18.440: Lecture 7

Bayes' formula and independence

Scott Sheffield

MIT

Outline

Bayes' formula

Independence

Outline

Bayes' formula

Independence

Recall definition: conditional probability

- Definition: $P(E|F) = P(EF)/P(F)$.

Recall definition: conditional probability

- Definition: $P(E|F) = P(EF)/P(F)$.
- Equivalent statement: $P(EF) = P(F)P(E|F)$.

Recall definition: conditional probability

- Definition: $P(E|F) = P(EF)/P(F)$.
- Equivalent statement: $P(EF) = P(F)P(E|F)$.
- Call $P(E|F)$ the "conditional probability of E given F" or "probability of E conditioned on F".

Dividing probability into two cases

- $$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$

Dividing probability into two cases

- $$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$
- In words: want to know the probability of E. There are two scenarios F and F^c. If I know the probabilities of the two scenarios and the probability of E conditioned on each scenario, I can work out the probability of E.

Dividing probability into two cases

-

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$

- In words: want to know the probability of E. There are two scenarios F and F^c. If I know the probabilities of the two scenarios and the probability of E conditioned on each scenario, I can work out the probability of E.
- Example: $D =$ "have disease", $T =$ "positive test."

Dividing probability into two cases

-

$$P(E) = P(EF) + P(EF^c)$$
$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

- In words: want to know the probability of E. There are two scenarios F and F^c. If I know the probabilities of the two scenarios and the probability of E conditioned on each scenario, I can work out the probability of E.
- Example: $D =$ "have disease", $T =$ "positive test."
- If $P(D) = p$, $P(T|D) = .9$, and $P(T|D^c) = .1$, then $P(T) = .9p + .1(1-p)$.

Dividing probability into two cases

- $$P(E) = P(EF) + P(EF^c)$$
$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

- In words: want to know the probability of E. There are two scenarios F and F^c. If I know the probabilities of the two scenarios and the probability of E conditioned on each scenario, I can work out the probability of E.
- Example: $D =$ "have disease", $T =$ "positive test."
- If $P(D) = p$, $P(T|D) = .9$, and $P(T|D^c) = .1$, then $P(T) = .9p + .1(1-p)$.
- What is $P(D|T)$?

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.
- So $P(A|B)$ is $\frac{P(B|A)}{P(B)}$ times $P(A)$.

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.
- So $P(A|B)$ is $\frac{P(B|A)}{P(B)}$ times $P(A)$.
- Ratio $\frac{P(B|A)}{P(B)}$ determines "how compelling new evidence is".

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.
- So $P(A|B)$ is $\frac{P(B|A)}{P(B)}$ times $P(A)$.
- Ratio $\frac{P(B|A)}{P(B)}$ determines "how compelling new evidence is".
- What does it mean if ratio is zero?

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.
- So $P(A|B)$ is $\frac{P(B|A)}{P(B)}$ times $P(A)$.
- Ratio $\frac{P(B|A)}{P(B)}$ determines "how compelling new evidence is".
- What does it mean if ratio is zero?
- What if ratio is $1/P(A)$?

Bayes' theorem

- Bayes' formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is often invoked as tool to guide intuition.

Bayes' theorem

- Bayes' formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is often invoked as tool to guide intuition.
- Example: A is event that suspect stole the $10,000$ under my mattress, B is event that suspect deposited several thousand dollars in cash in bank last week.

Bayes' theorem

- Bayes' formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is often invoked as tool to guide intuition.
- Example: A is event that suspect stole the $10,000$ under my mattress, B is event that suspect deposited several thousand dollars in cash in bank last week.
- Begin with subjective estimates of $P(A)$, $P(B|A)$, and $P(B|A^c)$. Compute $P(B)$. Check whether B occurred. Update estimate.

Bayes' theorem

- Bayes' formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is often invoked as tool to guide intuition.
- Example: A is event that suspect stole the $10,000$ under my mattress, B is event that suspect deposited several thousand dollars in cash in bank last week.
- Begin with subjective estimates of $P(A)$, $P(B|A)$, and $P(B|A^c)$. Compute $P(B)$. Check whether B occurred. Update estimate.
- Repeat procedure as new evidence emerges.

Bayes' theorem

- Bayes' formula $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ is often invoked as tool to guide intuition.
- Example: A is event that suspect stole the $10,000$ under my mattress, B is event that suspect deposited several thousand dollars in cash in bank last week.
- Begin with subjective estimates of $P(A)$, $P(B|A)$, and $P(B|A^c)$. Compute $P(B)$. Check whether B occurred. Update estimate.
- Repeat procedure as new evidence emerges.
- Caution required. My idea to check whether B occurred, or is a lawyer selecting the provable events $B_1, B_2, B_3, \ldots$ that maximize $P(A|B_1B_2B_3\ldots)$? Where did my probability estimates come from? What is my state space? What assumptions am I making?

“Bayesian” sometimes used to describe philosophical view

- Philosophical idea: we assign subjective probabilities to questions we can’t answer. Will Obama win election? Will Red Sox win world series? Will stock prices go up this year?

"Bayesian" sometimes used to describe philosophical view

- Philosophical idea: we assign subjective probabilities to questions we can't answer. Will Obama win election? Will Red Sox win world series? Will stock prices go up this year?
- Bayes essentially described probability of event as

$$\frac{\text{value of right to get some thing if event occurs}}{\text{value of thing}}.$$

“Bayesian” sometimes used to describe philosophical view

- Philosophical idea: we assign subjective probabilities to questions we can’t answer. Will Obama win election? Will Red Sox win world series? Will stock prices go up this year?
- Bayes essentially described probability of event as

$$\frac{\text{value of right to get some thing if event occurs}}{\text{value of thing}}.$$

- Philosophical questions: do we have subjective probabilities/hunches for questions we can’t base enforceable contracts on? Do there exist other universes? Are there other intelligent beings? Are there beings smart enough to simulate universes like ours? Are we part of such a simulation?...

“Bayesian” sometimes used to describe philosophical view

- Philosophical idea: we assign subjective probabilities to questions we can’t answer. Will Obama win election? Will Red Sox win world series? Will stock prices go up this year?
- Bayes essentially described probability of event as

$$\frac{\text{value of right to get some thing if event occurs}}{\text{value of thing}}.$$

- Philosophical questions: do we have subjective probabilities/hunches for questions we can’t base enforceable contracts on? Do there exist other universes? Are there other intelligent beings? Are there beings smart enough to simulate universes like ours? Are we part of such a simulation?...
- Do we use Bayes subconsciously to update hunches?

“Bayesian” sometimes used to describe philosophical view

- Philosophical idea: we assign subjective probabilities to questions we can’t answer. Will Obama win election? Will Red Sox win world series? Will stock prices go up this year?
- Bayes essentially described probability of event as

$$\frac{\text{value of right to get some thing if event occurs}}{\text{value of thing}}.$$

- Philosophical questions: do we have subjective probabilities/hunches for questions we can’t base enforceable contracts on? Do there exist other universes? Are there other intelligent beings? Are there beings smart enough to simulate universes like ours? Are we part of such a simulation?...
- Do we use Bayes subconsciously to update hunches?
- Should we think of Bayesian priors and updates as part of the epistemological foundation of science and statistics?

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.
- Is there nice way to describe ratio between odds and conditional odds?

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.
- Is there nice way to describe ratio between odds and conditional odds?
- $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} = ?$

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.
- Is there nice way to describe ratio between odds and conditional odds?
- $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} = ?$
- By Bayes $P(A|B)/P(A) = P(B|A)/P(B)$.

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.
- Is there nice way to describe ratio between odds and conditional odds?
- $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} =?$
- By Bayes $P(A|B)/P(A) = P(B|A)/P(B)$.
- After some algebra, $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} = P(B|A)/P(B|A^c)$

Updated "odds"

- Define "odds" of A to be $P(A)/P(A^c)$.
- Define "conditional odds" of A given B to be $P(A|B)/P(A^c|B)$.
- Is there nice way to describe ratio between odds and conditional odds?
- $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} = ?$
- By Bayes $P(A|B)/P(A) = P(B|A)/P(B)$.
- After some algebra, $\frac{P(A|B)/P(A^c|B)}{P(A)/P(A^c)} = P(B|A)/P(B|A^c)$
- Say I think A is 5 times as likely as A^c, and $P(B|A) = 3P(B|A^c)$. Given B, I think A is 15 times as likely as A^c.

$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i|F) = \sum P(E_i|F)$, if i ranges over a countable set and the E_i are disjoint.

$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i|F) = \sum P(E_i|F)$, if i ranges over a countable set and the E_i are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.

$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i|F) = \sum P(E_i|F)$, if i ranges over a countable set and the E_i are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.
- To get former from latter, we set probabilities of elements outside of F to zero and multiply probabilities of events inside of F by $1/P(F)$.

$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i|F) = \sum P(E_i|F)$, if i ranges over a countable set and the E_i are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.
- To get former from latter, we set probabilities of elements outside of F to zero and multiply probabilities of events inside of F by $1/P(F)$.
- It $P(\cdot)$ is the *prior* probability measure and $P(\cdot|F)$ is the *posterior* measure (revised after discovering that F occurs).

Outline

Bayes' formula

Independence

Outline

Bayes' formula

Independence

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H, H), (H, T), (T, H), (T, T)$.

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H, H), (H, T), (T, H), (T, T)$.
- Is event that first coin is heads independent of event that second coin heads.

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H, H), (H, T), (T, H), (T, T)$.
- Is event that first coin is heads independent of event that second coin heads.
- Yes: probability of each event is $1/2$ and probability of both is $1/4$.

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H,H),(H,T),(T,H),(T,T)$.
- Is event that first coin is heads independent of event that second coin heads.
- Yes: probability of each event is $1/2$ and probability of both is $1/4$.
- Is event that first coin is heads independent of event that number of heads is odd?

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H,H),(H,T),(T,H),(T,T)$.
- Is event that first coin is heads independent of event that second coin heads.
- Yes: probability of each event is $1/2$ and probability of both is $1/4$.
- Is event that first coin is heads independent of event that number of heads is odd?
- Yes: probability of each event is $1/2$ and probability of both is $1/4$...

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.
- Example: toss two coins. Sample space contains four equally likely elements $(H, H), (H, T), (T, H), (T, T)$.
- Is event that first coin is heads independent of event that second coin heads.
- Yes: probability of each event is $1/2$ and probability of both is $1/4$.
- Is event that first coin is heads independent of event that number of heads is odd?
- Yes: probability of each event is $1/2$ and probability of both is $1/4$...
- despite intuitive sense that oddness of the number of heads "depends" on the first coin.

Independence of multiple events

- Say $E_1 \dots E_n$ are independent if for each $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots n\}$ we have $P(E_{i_1} E_{i_2} \dots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \dots P(E_{i_k})$.

Independence of multiple events

- Say $E_1 \ldots E_n$ are independent if for each $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\}$ we have $P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \ldots P(E_{i_k})$.
- In other words, the product rule works.

Independence of multiple events

- Say $E_1 \dots E_n$ are independent if for each $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots n\}$ we have $P(E_{i_1} E_{i_2} \dots E_{i_k}) = P(E_{i_1})P(E_{i_2}) \dots P(E_{i_k})$.
- In other words, the product rule works.
- Independence implies $P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)P(E_6)}{P(E_4)P(E_5)P(E_6)} = P(E_1 E_2 E_3)$, and other similar statements.

Independence of multiple events

- Say $E_1 \ldots E_n$ are independent if for each $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\}$ we have $P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \ldots P(E_{i_k})$.
- In other words, the product rule works.
- Independence implies $P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)P(E_6)}{P(E_4)P(E_5)P(E_6)} = P(E_1 E_2 E_3)$, and other similar statements.
- Does pairwise independence imply independence?

Independence of multiple events

- Say $E_1 \dots E_n$ are independent if for each $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots n\}$ we have $P(E_{i_1} E_{i_2} \dots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \dots P(E_{i_k})$.
- In other words, the product rule works.
- Independence implies $P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)P(E_6)}{P(E_4)P(E_5)P(E_6)} = P(E_1 E_2 E_3)$, and other similar statements.
- Does pairwise independence imply independence?
- No. Consider these three events: first coin heads, second coin heads, odd number heads. Pairwise independent, not independent.

Independence: another example

- Shuffle 4 cards with labels 1 through 4. Let $E_{j,k}$ be event that card j comes before card k. Is $E_{1,2}$ independent of $E_{3,4}$?

Independence: another example

- Shuffle 4 cards with labels 1 through 4. Let $E_{j,k}$ be event that card j comes before card k. Is $E_{1,2}$ independent of $E_{3,4}$?
- Is $E_{1,2}$ independent of $E_{1,3}$?

Independence: another example

- Shuffle 4 cards with labels 1 through 4. Let $E_{j,k}$ be event that card j comes before card k. Is $E_{1,2}$ independent of $E_{3,4}$?
- Is $E_{1,2}$ independent of $E_{1,3}$?
- No. In fact, what is $P(E_{1,2}|E_{1,3})$?

Independence: another example

- Shuffle 4 cards with labels 1 through 4. Let $E_{j,k}$ be event that card j comes before card k. Is $E_{1,2}$ independent of $E_{3,4}$?
- Is $E_{1,2}$ independent of $E_{1,3}$?
- No. In fact, what is $P(E_{1,2}|E_{1,3})$?
- Generalize to $n > 7$ cards. What is $P(E_{1,7}|E_{1,2}E_{1,3}E_{1,4}E_{1,5}E_{1,6})$?