Probability II (Conditional Probability)

Sumanta Basu

Announcements

HW1 scores are out – only look at 'HW1 pdf' scores, not 'HW1 screenshot' and 'HW1 Rmarkdown'; points deducted for no/incorrect Rmarkdown submission

Lab solutions posted, HW3 posted (due Monday instead of the usual deadline of Saturday)

For HW/grading questions, email your lab TA first

It is fine to miss iClicker polls for one or two questions (or even lectures), we will only consider **long-run proportions** of your class participation patterns

- Reading: Textbook sections 2.1, 2.2
- Recommended Reading: Chapter 2 supplement (on blackboard) Section 2.3, 2.4
- Recommended exercise: 2.17, 2.23, 2.25

How to describe a random process

An **experiment** (or **trial**)

- generates a realization of a random **outcome**
- flip a coin; roll a die

Sample space - set of all possible outcomes

- flip a coin: $\Omega = \{H, T\}$
- roll a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Event

- a set of outcomes we can describe
- e.g., $A = \{\text{Roll an even number}\} = \{2, 4, 6\}$
- e.g., $B = \{\text{Roll higher than a } 5\} = \{6\}$

Probability as long-term frequency

- We can't know whether one flip will be heads.
- But we can talk about the **probability** of heads.

> **Probability** - proportion of times something would occur if we could repeat random process an infinite number of times.

$P(H) = \frac{1}{2}$ means we expect about 5000 Heads in 10000 flips.

Recall (Probability behaves like area…)

$P(\Omega) = 1$

$0 \leq P(A) \leq 1$

$P(A^C) = 1 - P(A)$

Union of events:

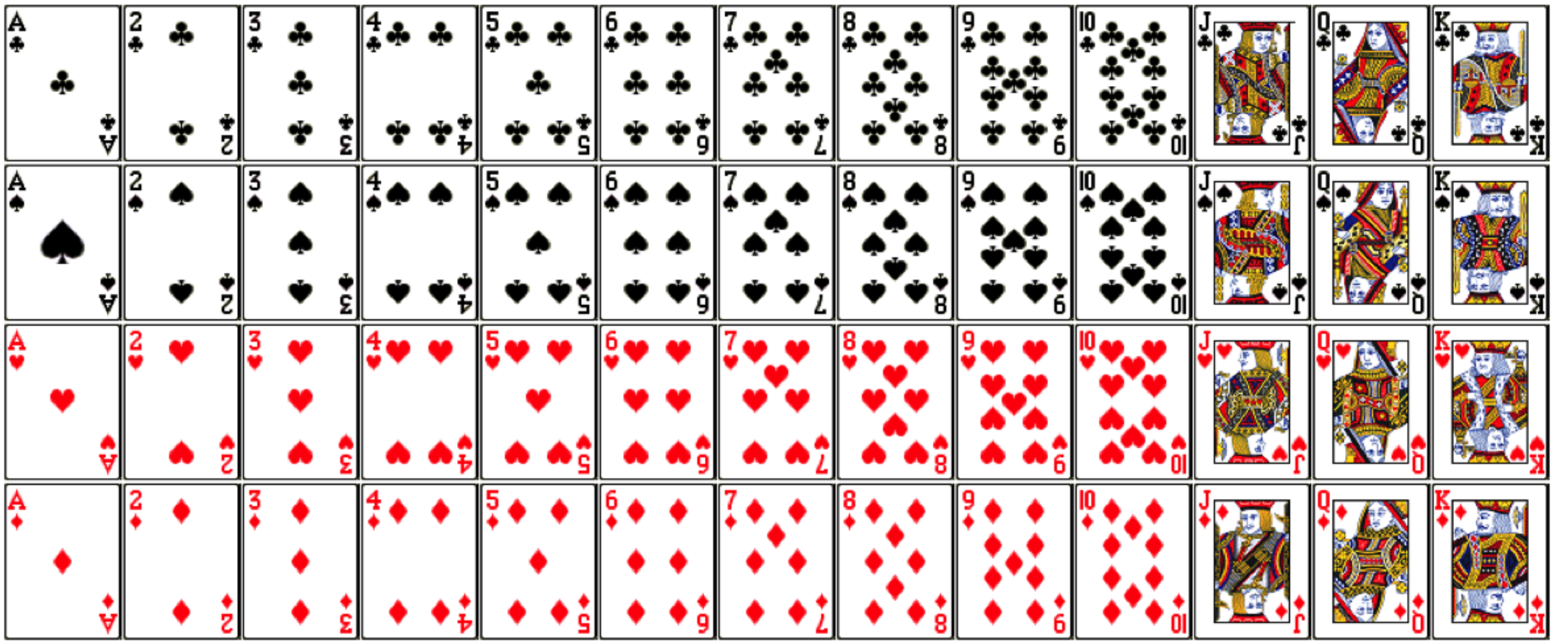
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Disjoint means $A \cap B = \emptyset$

Sometimes will write

- "A or B" to mean $A \cup B$
- "A and B" to mean $A \cap B$

Example: Playing cards

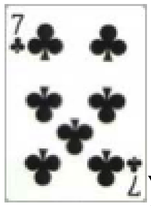


“Pick a card…”

don’t look at it yet

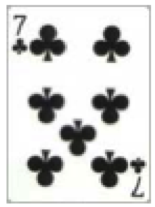
“Pick another card… (with first removed)”

“Pick a card…”

P(first picks )=?

P(second picks )=?

“Pick another card… (with first removed)”

P(second picks  **given**  removed)=?

this is a **conditional probability**

- imagine repeating experiment (of selecting two cards) many, many times. Restricting attention to only those experiments in which first card is 7C, what proportion of the time is second card QH?
- note:

$$P(\text{QH on second draw})$$
$$\neq P(\text{QH on second draw given 7C removed})$$

Conditional probability

For events A and B,

$$P(A|B)$$

denotes

- "Probability of A **given** B"
- "Probability of A **conditional on** B"
- Given that you know the outcome is in B, what's the probability that the outcome also lies in A?

Conditional probability: definition

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

- draw picture
- probability still behaves like area!

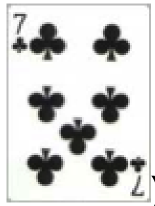
Multiplication Rule

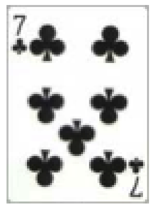
$$P(A \text{ and } B) = P(B)P(A|B)$$

- easiest proof ever?
- and how about this one?

$$P(A \text{ and } B) = P(A)P(B|A)$$

“Pick a card…”

P(first =  and second = )=?

=P(first = )P(second =  | first = )

$= \frac{1}{52} \times \frac{1}{51} = \frac{1}{2652} = 0.038\%$

Independence

Events A and B are **independent** if $P(A|B) = P(A)$.

- Implies $P(A \text{ and } B) = P(A)P(B)$
- Equivalently, $P(B|A) = P(B)$
- Knowing that B occurred doesn't make A any more or less likely
- Events are called **dependent** if they are not independent.

"Pick a card…"

$$A = \{\text{card is a Queen}\} \qquad B = \{\text{card is a Heart}\}$$

Independent?

$$P(A) = \frac{4}{52}$$

$$P(A|B) = \frac{1}{13}$$

Yes!

6010 Prelim Scores

$$A = \{\text{student gets above median on Prelim 1}\}$$
$$B = \{\text{student gets above median on Prelim 2}\}$$

In a previous year,

$$P(A \text{ and } B) = 36\%$$

Are A and B independent?

UC Berkeley graduate admissions (1973)

	Admitted	Rejected
Male	1198	1493
Female	557	1278

- 45% of males admitted
- 30% of females admitted

Evidence of sex bias?

> “The associate dean of the graduate school thought that the university might be sued,” Mr. Bickel says. -WSJ 2009

Is admission independent of sex?

```r
dat1
```

```
##          Admit
## Gender   Admitted Rejected
##   Male       1198     1493
##   Female      557     1278
```

$$A = \{\text{admitted}\} \qquad M = \{\text{male}\}$$

- $P(A)$:

```r
sum(dat1[,1]) / sum(dat1)
```

```
## [1] 0.3877596
```

Is admission independent of sex?

```
dat1
```

```
##          Admit
## Gender   Admitted Rejected
##   Male       1198     1493
##   Female      557     1278
```

$$A = \{\text{admitted}\} \qquad M = \{\text{male}\}$$

- $P(A|M)$:

```
numMale = sum(dat1[1,])
dat1[1,1] / numMale
```

```
## [1] 0.4451877
```

Is admission independent of sex?

Apparently, not.

$$P(A|M) \neq P(A)$$

In particular,

$$P(A|M) > P(A|F)$$

It's a more interesting story, however…

Admission by department

	Men			Women	
	Number applied	% admitted		Number applied	% admitted
A	825	62.1	A	108	82.4
B	560	63.0	B	25	68.0
C	325	36.9	C	593	34.1
D	417	33.1	D	375	34.9
E	191	27.7	E	393	23.9
F	373	5.9	F	341	7.0

These are of form $P(A \mid M \text{ and Dept. "C"})$

Observation

	Men			Women	
	Number applied	% admitted		Number applied	% admitted
A	825	62.1	A	108	82.4
B	560	63.0	B	25	68.0
C	325	36.9	C	593	34.1
D	417	33.1	D	375	34.9
E	191	27.7	E	393	23.9
F	373	5.9	F	341	7.0

51% of men applied to fields A and B.

7% of women applied to fields A and B.

What happened?

- data should be viewed as **3-dimensional** contingency table: **admission vs. major vs. gender**
- **aggregating** this into 2-dimensional table, **admission vs. gender** is **misleading**
- table "associations" can appear, disappear, or **change direction** when you collapse over important table dimensions
- phenomenon called **"Simpson's Paradox"**
- article on this appeared in Science, written by Berkeley professors!

> Unconditional probabilities can be very different from conditional probabilities

Law of Total Probability

Simple idea: Decompose B into part where A occurs and part where it doesn't:

$$B = (B \cap A) \cup (B \cap A^C)$$

- draw picture (probability still behaves like area!)
- Are $B \cap A$ and $B \cap A^C$ disjoint events?

$$P(B) = P(B \cap A) + P(B \cap A^C)$$

$$P(B) = P(A)P(B|A) + P(A^C)P(B|A^C)$$

Bayes Theorem

Recall:

$$P(B)P(A|B) = P(A \text{ and } B) = P(B \text{ and } A) = P(A)P(B|A)$$

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

Use law of total probability: >

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^C)P(B \mid A^C)}$$

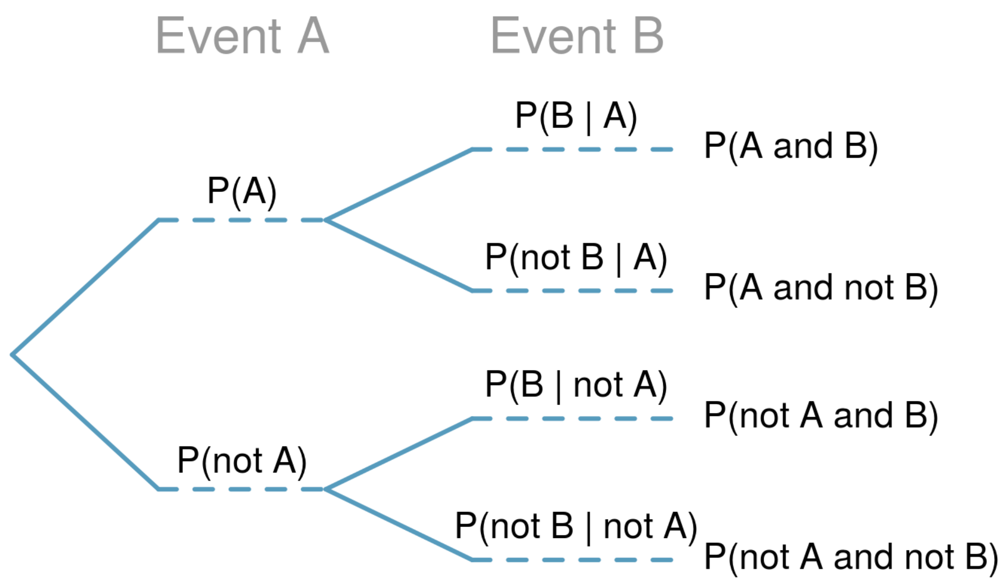
Bayes Theorem

Don’t memorize. Derive from two facts:

1. $$P(A \mid B) = \frac{P(B \text{ and } A)}{P(B)}$$

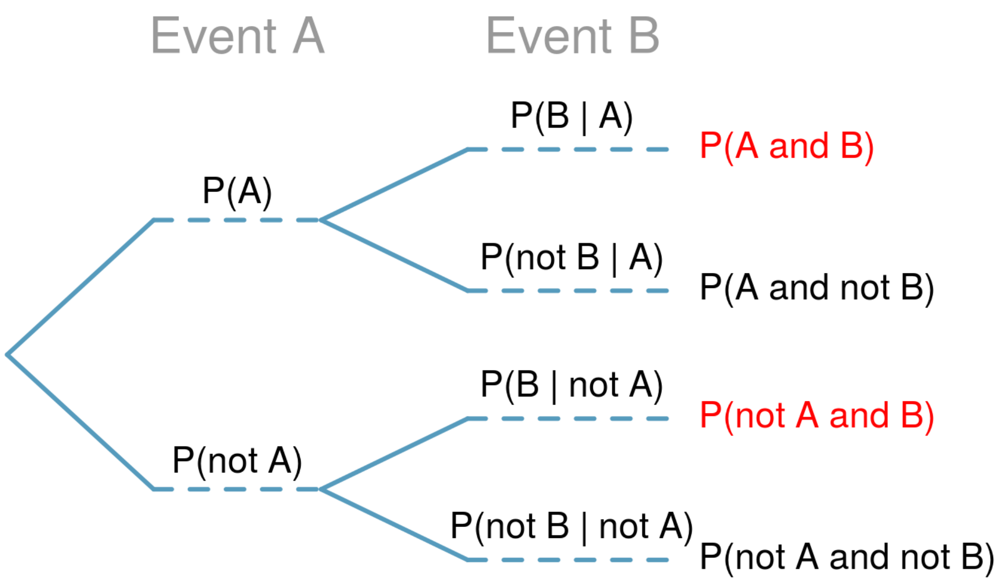
2. $$P(B) = P(A)P(B \mid A) + P(A^C)P(B \mid A^C)$$

Easier with tree diagram?



$$P(A \mid B) = ?$$

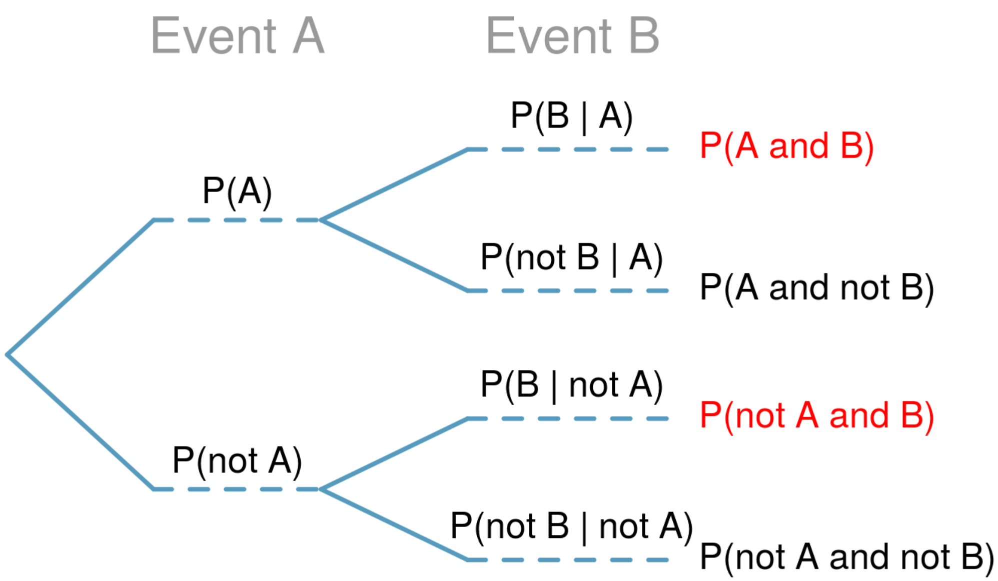
Easier with tree diagram?



$$P(A \mid B) = ?$$

"Given B" - where does B occur?

Easier with tree diagram?



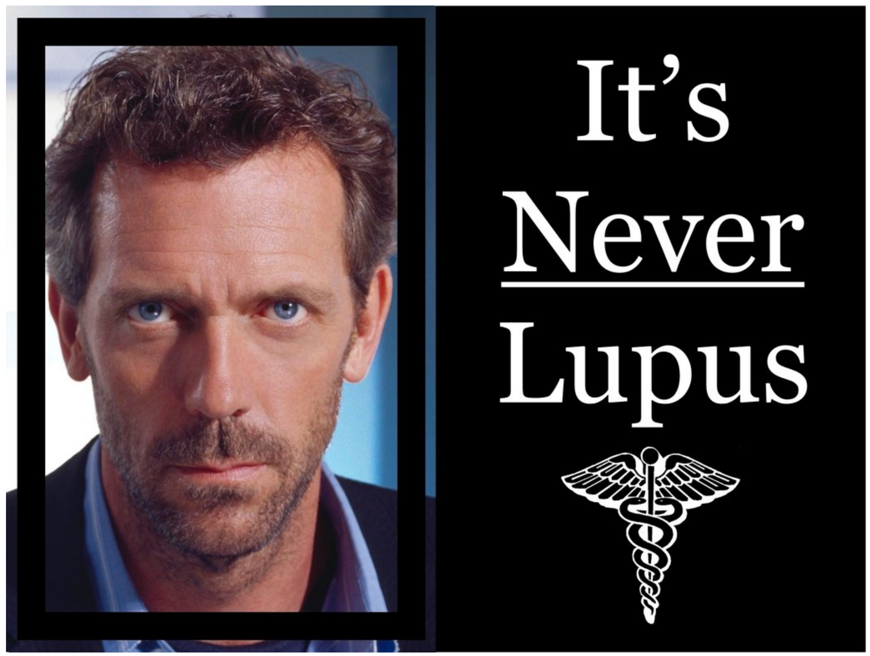
$$P(A \mid B) = ?$$

"**Given B**" - where does B occur?

$$\frac{P(A \text{ and } B)}{P(A \text{ and } B) + P(\text{not } A \text{ and } B)}$$

Diagnostic Test

Is it Lupus?



Try: Textbook Exercise 2.25 (or watch the youtube video at https://www.youtube.com/watch?v=59y8cOwb-xs&list=PLkIselvEzpM4AQEGTWF95pYy3ZqwTXkHc)

Diagnostic Test

Given:

- Test can detect disease when present: 95% of the time
- Test gives false positive: 1% of the time
- 0.5% of population has disease.

If test is positive, what is the chance that the person has the disease?

Translate…

$$\text{Events:} \quad D = \{\text{person has disease}\}, \quad T = \{\text{test is positive}\}$$

Given:

- $P(T \mid D) = 0.95$ …called **"sensitivity"**
- $P(T \mid D^C) = 0.01$
- $P(T^C \mid D^C) = 1 - P(T \mid D^C)$ called **"specificity"**
- $P(D) = 0.005$ …called **"prevalence"**

What is $P(D|T)$?

- rough guess before calculating?
- Bayes rule or draw tree…

Calculation

```
0.005 * 0.95 / (0.005 * 0.95 + (1 - 0.005) * 0.01)
```

```
## [1] 0.3231293
```

Intuition (Why is this roughly 1/3?)

- imagine 1000 people get test
- on average 5 of them have disease
- 95% of these (i.e. just about all 5) will get positives
- on average 995 of them are healthy
- about 1% of these (i.e., about 10) will get a positive

So there are about 15 people who get positive tests, 5 of which actually have the disease:

$P(D \mid T) \approx \frac{5}{15} = \frac{1}{3}.$

Article on conditional probability by Steven Strogatz

[“Chances Are” NY Times (2010)] - a good read!

More than two events…

- same idea
- Let $A_1, \dots, A_n$ be a partition of Ω

Law of Total Probability

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i)$$

- Partition means $\cup_{i=1}^{n} A_i = \Omega$ and A_i and A_j disjoint if $i \neq j$.

More than two events…

(Again assume $A_1, \ldots, A_n$ is a partition of Ω)

Bayes Rule

$$P(A_j \mid B) = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

- what happens if $n = 2$?