Methods of Mathematics

Kenneth A. Ribet

UC Berkeley

Math 10B
February 9, 2016

Welcome to our new home!

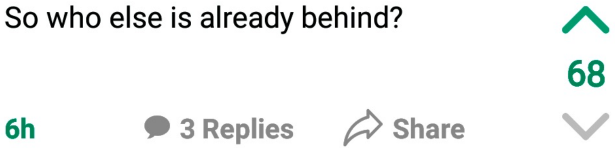
We are now in F295 Haas, a room without chalkboards.

MT #1

The first midterm exam will be one week from today, February 16. The exam will be “in class”—in this room.

Every student can bring in one two-sided page of notes (standard-size paper).

The exam “covers” everything through Bayes’s Rule: counting of all kinds, probability, conditional probability.



Don’t forget to consult the archive of past exams.

Ribet Office Hours

Monday 2:10–3:10, 885 Evans

Tuesday 10:30–noon, Student Learning Center

Thursday 10:30–11:30, 885 Evans

Sorry, there are no office hours on February 15 (UC holiday).

Upcoming breakfasts

8AM, Monday, February 22.

This will be at the Faculty Club (as usual). There are still some places available.

Because of last-minute cancellations, there are now one or two places available for this Friday's breakfast at 8:30AM. (The date is February 12.)

Send me email if you'd like to come.

Lunches

Let me know if you'd like to organize a DC or Faculty Club lunch one day in the future. A natural idea is to go to a DC after each SLC office hour.

Bayes' Rule

Today's class meeting concerns the deceptively simple formula

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

which follows directly from the definition of conditional probability.

The formula is known as *Bayes' Rule*. See the Wikipedia page https://en.wikipedia.org/wiki/Bayes%27_theorem for some background and examples.

As you will see from the examples, a recurrent theme is that we will consider B as the union of the two disjoint events $B \cap A$ and $B \cap A^c$.

Two lectures (available through bCourses)

Lisa Goldberg's MSRI lecture "What is Bayes' Rule and Why Does it Matter?"

My Math 55 slides on Bayes' Rule from March, 2015

Today

We will discuss several typical Bayes problems. In essence, they are all the same!

We begin with one from the “textbook”:

Identical twins always have the same sex, whereas fraternal twins are equally likely to have the same sex or not. About one third of all twins are identical. Given that ultrasound shows that an unborn pair of twins are both of the same sex, what is the probability that they are identical?

Following the textbook, we let A be the event that the twins are identical and B be the event that they have the same sex. We want to compute $P(A|B)$. We know:

- $P(B|A) = 1$, $P(B|A^c) = 1/2$;
- $P(A) = 1/3$; $P(A^c) = 2/3$;
- $P(B) = P(B \cap A) + P(B \cap A^c)$.
- $P(B \cap A) = P(B|A)P(A)$; $P(B \cap A^c) = P(B|A^c)P(A^c)$.

$$P(B) = 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} = 2/3$$

and

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1/3}{2/3} = 1/2.$$

The Tversky–Kahneman taxicab problem

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green?

[This slide is taken directly from Math 55 (spring, 2015).]

Let A be the event that the cab was blue. Let B be the event that the witness identifies the cab as blue. We seek $P(A|B)$ and use the formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

The denominator is $P(B \cap A) + P(B \cap A^c) = P(B)$.

We have $P(A) = 0.15$; this is the "prior" probability (calculated before the extra information). Thus $P(A^c) = 0.85$. Also,

$$P(B|A) = 0.80, \quad P(B|A^c) = 0.20$$

because the witness is right 80% of the time. We now know enough to calculate $P(A|B)$.

Notice that the numerator in the fraction that computes $P(A|B)$ is the first of the two terms in the denominator. The first denominator term is the contribution to $P(B)$ from a correct report of the witness, while the second term pertains to an incorrect report. The vlaue of $P(A|B)$ depends on the relative sizes of these two terms.

The first term is $0.8 \cdot 0.15 = 0.12$, while the second is $0.2 \cdot 0.85 = 0.17$. We see that the a "blue" report from the witness is *more* likely to come from a green cab than from a blue cab.

The conditional probability is $\frac{0.12}{0.29} \approx 0.41$. Despite the witness's apparent reliability, the probability that her report is accurate in this case is less than 50%.

The bad test result problem (NY Times)

Doctors in Germany and the US were asked to estimate the probability that a woman with a positive mammogram actually has breast cancer, even though she's in a low-risk group: 40 to 50 years old, with no symptoms or family history of breast cancer. To make the question specific, the doctors were told to assume the following statistics—couched in terms of percentages and probabilities—about the prevalence of breast cancer among women in this cohort, and also about the mammogram's sensitivity and rate of false positives:

The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram.

Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

This is $P(A|B)$, where A is the event of having the disease and B is the event of having a positive test result for the disease.

As in the previous problem, we have to tease out the values of $P(A)$, $P(A^c)$, $P(B|A)$ and $P(B|A^c)$ from the description that's provided to us.

We have $P(A) = 0.008$, $P(A^c) = 0.092$; this is the "prior" information.

Information about the test (true positives) gives $P(B|A) = 0.9$.

There are false positives: $P(B|A^c) = 0.07$.

Then $P(B|A)P(A) = 0.9 \cdot 0.008 = 0.0072$, $P(B|A^c)P(A^c) = 0.07 \cdot (1 - .008) = .6944$,

$$P(A|B) = \frac{0.0072}{0.0072 + 0.6944} \approx 0.09399 \ldots.$$

Thus there is less than a 10% chance that the woman with a positive mammogram actually has breast cancer.

This was very surprising to doctors who were asked to estimate the probability—http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/.