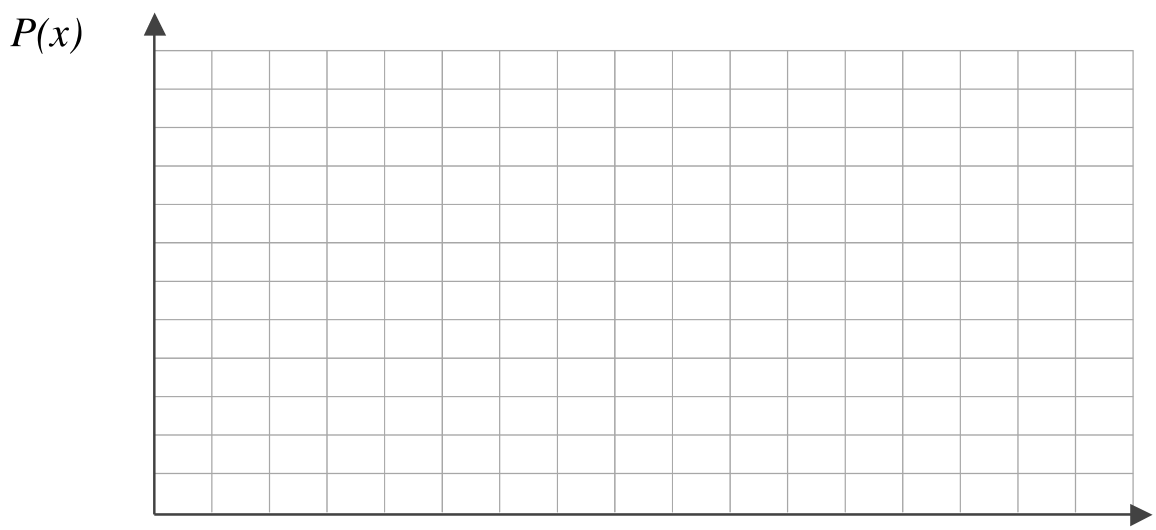
## Exercise 7

Data was collected over 208 nights tabulating the number of room calls in a night requiring a nurse.

a. Use the relative frequency to find P($x$). In words, what does each P($x$) represent?

| $x$ | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
|---|---|---|---|---|---|---|---|---|---|---|
| $f$ | 6 | 10 | 11 | 20 | 26 | 32 | 34 | 28 | 25 | 16 |
| $P(x)$ | | | | | | | | | | |

b. Graph the probability distribution. Completely annotate the graph.




*x = number of room calls per night requiring a nurse*

c. Estimate the probability that on a randomly selected night there will be between 39 and 43 (inclusive) room calls requiring a nurse.

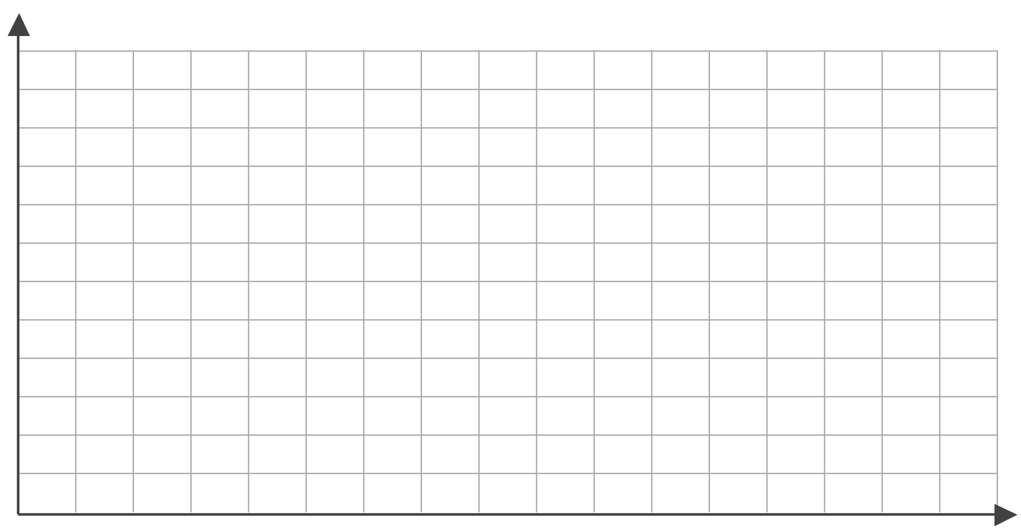
d. Find the expected number of room calls requiring a nurse. Find the standard deviation of the distribution.

## Exercise 8

In 1851 the percent age distribution of nurses (to the nearest year) in Great Britain was:

| **Age** | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80+ |
|---|---|---|---|---|---|---|---|
| ***x*** | 24.5 | 34.5 | 44.5 | 54.5 | 64.5 | 74.5 | 84.5 |
| **%** | 5.7 | 9.7 | 19.5 | 29.2 | 25 | 9.1 | 1.8 |

a. Use a histogram to graph the probability distribution. Completely annotate the graph.



b. Find the probability that a randomly selected British nurse in 1851 would be 60 years or older.

c. What is the expected value (the "balance point" on the graph) and standard deviation of the age of a British nurse in 1851?

**Ignore exercises 15-17 in section 5.1.**

## 5.2 Binomial Probabilities

A binomial experiment must have all of the following:

1. There are a fixed number of trials, $n$.
2. The $n$ trials are independent and repeated under the same conditions.
3. Each trial has two possible outcomes: success, $S$; failure, $F$.
4. For each trial, the probability of success is the same.
   Notation: $P(S) = p$ , $P(F) = q$ , and $p + q = 1$
5. The central problem of a binomial experiment is to find the probability of $r$ successes out of $n$ trials.

**Example A** *Show this is a binomial experiment.*

The registrar of a college noted that each semester the rate of withdrawal for chemistry has been 35%. She wondered if 80 students register for the course, what is the probability that 55 will complete it?

(a) What are the two outcomes $S$ and $F$?

(b) Are the trials independent? What is $n$?

(c) What is $p$, $q$ and $r$?

(d) How can the problem be stated in the form: "What is the probability of $r$ successes out of $n$ trials?"

**Example B**

Is the following a binomial experiment: A random sample of 30 men is asked to name his favorite TV show. Why / why not?

**Example C**

Is the following a binomial experiment? Why/why not?
A committee of 4 men and 6 women wish to select a chairperson and recorder. They do so by placing their names in a hat and drawing two names; the first will be the chairperson and the second the recorder. What is the probability that both offices will be held by women?

**Example D**

Determine $S, F, p, q, n$, and $r$ in the following binomial experiment.

Nine percent of the population has blood type B. If we choose 18 people at random, what is the probability that three will have blood type B?

## Example E The Binomial Distribution

Suppose you run out of exam time and guessed on the last three multiple-choice questions. Assuming there are four choices of answers, what is the probability that you guess 0, 1, 2, and all 3 questions correct?

a. What is a success and failure? What are $p$, $q$, and $n$?

b. What is $r$?

c. How many possible outcomes, in terms of $S$ and $F$, are possible? List them in a tree diagram. Find the probability of each outcome.

d. Find P($r$ successes) = P($r$), for $r = 0, 1, 2, 3$. Construct a probability distribution.

## Computation of the Binomial Probability Distribution

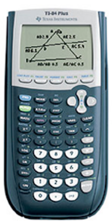
The probability of $r$ successes out $n$ trials in a binomial experiment is computed by the binomial probability density function:

number of combinations of $n$ things taken $r$ at a time

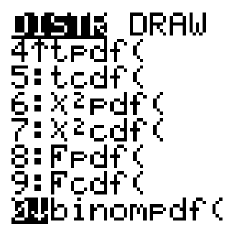
$$P(r) = C_{n,r}\, p^r q^{n-r} = binompdf(n, p, r)$$

probability of success raised to the number of successes

probability of failure raised to the number of failures



**TI-83/84**
DISTR / 0: binompdf($n$, $p$, $r$)




### Example F

Suppose that 65% of survey questionnaires sent to all faculty are completed and returned. If three faculty members are chosen at random, compute the probability that

a. exactly 2 will be completed and returned

b. all 3 will be completed and returned

## Steps to Show Your Work (SYW) for full credit

1. Setup: Define $S$. State $p$ and $n$.
2. Solve by showing the probability notation, the TI-83/84 function accessed along with its inputs and output.
   e.g. $P(5) = binompdf(12, 0.22, 5) = 0.0717$
3. Unless otherwise stated, write all probabilities to the nearest ten-thousandth (4 decimal places).

## Example 5

One survey showed that 59% of Internet users are somewhat concerned about the privacy of their e-mail. Based on this information, what is the probability that for a random sample of 10 Internet users, 6 are concerned about their e-mail privacy?

## Example 6

If 22% of U.S. households have a Nintendo game, compute the probability that

a. exactly 5 of 12 randomly chosen families will have Nintendo games.

b. at most 5 have Nintendo games.

| *r* | ***P(r)*** |
|---|---|
| 0 | binompdf(12, 0.22, 0) = 0.0507 |
| 1 | binompdf(12, 0.22, 1) = 0.1717 |
| 2 | binompdf(12, 0.22, 2) = 0.2663 |
| 3 | binompdf(12, 0.22, 3) = 0.2503 |
| 4 | binompdf(12, 0.22, 4) = 0.1589 |
| 5 | binompdf(12, 0.22, 5) = 0.0717 |
| | sum = 0.9696 |

## TI-83/84 Binomial Cumulative Density Function

$$P(r \le r_{\max}) = P(0) + P(1) + \ldots P(r_{\max}) = \text{binomcdf}(n, p, r_{\max})$$

c. Redo (b) using the **binomcdf(** function.

d. more than 5 will have Nintendo games.

e. less than 8 will have Nintendo games.

f. more than 2 will have Nintendo games.

**Example 6**

A biologist studying a hybrid tomato found that there is a probability of 0.70 that the seeds will germinate. If the biologist plants 10 seeds, compute the probability (SYW) that:

a. exactly 8 seeds will germinate.

b. exactly 5 seeds will germinate.

c. at most 7 seeds will germinate.

d. at most 5 seeds will germinate.

e. at least 8 seeds will germinate.

f. at least 4 seeds will germinate.

g. between 3 and 7 (inclusive, including 3 an 7) will germinate.

h. between 4 and 9 (exclusive) will germinate.

i. less than 5 will germinate.

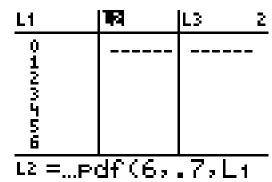
j. more than 3 will germinate

## 5.3 More on Binomial Distributions

A binomial distribution tells us the probability of $r$ successes out of $n$ trials for various values of $r$. The **graph of a binomial distribution** is a histogram where the discrete values of $r$ are on the horizontal axis and the corresponding probability values, P($r$), are on the vertical axis.

### Example 7

A waiter found that the probability a lone diner will leave a tip is 0.7. During one lunch he serves 6 lone diners. Make a graph of the binomial probability distribution which shows the probability that 0, 1, 2, 3, 4, 5, and 6 lone diners will leave tips.

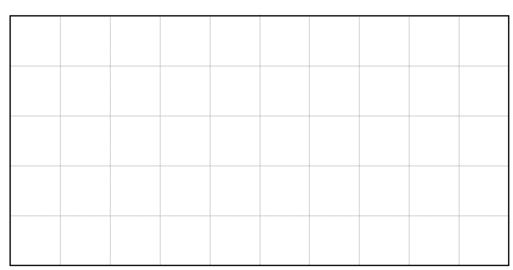



a. Let list $L_1$ be the $r$-values {0, 1, . . . , 6} and compute list $L_2$ by binompdf(6, 0.7, $L_1$) .

b. Fill-in the table:

| $r$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **P($r$)** | 0.0007 | 0.010 | 0.060 | 0.185 | 0.324 | 0.303 | 0.118 |

c. Construct the bar graph. Completely annotate the graph.



## Guided Exercise 6

Jim makes about 50% of the shots he attempts in basketball. Construct a probability distribution and histogram showing the probability that Jim will make 0, 1, 2, . . ., 6 shots out of 6 attempted shots.

| ***r*** | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **P(*r*)** | 0.0156 | 0.0938 | 0.2344 | 0.3125 | 0.2344 | 0.0938 | 0.0156 |

## Expected Value and Standard Deviation of a Binomial Distribution

For number of trials $n$, the probability of success $p$ and the probability of failure $q$:

**Expected Value** $\mu = np$ "balancing point"

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

## Example 8

Compute the expected value and standard deviation for guided exercise 6. Notice the expected value is the "balancing point" of a histogram.

## Example 9

In a Myers-Briggs Personality classification a rare personality type, called INFJ (introverted, intuitive, feeling, and judgmental), occurs in 2.1% of the population. Suppose that a high school graduating class has 167 students.

a. What is a success?

b. Find the probability that 0, 1, 2, 3, 4, 5 and 6 students in the class have an INFJ personality type.

| ***r*** | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| ***P(r)*** | 0.0289 | 0.1035 | 0.1842 | 0.2174 | 0.1912 | 0.1337 | 0.0774 |

c. Find the probability that the following number of students in the class have an INFJ personality type.

(1) more than 7.

(2) less than 8.

(3) between 2 and 5 (inclusive).

(4) between 2 and 5 (exclusive).

## Guided Exercise 8

Suppose a satellite requires 3 solar cells for its power, the probability that any one of these cells will fail is 0.15, and the cells operate and fail independently.

Part I: We want to find the least number of cells each satellite should have so that the expected number of working cells (the mean) is no smaller that 3. That is, $n$ is the total number of cells, and $r$ is the number of working cells, $p$ is probability that a cell will work, and $\mu \geq 3$.

a. Define $S$. What is the value of $q$ and $p$?

b. What is the minimum number of cells, $n$, required so that $\mu \geq 3$?

c. Write a summarizing statement of parts a-b.

If 4 cells are installed in each satellite, then we expect, on average, that 3.53 cells will be operational in each satellite. Of course each individual satellite has 0, 1, 2, 3, or 4 operational cells but, the mean number of operational cells over a large number of satellites is 3.53

Part II: Find the minimum number of cells each satellite must have so that there is a 99% probability that a randomly selected satellite will be operational.

a. If $r$ is the number of working cells what are we trying to find?

b. Find $n$ so that $P(r \geq 3) \geq 0.99$

c. Write a summarizing statement of parts a-b.

If ____ cells are installed in a satellite, then there is a 99% probability that the satellite will be operational. Or, 99% of all satellites with ____ cells in them will be operational.

## Exercise 18

The Denver Post reported that a recent audit of 911 calls showed that 88% were not emergencies. Suppose that the 911 operators have just received four calls.

a. What is the probability that all four calls are emergencies?

b. What is the probability that three or more calls are not emergencies?

c. Find the minimum number of calls *n* that 911 operators need to answer to be 97% sure that at least one call was in fact an emergency.

   i. What is *S*? In probability notation what do we want?

   ii. Restate the problem using the complements principle so that the binomcdf( function can be used.

   iii. Press the Y= key and set $Y_1 = 1 - binomcdf(\_\_\_,\_\_\_,\_\_\_)$

   iv. Set your table settings as shown.


   Then press TABLE (2nd GRAPH) to view the table. Scroll down until the probability is at least 0.97. Show your work by copying part of the table into the problem.

## 5.4 Exercise #1 Geometric Probability Distributions

Susan is taking Western Civilization on a pass/fail basis. Historically, the passing rate for this course has been 77% each term. Let $n = 1, 2, 3, \ldots$ represent the number of times a student takes this course until a passing grade is received. Assume the attempts are independent.

a. Construct a tree diagram for $n = 1, 2, 3$, and 4 attempts. List the outcomes and their associated probabilities.

b. What is the probability that Susan passes on the first try?

c. What is the probability that Susan passes on the second try?

d. What is the probability that Susan passes on the third try?

e. What is the probability that Susan passes on the *n*th try?

f. What is the probability that Susan needs three or more tries to pass?

## Geometric Probability Distribution

Suppose in an experiment there are two possible outcomes for each trial (success and failure), $n$ is the number of trials needed for the first success to occur, $p$ is the probability of success for each independent trial, and $p$ is the same for each trial. Then the following are true.

1. The probability of success on exactly the $n^{th}$ is given by
$$P(n) = p(1-p)^{n-1} = geometpdf(p,n)$$

2. The probability that a success will occur on or before trial number $n_{max}$ is given by
$$P(n \le n_{\max}) = P(1) + P(2) + \ldots + P(n_{\max}) = geometcdf(p, n_{\max})$$

3. The probability that a success will occur after the $n_{min}$ is given by
$$P(n > n_{\min}) = 1 - P(n \le n_{\min}) = 1 - geometcdf(p, n_{\min})$$

### Steps for Full Credit on Geometric Distribution Problems

1. Define what a success $S$ is. State what $p$ is.
2. State the problem in terms of proper probability notation, including the complements principle when appropriate.
3. Show all inputs, outputs, and functions accessed in the TI-83/84.
4. Unless otherwise stated, round all probabilities to 4 decimal places.

### Geometric Probability Distribution: Population Mean and Population Standard Deviation

Population Mean: $\mu = 1/p$

Population Standard Deviation: $\sigma = \dfrac{\sqrt{1-p}}{p}$

## 5.4 Exercise #2

Bob, a recent law school graduate, intends to take the bar exam which has a 57% passing rate. What is the probability that Bob

a. passes the exam on the second try?

b. needs four attempts to pass the exam?

c. needs more than four attempts to pass the exam?

d. needs at least seven attempts to pass the exam?

e. needs fewer than three attempts to pass the exam?

## 5.4 Poisson Distribution

The **Poisson Distribution** is a discrete probability distribution of a random variable *r* that has these characteristics.

i. The experiment consists of counting the number of times, *r*, an event occurs in a given interval. The interval can be and interval of time, space, area or volume.
ii. The probability of the event occurring is the same for each interval (of time, space, area, or volume).
iii. The number of occurrences of the event in one interval is independent of the number of occurrences in other intervals.
iv. The mean number of successes (i.e. expected number of success; expected value), denoted $\lambda$, is known over the interval (of time, volume, area, length,. That is, $\lambda$ is the expected value over the given interval (of time, space, area, or volume).ß

### Probabilities for a Poisson Distribution

1. The probability of exactly *r* successes ($r = 0, 1, 2, 3, \ldots$) over the same interval of time (volume, area, etc.) is

$$P(r) = \frac{e^{-\lambda} \cdot \lambda^r}{r!} = poissonpdf(\lambda, r)$$

The population mean and standard deviation are given by

$$\mu = \lambda \qquad \text{and} \qquad \sigma = \sqrt{\lambda}$$

2. The probability of at most $r_{\max}$ successes is

$$P(r \le r_{\max}) = poissoncdf(\lambda, r_{\max})$$

### Steps for Full Credit on Poisson Distribution Problems

1. Define *S*.
2. Find $\lambda$, the expected number of successes over the given interval of time, area, space, population, etc.
3. Answer the question showing the probability notation, the TI-83/84 function accessed along with its' input and output.
4. Unless otherwise stated, show all probabilities to the nearest ten-thousandth (4 decimal places).

**Example 10**

On average boat fishermen on Pyramid Lake catch 0.667 fish per hour. Suppose you decide to fish the lake on a boat for 7 hours.

a. What is *S*?

b. What is the expected number of fish caught over the 7-hour period? This is $\lambda$.

c. In the 7-hour period of time what is the probability that you will catch

i. 0, 1, 2, or 3 fish?

ii. construct a table for $r = 0, 1, 2, \ldots, 8$ fish.

| **r** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| **P(*r*)** | | | | | | | | | |

iii. at least 5 fish?

iv. more than 7 fish?

v. between 2 and 10 fish (inclusive)?

**Example 11**

A loom which produces plaid wool fabric is known to produce, on the average, one noticeable flaw per 20 yards of fabric.

a. What is the probability that there will be no flaws in a fifty-yard piece of the wool fabric?

b. What is the probability that there will be more than 3 flaws in a 30-yard piece of fabric.