Probability Distributions. Conditional Probability

Russell Impagliazzo and Miles Jones
Thanks to Janine Tiefenbruck

http://cseweb.ucsd.edu/classes/sp16/cse21-bd/
May 16, 2016
People have been interested in probability ever since gambling was invented, which was the day after “owning stuff” was invented.
Probability Spaces intuition

• In probability, we want to reason about the likelihood of complex events.
• To do this, we must model the chance of the underlying objects that contribute.
• We use the idea of a probability space to consider every possibility.
Probability Spaces Formal Definition

Sample space, $S$: (finite or countable) set of possible outcomes.

Probability distribution, $p$: assignment of probabilities to outcomes in $S$ so that

- $0 \leq p(s) \leq 1$ for each $s$ in $S$.

- Sum of probabilities is 1, $\sum_{s\in S} p(s) = 1$.  

Rosen p. 446. 453
Sample space, \( S \): (finite or countable) set of possible outcomes.

Probability distribution, \( p \): assignment of probabilities to outcomes in \( S \) so that

\[
0 \leq p(s) \leq 1 \quad \text{for each } s \in S
\]

- Sum of probabilities is 1, \( \sum_{s \in S} p(s) = 1 \).

Compare flipping a \textbf{fair coin} and a \textbf{biased coin}:

A. Have different sample spaces.
B. Have the same sample spaces but different probability distributions.
C. Have the same sample space and same probability distributions.

Rosen p. 446. 453
Sample space, $S$: (finite or countable) set of possible outcomes.

Probability distribution, $p$: assignment of probabilities to outcomes in $S$ so that

- $0 \leq p(s) \leq 1$ for each $s$ in $S$.
- Sum of probabilities is 1, $\sum_{s \in S} p(s) = 1$.

Event, $E$: subset of possible outcomes. $P(E) = \sum_{s \in E} p(s)$.
Uniform distribution

For sample space $S$ with $n$ elements, the uniform distribution assigns the probability $1/n$ to each element of $S$.

When flipping a fair coin successively three times:

A. The sample space is $\{H, T\}$
B. The empty set is not an event.
C. The event $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ has probability less than 1.
D. The uniform distribution assigns probability $1/8$ to each outcome.
E. None of the above.
Uniform distribution

For sample space S with n elements, **uniform distribution** assigns the probability 1/n to each element of S.

When flipping a fair coin successively three times:

A. The sample space is \{H, T\}
B. The empty set is not an event.
C. The event \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} has probability less than 1.
D. The uniform distribution assigns probability 1/8 to each outcome.
E. None of the above.
For sample space $S$ with $n$ elements, **uniform distribution** assigns the probability $1/n$ to each element of $S$.

When flipping a fair coin successively three times, what is the distribution of the number of $H$s that appear?

A. Uniform distribution.
B. $P(0 \ H) = P(3 \ H) = 3/8$ and $P(1 \ H) = P(2 \ H) = 1/8$.
C. $P(0 \ H) = P(1 \ H) = 1/8$ and $P(2 \ H) = P(3 \ H) = 1/8$.
D. $P(0 \ H) = P(3 \ H) = 1/8$ and $P(1 \ H) = P(2 \ H) = 1/8$.
E. None of the above.
For sample space $S$ with $n$ elements, uniform distribution assigns the probability $1/n$ to each element of $S$.

When flipping a fair coin successively three times, what is the distribution of the number of $H$s that appear?

A. Uniform distribution.
B. $P(0\ H) = P(3\ H) = 3/8$ and $P(1\ H) = P(2\ H) = 1/8$.
C. $P(0\ H) = P(1\ H) = 1/8$ and $P(2\ H) = P(3\ H) = 1/8$.
D. $P(0\ H) = P(3\ H) = 1/8$ and $P(1\ H) = P(2\ H) = 1/8$.
E. None of the above.

Not a uniform distribution!
If start with the uniform distribution on a set $S$, then the probability of an event $E$ is

$$P(E) = \frac{|E|}{|S|}$$

When flipping $n$ fair coins what is the **probability of getting exactly $k$ Hs**?

A. $1/n$
B. $k/n$
C. $1/2^n$
D. $C(n,k) / 2^n$
E. None of the above.
Binomial distribution

When flipping $n$ fair coins what is the probability of getting exactly $k$ Hs?

$$P(k \text{ Hs}) = \frac{\# \text{ coin toss sequences with } k \text{ Hs}}{\# \text{ possible coin toss sequences}}$$

Possible coin toss sequences: $\{ \text{HH..HH, HH..HT, ..., TT..TH, TT..TT} \}$

$$P(k \text{ Hs}) = \binom{n}{k} \cdot \frac{1}{2^n}$$
Binomial distribution

When flipping $n$ fair coins what is the **probability of getting exactly** $k$ Hs?

$$P(k \text{ Hs}) = \frac{\# \text{ coin toss sequences with } k \text{ Hs}}{\# \text{ possible coin toss sequences}}$$

Possible coin toss sequences: { HH..HH, HH..HT, ..., TT..TH, TT..TT }  

$$P(k \text{ Hs}) = \frac{\binom{n}{k}}{2^n}$$

What if the coin isn't fair?
Binomial distribution

Bernoulli trial: a performance of an experiment with two possible outcomes.  
*e.g. flipping a coin*

Binomial distribution: probability of exactly k successes in n independent Bernoulli trials, when probability of success is p.  
*e.g. # Hs in n coin flips when probability of H is p*

What is it?

A. $\frac{C(n,k)}{2^n}$  
B. $\frac{p^k}{2^n}$  
C. $C(n,k) \cdot p^k$  
D. $C(n,k) \cdot p^k (1-p)^{n-k}$  
E. None of the above.

*Rosen p. 480*
Randomness in the world

For the Nate-haters, here’s the 538 prediction and actual results side by side pic.twitter.com/jbny4pRX

Michael Cosentino
15 hours ago

Nate Silver: statistician famous for analyzing election predictions & baseball
When the input is random …

- data mining
  - elections
  - weather
  - stock prices
  - genetic markers

- analyzing experimental data

When is analysis valid? When are we overfitting to available data?
Randomness in Computer Science

When the desired **output** is random …

- picking a cryptographic key
- performing a scientific simulation
- programming a computer adversary in a game
When the **algorithm** uses randomness …

- Monte Carlo methods  *Rosen p. 463*

- search heuristics  *avoid local mins*

- randomized hashing

- quicksort
Dangers of probabilistic reasoning

"Intuitive probabilistic reasoning“ often goes wrong.
The Monty Hall Puzzle

Car hidden behind one of three doors.

Goats hidden behind the other two.

Player gets to choose a door.

Host opens another door, reveals a goat.

Player can choose whether to swap choice with other closed door or stay with original choice.

What's the player's best strategy?
A. Always swap. B. Always stay. C. Doesn't matter, it's 50/50.
Some history…

Puzzle introduced by Steve Selvin in 1975.

Marilyn vos Savant was a prodigy with record scores on IQ tests who wrote an advice column. In 1990, a reader asked for the solution to the Monty Hall puzzle.

• After she published the (correct) answer, thousands of readers (including PhDs and even a professor of statistics) demanded that she correct her "mistake".
• She built a simulator to demonstrate the solution so they could see for themselves how it worked.
The Monty Hall Puzzle … the solution

Pick a door at random to start
The Monty Hall Puzzle … the solution
The Monty Hall Puzzle … the solution

What's the probability of winning (C) if always switch ("Y")?

A. 1/3
B. 1/2
C. 2/3
D. 1
E. None of the above.
The Monty Hall Puzzle … the solution

What's the probability of winning (C) if always stay (“N”)?

A. 1/3
B. 1/2
C. 2/3
D. 1
E. None of the above.
What's wrong with the following argument?

"It doesn't matter whether you stay or swap because the host opened one door to show a goat so there are only two doors remaining, and both of them are equally likely to have the car because the prizes were placed behind the doors randomly at the start of the game"
Conditional probabilities

Probability of an event may change if we have additional information about outcomes.

Suppose E and F are events, and \( P(F) > 0 \). Then,

\[
P(E|F) = \frac{P(E \cap F)}{P(F)}
\]

i.e.

\[
P(E \cap F) = P(E|F)P(F)
\]

Rosen p. 456
Conditional probabilities

Are these probabilities equal?

The probability that two siblings are girls if know the oldest is a girl.

The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

A. They're equal.
B. They're not equal.
C. ???
Conditional probabilities

Are these probabilities equal?
The probability that \textbf{two siblings are girls} if know the oldest is a girl.
The probability that \textbf{two siblings are boys} if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

1. Sample space

2. Initial distribution on the sample space

3. What events are we conditioning on?
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

1. **Sample space**
   Possible outcomes: \{bb, bg, gb, gg\}  \textit{Order matters!}

2. **Initial distribution on the sample space**

3. **What events are we conditioning on?**
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

1. Sample space
Possible outcomes: \{bb, bg, gb, gg\} \hspace{1cm} \textit{Order matters!}

2. Initial distribution on the sample space
Uniform distribution, each outcome has probability \(\frac{1}{4}\).

3. What events are we conditioning on?
Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
A = { outcomes where oldest is a girl }       B = { outcomes where two are girls}
Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
A = \{ outcomes where oldest is a girl \} = \{ gg, gb \}
B = \{ outcomes where two are girls \} = \{ gg \}
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
A = \{ outcomes where oldest is a girl \}  \quad B = \{ outcomes where two are girls \}
   = \{ gg, gb \}  \quad = \{ gg \}
P(A) = \frac{1}{2}  \quad P(B) = \frac{1}{4} = P(A \cap B)
Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl.
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
A = { outcomes where oldest is a girl } = { gg, gb }
P(A) = ½
B = { outcomes where two are girls } = { gg }
P(B) = ¼ = P(A \cap B)

By conditional probability law: P(B | A) = P(A \cap B) / P(A) = (1/4) / (1/2) = ½.
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl. \( \frac{1}{2} \)
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

1. Sample space
   Possible outcomes: \{bb, bg, gb, gg\} \hspace{1cm} Order matters!

2. Initial distribution on the sample space
   Uniform distribution, each outcome has probability \( \frac{1}{4} \).

3. What events are we conditioning on?
Are these probabilities equal?
The probability that **two siblings are girls** if know the oldest is a girl. $$1/2$$
The probability that **two siblings are boys** if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. **What events are we conditioning on?**

C = { outcomes where one is a boy} = { bb, bg, gb }
P(C) = ¾

D = { outcomes where two are boys } = { bb }
P(D) = ¼ = P(C ∩ D)
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl. $\frac{1}{2}$
The probability that two siblings are boys if know that one of them is a boy.

Assume that each child being a boy or a girl is equally likely.

3. What events are we conditioning on?
$C = \{ \text{outcomes where one is a boy} \} = \{ \text{bb, bg, gb} \}$
$P(C) = \frac{3}{4}$

$D = \{ \text{outcomes where two are boys} \} = \{ \text{bb} \}$
$P(D) = \frac{1}{4} = P(C \cap D)$

By conditional probability law: $P(D \mid C) = \frac{P(C \cap D)}{P(C)} = \frac{1}{4} / \frac{3}{4} = \frac{1}{3}$. 
Conditional probabilities

Are these probabilities equal?
The probability that two siblings are girls if know the oldest is a girl. $\frac{1}{2}$
The probability that two siblings are boys if know that one of them is a boy. $\frac{1}{3}$

Assume that each child being a boy or a girl is equally likely.
# Conditional probabilities: Simpson's Paradox

Which is the better overall treatment?

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small stones</strong></td>
<td>81 successes / 87</td>
<td>234 successes / 270</td>
</tr>
<tr>
<td><strong>Large stones</strong></td>
<td>192 successes / 263</td>
<td>55 successes / 80</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td>273 successes / 350 (78%)</td>
<td>289 successes / 350 (83%)</td>
</tr>
</tbody>
</table>
### Conditional probabilities: Simpson's Paradox

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small stones</strong></td>
<td>81 successes / 87 (93%)</td>
<td>234 successes / 270 (87%)</td>
</tr>
<tr>
<td><strong>Large stones</strong></td>
<td>192 successes / 263 (73%)</td>
<td>55 successes / 80 (69%)</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td>273 successes / 350 (78%)</td>
<td>289 successes / 350 (83%)</td>
</tr>
</tbody>
</table>

Which treatment is better?

A. Treatment A for all cases.  
B. Treatment B for all cases.  
C. A for small and B for large.  
D. A for large and B for small.
Conditional probabilities: Simpson's Paradox

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small stones</td>
<td>81 successes / 87 (93%)</td>
<td>234 successes / 270 (87%)</td>
</tr>
<tr>
<td>Large stones</td>
<td>192 successes / 263 (73%)</td>
<td>55 successes / 80 (69%)</td>
</tr>
<tr>
<td>Combined</td>
<td>273 successes / 350 (78%)</td>
<td>289 successes / 350 (83%)</td>
</tr>
</tbody>
</table>

Simpson's Paradox

"When the less effective treatment is applied more frequently to easier cases, it can appear to be a more effective treatment."

A random variable assigns a real number to each possible outcome of an experiment.

The distribution of a random variable $X$ is the function

$$r \rightarrow P(X = r)$$

The expectation (average, expected value) of random variable $X$ on sample space $S$ is

$$E(X) = \sum_{s \in S} P(s)X(s)$$

$$= \sum_{r \in X(S)} P(X = r)r$$
The expectation (average, expected value) of random variable $X$ on sample space $S$ is

$$E(X) = \sum_{s \in S} P(s)X(s)$$

$$= \sum_{r \in X(S)} P(X = r)r$$

Calculate the expected number of boys in a family with two children.

A. 0
B. 1
C. 1.5
D. 2
Expected Value Examples

The **expectation** (average, expected value) of random variable $X$ on sample space $S$ is

$$E(X) = \sum_{s \in S} P(s)X(s)$$

$$= \sum_{r \in X(S)} P(X = r)r$$

Calculate the expected number of boys in a family with **three** children.

A. 0  
B. 1  
C. 1.5  
D. 2

*Rosen p. 460,478*
Expected Value Examples

The **expectation** (average, expected value) of random variable $X$ on sample space $S$ is

$$E(X) = \sum_{s \in S} P(s)X(s)$$

$$= \sum_{r \in X(S)} P(X=r)r$$

Calculate the expected number of boys in a family with three children.

A. 0  
B. 1  
C. 1.5  
D. 2

The expected value might not be a possible value of the random variable... like 1.5 boys!
The expectation (average, expected value) of random variable $X$ on sample space $S$ is

$$E(X) = \sum_{s \in S} P(s)X(s)$$

$$= \sum_{r \in \mathcal{X}(S)} P(X = r)r$$

Calculate the expected sum of two 6-sided dice.

A. 6  
B. 7  
C. 8  
D. 9  
E. None of the above.
Reminders

**Midterm 2**: this Monday, February 29 in class, covers material after the first midterm through Wednesday (not cumulative)

* Practice midterm on website/Piazza.
* Review sessions Thursday & Saturday: see website/Piazza.
* **Seating chart on website/Piazza.**
* One double-sided handwritten note sheet allowed.
* If you have AFA letter, see me as soon as possible.