Overview

- Conditional Probability
- Generalised Chain Rule
- Bayes Rule
- Examples

Conditional Probability

In 2011 Irish census[1]:

No. of children	0	1	2	3	> 3
No. of families	344,944	339,596	285,952	144,470	75,248
No. with all children < 15 years	-	178,012	92,826	30,010	8,327

Of the families with 1 child, pick one at random. What is probability that the child is less than 15 years old ?

[1]http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilies livingarrangementsinireland/

Conditional Probability

In 2011 Irish census[2]:

No. of children	0	1	2	3	> 3
No. of families	344,944	339,596	285,952	144,470	75,248
No. with all children < 15 years	-	178,012	92,826	30,010	8,327

Of the families with 1 child, pick one at random. What is probability that the child is less than 15 years old ?

- $178012/339596 = 0.52$
- Let F be event that family has 1 child, E be event that child is less than 15 years old.
- $P(E|F) = 0.52$

[2]http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilies livingarrangementsinireland/

Dice (again)

Roll two six-sided dice, one after the other. First dice comes up 2, call this event F. Roll next die. What is probability they sum to 4 ?

- $F = \{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}$ (6 possibilities)
- $E = \{(2,2)\}$
- $P(E \text{ given already observed first dice is } 2) = \frac{1}{6}$
- $P(E|F) = P(E \text{ given } F \text{ already observed}) = \frac{1}{6}$

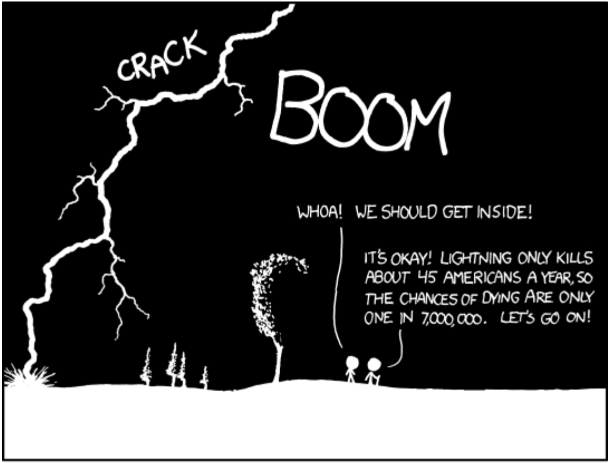
Dice (again)

Now roll the dice one after another again. First dice comes up 6. Roll next die. What is probability they sum to 4 ?

- $F = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ (36 possibilities)
- $E = \emptyset$
- $P(E|F) = P(E$ given already observed first dice is $6) = \frac{0}{6} = 0$

Observed events can may increase or decrease the probability of subsequent events.

Conditional Probability



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

http://xkcd.com/795/

Conditional Probability

Conditional probability is the probability that event E occurs given that event F has already occurred. Call this "conditioning" on F. Written as $P(E|F)$.

- Same meaning as $P(E$ given F already observed)
- Its a probability (satisfies all the axioms, will see this shortly) with:
  - Sample space S restricted to those outcomes consistent with F i.e. $S \cap F$.
  - Event space E restricted to those outcomes consistent with F i.e. $E \cap F$

Conditional Probability

With equally likely outcomes

$$P(E|F) = \frac{\#\text{outcomes in E consistent with F}}{\#\text{outcomes in S consistent with F}} = \frac{|E \cap F|}{|S \cap F|}$$

Note that $|S \cap F| = |F|$ so

$$P(E|F) = \frac{|E \cap F|}{|S \cap F|} = \frac{|E \cap F|}{|F|} = \frac{|E \cap F|}{|S|}\frac{|S|}{|F|} = \frac{\frac{|E \cap F|}{|S|}}{\frac{|F|}{|S|}}$$
$$= \frac{P(E \cap F)}{P(F)}$$

where $|S|$ is the number of elements in set S, etc.

Conditional Probability

General definition:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

where $P(F) > 0$. Implies

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

known as the chain rule – its important !
If $P(F) = 0$?

- $P(E|F)$ is undefined
- Can't condition on something that can't happen

Coins this time

Flip a coin twice. Observe first flip is heads. What is the probability of two heads ?

- Sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$.
- $F = \{(H, H), (H, T)\}$ is event that first flip is heads
- $E = \{(H, H)\}$ is the event that there are two heads
- $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$

Conditional Probability

$P(E|F)$ is a probability – it satisfies all the properties of ordinary probabilities.

- $0 \leq P(E|F) \leq 1$
  - $E \cap F \subset F$ so $P(E \cap F) \leq P(F)$ and $P(E|F) = \frac{P(E \cap F)}{P(F)} \leq 1$
- $P(S|F) = 1$
  - $P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$ (chain rule)
- If E_1, E_2 are mutually exclusive events then $P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F)$
  - $P(E_1 \cup E_2|F) = \frac{P((E_1 \cup E_2) \cap F)}{P(F)} = \frac{P((E_1 \cap F) \cup (E_2 \cap F))}{P(F)} = \frac{P(E_1 \cap F) + P(E_2 \cap F)}{P(F)}$

Marginalisation

In 2011 Irish census[3]:

No. of children	0	1	2	3	> 3
No. of families	344,944	339,596	285,952	144,470	75,248
No. with all children < 15 years	-	178,012	92,826	30,010	8,327

Pick a family at random from the population. What is probability that all of the children are less than 15 years old ?

[3] http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilies livingarrangementsinireland/

Marginalisation

In 2011 Irish census[4]:

No. of children	0	1	2	3	> 3
No. of families	344,944	339,596	285,952	144,470	75,248
No. with all children < 15 years	-	178,012	92,826	30,010	8,327

Pick a family at random from the population. What is probability that all of the children are less than 15 years old ?

- No. of families with all children < 15 years: 178,012+92,826+30,010+8,327=309,175
- Total no. of families: 344,944+339,596+285,952+144,470+75,248=1,190,210
- Ratio: 309,175/1,190,210 = 0.26

[4] http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilies-livingarrangementsinireland/

Marginalisation

Equivalently:

No. of children	0	1	2	3	> 3
Fraction of all 1,190,210 families with all children < 15 years	-	0.149	0.078	0.025	0.007

- Let F_i be the event that have i children and E be the event that all children are less than 15 years old.
-

$$\begin{aligned} P(E) &= P(E \cap F_1) + P(E \cap F_2) + \dots \\ &= 0.149 + 0.078 + 0.025 + 0.007 = 0.26 \end{aligned}$$

Marginalisation

Suppose we have mutually exclusive events F_1, F_2, ..., F_n such that $F_1 \cup F_2 \cup ... \cup F_n = S$. Then

$$P(E) = P(E \cap F_1) + P(E \cap F_2) + \cdots + P(E \cap F_n)$$

Proof. By the chain rule, $P(E \cap F_i) = P(F_i|E)P(E)$ so,

$$\begin{aligned} P(E \cap F_1) + P(E \cap F_2) + \cdots + P(E \cap F_n) \\ &= P(F_1|E)P(E) + P(F_2|E)P(E) + \cdots + P(F_n|E)P(E) \\ &= (P(F_1|E) + P(F_2|E) + \cdots + P(F_n|E))P(E) \\ &= P(E) \end{aligned}$$

since

$$P(F_1|E) + P(F_2|E) + \cdots + P(F_n|E) = P(F_1 \cup F_2 \cup \cdots \cup F_n|E) = F(S|E) = 1$$

Special case (remember this one):

$$\begin{aligned} P(E) &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned}$$

Example

Marginalisation is v handy. Example:

- Roll two coins. What is the probability that the first coin is heads ?
- Event E is first coin heads, F_1 is second coin heads, F_2 is second coin tails
- $P(E) = P(E \cap F_1) + P(E \cap F_2) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$

Another example:

- Suppose an HIV test identifies HIV 98% of the time but has a false positive rate of 1%.
- Approximately 0.2% of the population has HIV.
- What is the probability that you test positive ?
- E is event that test positive, F_1 that you have HIV, F_2 that you don't
- $P(E) = 0.002 \times 0.98 + (1 - 0.002) \times 0.01 = 0.012 = 1.2\%$

Example

Roll two dice. Let F_1 be the event that the first die comes up 1, F_2 that it comes up two, etc. Let E be the event that the sum of the dice is 6.

- $P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_6)P(F_6)$
- $P(F_1) = P(F_2) = \cdots P(F_6) = \frac{1}{6}$.
- $P(E|F_1) = \frac{1}{6}$ (second die comes up 5), $P(E|F_2) = \frac{1}{6}$ second die comes up 4) etc.
- But $P(E|F_6) = 0$ since second die must be at least 1.
- So $P(E) = 5 \times \frac{1}{6} \times \frac{1}{6} = \frac{5}{36}$
- Double check: event E that sum is 6 is $\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$, so $|E| = 5$. Since $|S| = 36$, $P(E) = \frac{5}{36}$

Generalised Chain Rule

For two events E and F have:

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

For events E_1, $E_2, \cdots, E_N$

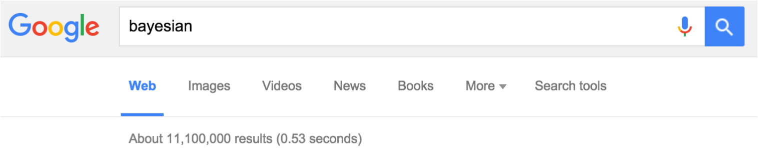
$$\begin{aligned} P(E_N|E_1 \cap E_2 \cdots \cap E_{N-1}) \cdots P(E_3|E1 \cap E_2)P(E_2|E_1)P(E_1) \\ = P(E_1 \cap E_2 \cdots \cap E_N) \end{aligned}$$

Bayes Rule

Thomas Bayes

Thomas Bayes (1702-1761) was a British mathematician and Presbyterian minister.

Adjective "Bayesian" is pretty popular:



Bayes Rule

Recall

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

Clearly, and also

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

But $P(E \cap F) = P(F \cap E)$, so

$$P(E|F)P(F) = P(F|E)P(E)$$

i.e.

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

This is **Bayes Rule** (or Bayes Theorem).

Terminology

$$\color{red}{P(E|F)} = \frac{\color{blue}{P(F|E)}\color{green}{P(E)}}{P(F)}$$

posterior *likelihood* *prior*

- Prior. Before seeing anything, what is our belief about event E.
- Likelihood. Probability of seeing event F given event E has occurred
- Posterior. After seeing event F, this is the probability of seeing event E.
- Evidence. The denominator is sometimes called the “evidence”.

Terminology

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

posterior *likelihood* *prior*

Suppose the event E is that it rains tomorrow, and F is the event that it is cloudy today.

- Prior. Our guess for the chance of rain tomorrow, with no extra info.
- Likelihood. The probability of a cloudy day before rain.
- Posterior. Our updated probability of rain tomorrow after observing clouds today
- Evidence $P(F)$ is the chance of a cloudy day, with no extra info.

HIV Testing

Bayes Rule is v useful as it lets us calculate probabilities that might otherwise be hard to calculate.
Suppose if you have HIV a test will identify that 98% of the time.

- However, test has a "false positive" rate of 1%
- Approx 0.2% of Irish population has HIV
- Event F = you test positive for HIV with this test
- Event E = you actually have HIV
- What is $P(E|F)$?

Apply Bayes Rule $P(E|F) = \frac{P(F|E)P(E)}{P(F)}$.

- Likelihood $P(F|E) = 0.98$ (probability that test is positive when you have HIV)
- Prior $P(E) = 0.002$ (probability that someone in population has HIV)
- $P(F) = P(F|E)P(E) + P(F|E^c)P(E^c) = 0.98 \times 0.002 + 0.01 \times (1 - 0.002)$

HIV Testing

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$
$$= \frac{0.98 \times 0.002}{0.98 \times 0.002 + 0.01 \times (1 - 0.002)} \approx 0.16$$

i.e. probability that do you <u>not</u> have HIV given that test is positive is 1-0.16=0.84.

- Surprising ?!

HIV Testing

Let's think about the numbers ...

- Irish population is about 4.5M. Approx 9000 (0.2% of 4.5M) are HIV positive.
- Test detects 98% of HIV positive people, so $9000 \times 0.98 = 8,820$ true positives
- Test has false positive rate of 1%, so expect $(4.5M - 9000) \times 0.01 = 44,910$ false positives
- So fraction of people who do not have HIV but who get a positive test is $44910/(44910 + 8820) = 0.84$

Simple Spam Detection

Say 60% of all email is spam,

- 90% of spam has a forged header
- 20% of non-spam has a forged header
- Event F = message contains a forged header
- Event E = message is spam.
- What is $P(E|F)$?

Apply Bayes Rule:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} = \frac{0.9 \times 0.6}{0.9 \times 0.6 + 0.2 \times (1 - 0.6)} \approx 0.87$$

- Likelihood $P(F|E) = 0.9$ (probability that spam has forged header)
- Prior $P(E) = 0.6$ (probability that email is spam)
- $P(F) = P(F|E)P(E) + P(F|E^c)P(E^c) = 0.9 \times 0.6 + 0.2 \times (1 - 0.6)$

Prosecutors Fallacy

In 1998 Sally Clark[5] was accused of killing ker first child at 11 weeks of age and then her second child at 8 weeks of age.

- Prosecution expert: probability of two children in same family from Sudden Infant Death Syndrome was 1 in 73 million
- Claim is then that 1 in 73M is the probability that Clark was innocent.
- Jury concluded guilty of murder, upheld on appeal, overturned on second appeal in 2003.
- "one of the great miscarriages of justice in modern British legal history"
- Where is the flaw in the reasoning ?

[5] see https://en.wikipedia.org/wiki/Sally_Clark

Prosecutors Fallacy

A prosecutor has evidence E against a suspect. Let I be the event that the suspect is innocent.

- Let E be the event of 2 SIDS deaths, I be the event that the suspect is innocent
- Suppose $P(E|I) = \frac{1}{73 \times 10^6}$ (this value is likely wrong too, but that's another story)
- But what we're really interested in is $P(I|E)$. By Bayes:

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)} = \frac{P(E|I)P(I)}{P(E|I)P(I) + P(E|I^c)(1 - P(I))}$$

- Suppose if guilty (event I^c) then $P(E|I^c) \approx 1$, then:

$$P(I|E) \approx \frac{P(E|I)P(I)}{P(E|I)P(I) + 1 - P(I)}$$

- We need to estimate $1 - P(I)$, the probability of a double murder (or one murder and one SIDS death). We can reasonably assume its small i.e. $P(I)$ close to 1. Then $P(I|E) \approx 1$ even when $P(E|I) \approx 0$.

Prosecutors Fallacy II

Here's another variant:

- The chance of winning the Irish lottery is roughly 1 in 2M.
- You've won the lottery. Its a super unlikely event, so you must have cheated ...