9/17: Conditioning and Bayes' Rule

Announcements

Exam 1: October 22 (in class)

Exam 2: November 19 (in class)

Final: Out on December 5, Back on December 6 (take-home)

Assignment 2 will be out today; due next Tuesday

If you're not on the course mailing list, please let me know.

Aside: σ-fields

A collection $\mathcal{F}$ of subsets of S is a *σ-field* in case it satisfies

1. $\emptyset \in \mathcal{F}$
2. if $A_1, A_2, \ldots \in \mathcal{F}$ then $\cup_i A_i \in \mathcal{F}$
3. if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

Probabilities are generally defined over a σ-field

Independence

Two events A and B are *independent* in case

$$\Pr(AB) = \Pr(A)\Pr(B)$$

A set of events $\{A_i\}_{i \in I}$ is independent in case

$$\Pr(\cap_{j \in J} A_j) = \prod_{j \in J} \Pr(A_j)$$

for all finite subsets $J \subset I$.

Example

Experiment: draw a card at random from a deck of 52 playing cards.

Are the suit and rank independent?

Example (cont)

$$
\begin{aligned}
\text{Pr(spade king)} &= \frac{1}{52} \\
&= \text{Pr(spade)Pr(king)} \\
&= \frac{1}{4} \cdot \frac{4}{52}
\end{aligned}
$$

Independence $\neq$ Disjoint

To disjoint events can only be independent if one of them has probability zero.

$$\text{Pr}(AB) = \text{Pr}(\emptyset) = 1 - \text{Pr}(S) = 0$$

Independence $\neq$ Pairwise Independence

Pairwise independence does not imply independence.

$$S = \{abc, acb, cab, cba, bca, bac, aaa, bbb, ccc\}$$
$$\text{Pr}(\omega_i) = \frac{1}{9}$$
$$A_i = (i\text{-th letter is } a)$$

These are pairwise independent but not independent events

Conditional Independence

Defined the same way:

Events A and B are *conditionally independent given C* in case

$$\Pr(AB \mid C) = \Pr(A \mid C) \Pr(B \mid C)$$

A set of events $\{A_i\}_{i \in I}$ is conditionally independent given C in case

$$\Pr(\cap_{j \in J} A_j \mid C) = \prod_{j \in J} \Pr(A_j \mid C)$$

for all finite subsets $J \subset I$.

Conditional Independence

Note: Conditional independence of A and B given C is neither implied nor implied by the independence of A and B (exercise).

Example

A court is investigating the possible occurrence of an unlikely event E. Two witnesses, Alice and Bob are called in–each known to tell the truth with probability $\frac{9}{10}$, and there is no collusion between them.

Suppose $\Pr(E) = 0.001$. What is the probability that E occurred given that both Alice and Bob say it did?

Example

Let A and B be the events that Alice and Bob say that E occurred.

$$\begin{aligned}
\Pr(E \mid AB) &= \frac{\Pr(EAB)}{\Pr(AB)} \\
&= \frac{\frac{1}{1000} \cdot \frac{9}{10} \cdot \frac{9}{10}}{\frac{1}{1000} \cdot \frac{9}{10} \cdot \frac{9}{10} + \frac{999}{1000} \cdot \frac{1}{10} \cdot \frac{1}{10}} \\
&= 0.075
\end{aligned}$$

Increases by factor of 75.

Example

A clinical drug trial has the following results

	Women		Men	
	Drug I	Drug II	Drug I	Drug II
Success	200	10	19	1000
Failure	1800	190	1	1000

Which drug is better?

View 1

	Women		Men	
	Drug I	Drug II	Drug I	Drug II
Success	200	10	19	1000
Failure	1800	190	1	1000

Drug I success rate: $\frac{219}{2020}$

Drug II success rate: $\frac{1010}{2200}$

Conclusion: Drug II better.

View 2

	Women		Men	
	Drug I	Drug II	Drug I	Drug II
Success	200	10	19	1000
Failure	1800	190	1	1000

Among women, Drug I vs. II success rates: $\frac{1}{10}$ vs. $\frac{1}{20}$

Among men, Drug I vs. II success rates: $\frac{19}{20}$ vs. $\frac{1}{2}$

Drug I better in both cases

"Simpson's Paradox"

Given three events A, B, C, it is possible to have

$$
\begin{aligned}
\Pr(A \mid B\,C) &> \Pr(A \mid B^c\,C) \\
\Pr(A \mid BC^c) &> \Pr(A \mid B^c\,C^c) \\
\Pr(A \mid B) &< \Pr(A \mid B^c)
\end{aligned}
$$

$\Pr(A \mid B)$ can behave very differently from $\Pr(B \mid A)$

Take care when reasoning probabilistically about events to the right of the bar!

Example

Berkeley graduate admission data:

	Business School	
	Admit	Deny
Men	480	120
Women	180	20

	Law School	
	Admit	Deny
Men	10	90
Women	100	200

Women admitted at higher rate in both schools

Example

Pooled data:

	Overall	
	Admit	Deny
Men	490	210
Women	280	220

Women admitted at lower rate overall

"Bias in Graduate Admissions: Data from Berkeley," P. Bickel et al., *Science*, 1975

Example

From a study of thyroid and heart disease:

	Age 55–64	
	Dead	Alive
Smokers	51	64
Non-smokers	40	81

	Age 65–74	
	Dead	Alive
Smokers	29	7
Non-smokers	101	28

Higher percentage of smokers die in each group

Example

Pooled data:

	Pooled	
	Dead	Alive
Smokers	80	71
Non-smokers	141	109

Appears smokers have a lower death rate. What's going on?

Bayes' Rule

Suppose that $B_1, B_2 \ldots, B_k$ form a partition of S; $B_i \cap B_j = \emptyset$ and $\cup_i B_i = S$. Suppose that $\Pr(B_i) > 0$ and $\Pr(A) > 0$. Then

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

Think of it as

$$\Pr(B_i \mid A) \;\propto\; \Pr(B_i)\Pr(A \mid B_i)$$

Bayes' Rule

B_iA are disjoint and satisfy $\cup_i B_iA = A$. So,

$$\begin{aligned}\Pr(A) &= \Pr(\cup_i B_iA)\\ &= \sum_i \Pr(B_iA)\\ &= \sum_i \Pr(B_i)\Pr(A \mid B_i)\end{aligned}$$

Bayes' Rule

Therefore

$$\begin{aligned}\Pr(B_i \mid A) &= \frac{\Pr(B_iA)}{\Pr(A)}\\ &= \frac{\Pr(B_iA)}{\sum_i \Pr(B_i)\Pr(A \mid B_i)}\\ &= \frac{\Pr(B_i)\Pr(A \mid B_i)}{\sum_i \Pr(B_i)\Pr(A \mid B_i)}\end{aligned}$$

Bayes' Rule

In spite of its simplicity, Bayes rule is very important.

Basis for reasoning probabilistically (in particular about events "on the other side of the bar")

Probabilities can be "updated" to take into account observed events

Conditional Form of Bayes' Rule

It should come as no surprise that the conditional form of Bayes rule follows directly:

Suppose that $B_1, B_2 \ldots, B_k$ form a partition of S; $B_i \cap B_j = \emptyset$ and $\cup_i B_i = S$. Suppose that $\Pr(B_i\, C) > 0$ and $\Pr(A\, C) > 0$. Then

$$\Pr(B_i \,|\, A\, C) = \frac{\Pr(B_i \,|\, C) \Pr(A \,|\, B_i\, C)}{\sum_{j=1}^{k} \Pr(B_j \,|\, C) \Pr(A \,|\, B_j\, C)}$$

Note: $\sum_{j=1}^{k} \Pr(B_j \,|\, C) \Pr(A \,|\, B_j\, C) = \Pr(A \,|\, C)$

Example

A rare disease affects 1 out of 100,000 people. A test shows positive with probability 0.99 when applied to an ill person, and with probability 0.01 when applied to a healthy person.

What is the probability that you have the disease given that you test positive?

Example

$$\begin{aligned}
\Pr(\text{ill} \,|\, \text{pos}) &= \frac{\Pr(\text{pos} \,|\, \text{ill}) \Pr(\text{ill})}{\Pr(\text{pos} \,|\, \text{ill}) \Pr(\text{ill}) + \Pr(\text{pos} \,|\, \text{healthy}) \Pr(\text{healthy})} \\
&= \frac{0.99 \times 10^{-5}}{0.99 \times 10^{-5} + 0.01 \cdot (1 - 10^{-5})} \\
&= \frac{99}{99 + 10^5 - 1} \\
&\approx \frac{1}{1011}
\end{aligned}$$

More likely the test was incorrect...

"Large Scale" Application of Bayes' Rule

Probability models of natural language are difficult to design and build

To translate from S to T we could build a probability model $\Pr(T \,|\, S)$ and translate as $T^* = \mathit{arg\,max}_T \Pr(T \,|\, S)$

It is sometimes better to estimate two models $\Pr(T)$ and $\Pr(S \,|\, T)$ and translate as

$$\begin{aligned}
T^* &= \mathop{\mathit{arg\,max}}_T \Pr(T \,|\, S) \\
&= \mathop{\mathit{arg\,max}}_T \frac{\Pr(S \,|\, T) \Pr(T)}{\Pr(S)} \\
&= \mathop{\mathit{arg\,max}}_T \Pr(S \,|\, T) \Pr(T)
\end{aligned}$$

"True Dogma"

Taken together, these simple relationships give a very powerful calculus for reasoning under uncertainty