STOR 435.001 Lecture 6

Conditional Probability and Independence - II

Jan Hannig

UNC Chapel Hill

Motivation: Conditional probability and mammograms

https://plus.maths.org/content/understanding-uncertainty-breast-screening-statistical-controversy

From the above link: Doctors were asked: “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?”

“When Gigerenzer asked 24 other German doctors the same question, their estimates whipsawed from 1 percent to 90 percent. Eight of them thought the chances were 10 percent or less, 8 more said 90 percent, and the remaining 8 guessed somewhere between 50 and 80 percent. Imagine how upsetting it would be as a patient to hear such divergent opinions.”

Correct answer: 9%

Important result:

$$P(E) = P(E \cap F) + P(E \cap F^c) = P(E|F)P(F) + P(E|F^c)P(F^c) \quad (1)$$

Useful formula

More generally: Suppose $S = \cup_{i=1}^{n} F_i$, where F_i are mutually disjoint. Then,

$$P(E) = \sum_{i=1}^{n} P(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i). \quad (2)$$

Equation (1) above is the case $n = 2$, $F_1 = F$, $F_2 = F^c$ for the general formula.

Example: Using equation (1)

Consider the breast cancer example. Compute the probability that a woman matching the description of the data gets a positive mammogram. ☼

Example: Using equation (2)

Example

Jan is teaching a small undergraduate class which has 24 students. He periodically calls on students to answer questions in class (for class participation) and he chooses these at random. He has already called on 6 of them (so 18 not yet called upon). Today he first selects a group of three students at random amongst all 24 of the students and asks them to work out a problem on the board (as a team). Then he selects another group of 3 students at random to work out the next set of problems. Find the probability that in this selection none of the students had been called before.

☼

Expressing $P(F|E)$ in terms of $P(E|F)$: Note that

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(E|F)P(F)}{P(E)} = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}.$$

More generally:[1]

Bayes formula

Suppose $S = \cup_{i=1}^{n} F_i$, where F_i are mutually disjoint. Then,

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}.$$

Above: $n = 2$, $F_1 = F$, $F_2 = F^c$.

[1] Thomas Bayes (1701-1761), English mathematician

Breast Cancer example

Consider the breast cancer and mammogram example. A woman fitting the description of the data goes in for a routine check up and gets a positive mammogram. What is the chance she actually has breast cancer?

☼

Bayes formula using Equation (2)

Jan is planning to goto grad school and has three universities to consider A, B, C. He thinks that if he gets into A, the chance of him having a great job (event = J) after is 95%, if he goes to B the chance of J is 75% and if gets into C, the chance of J is 5%. He also estimates his chance of getting into A to be 35%, chance of getting into B to be 65% and chance of getting into C to be 85%.

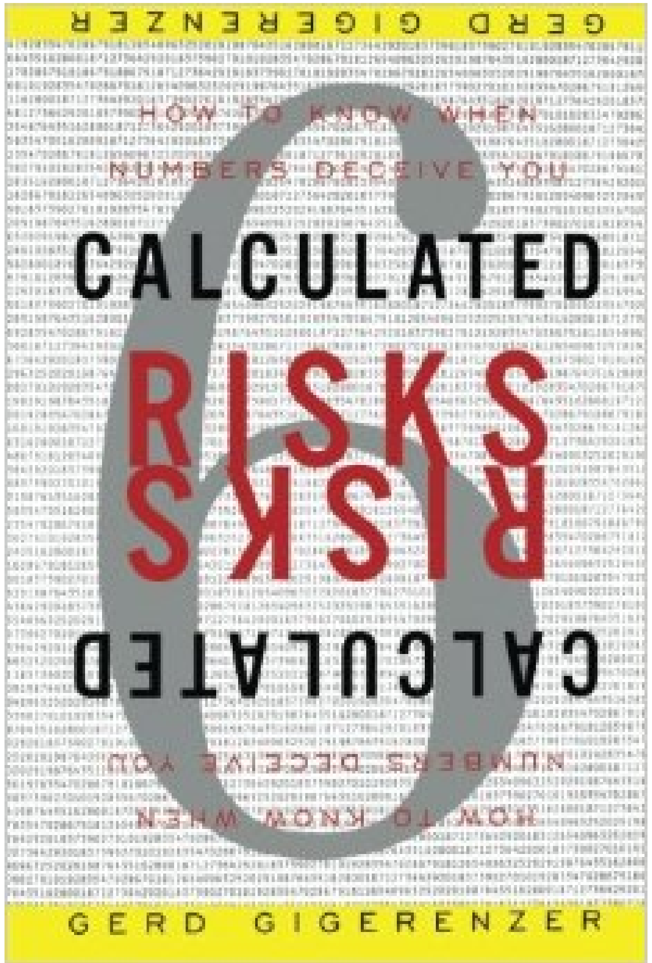
1. How likely is the event J?
2. If I tell you Jan had a great job and give you no other information about Jan, how likely is it for him to have attended school A?

☼

Solution continued

☼

Reference for the next case study on OJ Simpson trial



Ok let us get started.

Case 1

Figure: “O.J. Simpson 1990 DN-ST-91-03444 crop”. Main picture by Gerald Johnson - O.J.Simpson 1990. Licensed under Public Domain via Wikimedia Commons

OJ Simpson trial

- October 3, 1995: Verdict to be announced at 10:00 A.M.
- LA police on full alert with President Clinton informed of the arrangements.
- Long distance call volume dropped 58%. Trading volume on NYSE fell by 41%.
- Estimated 100 million people turned on news. Loss in productivity: $480 million.
- *Why?* To see how 12 jurors would judge OJ Simpson: accused of murdering his ex-wife Nicole Brown Simpson and her friend Ronald Graham in 1994.
- Jury delivered verdict:
  NOT GUILTY

Figure: “AlanDershowitz2” by Not given. Copyright held by Dershowitz. - Alan Dershowitz. [1] First uploaded by SlimVirgin in December 2006. Image released by Dershowitz by email on December 24, 2006. Licensed under Copyrighted free use via Wikimedia Commons

Prosecution

- One of the most dangerous pieces against OJ: past history of spousal abuse.
- Prosecution argued that a history of spousal abuse reflects a motive to kill.
- Alan Dershowitz: youngest full professor of law at Harvard and one of the advisors of OJ Simpson team in Book “Reasonable doubts: The criminal justice system and the OJ Simpson case” explains defense team great success in destroying prosecution’s argument that history of spousal abuse leads to murder. Made the case that Battery **not** a good predictor.

Data used

4 million women battered per year (~ 1993 figures)

Crucial argument

Data used

- 4 million women battered per year (~ 1993 figures) by husbands and boyfriends. In 1992 according to FBI crime reports, 913 women killed by husbands and 519 killed by boyfriends.
- Defense then concluded: there is less than 1 homicide per 2500 incidents of abuse.
- Thus spousal abuse not a good predictor.
- *Sounds super convincing right?*

We will learn that more careful calculation gives much higher probability! Note: not claiming anything about the particular case in question and the innocence or lack thereof of OJ as that trial had a ton of other factors and pieces of evidence that the jury had to consider. Rather this claim is only about the probability calculation above.

What do we want to calculate?

- **Want:** Probability that a woman who was battered by partner earlier is then killed by her partner *conditional or given that* she was battered by her partner before and was eventually found murdered.
- Let us simplify by using some notation by defining a few events.
- Let B be the event that a woman gets battered by partner.
- M be the event that woman is murdered.
- M_p be the event that woman is murdered by partner.

We want:

$$\mathbb{P}\left(M_p \mid B \text{ and } M\right)$$

Let us use what we know from conditional probability

$$\mathbb{P}\left(M_p \mid B \text{ and } M\right) = \frac{\mathbb{P}(M_p \text{ and } B \text{ and } M)}{\mathbb{P}(B \text{ and } M)}$$
$$= \frac{\mathbb{P}(M_p \text{ and } B)}{\mathbb{P}(B \text{ and } M)}$$
$$= \frac{\mathbb{P}(M_p \text{ and } B)}{\mathbb{P}(M_p \text{ and } B) + \mathbb{P}(M^* \text{ and } B)}$$

where M^* is the event that women is murdered by someone **OTHER** than her partner.

Again using conditional probability definition

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

Similarly

$$\mathbb{P}(M_p \text{ and } B) = \mathbb{P}(B)\mathbb{P}(M^*|B)$$

Punchline

$$\mathbb{P}\left(M_p \mid B \text{ and } M\right) = \frac{\mathbb{P}(M_p|B)}{\mathbb{P}(M_p|B) + \mathbb{P}(M^*|B)}$$

We will estimate these probabilities by

$$\mathbb{P}(M_p|B) = \frac{1}{2500} \text{ as estimated by the defense}$$

$$\mathbb{P}(M^*|B) = \text{ Chance woman is murdered } \frac{5}{100,000},$$

using FBI crime reports stats for 1993.

$$\mathbb{P}\left(M_p \mid B \text{ and } M\right) = \frac{1/2500}{1/2500 + 5/100000} = \frac{8}{9} = 88\%$$

Again not claiming this is the chance OJ is guilty. Just what the right calculation about the proportion of murdered women who were previously battered by partner and were actually murdered by partner.

Motivation: Case of Sally Clark

Case of Sally Clark

- 1996: first son died suddenly within few weeks of birth
- 1998: second son died in a similar manner. She was subsequently arrested
- Pediatrician testified that chance of two children from affluent family suffering SIDS was

$$\frac{1}{8500} * \frac{1}{8500} = \frac{1}{72,250,000}$$

- 1999: Convicted, life imprisonment
- **Why should the two deaths independent of each other?**
- January 2003: Released
- March 2007: Died of alcohol intoxication

Independent events:

Events E and F are independent if

$$P(E|F) = P(E), \quad P(F|E) = P(F). \tag{3}$$

In other words, knowing that one of them occurs does not change the probability that the other occurs. Note also that this is effectively a consequence of the model. If we believe or data show that there is independence, the model has to incorporate this.

Definition (Independence)

Equation (3) gave one definition of independence. It turns out this is equivalent to the following: Events E and F are independent if

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

Example:

Toss a coin twice, $E = \{\text{first toss is } H\}$, $F = \{\text{second toss is } T\}$. Are E and F independent?

☼

Example

You select a card randomly from a deck. Let E be the event that it is a $\clubsuit$ and F be the event it is a 6. Are these two events independent?

☼

Definition (Independence)

Recall from previous slide: events E and F are independent if

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

Example 4c: Suppose that we toss 2 fair dice. (a) Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Are E_1 and F independent? (b) Now, suppose that we let E_2 be the event that the sum of the dice equals 7. Is E_2 independent of F?

☼

Note 1: If E and F are independent, then so are E and F^c.

☼

Note 2: Do not confuse independence and disjointness.

☼

Definition (Independence of 3 events)

Events E, F and G are independent if

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

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

Note: If E, F and G are independent, then, for example, E, F^c and G are independent, E^c, F^c and G^c are independent, E, $F \cap G$ are independent, E, $F \cup G$ are independent, etc.

⚠

Definition (Independence of n events)

Events $E_1, E_2, \dots, E_n$ are independent if, for every subset $E_{1'}, E_{2'}, \dots, E_{r'}$, $r' \le n$:

$$P(E_{1'} \cap E_{2'} \cap \ldots \cap E_{r'}) = P(E_{1'})P(E_{2'}) \ldots P(E_{r'}).$$

Note: Events involving disjoint collections (across indices) of E_i's are independent, for example, $E_1 \cap E_2$ and $E_3^c \cap E_4$ are independent, E_1^c, $E_2 \cup E_3$ and E_3 are independent, etc.

Independent trials: An experiment may consist of a sequence of identical subexperiments (same outcomes and same probabilities), called trials. E.g. tossing a coin many times. Moreover, one can assume that trials are independent, that is, $E_1, E_2, \ldots, E_n$ are independent whenever E_i is determined by the ith trial.

Example:

After the weekend, four friends Donatello, Michelangelo, Raphael and Leonardo were supposed to attend STOR 435 but ended up at the comic book store and so missed out on some amazing knowledge. They told the instructor that all 4 of them were returning from Charlotte after taking care of an evil gang called the foot clan but then one tire of their car had a flat and so were unable to come to class. The instructor took each aside **separately** and asked them **which tire**? The friends did not expect this. Assume the guesses of each of the above 4 are independent of each other and further:

1. Donatello and Michelangelo are unbiased and pick one of the four tires (FL “Front Left”, FR “Front Right”, BL, BR) at random (with equal probability).
2. Leonardo leans left and picks the left tires with double the probability of the right tires.
3. Raphael leans right and picks the right tires with double the probability of the left tires.

What is the probability that the answers of **all** the four friends match and thus they are able to get away with their fib?

☼

Solution continued

☼

Example:

You have a sequence of Independent trials consisting of rolling a pair of fair 6-faced dice. What is the probability that an outcome of 5 appears before an outcome of 7 when the outcome of a roll is the sum of the dice?

Solution continued

☼