

## 3. Probability Measure

### Definitions and Interpretations

Suppose that we have a [random experiment](#) with [sample space](#)  $S$ . Intuitively, the probability of an event is a measure of how likely the event is to occur when we run the experiment.

#### Axioms

Mathematically, a **probability measure** (or **distribution**)  $\mathbb{P}$  for a random experiment is a real-valued function, defined on the collection of events, and satisfying the following axioms:

1.  $\mathbb{P}(A) \geq 0$  for every event  $A$ .
2.  $\mathbb{P}(S) = 1$
3. If  $\{A_i : i \in I\}$  is a [countable](#), pairwise disjoint collection of events then

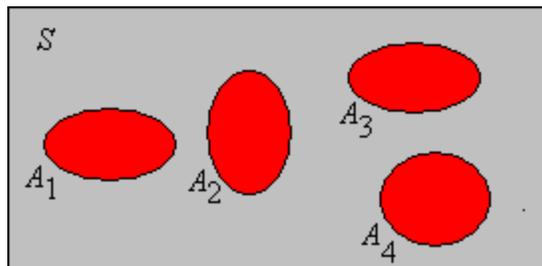
$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$$

Axiom 3 is known as **countable additivity**, and states that the probability of a union of a finite or countably infinite collection of disjoint events is the sum of the corresponding probabilities. The axioms are known as the **Kolmogorov axioms**, in honor of [Andrei Kolmogorov](#).

Axioms 1 and 2 are really just a matter of convention; we choose to measure the probability of an event with a number between 0 and 1 (as opposed, say, to a number between  $-5$  and  $7$ ). Axiom 3 however, is fundamental and inescapable. It is required for probability for precisely the same reason that it is required for other measures of the “size” of a set, such as

- cardinality for finite sets,
- length for subsets of  $\mathbb{R}$ ,
- area for subsets of  $\mathbb{R}^2$ ,
- volume for subsets of  $\mathbb{R}^3$ .

In all these cases, the size of a set that is composed of countably many disjoint pieces is the sum of the sizes of the pieces. For more on general measures, see the section on [Measure Theory](#).



On the other hand, *uncountable additivity* (the extension of axiom 3 to an uncountable index set  $I$ ) is unreasonable for probability, just as it is for other measures. For example, an interval of positive length in  $\mathbb{R}$  is a union of uncountably

many points, each of which has length 0.

We now have defined the three essential ingredients for the model a random experiment:

1. A sample space  $S$ .
2. A collection of events  $\mathcal{S}$ .
3. A probability measure  $\mathbb{P}$ .

Together these define a **probability space**  $(S, \mathcal{S}, \mathbb{P})$

### The Law of Large Numbers

Intuitively, the probability of an event is supposed to measure the long-term relative frequency of the event--in fact, this concept was taken as the definition of probability by **Richard Von Mises**. Specifically, suppose that we repeat the experiment indefinitely. (Note that this actually creates a new, compound experiment.) For an event  $A$  in the basic experiment, let  $N_n(A)$  denote the number of times  $A$  occurred (the **frequency** of  $A$ ) in the first  $n$  runs. (Note that this is a random variable in the compound experiment.) Thus,

$$P_n(A) = \frac{N_n(A)}{n}$$

is the **relative frequency** of  $A$  in the first  $n$  runs (it is also a random variable in the compound experiment). If we have chosen the correct probability measure for the experiment, then in some sense we expect that the relative frequency of each event should converge to the probability of the event:

$$P_n(A) \rightarrow \mathbb{P}(A) \text{ as } n \rightarrow \infty$$

The precise statement of this is the **law of large numbers** or **law of averages**, one of the fundamental theorems in probability. To emphasize the point, note that in general there will be lots of *possible* probability measures for an experiment, in the sense of the axioms. However, only the true probability measure will satisfy the law of large numbers.

It follows that if we have the data from  $n$  runs of the experiment, the observed relative frequency  $P_n(A)$  can be used as an approximation for  $\mathbb{P}(A)$ ; this approximation is called the **empirical probability** of  $A$ .

1. Show that  $P_n$  satisfies the **axioms of a probability measure** (given the data from  $n$  runs of the experiment)

### The Distribution of a Random Variable

Suppose that  $X$  is a **random variable** for the experiment, taking values in a set  $T$ .

2. Show that  $B \mapsto \mathbb{P}(X \in B)$  defines a probability measure on  $T$ .

*Hint:* Recall that the **inverse image** preserves all set operations.





The probability measure in the previous exercise is called the **probability distribution** of  $X$ . Thus, any random variable  $X$  for an experiment defines a new probability space:

1. A set of outcomes  $T$  (a set that includes the possible values of  $X$ ).
2. A collection of events  $\mathcal{F}$  (the admissible subsets of  $T$ ).
3. A probability measure on these events (the probability distribution of  $X$ ).

Moreover, recall that the outcome of the experiment itself can be thought of as a random variable. Specifically, if we take  $X$  to be the identity function on  $S$ , then  $X$  is a random variable and

$$\mathbb{P}(X \in A) = \mathbb{P}(A)$$

Thus, *any* probability measure can be thought of as the distribution of a random variable.

## Constructions

### Measures

How can we construct probability measures? As noted briefly above, there are other **measures** of the “size” of sets; in many cases, these can be converted into probability measures. First, a (nonnegative) **measure**  $\mu$  on  $S$  is a real-valued function defined on the collection of events  $\mathcal{S}$  that satisfies axioms 1 and 3 above. In general,  $\mu(A)$  is allowed to be infinite for a subset  $A$ . However, if  $\mu(S)$  is positive and finite, then  $\mu$  can easily be re-scaled into a probability measure.

3. Show that if  $\mu$  is a measure on  $S$  with  $0 < \mu(S) < \infty$  then  $\mathbb{P}$  defined below is a probability measure on  $S$ .

$$\mathbb{P}(A) = \frac{\mu(A)}{\mu(S)}, \quad A \in \mathcal{S}$$

In the context of [Exercise 3](#),  $\mu(S)$  is called the **normalizing constant**. In the next two subsections, we consider some very important special cases.

### Discrete Distributions

Suppose that  $S$  is a finite, nonempty set. Clearly, **counting measure**  $\#$  is a finite measure on  $S$ :

$$\#(A) = \text{the number of elements in } A, \quad A \subseteq S$$

The corresponding probability measure is called the **discrete uniform distribution** on  $S$ , and is particularly important in combinatorial and sampling experiments:

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

We can give a more general construction for countable sample spaces that can be used to define many probability

measures.

4. Suppose that  $S$  is nonempty and countable, and that  $g$  is a nonnegative real-valued function defined on  $S$ . Show that  $\mu$  defined below is a measure on  $S$ :

$$\mu(A) = \sum_{x \in A} g(x), \quad A \subseteq S$$

Thus, if  $0 < \mu(S) < \infty$  then  $\mathbb{P}(A) = \frac{\mu(A)}{\mu(S)}$  defines a probability measure by [Exercise 3](#). Distributions of this type are said to be **discrete**. [Discrete distributions](#) are studied in detail in the chapter on [Distributions](#).

5. In the setting of previous exercise, show that if  $S$  is finite and  $g$  is a constant function, then the corresponding probability measure  $\mathbb{P}$  is the discrete uniform distribution on  $S$ .

## Continuous Distributions

We define the standard  **$n$ -dimensional measure**  $\lambda_n$  on  $\mathbb{R}^n$  (called **Lebesgue measure**, in honor of [Henri Lebesgue](#)) by

$$\lambda_n(A) = \int_A 1 dx$$

Technically, the integral is more general than the one defined in calculus. However, the standard calculus integral will suffice for most of this project. In particular, we assume that set  $A$  is nice enough that the integral exists; see the section on [Measurability](#) for more details. Note that if  $n > 1$ , the integral above is a multiple integral;  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$ . The countable additivity axiom holds because of a basic property of integrals. In particular, note from calculus that

1.  $\lambda_1(A)$  is the length of  $A$  for  $A \subseteq \mathbb{R}$ .
2.  $\lambda_2(A)$  is the area of  $A$  for  $A \subseteq \mathbb{R}^2$ .
3.  $\lambda_3(A)$  is the volume of  $A$  for  $A \subseteq \mathbb{R}^3$ .

Now, if  $S \subseteq \mathbb{R}^n$  with  $0 < \lambda_n(S) < \infty$ , then

$$\mathbb{P}(A) = \frac{\lambda_n(A)}{\lambda_n(S)}, \quad A \subseteq S$$

is a probability measure on  $S$  by [Exercise 3](#), called the **continuous uniform distribution** on  $S$ .

We can generalize this construction to produce many other distributions. Suppose that  $g$  is a nonnegative real valued function defined on  $S$ . Define

$$\mu(A) = \int_A g(x) dx, \quad A \subseteq S$$

Then  $\mu$  is a measure on  $S$ . Thus if  $0 < \mu(S) < \infty$ , then  $\mathbb{P}(A) = \frac{\mu(A)}{\mu(S)}$  defines a probability measure as in [Exercise 3](#).

Distributions of this type are said to be **continuous**. [Continuous distributions](#) are studied in detail in the chapter on [Distributions](#).

It is important to note again that, unlike many other areas of mathematics, the low-dimensional spaces ( $n \in \{1, 2, 3\}$ ) do not play a special role, except for exposition. For example in the [Cicada data](#), some of the variables recorded are body

weight, body length, wing width, and wing length. A probability model for these variables would specify a distribution on a subset of  $\mathbb{R}^4$ .

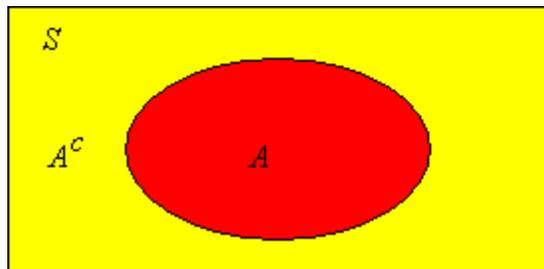
## Rules of Probability

### Basic Rules

Suppose that we have a random experiment with sample space  $S$  and probability measure  $\mathbb{P}$ . In the following exercises,  $A$  and  $B$  are events. Prove the results using the [axioms of probability](#).

6. Show that  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .

*Hint:*  $A$  and  $A^c$  are disjoint and their union is  $S$ .

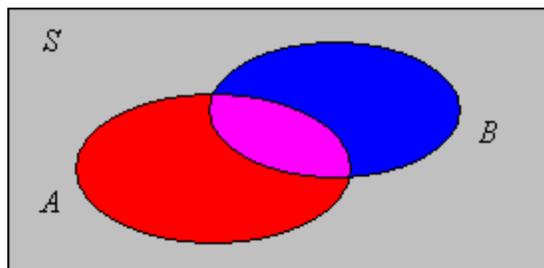


7. Show that  $\mathbb{P}(\emptyset) = 0$ .

*Hint:* Apply the [complement rule](#) to  $A = S$ .

8. Show that  $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

*Hint:*  $A \cap B$  and  $B \setminus A$  are disjoint and their union is  $B$ .

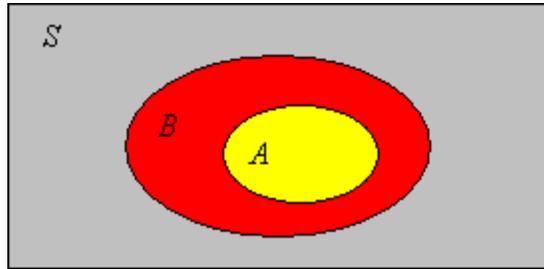


9. Show that if  $A \subseteq B$  then  $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ .

*Hint:* Apply the [difference rule](#) and note that  $A \cap B = A$

10. Show that if  $A \subseteq B$  then  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

Thus,  $\mathbb{P}$  is an **increasing function**, relative to the subset partial order on the collection of events, and the ordinary order on  $\mathbb{R}$ . In particular, it follows that  $\mathbb{P}(A) \leq 1$  for any event  $A$ .



11. Suppose that  $A \subseteq B$ . Show that

- If  $\mathbb{P}(B) = 0$  then  $\mathbb{P}(A) = 0$ .
- If  $\mathbb{P}(A) = 1$  then  $\mathbb{P}(B) = 1$ .

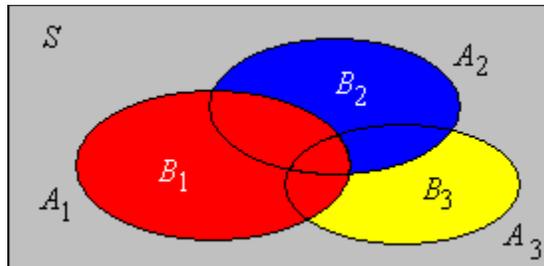
### Boole's Inequality

12. Suppose that  $\{A_i : i \in I\}$  is a countable collection of events. Prove **Boole's inequality** (named after **George Boole**):

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbb{P}(A_i)$$

- Argue that without loss of generality, we can assume that  $I = \{1, 2, \dots\}$ .
- Define  $B_1 = A_1$  and  $B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1})$  for  $n \in \{2, 3, \dots\}$ . Show that  $\{B_1, B_2, \dots\}$  are pairwise disjoint and have the same union as  $\{A_1, A_2, \dots\}$ .
- Use the **additivity axiom**
- Finally, use the **increasing property**.

Intuitively, Boole's inequality holds because parts of the union have been measured more than once in the expression on the right.



13. Suppose that  $\{A_i : i \in I\}$  is a countable collection of events with  $\mathbb{P}(A_i) = 0$  for  $i \in I$ . Use **Boole's inequality** to show that

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = 0$$

An event  $A$  with  $\mathbb{P}(A) = 0$  is said to be **null**. Thus, a countable union of null events is still a null event.

### Bonferroni's Inequality

14. Suppose that  $\{A_i : i \in I\}$  is a countable collection of events. Prove **Bonferroni's inequality** (named after **Carlo Bonferroni**):

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) \geq 1 - \sum_{i \in I} (1 - \mathbb{P}(A_i))$$

*Hint:* Apply **Boole's inequality** to  $\{A_i^c : i \in I\}$

15. Suppose that  $\{A_i : i \in I\}$  is a countable collection of events with  $\mathbb{P}(A_i) = 1$  for each  $i \in I$ . Use **Bonferroni's inequality** to show that

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = 1$$

An event  $A$  with  $\mathbb{P}(A) = 1$  is sometimes called **almost sure** or **almost certain**. Thus, a countable intersection of almost sure events is still almost sure.

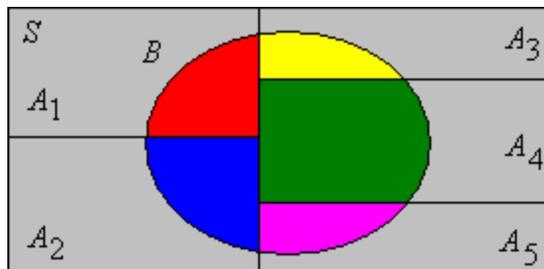
16. Suppose that  $A$  and  $B$  are events in an experiment. Prove the following:

- a. If  $\mathbb{P}(A) = 0$ , then  $\mathbb{P}(A \cup B) = \mathbb{P}(B)$ .
- b. If  $\mathbb{P}(A) = 1$ , then  $\mathbb{P}(A \cap B) = \mathbb{P}(B)$ .

### Computing Probabilities via a Partition

17. Suppose that  $\{A_i : i \in I\}$  is a countable collection of events that **partition** the sample space  $S$ . Show that for any event  $B$ ,

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i \cap B)$$



Naturally, this result is useful when the probabilities of the intersections are known. Partitions usually arise in connection with a random variable. Suppose that  $X$  is a random variable taking values in a countable set  $T$ , and that  $B$  is an event. Then

$$\mathbb{P}(B) = \sum_{x \in T} \mathbb{P}(X = x, B)$$

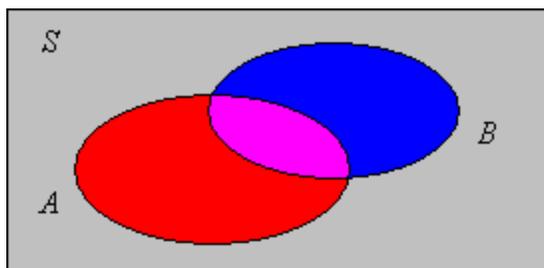
In this formula, note that the comma acts like the intersection symbol in the previous formula.

### The Inclusion-Exclusion Formula

The inclusion-exclusion formulas provide a method for computing the probability of a union of events in terms of the probabilities of the intersections of the events.

18. Show that  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$  for any events  $A$  and  $B$ .

- Note first that  $A \cup B = A \cup (B \setminus A)$  and the latter two sets are disjoint.
- Now use the [additivity axiom](#) and the [difference rule](#).



19. Show that  $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$  for any events  $A$ ,  $B$ , and  $C$ .

*Hint:* Use the [inclusion-exclusion rule for two events](#). You will use this rule three times.



The last two exercises can be generalized to a union of  $n$  events; the generalization is known as the [inclusion-exclusion formula](#).

20. Suppose that  $A_i$  is an event for each  $i \in I$  where  $\#(I) = n$ . Show that

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{(J \subseteq I) \text{ and } (\#(J)=k)} \mathbb{P}\left(\bigcap_{j \in J} A_j\right)$$

*Hint:* Use induction on  $n$ .

The [general Bonferroni inequalities](#) state that if sum on the right in Exercise 20 is truncated after  $k$  terms ( $k < n$ ) then the truncated sum is an *upper* bound for the probability of the union if  $k$  is odd (so that the last term has a *positive* sign) and is a *lower* bound for the probability of the union if  $k$  is even (so that the last terms has a *negative* sign).

If you go back and look at your proofs of the basic properties in Exercises 6-20, you will see that they hold for any finite measure  $\mu$ , not just probability. The only change is that the number 1 is replaced by  $\mu(S)$ . In particular, the inclusion-exclusion rule is as important in [combinatorics](#) (the study of counting measure) as it is in probability.

## Equivalence

Intuitively, equivalent events or random variables are those that are indistinguishable from a probabilistic point of view. The purpose of this subsection is to make this idea precise.

Events  $A$  and  $B$  in a random experiment are said to be **equivalent** if the probability of the symmetric difference is 0:

$$\mathbb{P}((A \setminus B) \cup (B \setminus A)) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) = 0$$

21. Show that equivalence really is an **equivalence relation** on the collection of events of a random experiment. Thus, the collection of events is partitioned into disjoint classes of mutually equivalent events.

- $A$  is equivalent to  $A$  for any event  $A$  (the **reflexive property**).
- If  $A$  is equivalent to  $B$  then  $B$  is equivalent to  $A$  (the **symmetric property**).
- If  $A$  is equivalent to  $B$  and  $B$  is equivalent to  $C$  then  $A$  is equivalent to  $C$  (the **transitive property**).

22. Show that equivalent events have the same probability: if  $A$  and  $B$  are equivalent then  $\mathbb{P}(A) = \mathbb{P}(B)$

23. The converse fails with a passion. Consider the simple experiment of tossing a fair coin. Show that the event that the coin lands heads and the event that the coin lands tails have the same probability, but are not equivalent.

24. However, the null and almost sure events do form equivalence classes.

- Suppose that  $A$  is an event with  $\mathbb{P}(A) = 0$ . Show that  $B$  is equivalent to  $A$  if and only if  $\mathbb{P}(B) = 0$ .
- Suppose that  $A$  is an event with  $\mathbb{P}(A) = 1$ . Show that  $B$  is equivalent to  $A$  if and only if  $\mathbb{P}(B) = 1$ .

Now suppose that  $X$  and  $Y$  are random variables for an experiment, each taking values in a set  $T$ . Then  $X$  and  $Y$  are said to be **equivalent** if

$$\mathbb{P}(X = Y) = 1$$

25. Show that equivalence really is an **equivalence relation** on the collection of random variables that take values in  $T$ . Thus, the collection of such random variables is partitioned into disjoint classes of mutually equivalent variables.

- $X$  is equivalent to  $X$  for any random variable  $X$  (the **reflexive property**).
- If  $X$  is equivalent to  $Y$  then  $Y$  is equivalent to  $X$  (the **symmetric property**).
- If  $X$  is equivalent to  $Y$  and  $Y$  is equivalent to  $Z$  then  $X$  is equivalent to  $Z$  (the **transitive property**).

26. Suppose that  $X$  and  $Y$  are equivalent random variables, taking values in a set  $T$ . Show that for any  $B \subseteq T$ , the events  $\{X \in B\}$  and  $\{Y \in B\}$  are equivalent. Conclude that  $X$  and  $Y$  have the same distribution.

27. Suppose that  $A$  and  $B$  are events for a random experiment. Show that  $A$  and  $B$  are equivalent if and only if the indicator random variables  $\mathbf{1}(A)$  and  $\mathbf{1}(B)$  are equivalent.

28. Suppose that  $X$  and  $Y$  are equivalent random variables, taking values in a set  $T$ , and that  $g$  is a function from  $T$  into a set  $U$ . Show that  $g(X)$  and  $g(Y)$  are equivalent.

## Examples and Applications

29. Suppose that  $A$  and  $B$  are events in an experiment with  $\mathbb{P}(A) = \frac{1}{3}$ ,  $\mathbb{P}(B) = \frac{1}{4}$ ,  $\mathbb{P}(A \cap B) = \frac{1}{10}$ . Express each of the following events in the language of the experiment and find its probability:

- $A \setminus B$
- $A \cup B$
- $A^c \cup B^c$
- $A^c \cap B^c$
- $A \cup B^c$



30. Suppose that  $A$ ,  $B$ , and  $C$  are events in an experiment with  $\mathbb{P}(A) = 0.3$ ,  $\mathbb{P}(B) = 0.2$ ,  $\mathbb{P}(C) = 0.4$ ,  $\mathbb{P}(A \cap B) = 0.04$ ,  $\mathbb{P}(A \cap C) = 0.1$ ,  $\mathbb{P}(B \cap C) = 0.1$ ,  $\mathbb{P}(A \cap B \cap C) = 0.01$ . Express each of the following events in set notation and find its probability:

- At least one of the three events occurs.
- None of the three events occurs.
- Exactly one of the three events occurs.
- Exactly two of the three events occur.



31. Suppose that  $A$  and  $B$  are events in an experiment with  $\mathbb{P}(A \setminus B) = \frac{1}{6}$ ,  $\mathbb{P}(B \setminus A) = \frac{1}{4}$ , and  $\mathbb{P}(A \cap B) = \frac{1}{12}$ . Find the probability of each of the following events:

- $A$
- $B$
- $A \cup B$
- $A^c \cup B^c$
- $A^c \cap B^c$



32. Suppose that  $A$  and  $B$  are events in an experiment with  $\mathbb{P}(A) = \frac{2}{5}$ ,  $\mathbb{P}(A \cup B) = \frac{7}{10}$ , and  $\mathbb{P}(A \cap B) = \frac{1}{5}$ . Find the probability of each of the following events:

- $B$
- $A \setminus B$
- $B \setminus A$
- $A^c \cup B^c$
- $A^c \cap B^c$



## Coins and Dice

33. Recall that the **coin experiment** consists of tossing  $n$  coins and recording the sequence of scores  $X = (X_1, X_2, \dots, X_n)$  (where 1 denotes heads and 0 denotes tails). Let  $Y$  denote the number of heads.
- Note that the sample space  $S$  of the experiment is  $\{0, 1\}^n$
  - Argue that if the coins are **fair** then  $X$  has the uniform distribution on  $S$ .
  - Show that if the coins are fair then for  $\mathbb{P}(Y = k) = \binom{n}{k} \left(\frac{1}{2}\right)^n$  for  $k \in \{0, 1, \dots, n\}$ ,
  - Explicitly compute the probabilities in part (c) with 5 fair dice.



The experiment in the previous exercise is a special case of **Bernoulli trials**, named for **Jacob Bernoulli**. The number of heads  $Y$  has a **binomial distribution**

34. Consider the coin experiment with 3 fair coins. Let  $A$  be the event that the first coin is heads and  $B$  the event that there are exactly 2 heads. Find each of the following probabilities:
- $\mathbb{P}(A)$ .
  - $\mathbb{P}(B)$ .
  - $\mathbb{P}(A \cap B)$ .
  - $\mathbb{P}(A \cup B)$ .
  - $\mathbb{P}(A^c \cup B^c)$ .
  - $\mathbb{P}(A^c \cap B^c)$ .
  - $\mathbb{P}(A \cup B^c)$ .



35. In the **Coin experiment**, select 3 coins. Run the experiment 1000 times, updating after every run, and compute the empirical probability of each event in the previous exercise.

36. Recall that the **dice experiment** consists of throwing  $n$  distinct,  $k$ -sided dice (with sides numbered from 1 to  $k$ ) and recording the sequence of scores  $X = (X_1, X_2, \dots, X_n)$ . Recall that this experiment serves as a generic example of multinomial trials and as a generic example of sampling with replacement from a finite population. The special case  $k = 6$  corresponds to **standard dice**.
- Note that the sample space  $S$  of the experiment is  $\{1, 2, \dots, k\}^n$  and that  $\#(S) = k^n$
  - Argue that if the dice are **fair** then  $X$  has the uniform distribution on  $S$ .

37. Suppose that 2 fair, standard dice are rolled and the sequence of scores  $X = (X_1, X_2)$  recorded. Let  $A$  denote the event that the first die score is less than 3 and  $B$  the event that the sum of the dice scores is 6.
- Express  $A$  in terms of the basic outcome variables.
  - Express  $B$  in terms of the basic outcome variables.
  - Find  $\mathbb{P}(A)$ .
  - Find  $\mathbb{P}(B)$ .

- e. Find  $\mathbb{P}(A \cap B)$ .
- f. Find  $\mathbb{P}(A \cup B)$ .
- g. Find  $\mathbb{P}(B \setminus A)$ .



38. In the **dice experiment**, set  $n = 2$ . Run the experiment 100 times and compute the empirical probability of each event in the previous exercise.

39. Suppose that 2 fair, standard dice are rolled and the sequence of scores  $X = (X_1, X_2)$  recorded. Let  $Y$  denote the sum of the scores,  $U$  the minimum score, and  $V$  the maximum score.

- a. Find  $\mathbb{P}(Y = y)$  for each  $y \in \{2, 3, \dots, 12\}$ .
- b. Find  $\mathbb{P}(U = u)$  for each  $u \in \{1, 2, \dots, 6\}$ .
- c. Find  $\mathbb{P}(V = v)$  for each  $v \in \{1, 2, \dots, 6\}$ .
- d. Find  $\mathbb{P}(U = u, V = v)$  for each  $(u, v) \in \{1, 2, \dots, 6\}^2$
- e. Note that  $(U, V)$  could serve as the outcome vector for the experiment of rolling two standard, fair dice if we do not bother to distinguish the dice. Note that this random vector does *not* have a uniform distribution.



40. A pair of fair, standard dice are thrown repeatedly until the sum of the scores is either 5 or 7. Let  $A$  denote the event that the sum of the scores on the last throw is 5 rather than 7. Events of this type are important in the game of **craps**.

- a. Suppose that we record the pair of scores on each throw. Define an appropriate sample space.
- b. Represent  $A$  in terms of the sample space in (a).
- c. Compute the probability of  $A$  in the setting of parts (a) and (b).
- d. Now suppose that we just record the pair of scores on the last throw. Define an appropriate sample space.
- e. Argue that the sample space in part (d) should be given the uniform distribution.
- f. Compute the probability of  $A$  in the setting of parts (d) and (e).



## 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). Sometimes we represent a card as a *string* rather than an ordered pair (for example  $q\heartsuit$ ).

41. Recall that the **card experiment** consists of dealing  $n$  cards from a well-shuffled deck and recording the sequence of cards  $X = (X_1, X_2, \dots, X_n)$ , where  $X_i \in D$  is the  $i^{\text{th}}$  card. Let  $W = \{X_1, X_2, \dots, X_n\}$  denote the unordered set of cards in the hand.

- Note that the sample space  $S$  consists of all permutations of size  $n$  from  $D$  and that  $\#(S) = 52^{(n)}$
- Argue that if the deck is well-shuffled, then  $X$  is uniformly distributed on  $S$ .
- Note that  $W$  takes values in the set  $T$  of all combinations of size 5 chosen from  $D$ , and that  $\#(T) = \binom{52}{n}$
- Show that  $W$  is uniformly distributed on  $T$ .

Recall also that the special case  $n = 5$  is the **poker experiment** and the special case  $n = 13$  is the **bridge experiment**. The **poker experiment** is studied in more detail in the chapter on **Games of Chance**.

42. Consider the card experiment with  $n = 2$  cards. For  $i \in \{1, 2\}$ , let  $H_i$  denote the event that card  $i$  is a heart.

- Find  $\mathbb{P}(H_1)$
- Find  $\mathbb{P}(H_1 \cap H_2)$
- Find  $\mathbb{P}(H_2 \setminus H_1)$
- Find  $\mathbb{P}(H_2)$
- Find  $\mathbb{P}(H_1 \cup H_2)$



43. In the **card experiment**, set  $n = 2$ . Run the experiment 100 times and compute the empirical probability of each event in the previous exercise

44. In the poker experiment, find the probability of each of the following events:

- The hand is a **full house** (3 cards of one kind and 2 cards of another kind).
- The hand has **four of a kind** (4 cards of one kind and 1 of another kind).
- The cards are all in the same suit (thus, the hand is either a **flush** or a **straight flush**).



45. Run the **poker experiment** 10000 times, updating every 10 runs. Compute the empirical probability of each event in the previous problem.

46. 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**.



47. Find the probability that a bridge hand will contain

- Exactly 4 hearts.
- Exactly 4 hearts and 3 spades.
- Exactly 4 hearts, 3 spades, and 2 clubs.



48. A card hand that contains no cards in a particular suit is said to be **void** in that suit. Use the **inclusion-exclusion rule** to find the probability of each of the following events:

- a. A poker hand is void in at least one suit.
- b. A bridge hand is void in at least one suit.



### Buffon's Coin Experiment

49. Recall that in **Buffon's coin experiment**, a coin with radius  $r \leq \frac{1}{2}$  is tossed "randomly" on a floor with square tiles of side length 1, and the coordinates  $(X, Y)$  of the center of the coin are recorded, relative to the center of the square in which the coin lands. Let  $A$  denote the event that the coin does not touch the sides of the square, and let  $Z$  denote the random variable that gives the distance from the coin center to the origin.

- a. Define the sample space  $S$  mathematically.
- b. Argue that  $(X, Y)$  is uniformly distributed on  $S$ .
- c. Express  $A$  in terms of the outcome variables  $(X, Y)$ .
- d. Find  $\mathbb{P}(A)$
- e. Find  $\mathbb{P}(A^c)$
- f. Find  $\mathbb{P}\left(Z < \frac{1}{2}\right)$



50. In **Buffon's coin experiment**, set  $r = 0.2$ . Run the experiment 100 times and compute the empirical probability of each event in the previous exercise.

### Urn Models

51. Recall the basic **urn model** in which an urn contains  $m$  balls with distinct labels. A random sample of  $n$  balls is chosen from the urn without replacement, and the sequence of ball labels  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is recorded. Let  $\mathbf{W} = \{X_1, X_2, \dots, X_n\}$  denote the unordered sample. Recall that this experiment serves as a generic example of sampling without replacement from a finite population.

- a. Note that the sample space  $S$  consists of all permutations of size  $n$  from the population of  $m$  balls, and that  $\#(S) = m^{(n)}$
- b. Argue that  $\mathbf{X}$  has the uniform distribution on  $S$ .
- c. Note that  $\mathbf{W}$  takes values in the set  $T$  of all combinations of size  $n$  chosen from  $D$ , and that  $\#(T) = \binom{m}{n}$
- d. Show that  $\mathbf{W}$  is uniformly distributed on  $T$ .

52. Suppose that an urn contains  $m$  balls;  $r$  are red and the remaining  $m - r$  are green. A random sample of  $n$  balls is chosen without replacement. Let  $Y$  denote the number of red balls in the sample. This urn model serves as a generic example of sampling without replacement from a finite, **dichotomous population**. Random variable  $Y$  has a **hypergeometric distribution**. Determine the set of possible values of  $Y$  and show that

$$\mathbb{P}(Y = k) = \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} \text{ for } k \in \{0, 1, \dots, n\}$$

53. Consider an urn with 30 balls; 10 are red and 20 are green. A sample of 5 balls is chosen at random. Explicitly compute the probabilities in the last exercise.



54. In the simulation of the **ball and urn experiment**, select 30 balls with 10 red and 20 green, and select sample size 5. Run the experiment 1000 times and compare the empirical probabilities with the ones that you computed in the previous exercise.

55. An urn contains 12 balls: 5 are red, 4 are green, and 3 are blue. Three balls are chosen at random, without replacement.

- Define a sample space for which the outcomes are equally likely.
- Find the probability that the chosen balls are all the same color.
- Find the probability that the chosen balls are all different colors.



56. Repeat the last exercise under the assumption that the balls are chosen with replacement.

## Genetics

First, let's consider an overly simplified model of an **inherited trait** that has two possible states, for example a pea plant whose pods are either green or yellow. A plant has two genes for the trait (one from each parent), so the possible **genotypes** are

- $gg$ , a gene for green pods from each parent.
- $gy$ , a gene for green pods from one parent and a gene for yellow pods from the other (we usually cannot observe which parent contributed which gene).
- $yy$ , a gene for yellow pods from each parent.

The genotypes  $gg$  and  $yy$  are called **homozygotes**, while the genotype  $gy$  is called **heterozygote**. Typically, one of the states of the inherited trait is **dominant** and the other **recessive**. Thus, for example, if green is the dominant state for pod color, then a plant with genotype  $gg$  or  $gy$  has green pods, while a plant with genotype  $yy$  has yellow pods. Finally, the gene passed from a parent to a child is randomly selected from the parent's two genes. The inheritance of pea pod color was studied by **Gregor Mendel**, the father of modern genetics.

57. Let  $A$  be the event that a child plant has genotype  $gg$ ,  $B$  the event that the plant has genotype  $gy$ , and  $C$  the event that the plant has genotype  $yy$ . Find  $\mathbb{P}(A)$ ,  $\mathbb{P}(B)$ , and  $\mathbb{P}(C)$  in each of the following cases:

- Both parents are type  $gg$ .
- Both parents are type  $yy$ .
- Both parents are type  $gy$ .
- One parent is type  $gg$  and the other is type  $yy$ .

- e. One parent is type  $gg$  and the other is type  $gy$ .
- f. One parent is type  $yy$  and the other is type  $gy$ .



A **sex-linked** hereditary disorder is a disorder due to a defect on the **X chromosome** (one of the two chromosomes that determine gender). Suppose that  $n$  denotes the normal gene and  $d$  the defective gene linked to the disorder. Women have two X chromosomes, and  $d$  is **recessive**. Thus, a woman with genotype  $nn$  is completely normal with respect to the condition; a woman with genotype  $nd$  does not have the disorder, but is a **carrier**, since she can pass the defective gene to her children; and a woman with genotype  $dd$  has the disorder. A man has only one X chromosome (his other sex chromosome, the **Y chromosome**, typically plays no role in the disorder). A man with genotype  $n$  is normal and a man with genotype  $d$  has the disorder. Examples of sex-linked hereditary disorders are **dichromatism**, the most common form of color-blindness, and the most common form of **hemophilia**, a bleeding disorder. The following exercise explore the transmission of a sex-linked hereditary disorder.

58. Let  $B$  be the event that a son has the disorder,  $C$  the event that a daughter is a carrier, and  $D$  the event that a daughter has the disease. Find  $\mathbb{P}(B)$ ,  $\mathbb{P}(C)$  and  $\mathbb{P}(D)$  in each of the following cases:

- a. The mother and father are normal.
- b. The mother is a carrier and the father is normal.
- c. The mother is normal and the father has the disorder.
- d. The mother is a carrier and the father has the disorder.
- e. The mother has the disorder and the father is normal.
- f. The mother and father both have the disorder.



From this exercise, note that transmission of the disorder to a daughter can only occur if the mother is at least a carrier and the father has the disorder. In ordinary large populations, this is a unusual intersection of events, and thus sex-linked hereditary disorders are typically much less common in women than in men. In brief, women are protected by the extra X chromosome.

### Radioactive Emissions

59. Suppose that  $T$  denotes the time between emissions (in milliseconds) for a certain type of radioactive material, and that  $T$  has the following **exponential distribution**:

$$\mathbb{P}(T \in A) = \int_A e^{-t} dt, \quad A \subseteq (0, \infty)$$

- a. Show that this really does define a probability distribution.
- b. Find  $\mathbb{P}(T > 3)$
- c. Find  $\mathbb{P}(2 < T < 4)$



60. Suppose that  $N$  denotes the number of emissions in a one millisecond interval for a certain type of radioactive material, and that  $N$  has the following **Poisson distribution**:

$$\mathbb{P}(N \in A) = \sum_{n \in A} \frac{e^{-1}}{n!}, \quad A \subseteq \mathbb{N}$$

- Show that this really does define a probability distribution.
- Find  $\mathbb{P}(N \geq 3)$
- Find  $\mathbb{P}(2 \leq N \leq 4)$



## Matching

Suppose that a secretary prepares  $n$  letters and corresponding envelopes to send to  $n$  different people, but then stuffs the letters in the envelopes randomly. We are interested in the event  $M$  that at least one letter is inserted into the proper envelope.

61. Use the steps below and the inclusion-exclusion rule to show that

$$\mathbb{P}(M) = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}$$

- Let  $M_i$  denote the event that letter  $i$  is in envelope  $i$ . Show that  $M = \bigcup_{i=1}^n M_i$
- Suppose that  $J \subseteq \{1, 2, \dots, n\}$  and that  $\#(J) = k$ . Show that  $\mathbb{P}\left(\bigcap_{j \in J} M_j\right) = \frac{(n-k)!}{n!}$

62. Use the previous result to show that

$$\mathbb{P}(M) \rightarrow 1 - e^{-1} \text{ as } n \rightarrow \infty$$

63. Explicitly compute  $\mathbb{P}(M)$  when  $n = 5$  and  $n = 10$ . Compare with the limiting probability in the previous exercise.



A complete analysis of the [matching experiment](#) is given in the chapter on [Finite Sampling Models](#).

## Data Analysis Exercise

64. For the [M&M data set](#), let  $R$  denote the event that a bag has at least 10 red candies,  $T$  the event that a bag has at least 57 candies total, and  $W$  the event that a bag weighs at least 50 grams. Find the empirical probability the following events:

- $R$
- $T$
- $W$
- $R \cap T$
- $T \setminus W$



65. For the **cicada data**, let  $W$  denote the event that a cicada weighs at least 0.20 grams,  $F$  the event that a cicada is female, and  $T$  the event that a cicada is type tredecula. Find the empirical probability of each of the following:

- a.  $W$
- b.  $F$
- c.  $T$
- d.  $W \cap F$
- e.  $F \cup T \cup W$

