Math Camp

Justin Grimmer

Associate Professor
Department of Political Science
University of Chicago

September 11th, 2017

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Intuition:

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Intuition:

- Some event has occurred: an outcome was realized

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Intuition:

- Some event has occurred: an outcome was realized
- And with the knowledge that this outcome has already happened

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Intuition:

- Some event has occurred: an outcome was realized
- And with the knowledge that this outcome has already happened
- What is the probability that something in another set happens?

Conditional Probability

Social scientists almost always examine conditional relationships

- Given opposite Party ID, probability of date
- Given low-interest rates probability of high inflation
- Given "economic anxiety" probability of voting for Trump

Intuition:

- Some event has occurred: an outcome was realized
- And with the knowledge that this outcome has already happened
- What is the probability that something in another set happens?

Let's formalize this idea.

Conditional Probability: Definition

Definition

Suppose we have two events, E and F, and that $P(F) > 0$. Then,

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

Conditional Probability: Definition

Definition

Suppose we have two events, E and F, and that $P(F) > 0$. Then,

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

- $P(E \cap F)$: Both E and F must occur

Conditional Probability: Definition

Definition

Suppose we have two events, E and F, and that $P(F) > 0$. Then,

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

- $P(E \cap F)$: Both E and F must occur
- $P(F)$ normalize: we know $P(F)$ already occurred

Some Examples

Some Examples

Example 1:

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Some Examples

Example 1:

- $F = \{$All Democrats Win$\}$
- $E = \{$Nancy Pelosi Wins (D-CA)$\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$

Some Examples

Example 1:

- $F = \{$All Democrats Win$\}$
- $E = \{$Nancy Pelosi Wins (D-CA)$\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{$All Democrats Win$\}$
- $E = \{$Louie Gohmert Wins (R-TX) $\}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Example 3: Incumbency Advantage

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Example 3: Incumbency Advantage

- $I = \{\text{Candidate is an incumbent}\}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Example 3: Incumbency Advantage

- $I = \{\text{Candidate is an incumbent}\}$
- $D = \{\text{Candidate Defeated}\}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Example 3: Incumbency Advantage

- $I = \{\text{Candidate is an incumbent}\}$
- $D = \{\text{Candidate Defeated}\}$
- $P(D|I) = \frac{P(D \cap I)}{P(I)}$

Some Examples

Example 1:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Nancy Pelosi Wins (D-CA)}\}$
- If F occurs then E most occur, $P(E|F) = 1$

Example 2:

- $F = \{\text{All Democrats Win}\}$
- $E = \{\text{Louie Gohmert Wins (R-TX) }\}$
- $F \cap E = \emptyset \Rightarrow P(E|F) = \frac{P(F \cap E)}{P(F)} = \frac{P(\emptyset)}{P(F)} = 0$

Example 3: Incumbency Advantage

- $I = \{\text{Candidate is an incumbent}\}$
- $D = \{\text{Candidate Defeated}\}$
- $P(D|I) = \frac{P(D \cap I)}{P(I)}$
- In words?

Conditional Probabilities are Probabilities

Everything we proved previously holds for $P(\cdot|B)$.

Conditional Probabilities are Probabilities

Everything we proved previously holds for $P(\cdot|B)$.

- $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

Conditional Probabilities are Probabilities

Everything we proved previously holds for $P(\cdot|B)$.

- $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$
- Suppose $E_1, E_2, \ldots, E_N$ are mutually exclusive.
  Recall: $(\cup_{i=1}^N E_i) \cap B = \cup_{i=1}^N E_i \cap B$

$$\begin{aligned}
P(\cup_{i=1}^N E_i | B) &= \frac{P(\cup_{i=1}^N E_i \cap B)}{P(B)} \\
&= \frac{\sum_{i=1}^N P(E_i \cap B)}{P(B)} \\
&= \sum_{i=1}^N P(E_i|B)
\end{aligned}$$

Conditional Probabilities are Probabilities

Everything we proved previously holds for $P(\cdot|B)$.

- $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$
- Suppose $E_1, E_2, \ldots, E_N$ are mutually exclusive.
  Recall: $(\cup_{i=1}^{N} E_i) \cap B = \cup_{i=1}^{N} E_i \cap B$

$$\begin{aligned} P(\cup_{i=1}^{N} E_i|B) &= \frac{P(\cup_{i=1}^{N} E_i \cap B)}{P(B)} \\ &= \frac{\sum_{i=1}^{N} P(E_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{N} P(E_i|B) \end{aligned}$$

We are calculating probabilities in the new "universe" B

$P(A|B)$ and $P(B|A)$ are usually different

$P(A|B)$ and $P(B|A)$ are usually different

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(A|B)$ and $P(B|A)$ are usually different

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(B|A) &= \frac{P(A \cap B)}{P(A)}
\end{aligned}
$$

$P(A|B)$ and $P(B|A)$ are usually different

Numerous serious examples: Why American Hate Welfare (Gilens 1995), Police Involved Shootings (Streeter 2017)

$P(A|B)$ and $P(B|A)$ are usually different

Numerous serious examples: Why American Hate Welfare (Gilens 1995), Police Involved Shootings (Streeter 2017)
Less Serious Example

$P(A|B)$ and $P(B|A)$ are usually different

Numerous serious examples: Why American Hate Welfare (Gilens 1995), Police Involved Shootings (Streeter 2017)
Less Serious Example$\leadsto$ type of person who flies to vegas on Southwest Airlines

$P(A|B)$ and $P(B|A)$ are usually different

Numerous serious examples: Why American Hate Welfare (Gilens 1995), Police Involved Shootings (Streeter 2017)
Less Serious Example$\rightsquigarrow$ type of person who flies to vegas on Southwest Airlines

$$P(\text{Cutoff Shirt}|\text{Southwest Airlines}) = 0.2$$

$P(A|B)$ and $P(B|A)$ are usually different

Numerous serious examples: Why American Hate Welfare (Gilens 1995), Police Involved Shootings (Streeter 2017)
Less Serious Example$\rightsquigarrow$ type of person who flies to vegas on Southwest Airlines

$$\begin{aligned} P(\text{Cutoff Shirt}|\text{Southwest Airlines}) &= 0.2 \\ P(\text{Southwest Airlines}|\text{Cutoff Shirt}) &\approx 1 \end{aligned}$$

Proposition

Multiplication Rule: Suppose $E_1, E_2, \ldots, E_N$ is a sequence of events.

$$P(E_1 \cap E_2 \cap \cdots \cap E_N) = P(E_1)P(E_2|E_1)P(E_3|E_2, E_1) \times \cdots \times P(E_N|E_{N-1}, E_{N-2}, \ldots, E_1)$$

Proof.

$\square$

Proposition

Multiplication Rule: Suppose $E_1, E_2, \ldots, E_N$ is a sequence of events.

$$P(E_1 \cap E_2 \cap \cdots \cap E_N) = P(E_1)P(E_2|E_1)P(E_3|E_2, E_1) \times \cdots \times P(E_N|E_{N-1}, E_{N-2}, \ldots, E_1)$$

Proof.

$$\begin{aligned} P(E_1)P(E_2|E_1) &= P(E_1)\frac{P(E_2 \cap E_1)}{P(E_1)} \\ &= P(E_1 \cap E_2) \end{aligned}$$

Proposition

Multiplication Rule: Suppose $E_1, E_2, \ldots, E_N$ is a sequence of events.

$$\begin{aligned} P(E_1 \cap E_2 \cap \cdots \cap E_N) &= \\ P(E_1)P(E_2|E_1)P(E_3|E_2, E_1) \times \cdots \times P(E_N|E_{N-1}, E_{N-2}, \ldots, E_1) \end{aligned}$$

Proof.

$$\begin{aligned} P(E_1)P(E_2|E_1) &= P(E_1)\frac{P(E_2 \cap E_1)}{P(E_1)} \\ &= P(E_1 \cap E_2) \\ P(E_1 \cap E_2)P(E_3|E_1, E_2) &= P(E_1 \cap E_2)\frac{P(E_3 \cap E_2 \cap E_1)}{P(E_2 \cap E_1)} \\ &= P(E_3 \cap E_2 \cap E_1) \end{aligned}$$

$\square$

Proposition

Multiplication Rule: Suppose $E_1, E_2, \ldots, E_N$ is a sequence of events.

$$\begin{aligned} P(E_1 \cap E_2 \cap \cdots \cap E_N) &= \\ P(E_1)P(E_2|E_1)P(E_3|E_2, E_1) \times \cdots \times P(E_N|E_{N-1}, E_{N-2}, \ldots, E_1) \end{aligned}$$

Proof.

$$\begin{aligned} P(E_1)P(E_2|E_1) &= P(E_1)\frac{P(E_2 \cap E_1)}{P(E_1)} \\ &= P(E_1 \cap E_2) \\ P(E_1 \cap E_2)P(E_3|E_1, E_2) &= P(E_1 \cap E_2)\frac{P(E_3 \cap E_2 \cap E_1)}{P(E_2 \cap E_1)} \\ &= P(E_3 \cap E_2 \cap E_1) \end{aligned}$$

Repeating for all probabilities proves the proposition

□

Law of Total Probability

Proposition

Suppose that we have a set of events $F_1, F_2, \ldots, F_N$ *such that the events are mutually exclusive and together comprise the entire sample space* $\cup_{i=1}^{N} F_i = \textit{Sample Space}$*. Then, for any event* E

$$P(E) \;=\; \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Law of Total Probability

$$P(E) = \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Law of Total Probability

$$P(E) \;=\; \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Suppose $F_1, F_2, \ldots, F_N$ are mutually exclusive and $\cup_{i=1}^{N} F_i = S$. Then we can write E as:

Law of Total Probability

$$P(E) = \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Suppose $F_1, F_2, \ldots, F_N$ are mutually exclusive and $\cup_{i=1}^{N} F_i = S$. Then we can write E as:

$$E = (E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N)$$

Law of Total Probability

$$P(E) = \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Suppose $F_1, F_2, \ldots, F_N$ are mutually exclusive and $\cup_{i=1}^{N} F_i = S$. Then we can write E as:

$$\begin{aligned} E &= (E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N) \\ P(E) &= P\left((E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N)\right) \end{aligned}$$

Law of Total Probability

$$P(E) = \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Suppose $F_1, F_2, \ldots, F_N$ are mutually exclusive and $\cup_{i=1}^{N} F_i = S$. Then we can write E as:

$$\begin{aligned}
E &= (E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N) \\
P(E) &= P\left((E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N)\right) \\
&= \sum_{i=1}^{N} P(E \cap F_i)
\end{aligned}$$

Law of Total Probability

$$P(E) = \sum_{i=1}^{N} P(E|F_i) \times P(F_i)$$

Proof.

Suppose $F_1, F_2, \ldots, F_N$ are mutually exclusive and $\cup_{i=1}^{N} F_i = S$. Then we can write E as:

$$\begin{aligned}
E &= (E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N) \\
P(E) &= P\left((E \cap F_1) \cup (E \cap F_2) \ldots \cup (E \cap F_N)\right) \\
&= \sum_{i=1}^{N} P(E \cap F_i) \\
&= \sum_{i=1}^{N} P(E|F_i)P(F_i)
\end{aligned}$$

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$
- $P(\text{mobilized}) = 0.6; P(\text{not mobilized}) = 0.4$

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$
- $P(\text{mobilized}) = 0.6; P(\text{not mobilized}) = 0.4$
- What is $P(\text{vote})$?

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$
- $P(\text{mobilized}) = 0.6; P(\text{not mobilized}) = 0.4$
- What is $P(\text{vote})$?

Sample space (one person) $=$
{ (mobilized, vote), (mobilized, not vote), (not mobilized, vote) , (not mobilized, not vote) }
Mobilization partitions the space (mutually exclusive and exhaustive)

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$
- $P(\text{mobilized}) = 0.6; P(\text{not mobilized}) = 0.4$
- What is $P(\text{vote})$?

Sample space (one person) =
{ (mobilized, vote), (mobilized, not vote), (not mobilized, vote) , (not mobilized, not vote) }
Mobilization partitions the space (mutually exclusive and exhaustive)
We can use the law of total probability

Example: Law of Total Probability

Infer $P(\text{vote})$ after mobilization campaign

- $P(\text{vote}|\text{mobilized}) = 0.75$
- $P(\text{vote}|\text{not mobilized}) = 0.25$
- $P(\text{mobilized}) = 0.6; P(\text{not mobilized}) = 0.4$
- What is $P(\text{vote})$?

Sample space (one person) =
{ (mobilized, vote), (mobilized, not vote), (not mobilized, vote) , (not mobilized, not vote) }
Mobilization partitions the space (mutually exclusive and exhaustive)
We can use the law of total probability

$$
\begin{aligned}
P(\text{vote}) &= P(\text{mob.}) \times P(\text{vote}|\text{mob.}) + P(\text{not mob} \times P(\text{vote}|\text{not mob}) \\
&= 0.6 \times 0.75 + 0.4 \times 0.25 \\
&= 0.55
\end{aligned}
$$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.

Example Law of Total Probability

Mixture Models: flexible modeling strategy.

Two coins:

Example Law of Total Probability

Mixture Models: flexible modeling strategy.

Two coins:

- Fair: $P(H) = 1/2$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Draw a coin from urn ($P(\text{fair}) = 1/2$) and then flip.
$P(H)$?

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Draw a coin from urn ($P(\text{fair}) = 1/2$) and then flip.
$P(H)$?
$S = \{(\text{fair}, H), (\text{fair}, T), (\text{bias}, H), (\text{bias}, T)\}$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Draw a coin from urn ($P(\text{fair}) = 1/2$) and then flip.
$P(H)$?
$S = \{(\text{fair}, H), (\text{fair}, T), (\text{bias}, H), (\text{bias}, T)\}$

$$P(H) = P(\text{fair}) \times P(H|\text{fair}) + P(\text{bias}) \times P(H|\text{bias})$$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Draw a coin from urn ($P(\text{fair}) = 1/2$) and then flip.
$P(H)$?
$S = \{(\text{fair}, H), (\text{fair}, T), (\text{bias}, H), (\text{bias}, T)\}$

$$\begin{aligned} P(H) &= P(\text{fair}) \times P(H|\text{fair}) + P(\text{bias}) \times P(H|\text{bias}) \\ &= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{4} \\ &= \frac{5}{8} \end{aligned}$$

Example Law of Total Probability

Mixture Models: flexible modeling strategy.
Two coins:

- Fair: $P(H) = 1/2$
- Biased $P(H) = 3/4$

Draw a coin from urn ($P(\text{fair}) = 1/2$) and then flip.
$P(H)$?
$S = \{(\text{fair}, H), (\text{fair}, T), (\text{bias}, H), (\text{bias}, T)\}$

$$\begin{aligned}
P(H) &= P(\text{fair}) \times P(H|\text{fair}) + P(\text{bias}) \times P(H|\text{bias}) \\
&= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{4} \\
&= \frac{5}{8}
\end{aligned}$$

Mixture of two coins

Bayes' Rule

- $P(B|A)$ may be easy to obtain
- $P(A|B)$ may be harder to determine
- Bayes' rule provides a method to move from $P(B|A)$ to $P(A|B)$.

Definition

Bayes' Rule: For two events A and B,

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Proof.

$\square$

Definition

Bayes' Rule: For two events A and B,

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Proof.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

□

Definition

Bayes' Rule: For two events A and B,

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Proof.

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

□

Bayes' Rule: Example

Enos (2011), Fraga (2015), Imai and Khanna (2015): how do we identify racial groups from lists of names?

Bayes' Rule: Example

Enos (2011), Fraga (2015), Imai and Khanna (2015): how do we identify racial groups from lists of names?

Census Bureau collects information on distribution of names by race.

Bayes' Rule: Example

Enos (2011), Fraga (2015), Imai and Khanna (2015): how do we identify racial groups from lists of names?

Census Bureau collects information on distribution of names by race.

For example, Washington is the "blackest" name in America.

Bayes' Rule: Example

Enos (2011), Fraga (2015), Imai and Khanna (2015): how do we identify racial groups from lists of names?

Census Bureau collects information on distribution of names by race.

For example, Washington is the "blackest" name in America.

- P(black)$=$ 0.126.
- P(not black) $=$ 1 - P(black) $=$ 0.874.
- P(Washington$|$ black) $=$ 0.00378.
- P(Washington$|$nb) $=$ 0.000060615.

Bayes' Rule: Example

Enos (2011), Fraga (2015), Imai and Khanna (2015): how do we identify racial groups from lists of names?

Census Bureau collects information on distribution of names by race.

For example, Washington is the "blackest" name in America.

- P(black)= 0.126.
- P(not black) = 1 - P(black) = 0.874.
- P(Washington| black) = 0.00378.
- P(Washington|nb) = 0.000060615.

$$
\begin{aligned}
P(\text{black}|\text{Wash}) &= \frac{P(\text{black})P(\text{Wash}|\text{black})}{P(\text{Wash})} \\
&= \frac{P(\text{black})P(\text{Wash}|\text{black})}{P(\text{black})P(\text{Wash}|\text{black}) + P(\text{nb})P(\text{Wash}|\text{nb})} \\
&= \frac{0.126 \times 0.00378}{0.126 \times 0.00378 + 0.874 \times 0.000060616} \\
&\approx 0.9
\end{aligned}
$$


MAKE
A
DEAL


LET'S MAK
A DEAL


MARILYN vos SAVANT
Columnist Parade Magazine

"You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!" Scott Smith, Ph.D. University of Florida (From Wikipedia)

Monty Hall Problem

Suppose we have three doors. A, B, C.

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch
- Should the contestant switch?

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch
- Should the contestant switch?

Contestant guesses A

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch
- Should the contestant switch?

Contestant guesses A
$P(A) = 1/3 \rightsquigarrow$ chance of winning without switch

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch
- Should the contestant switch?

Contestant guesses A
$P(A) = 1/3 \rightsquigarrow$ chance of winning without switch
If C is revealed to not have a car:

Monty Hall Problem

Suppose we have three doors. A, B, C.
Behind one door there is a car. Behind the others is a goat (you don't want a goat)

- A contestant guesses a door.
- The host opens a different door and then contestant has option to switch
- Should the contestant switch?

Contestant guesses A
$P(A) = 1/3 \rightsquigarrow$ chance of winning without switch
If C is revealed to not have a car:

Monty Hall Problem

Monty Hall Problem

$$P(B|C \text{ revealed}) = \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)}$$

Monty Hall Problem

$$\begin{aligned}
P(B|C \text{ revealed}) &= \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1}{1/3 + 1/3 \times 1/2} = \frac{1/3}{1/2} = \frac{2}{3}
\end{aligned}$$

Monty Hall Problem

$$
\begin{aligned}
P(B|C \text{ revealed}) &= \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1}{1/3 + 1/3 \times 1/2} = \frac{1/3}{1/2} = \frac{2}{3} \\
P(A|C \text{ revealed}) &= \frac{P(A)P(C \text{ revealed}|A)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)}
\end{aligned}
$$

Monty Hall Problem

$$
\begin{aligned}
P(B|C \text{ revealed}) &= \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1}{1/3 + 1/3 \times 1/2} = \frac{1/3}{1/2} = \frac{2}{3} \\
P(A|C \text{ revealed}) &= \frac{P(A)P(C \text{ revealed}|A)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1/2}{1/3 + 1/3 \times 1/2} = \frac{1}{3}
\end{aligned}
$$

Monty Hall Problem

$$\begin{aligned}
P(B|C \text{ revealed}) &= \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1}{1/3 + 1/3 \times 1/2} = \frac{1/3}{1/2} = \frac{2}{3} \\
P(A|C \text{ revealed}) &= \frac{P(A)P(C \text{ revealed}|A)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1/2}{1/3 + 1/3 \times 1/2} = \frac{1}{3}
\end{aligned}$$

Double chances of winning with switch

Monty Hall Problem

$$\begin{aligned}
P(B|C \text{ revealed}) &= \frac{P(B)P(C \text{ revealed}|B)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1}{1/3 + 1/3 \times 1/2} = \frac{1/3}{1/2} = \frac{2}{3} \\
P(A|C \text{ revealed}) &= \frac{P(A)P(C \text{ revealed}|A)}{P(B)P(C \text{ revealed}|B) + P(A)P(C \text{ revealed}|A)} \\
&= \frac{1/3 \times 1/2}{1/3 + 1/3 \times 1/2} = \frac{1}{3}
\end{aligned}$$

Double chances of winning with switch

`R Code!`

Testing for a Rare Disease

Suppose there is a medical test

$$
\begin{aligned}
P(\text{positive}|\text{disease}) &= 0.99 \\
P(\text{positive}|\text{not disease}) &= 0.10 \\
P(\text{disease}) &= 0.0001
\end{aligned}
$$

Testing for a Rare Disease

Suppose there is a medical test

$$\begin{aligned} P(\text{positive}|\text{disease}) &= 0.99 \\ P(\text{positive}|\text{not disease}) &= 0.10 \\ P(\text{disease}) &= 0.0001 \end{aligned}$$

After a positive test, how worried should we be?

Testing for a Rare Disease

Suppose there is a medical test

$$\begin{aligned} P(\text{positive}|\text{disease}) &= 0.99 \\ P(\text{positive}|\text{not disease}) &= 0.10 \\ P(\text{disease}) &= 0.0001 \end{aligned}$$

After a positive test, how worried should we be?

$$P(\text{disease}|\text{pos.}) = \frac{P(\text{dis.})P(\text{pos}|\text{dis.})}{P(\text{dis.})P(\text{pos}|\text{dis.}) + P(\text{not dis.})P(\text{pos}|\text{not dis.})}$$

Testing for a Rare Disease

Suppose there is a medical test

$$\begin{aligned} P(\text{positive}|\text{disease}) &= 0.99 \\ P(\text{positive}|\text{not disease}) &= 0.10 \\ P(\text{disease}) &= 0.0001 \end{aligned}$$

After a positive test, how worried should we be?

$$\begin{aligned} P(\text{disease}|\text{pos.}) &= \frac{P(\text{dis.})P(\text{pos}|\text{dis.})}{P(\text{dis.})P(\text{pos}|\text{dis.}) + P(\text{not dis.})P(\text{pos}|\text{not dis.})} \\ &= \frac{0.0001 \times 0.99}{0.0001 \times 0.99 + 0.9999 \times 0.1} \end{aligned}$$

Testing for a Rare Disease

Suppose there is a medical test

$$\begin{aligned} P(\text{positive}|\text{disease}) &= 0.99 \\ P(\text{positive}|\text{not disease}) &= 0.10 \\ P(\text{disease}) &= 0.0001 \end{aligned}$$

After a positive test, how worried should we be?

$$\begin{aligned} P(\text{disease}|\text{pos.}) &= \frac{P(\text{dis.})P(\text{pos}|\text{dis.})}{P(\text{dis.})P(\text{pos}|\text{dis.}) + P(\text{not dis.})P(\text{pos}|\text{not dis.})} \\ &= \frac{0.0001 \times 0.99}{0.0001 \times 0.99 + 0.9999 \times 0.1} \\ &\approx 0.0009891 \end{aligned}$$

Independence and Information

Does one event provide information about another event?

Independence and Information

Does one event provide information about another event?

Definition

Independence: Two events E and F are independent if

$$P(E \cap F) = P(E)P(F)$$

If E and F are not independent, we'll say they are dependent

Independence and Information

Does one event provide information about another event?

Definition

Independence: Two events E and F are independent if

$$P(E \cap F) \quad = \quad P(E)P(F)$$

If E and F are not independent, we'll say they are dependent

- Independence is symetric: if F is independent of E, then E is indepenent of F

Example Independence Relationship

Flip a fair coin twice.

$E =$ first flip heads

$F =$ second flip heads

$$
\begin{aligned}
P(E \cap F) &= P(\{(H,H),(H,T)\} \cap \{(H,H),(T,H)\}) \\
&= P(\{(H,H)\}) \\
&= \frac{1}{4} \\
P(E) &= \frac{1}{2} \\
P(F) &= \frac{1}{2} \\
P(E)P(F) &= \frac{1}{2}\frac{1}{2} = \frac{1}{4} = P(E \cap F)
\end{aligned}
$$

Independence: No Information

Suppose E and F are independent. Then,

$$\begin{aligned} P(E|F) &= \frac{P(E \cap F)}{P(F)} \\ &= \frac{P(E)P(F)}{P(F)} \\ &= P(E) \end{aligned}$$

Conditioning on the event F does not modify the probability of E.
No information about E in F

Independence: No Information

Suppose E and F are independent. Then,

$$
\begin{aligned}
P(E|F) &= \frac{P(E \cap F)}{P(F)} \\
&= \frac{P(E)P(F)}{P(F)} \\
&= P(E)
\end{aligned}
$$

Conditioning on the event F does not modify the probability of E.
No information about E in F
Mutually exclusive $\neq$ Independent
Suppose E and F are mutually exclusive events:

$E = \{(H,H),(H,T)\}$; $F = \{(T,H),(T,T)\}$

$E \cap F = \emptyset$

$P(E|F) = 0$; $P(E) = \frac{1}{2}$.

Independence and Complements

Proposition

Suppose A and B are independent events. Then the events A and B^c are also independent.

Proof.

Independence and Complements

Proposition

Suppose A and B are independent events. Then the events A and B^c are also independent.

Proof.

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

Independence and Complements

Proposition

Suppose A and B are independent events. Then the events A and B^c are also independent.

Proof.

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \end{aligned}$$

Independence and Complements

Proposition

Suppose A and B are independent events. Then the events A and B^c are also independent.

Proof.

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \end{aligned}$$

□

Independence and Complements

Proposition

Suppose A and B are independent events. Then the events A and B^c are also independent.

Proof.

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c) \end{aligned}$$

$\square$

Example: Independence and Causal Inference

Selection and Observational Studies

- We often want to infer the effect of some treatment
  - Incumbency on vote return
  - Democracy on war
- Observational studies: observe what we see to make inference
- Problem: units select into treatment
  - Simple example: enroll in job training if I think it will help
  - $P(\text{job}|\text{training in study}) \neq P(\text{job}|\text{forced training})$
- Background characteristic: difference between treatment and control groups
- Experiments (second greatest discovery of 20th century): make background characteristics and treatment status independent

Conditional Probability

Definition

Let E_1 and E_2 be two events. We will say that the events are conditionally independent given E_3 if

$$P(E_1 \cap E_2 | E_3) = P(E_1|E_3)P(E_2|E_3)$$

Proposition

Suppose E_1 and E_2 and E_3 are events such that $P(E_1 \cap E_2) > 0$ and $P(E_2 \cap E_3) > 0$. Then E_1 and E_2 are conditionally independent given E_3 if and only if $P(E_1|E_2 \cap E_3) = P(E_1|E_3)$.

Proof.

Suppose E_1 and E_2 are conditionally independent given E_3. Then

$$\begin{aligned}
P(E_1 \cap E_2|E_3) &= \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_3)} \\
&= \frac{P(E_3)P(E_2|E_3)P(E_1|E_2 \cap E_3)}{P(E_3)} \\
P(E_1|E_3)P(E_2|E_3) &= P(E_2|E_3)P(E_1|E_2 \cap E_3) \\
P(E_1|E_3) &= P(E_1|E_2 \cap E_3)
\end{aligned}$$

□

Proof.

Suppose $P(E_1|E_2 \cap E_3) = P(E_1|E_3)$

$$\begin{aligned} P(E_1 \cap E_2|E_3) &= P(E_2|E_3)P(E_1|E_2 \cap E_3) \\ &= P(E_2|E_3)P(E_1|E_3) \end{aligned}$$

□

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

$E_1 =$ High Quality selected

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

$E_1 =$ High Quality selected
$H_i =$ Event NFU on day i

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

$E_1 =$ High Quality selected
$H_i =$ Event NFU on day i

$$P(H_1 \cap H_2 | E_1) \quad = \quad P(H_1|E_1)P(H_2|E_2)$$

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

$E_1 =$ High Quality selected
$H_i =$ Event NFU on day i

$$P(H_1 \cap H_2 | E_1) \quad = \quad P(H_1|E_1)P(H_2|E_2)$$

But

Conditional Independence

Suppose we want to hire an employee, but applicants have variable quality.

- $1/2$ low quality (LQ): $P(\text{NFU}) = 0.01$ each day
- $1/2$ high quality (HQ) : $P(\text{NFU}) = 0.99$ each day

$E_1 =$ High Quality selected
$H_i =$ Event NFU on day i

$$P(H_1 \cap H_2 | E_1) = P(H_1|E_1)P(H_2|E_2)$$

But

$$\begin{aligned}
P(H_1) &= P(E_1)P(H_1|E_1) + P(E_1^c)P(H_1|E_1^c) = 1/2(0.99) + 1/2(0.01 \\
P(H_2) &= 1/2 \\
P(H_1 \cap H_2) &= P(E_1)P(H_1 \cap H_2|E_1) + P(E_1^c)P(H_1 \cap H_2|E_1^c) \\
&= 0.5(0.99 \times 0.99) + 0.5(0.01 \times 0.01) \approx 0.5
\end{aligned}$$

Definition

Suppose we have a sequence of events $E_1, E_2, \ldots, E_n$*. We say the sequence of events is* *mutually indepenent* *if for each subset of the sequence,* $E_{i_1}, E_{i_2}, \ldots, E_{i_j}$

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_j}) \quad = \quad \prod_{m=1}^{j} P(E_{i_m})$$

For a sequence to be independent, every subset is independent

Definition

Define the odds of some event E as

$$odds_E = \frac{P(E)}{1 - P(E)}$$

Suppose F is another event. Define the odds ratio of E to F as

$$\begin{aligned} odds\ ratio_{E:F} &= \frac{odds_E}{odds_F} \\ &= \frac{\frac{P(E)}{1-P(E)}}{\frac{P(F)}{1-P(F)}} \end{aligned}$$

- Big: implies E is very likely
- Small: implies E is unlikely
- Problem: big changes in odd ratio may correspond to very small changes in chance something will happen $\rightsquigarrow$ baseline problem

Where we're going

Today

- Conditional probability
- Bayes' Rule
- Independence

Next lecture: Random variables (discrete and continuous)