Probability and Inference

Revisions

You can find the latest versions of the slides in their GitHub repo. If you have any suggestions, feel free to open an issue ticket.

Section 1

Basics

Supplemental Readings for Section 1

The following readings expand on the ideas in these notes. FPP provides a nice conceptual discussion. Aronow and Miller give a more technical presentation, but remains gentle and compact. DeGroot and Schervish offer a thorough introduction.

- FPP, chs. 13 and 14 (pp. 221-224)
- Aronow and Miller, ch. 1, section 1.1 (pp. 3-15)
- DeGroot and Schervish, ch. 1, sections 1.1-1.7, 1.10, 2.1-2.3.

Definition 1

*An **experiment** is a repeatable procedure to obtain an observation from a defined set of outcomes.*

Definition 2

*The **sample space** S is the collection of all possible outcomes of the experiment.*

Definition 3

*A **realization** of the experiment produces an outcome from the sample space.*

Definition 4

*An **event** A is a subset of the sample space.*

Axioms of Probability (Kolmogorov Axioms)

Axiom 1

For every event A, $\Pr(A) \geq 0$.

Axiom 2

$\Pr(S) = 1$.

Axiom 3

For every infinite sequence of disjoint events $A_1, A_2, \ldots,$

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i).$$

Notes on Axiom 3

Examples of an infinite sequence of disjoint events? For $S = \mathbb{R}^+$? For $S = \{0, 1\}$?

An infinite sequence of disjoint events is difficult to conceptualize. For $S = \mathbb{R}^+$, *one* such sequence would be $[0, 1), [1, 2), [2, 3), \ldots.$ For $S = \{0, 1\}$, *one* such sequence would be $\{0\}, \{1\}, \emptyset, \emptyset, \emptyset, \ldots.$

Definition 5

For a sample space S, a ***probability*** *is a collection of real numbers assigned to all events A consistent with Axioms 1, 2, and 3*

Interpretations of "Probability"

- We can interpret this probability of A as the fraction of time that A occurs in this long-run when we repeat the experiment over-and-over. This is known as the **frequentist** interpretation of probability.
- We can also interpret the probability as your beliefs about A. This is known as the **Bayesian** interpretation.

Coin Toss example.

Theorem 1

$\Pr(\emptyset) = 0$.

Exercise 1

Prove Theorem 1. Hint: Use Axiom 3.

Solution

$\Pr(\emptyset) = \Pr(\cup_{i=1}^{\infty}\emptyset) = \sum_{i=1}^{\infty} \Pr(\emptyset)$. $\Pr(\emptyset) = \sum_{i=1}^{\infty} \Pr(\emptyset)$ iff $\Pr(\emptyset) = 0$.

Theorem 2

For every finite sequence of n disjoint events $A_1, A_2, ..., A_n$,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i).$$

Corollary 1 (Addition Rule for Two Disjoint Events)

For disjoint events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Proof.

This follows directly from Theorem 2. $\square$

Theorem 3

If events $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$.

Exercise 2

Prove Theorem 3. Hint: Notice that $B = A \cup (B \cap A^c)$. Then use the Additional Rule for Two Disjoint Events (Corollary 1).

Solution

$B = A \cup (B \cap A^c)$. By the addition rule for disjoint events, $\Pr(B) = \Pr(A) + \Pr(B \cap A^c)$. By Axiom 1, $\Pr(B \cap A^c) \geq 0$, so $\Pr(A) \leq \Pr(B)$.

Theorem 4

For event A, $0 \leq \Pr(A) \leq 1$.

Exercise 3

Prove Theorem 4. Hint: Axiom 1 established that $0 \leq \Pr(A)$. Now show that $\Pr(A) \leq 1$. To do this, use Axiom 2 and Theorem 3.

Solution

$A \subseteq S$. $\Pr(A) \leq \Pr(S)$. By Axiom 1, $\Pr(S) = 1$, so $\Pr(A) \leq 1$.

Theorem 5 (Addition Rule for Two Events)

For any events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

Theorem 6 (Addition Rule for Three Events)

For any events A, B, and C,

$$\begin{aligned}\Pr(A \cup B \cup C) &= \Pr(A) + \Pr(B) + \Pr(C)\\ &\quad - [\Pr(A \cap B) + \Pr(A \cap C) + \Pr(B \cap C)]\\ &\quad + \Pr(A \cap B \cap C).\end{aligned}$$

Exercise 4

If they answer honestly and accurately, a randomly selected survey respondent will report voting in the 2016 presidential election with probability 0.6. What's the probability that a randomly selected survey respondent will report not voting. Make sure to connect your answer to the results above. Hint: show that $\Pr(A^c) = 1 - \Pr(A)$. We haven't established this simple, intuitive result. Then use this result to answer the question.

Solution

$\Pr(S) = \Pr(A) + \Pr(A^c)$. Then $1 = \Pr(A) + \Pr(A^c)$ and $\Pr(A^c) = 1 - \Pr(A)$. Then $\Pr(\text{non-voter}) = 0.4$.

Exercise 5

Suppose events A and B, where $\Pr(A) = 0.5$ and $\Pr(B) = 0.8$. Without more information, you can't figure out $\Pr(A \cap B)$, but you can bound it. What is the largest possible value of $\Pr(A \cap B)$? What's the smallest?

Solution

Remember that $\Pr(A \cap B) = \Pr(A) \Pr(B \mid A)$. Because A is smaller than B, it could be that all of A is inside B, in which case $\Pr(B \mid A) = 1$ and $\Pr(A \cap B) = \Pr(A) = 0.5$. Or it could be that $\Pr(B \mid A^c) = 1$, in which case $\Pr(B \cap A^c) = \Pr(B \mid A^c) \Pr(A^c) = 1 \times 0.5 = 0.5$. Then $\Pr(B) = \Pr(B \cap A^c) + \Pr(B \cap A) = 0.8$, according to the Addition Rule for Disjoint Events (that's a "partition, see below). So we have $.5 + \Pr(B \cap A) = 0.8$ or $\Pr(B \cap A) = 0.3$.

Definition 6 (Conditional Probability)

$\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$ *for* $\Pr(B) > 0$. *If* $\Pr(B) = 0$, *then* $\Pr(A \mid B)$ *is undefined.*

A Note on Conditional Probability

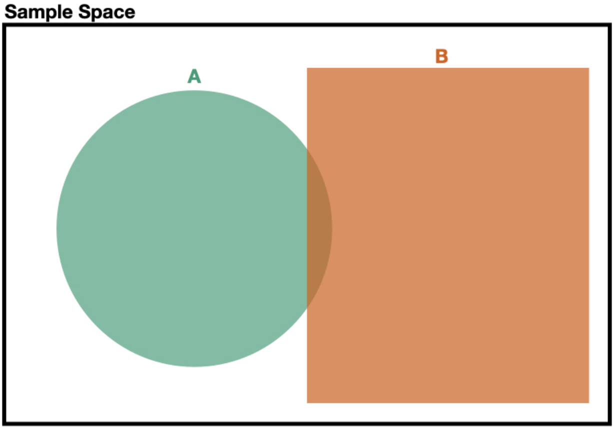
We interpret the conditional probability $\Pr(A \mid B)$ as the probability of A given that B happens (or has already happened). Suppose a bag with two green marbles and two red marbles. I draw two marbles without replacement and see that the first is green. Then the probability that the second is green, given that the first is/was green, is

$$\Pr(\text{second is green} \mid \text{first is green}) = \frac{\Pr(\text{second is green AND first is green})}{\Pr(\text{first is green})}.$$

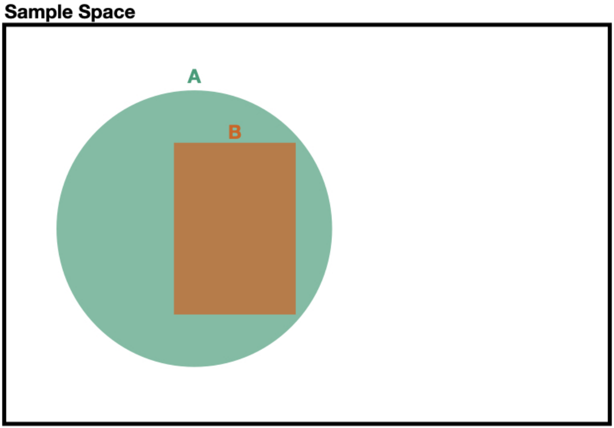
Carlisle's Happy-Sad Principle

If event A happens, you win \$100. You see that event B happens. Are you now **happy** (more likely to win), **sad** (less likely to win), or **indifferent** (equally likely to win)?

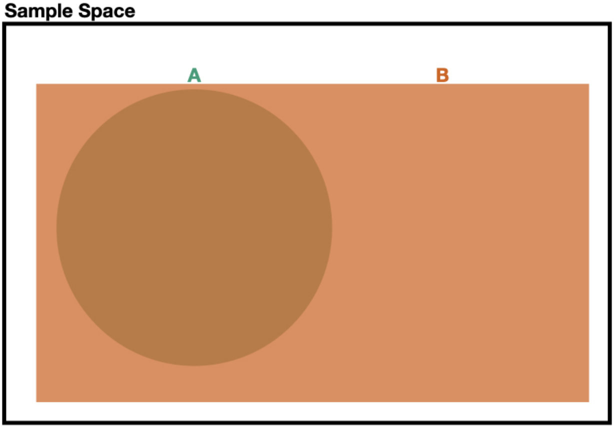
A Visualization of Conditional Probability



A Visualization of Conditional Probability



A Visualization of Conditional Probability



Theorem 7 (Multiplication Rule for Two Events)

For events A and B, $\Pr(A \cap B) = \Pr(B)\Pr(A|B)$ *if* $\Pr(B > 0)$*.*
Similarly, $\Pr(A \cap B) = \Pr(A)\Pr(B|A)$ *if* $\Pr(A > 0)$*.*

Theorem 8 (Multiplication Rule for n Events)

For events $A_1, A_2, ..., A_n$ where $\Pr(A_1 \cap A_2 \cap ... \cap A_{n-1}) > 0$,

$$\begin{aligned}
\Pr(A_1 \cap A_2 \cap ... \cap A_n) &= \Pr(A_1) \\
&\times \Pr(A_2 \mid A_1) \\
&\times \Pr(A_3 \mid A_2, A_1) \\
&\times ... \\
&\times \Pr(A_n \mid A_{n-1}, ..., A_2, A_1)
\end{aligned}$$

Exercise 6

Simplify $\Pr(A \mid B)$ for the following scenarios. Connect your answers to the results above.

1. $A \subset B$ and $\Pr(B) > 0$.
2. A and B are disjoint and $\Pr(B) > 0$.
3. B is the empty set (tricky!).
4. B is the sample space S.

Solution

(1) $\Pr(A)/\Pr(B)$, (2) 0, (3) not defined, (4) $\Pr(A)\Pr(B)$

Exercise 7

Voting is a habit. Suppose that elections occur every four years and a hypothetical voter can vote in their first election at 18 in 2020. They vote in their first election with probability 0.5. If they voted in the last election, then they vote in the next election with probability 0.8. If they abstained in the last election, then they vote in the next election with probability 0.3. What is the probability that they vote in their first three election (i.e., vote, then vote, then vote)? What the probability that they abstain in their first three elections (i.e., abstain, then abstain, then abstain)? What's the probability that they either vote or abstain in their first three elections (i.e., [vote, then vote, then vote] *OR* [abstain, then abstain, then abstain])?

Solution

$0.5 \times 0.8 \times 0.8$; $0.5 \times 0.7 \times 0.7$; the sum of the two previous

Definition 7 (Independence of Two Events)

*Events A and B are **independent** if* $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

If $\Pr(A) > 0$ and $\Pr(B) > 0$, then Definitions 6 and 7 imply that two events are independent if and only if their conditional probabilities equal their unconditional probabilities so that $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$.

Definition 8 (Independence of *n* Events)

*Events $A_1, A_2, ..., A_n$ are **independent** if for every subset $A_a, ..., A_m$ with at least two events,* $\Pr(A_a \cap ... \cap A_m) = \Pr(A_a) ... \Pr(A_m)$.

The "every subset"' part of Definition 8 is subtle, so let's create a specific example. "Every subset" of A, B, and C with at least two events includes the following: $\{A, B\}$, $\{A, C\}$, $\{B, C\}$, and $\{A, B, C\}$.

Exercise 8

Suppose A and B are independent and $\Pr(B) < 1$. Find $\Pr(A^c|B^c)$ in terms of A and B. Prove that A^c and B^c are independent.

Solution

First, show that A^c and B^c are independent.

$$\begin{aligned}
\Pr(A^c \cap B^c) &= \Pr([A \cup B]^c) \\
&= 1 - \Pr(A \cup B) \\
&= 1 - [\Pr(A) + \Pr(B) - \Pr(A \cap B)] \\
&= 1 - \Pr(A) - \Pr(B) + \Pr(A)\Pr(B) \\
&= [1 - \Pr(A)] \times [1 - \Pr(B)] \\
&= \Pr(A^c) \times \Pr(B^c)
\end{aligned}$$

Then, by independence, we know that $\Pr(A^c|B^c) = \Pr(A^c)$ and that $\Pr(A^c) = 1 - \Pr(A)$.

Exercise 9

Suppose A and B are events and $\Pr(B) = 0$. (A is any event.) Find $\Pr(A \cap B)$. Prove that A and B are independent.

Solution

If A and B are independent, then $\Pr(A \cap B) = \Pr(A)\Pr(B)$. Well, if $\Pr(B) = 0$, then $\Pr(A \cap B)$ must be zero. If the chance of B is zero, then the chance of A and B is zero. Thus the two sides are equal and therefore independent.

Exercise 10

Suppose a six-die is rolled 10 times. What's the probability of...

1. all sixes?
2. not all-sixes?
3. all not-sixes?

Solution

(1) $(1/6)^{10}$, (2) $1 - (1/6)^{10}$, (3) $(5/6)^{10}$

Definition 9

To create a ***partition*** $B_1, B_2, ..., B_k$ *of the sample space* S*, divide* S *into* k *disjoint events* $B_1, B_2, ..., B_k$ *so that* $\bigcup_{i=1}^{n} B_i = S$*.*

Theorem 9 (Law of Total Probability)

Suppose a partition (see Definition 9) $B_1, B_2, ..., B_k$ *of the sample space* S *where* $\Pr(B_j) > 0$ *for* $j = 1, 2, ..., k$*. Then*

$$\Pr(A) = \sum_{j=1}^{k} \Pr(B_j) \Pr(A \mid B_j).$$

Theorem 10 (Bayes' Rule)

Suppose a partition (see Definition 9) $B_1, B_2, ..., B_k$ of the sample space S where $\Pr(B_j) > 0$ for $j = 1, 2, ..., k$. Suppose an event A, where $\Pr(A) > 0$. Then

$$\Pr(B_i \mid A) = \frac{\Pr(B_i)\Pr(A \mid B_i)}{\sum_{j=1}^{k}\Pr(B_j)\Pr(A \mid B_j)}.$$

We can simplify the rule a bit by assuming the partition B and B^c. In applications, this partition is usually sufficient (see Exercise 11).

Theorem 11 (Bayes' Rule for a simpler partition)

Suppose the simple partition B and B^c of the sample space S where $\Pr(B) > 0$ and $\Pr(B^c) > 0$. Suppose an event A, where $\Pr(A) > 0$. Then

$$\Pr(B \mid A) = \frac{\Pr(B)\Pr(A \mid B)}{\Pr(B)\Pr(A \mid B) + \Pr(B^c)\Pr(A \mid B^c)}.$$

Exercise 11

You're considering getting tested for a rare disease that 1 in 100,000 people have. If given to a person with the disease, the test will produce a positive result 99% of the time. If given to a person without the disease, the test will produce a positive result 0.1% of the time (i.e., 1 in 1,000). You are given the test and the result comes back positive. Use Bayes' rule to compute the chance that you have the disease.

Section 2

Random Variables

Supplemental Readings

- Aronow and Miller, section 1.2, pp. 15-30
- DeGroot and Schevish, sections 3.1 to 3.3, pp. 93-117

Definition 10

A ***random variable*** *is a real-valued function defined on the sample space S.*

Consider tossing a coin twice. The sample space S consists of the outcomes $\{HH, TH, HT, TT\}$. We can create the random variable X by mapping HH to 2, TH and HT to 1, and TT to zero—we count the number of heads.

Outcome or o	Value x of X or $X(o)$
HH	2
HT	1
TH	0

We tend to equate the function X with the outcomes of the experiment. In this case, we would imagine 2 rather than HH, e.g., as the outcome of the experiment so that $X(x) = I(x) = x$. That's fine. We simply require that a random variable X take on real-values that map from the sample space S and possible just the real-values in the sample space.

Definition 11

*The **distribution** of random variable X is is the collection of all probabilities $\Pr(X \in C)$ for all sets C of real numbers such that $X \in C$ is an event.*

Discrete Distributions

Definition 12

Random variable X has a ***discrete distribution*** *(or "is discrete") if either (a) X can take on only a finite number of values $\{x_1, x_2, ..., x_n\}$ or (b) an infinite sequence of different values $\{x_1, x_2, x_3, ...\}$.*

Part (b) of Definition 12 is subtle. It is crucial that you can arrange the values that X can take on into a *sequence*. That is, a first value, a second value, etc., that reaches all the possibilities. We can arrange all the positive integers into the sequence $\{1, 2, 3, ...\}$, all the integers into the sequence $\{0, -1, 1, -2, 2, -3, 3, ...\}$. Perhaps somewhat surprising, the rational numbers can be arranged into a sequence as well. One cannot arrange the real numbers $\mathbb{R}$ into a sequence or arrange numbers in the interval $[0, 1]$ into sequence.

Discrete Distributions

Definition 13 (pmf)

For a discrete random variable X, the ***probability mass function*** *(pmf) is the function $f(x) = \Pr(X = x)$ for every real number x. As a short hand for "X has pmf $f(x)$, we write $X \sim f(x)$.*

Some authors refer to the pmf as the "probability function."

Discrete Distributions

Theorem 12

Let f represent the pmf of a discrete random variable X. If the set of possible values of X does not include x, then $f(x) = 0$. If the sequence $x_1, x_2, ..., x_n$ includes all the possible values of X, then $\sum_{i=1}^{n} f(x_i) = 1$.

Theorem 13

Let f represent the pmf of a discrete random variable X. The probability of each subset C of the real line is $\Pr(X \in C) = \sum_{x_i \in C} f(x_i)$.

Example 1 (Bernoulli Distribution)

A Bernoulli random variable X (or "a random variable X with the Bernoulli distribution) has support $\{0, 1\}$ and $Pr(X = 1) = \pi$. We can write the pmf as

$$f(x) = \begin{cases} \pi & \textit{if } x = 1 \\ 1 - \pi & \textit{if } x = 0 \\ 0 & \textit{otherwize} \end{cases},$$

or equivalently as $f(x) = \pi^x (1 - \pi)^{(1-x)}$ for $x \in \{0, 1\}$ (and 0 otherwise). Note that we often leave the "and 0 otherwise" implicit, but it matters. Understanding the Bernoulli distribution, we can now write a Bernoulli random variable X as $X \sim Bernoulli(\pi)$.

Discrete Distributions

Each "distribution" (the Bernoulli, binomial, Poisson, geometric, normal, exponential, Cauchy, etc) is actually a *family* of distributions. It's a "family" because it has parameters that can vary.

For example, the Bernoulli distribution has the parameter $\pi \in [0, 1]$ that allows $\Pr(X = 1)$ to vary. If we set $\pi = 0.3$, then $\Pr(X = 1) = 0.3$. The ideas is that there are *lots of different Bernoullis*, one for each value of π. Depending on the context, you might choose a particular value (e.g., let $\pi = 0.3$), leave it arbitrary, or try to estimate it.

The key point is this: **distributions have parameters that allow them to represent different random variables**. By varying the parameters, you vary the distribution.

Discrete Distributions

Example 2 (Uniform Distribution on the Integers a to b)

A uniform random variable X on the integers $a, ..., b$ has the pmf $f(x) = \frac{1}{b-a+1}$ for $x = a, ..., b$ (and 0 otherwise).

The parameters a and b vary the integers over which the distribution varies. We can make it vary from -10 to 10, 1 to 6, or 1 to 100... whatever we like.

Discrete Distributions

Example 3 (Geometric Distribution)

Imagine tossing a possibly-biased coin until you obtain a head. Let the π *represent the probability of head for on each toss. Let* X *represent the number of tosses. (In this case, we can understand each toss as* $Z \sim$ *Bernoulli*(π)*). The table below shows the probability of several valures of* x *computed using the multiplication rule.*

Outcome	x	*Probability*
H	1	π
TH	2	$(1-\pi)\pi$
TTH	3	$(1-\pi)^2\pi$
$TTTH$	4	$(1-\pi)^3\pi$
$TTTTH$	5	$(1-\pi)^4\pi$
	$\vdots$	
$\underbrace{TT...TT}_{x-1 \text{ tails}}H$	x	$(1-\pi)^{(x-1)}\pi$

We can see that $\Pr(X=x) = f(x) = (1-\pi)^{(x-1)}\pi$ *for* $x \in \{1, 2, ...\}$ *and 0 otherwise. We refer to this distribution as the geometric distribution, so we can now write* $X \sim$ *geometric*(π)*.*

Exercise 12

X has a uniform distribution on the integers $10, ..., 20$. Write the pmf of X. What's the probability that X is even? Odd? Prime? Not 20? Represent your answers with summation notation before finding the numeric probability.

Exercise 13

Suppose I have a box with the tickets $\{-17, 1, 1, 1, 2, 3, 23.3, \pi\}$. I draw one ticket at random and record the number on the ticket. Write down the pmf that corresponds to this experiment.

Exercise 14

Verify that the Bernoulli pmf $f(x) = \pi^x(1-\pi)^{(1-x)}$ for $x \in \{0, 1\}$ (and 0 otherwize) is a pmf. This requires two things. First, show that $f(x) \geq 0$ for $x \in \mathbb{R}$. Second, show that $\sum_{\text{support of } X} f(x) = 1$.

Exercise 15

(Hard) Verify that the geometric pmf $f(x) = (1-\pi)^{(x-1)}\pi$ for $x \in \{1, 2, ...\}$ (and 0 otherwise)is a pmf. It's easy to show that $f(x) \geq 0$ for $x \in \mathbb{R}$. It's much trickier to show that $\sum_{\text{support of } X} f(x) = 1$. You need to know something about a geometric series.

Exercise 16

X has the pmf $f(x) = cx$ for $x \in \{1, 2, 3, 4, 5\}$ (and 0 otherwise). Find c. Hint: Use Theorem 12 and make the pmf sum to 1.

Discrete Distributions in R

R offers a collection of pmfs to use. These always/usually start with a `d`. As an example, here's the pmf for the geometric distribution, with the parameter π (i.e., the probability of success) equal to 0.5.

Unfortunately, there are a couple of different parameterizations of the geometric. The version we discussed above treats *the number of trials until a success* as the outcome. The other treats *number of failures until* until a success. R uses the second parameterization Fortunately, the former is always one plus the latter.

```r
# the chance of it taking 5 tosses to obtain a head
# (alternatively, the chance of getting 4 tails
# before the first head)
dgeom(4, prob = 0.5)
```

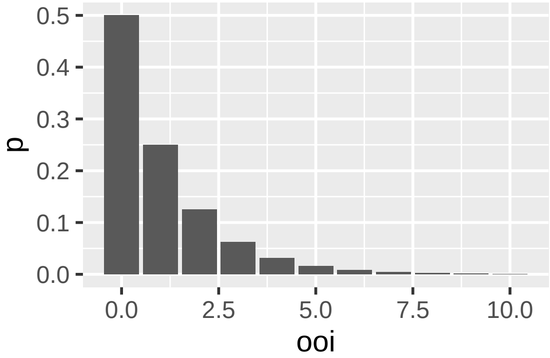
```
## [1] 0.03125
```

Discrete Distributions in R

```
ooi <- 0:10                           # outcomes of interest
p <- dgeom(ooi, prob = 0.5)  # probability of ooi

round(p, 3)
```

```
##  [1] 0.500 0.250 0.125 0.063 0.031 0.016 0.008 0.004 0.002 0.
```

```r
qplot(x = ooi, y = p, geom = "col")
```



Discrete Distributions in R

The binomial distribution models the sum of a series of n Bernoulli trials. It's like the number of heads if you toss a coin 10 times. Or the number of sixes if you roll a die 20 times.

```
# binomial(10, 0.5); number of H if you toss a fair coin 10x
dbinom(7, size = 10, prob = 0.5)  # Pr(7)
```

```
## [1] 0.1171875
```

```
dbinom(8, size = 10, prob = 0.5)  # Pr(8)
```

```
## [1] 0.04394531
```

```
dbinom(9, size = 10, prob = 0.5)  # Pr(8)
```

```
## [1] 0.009765625
```

If you set the number of trials to `size = 1`, the you have a Bernoulli distribution

```
# for size = 1, we have a Bernoulli
dbinom(1, size = 1, prob = 0.7)   # Pr(1)
```

```
## [1] 0.7
```

Discrete Distributions in R

We can also use the `r` version of the distribution functions to *simulate*.

```r
# 5 values from geometric(0.3)
rgeom(5, prob = 0.3)
```

```
## [1] 2 5 4 0 3
```

```r
# 3 values from binomial(100, 0.5)
rbinom(5, size = 100, prob = 0.3)
```

```
## [1] 28 27 43 29 32
```

```r
# 10 values from Bernoulli(0.5)
rbinom(10, size = 1, prob = 0.5)
```

```
##  [1] 0 1 1 1 1 1 0 1 0 1
```

Continuous Distributions

Definition 14

Random variable X has a ***continuous distribution*** *(or "is continuous") if there exists a non-negative, real function f such that $\Pr(X \in C) \int_C f(x)$ for every interval of real-valued numbers C.*

To help make sense of Definition 14, the table below shows a few sets C and its associated interval.

C	Probability	Integral
$[a, b]$ for $a < b$	$\Pr(a \leq X \leq b)$	$\int_a^b f(x)dx$
$[\infty, b]$	$\Pr(X \leq b)$	$\int_\infty^b f(x)dx$
$[a, \infty]$	$\Pr(X \geq a)$	$\int_a^\infty f(x)dx$

Definition 15

We refer to the function f from Definition 14 as the ***probability density function*** *(pdf) of X.*

Continuous Distributions

Theorem 14

Let f represent the pdf of X. $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Definition 16

For a pmf or pdf f, refer to the set of values x such that $f(x) > 0$ as the ***support*** *of X.*

The support refers to the possible values of X. For example, the support of a geometric random variable is $\{1, 2, 3, ...\}$. The support of a uniform distribution on the interval $[a, b]$ is $[a, b]$.

Example 4 (Uniform Distribution on the Interval a to b)

A uniform random variable X on the interval $[a, b]$ has the pdf $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$ (and 0 otherwise).

Continuous Distributions

Example 5 (Logistic Distribution)

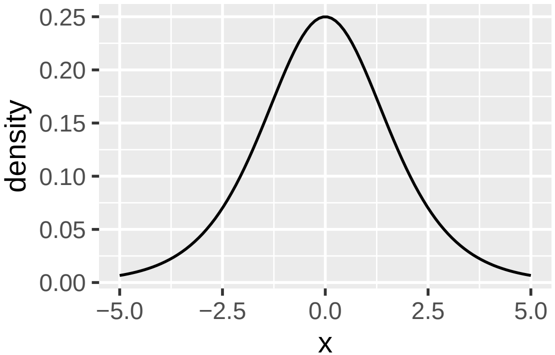
The (standard) logistic distribution has pdf $f(x) = \frac{e^{-x}}{(1-e^{-x})^2}$. *For a logistic random variable* X,

$$\begin{aligned}
\Pr(0 \leq x \leq 1) &= \int_0^1 f(x)dx \\
&= \int_0^1 \frac{e^{-x}}{(1-e^{-x})^2} dx \\
&= \left. \frac{1}{1+e^{-x}} \right|_0^1 \\
&= \frac{1}{1+e^{-1}} - \frac{1}{1+e^0} \\
&= \frac{1}{1+e^{-1}} - \frac{1}{2} \\
&\approx 0.23
\end{aligned}$$

Continuous Distributions: Plotting with R

We can use R to plot the pdf of the standard logistic distribution.

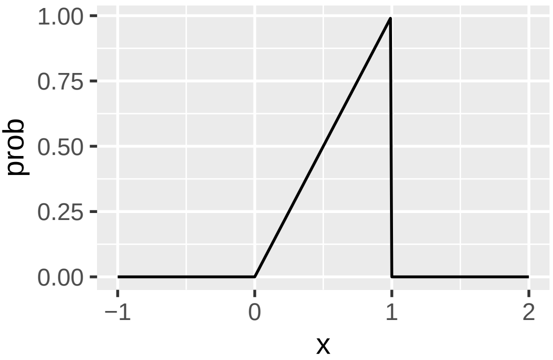
```
from <- -5
to <- 5

data <- data.frame(x = seq(from, to, length.out = 100)) %>%
  mutate(density = dlogis(x))

ggplot(data, aes(x = x, y = density)) +
  geom_line()
```



\begin{Exercise} The exponential distribution has pdf $f(x) = \lambda e^{-\lambda x}$. We can think of the lifespan L of a light bulb (in years) as following an exponential distribution with $\lambda = 0.25$. (Or we might write $L \sim$ exponential(0.25)}). Find $\Pr(1 < L < 5)$ (i.e., the probability that a light bulb last between one and five years). Hints: First, don't worry about whether the inequalities are strict or not–these point have probability zero, so it doesn't matter. Second, remember that $\Pr(1 < L < 5) = \int_1^5 0.25 e^{-0.25x}$. \end{Exercise}

Exercise 17

I've invented a new distribution called the "triangle distribution." It's $f(x) = x$ for $x \in [0, 1]$ (and 0 otherwise), so it's shaped like a right-triangle. Unfortunately, the area under the pdf is 0.5, not 1. So I need to add a normalizing constant c, so that $f(x) = cx$. Find c so that my triangle distribution is actually a pdf (integrates to one.) Be sure to show that $\int_{-\infty}^{\infty} f(x)dx = 1$ for your chosen value of c.



Exercise 18

For the triangle distribution in the previous question (the corrected version that integrates to one), find $\Pr(X < 0.25)$. Find $\Pr(X > 0.5)$.

Exercise 19

The beta distribution is a weird little distribution defined over the $[0, 1]$ interval. The pdf for the beta distribution is $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$. (Here, $B()$ is the "beta function." It's "just" a normalizing constant here, so not critical to know, but feel free to Google it). Use the `dbeta()` function in R to plot the pdf for different values of α (`shape1` argument) and β (`shape2` argument). What happens when you make both parameters large (> 10)? What about when you make both small (< 1)?

CDFs

Definition 17 (cdf)

The ***cumulative distribution function*** *(cdf) F of a random variable X is $F(x) = \Pr(X \leq x)$ for $-\infty < x < \infty$.*

Definition 18 (cdf of a Discrete Distribution)

For a discrete random variable X with pmf $f(x)$, $F(x) = \sum_{t \leq x} f(t)$.

Definition 19 (cdf of a Continuous Distribution)

For a continuous random variable X with pdf $f(x)$, $F(x) = \int_{-\infty}^{x} f(t)dt$.

CDFs

Notice four things:

1. $F(x)$ is non-decreasing.
2. $\lim_{x \to -\infty} F(x) = 0$.
3. $\lim_{x \to \infty} F(x) = 1$.
4. $F(x)$ is continuous from the right.

Theorem 15

$\Pr(X > x) = 1 - F(x)$ *for all* x.

Theorem 16

$\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$ *for all* $x_2 > x_1$.

Exercise 20

Sketch the cdf of the Bernoulli distribution.

Exercise 21

Sketch the cdf of the geometric distribution. (You don't need to go all the way to $x = \infty$, but get $\Pr(X \leq x)$ close to one.)

Exercise 22

Suppose I'm going to sample once from a box with five tickets. The ticket are numbered with 1, 2, 3, 3, and 7. Write the pmf as a stepwise function. Sketch the cdf.

\begin{Find the cdf for the triangle distribution we derived above. Use the cdf to find: (a) $\Pr(X \leq 0.25)$, $\Pr(X > 0.5)$, and $\Pr(0.25 \leq X)$}

Exercise 23

Browse Wikipedia's list of probability distributions. Find a *discrete distribution with finite support* that seems interesting.

1. Summarize it verbally (i.e., "it's like the number of tosses required to get a head").
2. Write down the pmf. Are there different versions or parameterizations?
3. Does R have the `d` function to compute probabilities? If so, compute a few to test it out. You might need to find a package. Or the function might not exist.
4. Write down the cdf.
5. Does R have the `p` function to compute (cumulative) probabilities? If so, compute a few to test it out. You might need to find a package. Or the function might not exist.
6. Does R have the `r` function to simulate from the distribution? If so, simulate 1,000 draws for a particular set of parameters and use `geom_bar()` to draw a bar plot of the simulations. Repeat for a different set of parameters.

Repeat for a *discrete distribution with infinite support*.

Exercise 24

Browse Wikipedia's list of probability distributions. Find a *continuous distribution* that seems interesting.

1. Summarize it verbally, if you can. Continuous distributions might not have a good verbal description, but you can at least describe the shape of the pdf.
2. Write down the pdf. Are there different versions or parameterizations?
3. Does R have the `d` function to compute the density for particular outcomes? If so, compute a few to test it out. You might need to find a package. Or the function might not exist. Realize that these are not probabilities—in order to turn the densities into probabilities, you must integrate them over and interval.
4. Write down the cdf.
5. Does R have the `p` function to compute (cumulative) probabilities? If so, compute a few to test it out. You might need to find a package. Or the function might not exist. These are probabilities—the integration is already done (that's where the cdf comes from).
6. Does R have the `r` function to simulate from the distribution? If so, simulate 1,000 draws for a particular set of parameters and use `geom_histogram()` to draw a histogram of the simulations. Repeat for a different set of parameters.