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.
- Casella and Berger, ch. 1, sections 1.1-1.3.

Definition 1

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

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 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 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, $\Pr(A) \geq 0$*.*

Axioms of Probability (Kolmogorov Axioms)

Axiom 1

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

Axiom 2

$\Pr(S) = 1$.

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 $S = \{0, 1\}$?

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.

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 $[0, 1), [1, 2), [2, 3), \ldots$.

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 $[0, 1), [1, 2), [2, 3), ...$
- For $S = \{0, 1\}$, *one* such sequence would be $\{0\}, \{1\}, \emptyset, \emptyset, \emptyset, ...$

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.

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.

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

Theorem 2

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

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

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

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

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 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

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

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

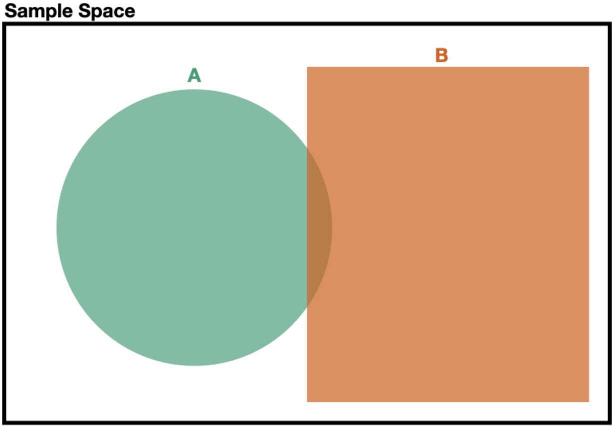
We interpret the conditional probability $\Pr(A \mid B)$ as the probability of A given that B happens (or has already happened). Suppose a bag with two green marbles and two red marbles. I draw two marbles without replacement and see that the first is green. Then the probability that the second is green,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

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

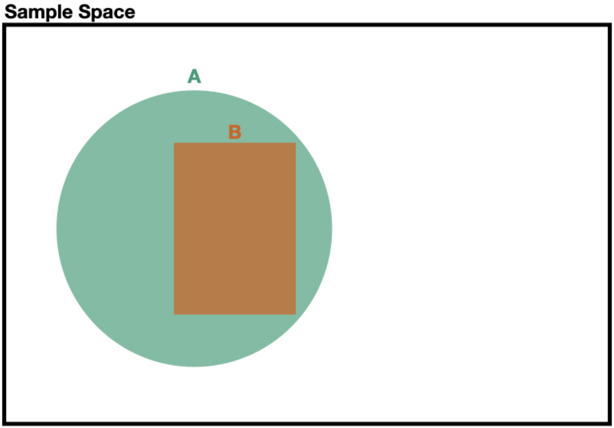
Carlisle's Happy-Sad Principle

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)?

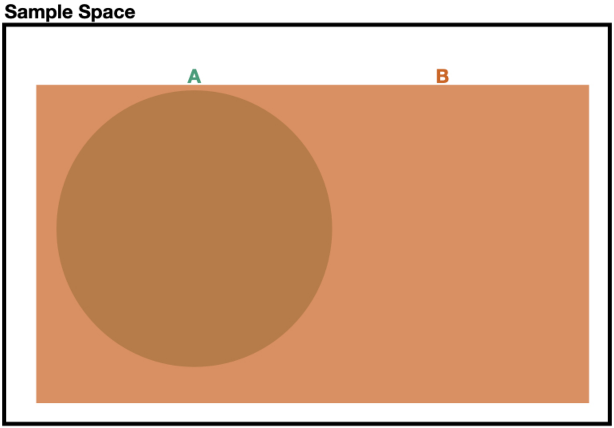
A Visualization of Conditional Probability



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

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

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

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

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

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.