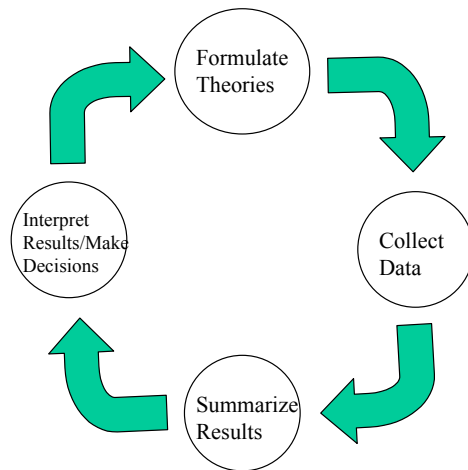# Chapter Goals

We now consider the problem of making inference. That is, using sample information to infer about the population with a certain measure of reliability.

Meaning of Probability

Objective Interpretation: The probability value of an event is equated with the relative frequency of occurrence of the event in the long run under constant causal conditions. The objective interpretation of probability applies only to repeatable events and not to unique events.

Subjective Interpretation: The subjective interpretation of probability (also called the personal interpretation) relates probability to degree of personal belief.

Simulating Probabilities

Definition *The probability that an outcome will occur is the proportion of time it occurs over the long run; that is, the relative frequency with which that outcome occurs.*

Example *What is probability of a tail when you toss a coin?*

Definition *A **random process** is a repeatable process whose set of possible outcomes is known, but the exact outcome cannot be predicted with certainty.*

Definition *A **simulation** is the imitation of random or chance behavior using random devices such as random number generators or a table of random numbers.*

Steps for finding probabilities using simulation:

1. Specify a model for the individual outcomes of the underlying random phenomenon.
2. Outline how to simulate an individual outcome and how to represent a single repetition of the random process.
3. Simulate many repetitions and estimate the probability of an event by its relative frequency.

Let's do it! 8.1

A couple plans to have children. They would like to have a boy to be able to pass on the family name. After some discussion, they decide to continue to have children until they have a boy or until they have three children, whichever comes first. What is the probability that they will have a boy among their children?

1. Specify a model for the individual outcomes.

2. Simulate individual outcomes.

3. Perform several repetitions, say 10.

4. Estimate the probability.
   a. Out of the first 10 repetitions, how many times did the couple have a boy?
   b. What is your group's estimate of the probability that this strategy will produce a boy?

5. Let us now combine the results from all groups to get a better estimate.

Let's do it! 8.4

There are three doors. Behind one door is a car. Behind each of the other two doors is a goat. As a contestant, you are asked to select a door, with the idea that you will receive the prize that is behind that door. The game host knows what is behind each door. After you select a door, the host opens one of the remaining doors that has a goat behind it. Note that no matter which door you select, at least one of the remaining doors has a goat behind it for the host to open. The host then gives you the following two options:

1. Stay with the door you originally selected and receive the prize behind it.

2. Switch to the other remaining closed door and receive the prize behind it.

What is the probability of winning the car if you stay? What is the probability of winning the car if you switch? Will switching increase your chance of winning the car?

Random Variables

Definition *A **random variable** is a rule that assigns one (and only one) numerical value to each simple event of an experiment*

Example *Consider the experiment: Toss two coins Let X be the random variable denoting the number of tails*

Simple Event   Numerical Value

Discrete and Continuous Random Variables

● Random variables that assume only a countable number of values are called discrete.

Example *The number of customers served in a restaurant on any particular day.*

● Random variables that can take on any value on a continuum are called continuous.

Example *The weight of a food brought to a checkout counter.*

Discrete Random Variables

Definition *The **probability distribution** of a discrete random variable is a graph, table, or formula that specifies the probability associated with each possible value the random variable can take.*

Example *Let X be the number of people in a household for a certain community. Consider the following probability distribution for X, which assumes that there are no more than seven people in a household.*

$X = x$	1	2	3	4	5	6	7
$P(X = x)$	0.20	0.32	0.18	0.15	0.07	0.03	

Comment: Any probability distribution for a discrete random variable X has the usual properties of a probability distribution:

1. $0 \leq P(X = x_i) \leq 1$ for $i = 1, 2, \ldots, k$
2. $\sum_{i=1}^{k} P(X = x_i) = 1$

Definition *The probability distribution for a discrete random variable is also called **probability mass function** or simply **probability function**.*

1. What must be the probability of 7 people in a household for this to be a legitimate discrete distribution?

2. What is the probability that a randomly chosen household contains more than 5 people?

3. What is the probability that a randomly chosen household contains no more than 2 people?

4. What is $P(2 \leq X \leq 4)$?

Definition *The **cumulative probability distribution** for a discrete random variable X provides the cumulative probabilities $P(X \leq x)$ for all values of x. Two important properties of any cumulative probability distribution are as follows:*

1. $P(X \leq x)$ *is always a value between 0 and 1.*

2. *The cumulative distribution never decreases as x increases.*

The cumulative probability distribution is also called cumulative probability function.

Example *Find $P(X \leq 4)$ from the previous example.*

Expected Value and Variance of Random Variables

Definition *The **mean**, or **expected value** of a discrete random variable X is denoted by E(X) and defined as*

$$E(X) = \sum_{i=1}^{k} x_i P(x_i),$$

where, $P(x_i)$ denotes $P(X = x_i)$.

Definition *The **variance** of a discrete random variable X is denoted by $\sigma^2(X)$ and defined as*

$$\sigma^2(X) = \sum_{i=1}^{k} (x_i - E(X))^2 P(x_i),$$

Definition *The positive square root of the variance of X is called the **standard deviation** of X and is denoted by*
$\sigma(X) = \sqrt{\sigma^2(X)}$

Continuous Random Variables

Definition *The **probability distribution** of a continuous random variable X is a function denoted by f(x), such that*
P(X takes on values in a set A) = the area under the function f(x) above the set A. The function f(x) must satisfy

1. $f(x) \geq 0, \forall x$ *(for all x)*

2. *The total area under* $f(x) = 1$.

Example *Let X be the length of a pregnancy in days, with X~N(266, 256). What is the probability that a pregnancy lasts at least 310 days?*