

Virtual Laboratories > 12. Finite Sampling Models > **1** 2 3 4 5 6 7 8 9

1. Introduction

Basic Theory

Sampling Models

Suppose that we have a **population** D of m objects. The population could be a deck of cards, a set of people, an urn full of balls, or any number of other collections. In many cases, we simply label the objects from 1 to m , so that $D = \{1, 2, \dots, m\}$. In other cases (such as the card experiment), it may be more natural to label the objects with vectors. In any case, D is usually a subset of \mathbb{R}^k for some k .

Our basic experiment consists of selecting n objects from the population D at random and recording the sequence of objects chosen:

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

where $X_i \in D$ is the i^{th} object chosen. If the sampling is **with replacement**, the sample size n can be any positive integer. In this case, the sample space S is

$$S = D^n = \{(x_1, x_2, \dots, x_n) : x_i \in D \text{ for all } i \in \{1, 2, \dots, n\}\}$$

If the sampling is **without replacement**, the sample size n can be no larger than the population size m . In this case, the sample space S consists of all **permutations** of size n chosen from D :

$$S = D_n = \{(x_1, x_2, \dots, x_n) : x_i \in D \text{ for all } i \text{ and } x_i \neq x_j \text{ for all } i \neq j\}$$

1. Show that

- $\#(D^n) = m^n$
- $\#(D_n) = m^{(n)} = m(m-1) \cdots (m-n+1)$

With either type of sampling, we assume that the samples are equally likely and thus that the outcome variable \mathbf{X} is **uniformly distributed** on S ; this is the meaning of the phrase **random sample**:

$$\mathbb{P}(\mathbf{X} \in A) = \frac{\#(A)}{\#(S)}, \quad A \subseteq S$$

The Exchangeable Property

Suppose again that we select n objects at random from the population D , either with or without replacement.

2. Show that any permutation of $\mathbf{X} = (X_1, X_2, \dots, X_n)$ has the same distribution as \mathbf{X} itself, namely the uniform distribution on the appropriate sample space:
- D^n if the sampling is with replacement.
 - D_n if the sampling is without replacement.

A sequence of random variables with the property in the last exercise is said to be **exchangeable**. Although this property is very simple to understand, both intuitively and mathematically, it is nonetheless very important. We will use the exchangeable property often in this chapter.

3. More generally, show that any sequence of k of the n outcome variables is uniformly distributed on the appropriate sample space:
- D^k if the sampling is with replacement.
 - D_k if the sampling is without replacement.

In particular, for either sampling method, X_i is uniformly distributed on D for each i .

4. Show that if the sampling is with replacement, (X_1, X_2, \dots, X_n) is a sequence of **independent** random variables.

Thus, when the sampling is with replacement, the sample variables form a **random sample** from the uniform distribution, in the technical sense.

5. Show that if the sampling is without replacement, then the conditional distribution of a sequence of k of the outcome variables, given the values of a sequence of j other outcome variables, is the uniform distribution on the set of permutations of size k chosen from the population when the j known values are removed (of course, $k + j \leq n$).

In particular, X_i and X_j are dependent for any distinct i and j when the sampling is without replacement.

The Unordered Sample

In many cases, particularly when the sampling is without replacement, the *order* in which the objects are chosen is not important; all that matters is the (unordered) *set* of objects:

$$\mathbf{W} = \{X_1, X_2, \dots, X_n\}$$

Suppose first that the sampling is without replacement. In this case, \mathbf{W} takes values in the set of **combinations** of size n chosen from D :

$$T = \left\{ \{x_1, x_2, \dots, x_n\} : x_i \in D \text{ for all } i \text{ and } x_i \neq x_j \text{ for all } i \neq j \right\}$$

6. Show that $\#(T) = \binom{m}{n}$.

7. Show that \mathbf{W} is uniformly distributed over T :

$$\mathbb{P}(\mathbf{W} \in B) = \frac{\#(B)}{\#(T)} = \frac{\#(B)}{\binom{m}{n}}, \quad B \subseteq T$$

Hint: For any combination of size n from D , there are $n!$ permutations of size n .

If the sampling is with replacement, \mathbf{W} takes values in the collection of subsets of D , of size from 1 to n :

$$T = \{\{x_1, x_2, \dots, x_n\} : x_i \in D \text{ for all } i\}$$

8. Show that $\#(T) = \binom{m+n-1}{n}$.

9. Show that \mathbf{W} is *not* uniformly distributed on T .

Summary of Sampling Formulas

The following table summarizes the formulas for the number of samples of size n chosen from a population of m elements, based on the criteria of order and replacement.

Sampling Formulas

	With Order	Without Order
With Replacement	m^n	$\binom{m+n-1}{n}$
Without Replacement	$m^{(n)}$	$\binom{m}{n}$

Examples and Applications

10. Suppose that a sample of size 2 is chosen from the population $\{1, 2, 3, 4\}$. Explicitly list all samples in the following cases:

- Ordered samples, with replacement.
- Ordered samples, without replacement.
- Unordered samples, with replacement.
- Unordered samples, without replacement.



Multi-type Populations

A **dichotomous population** consists of two types of objects.

11. Suppose that a batch of 100 components includes 10 that are defective. A random sample of 5 components is selected without replacement. Compute the probability that the sample contains at least one defective component.



12. An urn contains 50 balls, 30 red and 20 green. A sample of 15 balls is chosen at random. Find the probability that the sample contains 10 red balls in each of the following cases:

- The sampling is without replacement
- The sampling is with replacement



13. In the **ball and urn experiment** select 50 balls with 30 red balls, and sample size 15. Run the experiment 100 times, updating after each run. Compute the relative frequency of the event that the sample has 10 red balls in each of the following cases, and compare with the respective probability in the previous exercise:

- The sampling is without replacement
- The sampling is with replacement

14. Suppose that a club has 100 members, 40 men and 60 women. A committee of 10 members is selected at random (and without replacement, of course).

- Find the probability that both genders are represented on the committee.
- If you observed the experiment and in fact the committee members are all of the same gender, would you believe that the sampling was random?



15. Suppose that a small pond contains 500 fish, 50 of them tagged. A fisherman catches 10 fish. Find the probability that the catch contains at least 2 tagged fish.



The basic distribution that arises from sampling without replacement from a dichotomous population is studied in the section on the [hypergeometric distribution](#). More generally, a **multi-type population** consists of objects of k different types.

16. Suppose that a legislative body consists of 60 republicans, 40 democrats, and 20 independents. A committee of 10 members is chosen at random. Find the probability that at least one party is not represented on the committee. *Hint:* Use the [inclusion-exclusion](#) law.



The basic distribution that arises from sampling without replacement from a multi-type population is studied in the section on the [multivariate hypergeometric distribution](#).

Cards

Recall that a standard **card deck** can be modeled by the product set

$$D = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, j, q, k\} \times \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$$

where the first coordinate encodes the **denomination** or **kind** (ace, 2-10, jack, queen, king) and where the second coordinate encodes the **suit** (clubs, diamonds, hearts, spades). The general **card experiment** consists of drawing n cards at random and without replacement from the deck D . Thus, the i^{th} card is $X_i = (Y_i, Z_i)$ where Y_i is the denomination and Z_i is the suit. The special case $n = 5$ is the **poker experiment** and the special case $n = 13$ is the **bridge experiment**. Note that with respect to the denominations or with respect to the suits, a deck of cards is a multi-type population as discussed above.

17. In the card experiment with $n = 5$ cards (poker), show that there are

- 311,875,200 ordered hands
- 2,598,960 unordered hands

18. In the card experiment with $n = 13$ cards (bridge), show that there are

- 3,954,242,643,911,239,680,000 ordered hands
- 635,013,559,600 unordered hands

19. In the **card experiment**, set $n = 5$. Run the simulation 5 times and on each run, list all of the (ordered) sequences of cards that would give the same unordered hand as the one you observed.

20. In the card experiment, show that

- Y_i is uniformly distributed on $\{1, 2, \dots, 10, j, k, q\}$ for each i .
- Z_i is uniformly distributed on $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$ for each i .

21. In the card experiment, show that Y_i and Z_j are independent for any i and j .

22. In the card experiment, show that (Y_1, Y_2) and (Z_1, Z_2) are dependent. Compare this result with the previous exercise.

23. Suppose that a sequence of 5 cards is dealt.

- Find the probability that the third card is a spade.
- Find the probability that the second and fourth cards are queens.
- Find the conditional probability that the second card is a heart given that the fifth card is a heart.
- Find the probability that the third card is a queen and the fourth card is a heart.



24. Run the **card experiment** 500 times, updating after each run. Compute the relative frequency corresponding to each probability in the previous exercise.

25. Find the probability that a bridge hand will contain no **honor cards** that is, no cards of denomination 10, jack, queen, king, or ace. Such a hand is called a **Yarborough**, in honor of the **second Earl of Yarborough**.



Dice

Rolling n fair, six-sided **dice** is equivalent to choosing a random sample of size n with replacement from the population $D = \{1, 2, 3, 4, 5, 6\}$. Generally, selecting a random sample of size n with replacement from $D = \{1, 2, \dots, m\}$ is equivalent to rolling n fair, m -sided dice.

26. In the game of **poker dice**, 5 standard, fair dice are thrown.
- Find the probability that all dice show the same score.
 - Find the probability that the scores are distinct.
 - Find the probability that 1 occurs twice and 6 occurs 3 times.



27. Run the **poker dice experiment** 500 times, updating after each run. Compute the relative frequency of each event in the previous exercise and compare with the corresponding probability.

The game of **poker dice** is treated in more detail in the chapter on [Games of Chance](#).

Birthdays

Supposes that we select n persons at random and record their **birthdays**. If we assume that birthdays are uniformly distributed throughout the year, and if we ignore leap years, then this experiment is equivalent to selecting a sample of size n with replacement from $D = \{1, 2, \dots, 365\}$. Similarly, we could record birth months or birth weeks.

28. Suppose that a mathematics class has 30 students.
- Find the probability that the birthdays are distinct.
 - Find the probability that there is at least one duplicate birthday.



29. In the **birthday experiment**, set $m = 365$ and $n = 30$. Run the experiment 100 times, updating after each run. Compute the relative frequency of each event in the previous exercise and compare to the corresponding probability.

The **birthday problem** is treated in more detail later in this chapter.

Balls into Cells

Suppose that we distribute n distinct **balls** into m distinct **cells** at random. This experiment also fits the basic model, where D is the population of cells and X_i is the cell containing the i^{th} ball. Sampling with replacement means that a cell may contain more than one ball; sampling without replacement means that a cell may contain at most one ball.

30. Suppose that 5 balls are distributed into 10 cells (with no restrictions).

- Find the probability that the balls are all in different cells.
- Find the probability that the balls are all in the same cell.



Coupons

Suppose that when we purchase a certain product (bubble gum, or cereal for example), we receive a **coupon** (a baseball card or small toy, for example), which is equally likely to be any one of m types. We can think of this experiment as sampling with replacement from the population of coupon types; X_i is the coupon that we receive on the i^{th} purchase.

31. Suppose that a kid's meal at a fast food restaurant comes with a toy. The toy is equally likely to be any of 5 types. Suppose that a mom buys a kid's meal for each of her 3 kids.

- Find the probability that the toys are all the same.
- Find the probability that the toys are all different..



The **coupon collector problem** is studied in more detail later in this chapter.

The Key Problem

Suppose that a person has n keys, only one of which opens a certain door. The person tries the keys at random. We will let N denote the trial number when the person finds the correct key.

32. Suppose that unsuccessful keys are discarded (the rational thing to do, of course). Show that

- $\mathbb{P}(N = i) = \frac{1}{n}$ for $i \in \{1, 2, \dots, n\}$ Thus, N has the **uniform distribution** on $\{1, 2, \dots, n\}$.
- $\mathbb{E}(N) = \frac{n+1}{2}$.
- $\text{var}(N) = \frac{n^2-1}{12}$.

33. Suppose that unsuccessful keys are not discarded (perhaps the person has had a bit too much to drink). Show that

- a. $\mathbb{P}(N = i) = \frac{1}{n} \left(\frac{n-1}{n}\right)^{i-1}$ for $i \in \mathbb{N}_+$. Thus, N has the [geometric distribution](#) on \mathbb{N}_+ .
- b. $\mathbb{E}(N) = n$.
- c. $\text{var}(N) = n(n-1)$.

Simulating a Random Samples

It's very easy to simulate a random sample of size n , with replacement from $D = \{1, 2, \dots, m\}$. Recall that the [ceiling function](#) $\lceil x \rceil$ gives the smallest integer that is at least as large as x .

34. Let $\mathbf{U} = (U_1, U_2, \dots, U_n)$ be a sequence of be a random numbers. Recall that these are independent random variables, each uniformly distributed on the interval $[0, 1]$. Show that $X_i = \lceil m U_i \rceil$ for $i \in \{1, 2, \dots, n\}$ simulates a random sample, with replacement, from D .

It's a bit harder to simulate a random sample of size n , without replacement, since we need to remove each sample value before the next draw.

35. Show that the following algorithm generates a random sample of size n , without replacement, from D .
1. For $i = 1$ to m , let $b_i = i$.
 2. For $i = 1$ to n ,
 - a. let $j = m - i + 1$
 - b. let U_i be a random number
 - c. let $J = \lceil j U_i \rceil$
 - d. let $X_i = b_j$
 - e. let $k = b_j$
 - f. let $b_j = b_J$
 - g. let $b_J = k$
 3. Return $\mathbf{X} = (X_1, X_2, \dots, X_n)$

Virtual Laboratories > 12. Finite Sampling Models > [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#)

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Key words](#) | [Feedback](#) | ©