Bayes Theorem

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

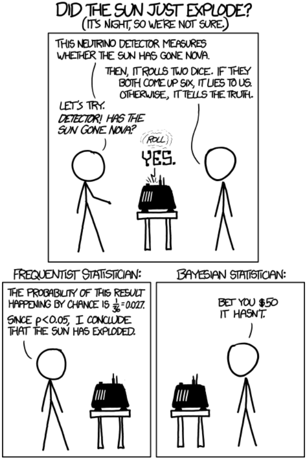
posterior *likelihood* *prior*

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.

Bayes Theorem

- If repeat an experiment many times, can think of probability of an event as being the fraction of times event occurs $\frac{n(E)}{n}$
- What if we can't repeat the experiment ?
  - Axioms of probability still all work fine
  - Probability as frequency doesn't work ...
  - ... interpret probability as belief
  - "Bayesian" vs "frequentist"



http://xkcd.com/1132/

Overview

- Independence
- Examples
- Conditional Independence

Independence

In English: two events E and F are independent if the order in which they occur doesn't matter. Alternatively, if observing one doesn't affect the other.

- Event E is survive parachute jump, event F is event that put a parachute on. We expect the order to matter: jumping and then putting parachute on (event E then F) is not the same as putting parachute on and jumping (F then E).
- Draw a ball from a bag with green balls and orange balls. Then draw another. For second ball there are fewer balls left in bag (since have taken one out), so expect chance of drawing a green ball to have changed.
- Toss a coin twice. We expect that the outcome of the second toss does not depend on the outcome of the first.

Independence

Definition. Two events E and F are **independent** if

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

When events E and F are independent then $P(E|F) = P(E)$ (recall chain rule: $P(E \cap F) = P(E|F)P(F)$). Note: $P(E|F) = P(E)$ is not used as the definition however.

- Otherwise E and F are **dependent** events

Quick Examples

- Pick a random leaving cert student – are the events “applied to TCD” and “applied to UCD” independent ?
- Probably not – if you apply to one you’re more likely to apply to the other
- Pick a random person in Ireland – are the events “are a TCD student” and “have brown eyes” independent ?
- Probably yes – colour of eyes probably not related to whether you’re at TCD or not.
- Gambler’s Fallacy – a run of heads when flipping a coin doesn’t make you “due for a tails”.

Quick Examples

Roll two 6-sided dice. Let E be the event that the first dice is 1 and F the event that the second dice is 1.

- $P(E) = \frac{1}{6}$, $P(F) = \frac{1}{6}$, $P(E \cap F) = \frac{1}{36}$
- $P(E \cap F) = P(E)P(F)$ for E and F are <u>independent</u>.

Let G be the event that the dice sum to 5 (outcomes are $\{(1,4),(2,3),(3,2),(4,1)\}$.

- $P(E) = \frac{1}{6}$, $P(G) = \frac{1}{9}$, $P(E \cap G) = \frac{1}{36}$
- $P(E \cap G) \neq P(E)P(G)$ for E and G are <u>dependent</u>.

Independence

- Three events E, F and G are independent if they are pairwise independent <u>and</u> triply independent

$$
\begin{aligned}
P(E \cap F \cap G) &= P(E)P(F)P(G) \\
P(E \cap F) &= P(E)P(F) \\
P(E \cap G) &= P(E)P(G) \\
P(F \cap G) &= P(F)P(G)
\end{aligned}
$$

- Pairwise independence is <u>not</u> enough.

Independence

Are three events independent if they are pairwise independent ?

- Four balls in an urn numbered 110, 101, 011, 000
- Let A_k be the event of a 1 in the kth place.
- $P(A_k) = \frac{1}{2}$, $P(A_l \cap A_k) = \frac{1}{4}$, $P(A_1 \cap A_2 \cap A_3) = 0$!

Generating Random Bits

- A computer produces a series of random bits, with probabilty p of producing a 1
- Each bit generated is an independent trial
- Event E is that the first n bits are 1s, followed by a single 0
- What is $P(E)$?

Solution:

$$
\begin{aligned}
P(first\ n\ 1's) &= P(1st\ bit = 1)P(2nd\ bit = 1)\cdots P(nth\ bit = 1) \\
&= p^n \\
P(n+1\ bit = 0) &= (1-p) \\
P(E) &= P(first\ n\ 1's)P(n+1\ bit = 0) = p^n(1-p)
\end{aligned}
$$

Coin flips

- Say a coin comes up heads with probability p (need not be $\frac{1}{2}$)
- Each coin flip is an independent trial
- $P(n\ heads\ on\ n\ coin\ flips) = p^n$
- $P(n\ tails\ on\ n\ coin\ flips) = (1-p)^n$
- $P(first\ k\ heads,\ then\ n-k\ tails) = p^k(1-p)^{n-k}$
- $P(exactly\ k\ heads\ on\ n\ flips) = \binom{n}{k}p^k(1-p)^{n-k}$

Sending Messages Through a Network

A mobile handset has both an LTE and a WiFi interface.

- The probability that the LTE interface is functioning is p_1
- The probability that the WiFi interface is functioning is p_2
- E is the event that there is at least one functioning interface.
- What is $P(E)$?

Solution:

$$\begin{aligned} P(E) &= 1 - P(\textit{both interfaces fail}) = 1 - P(\textit{LTE fails})P(\textit{WiFi fails}) \\ &= 1 - (1 - p_1)(1 - p_2) \end{aligned}$$

Hash Tables

- m strings are hashed (equally randomly) into a hash table with n buckets
- Each string hashed is an independent trial
- Event E is that at least one string is hashed to the first bucket
- What is $P(E)$?

Solution:

- Event F_i is that string i is not hashed into first bucket, $i = 1, 2, \cdots, m$
- $P(F_i) = 1 - \frac{1}{n} = \frac{n-1}{n}$
- Event $F_1 \cap F_2 \cap \cdots F_m$ is that no strings hashed to first bucket

$$\begin{aligned} P(E) &= 1 - P(F_1 \cap F_2 \cap \cdots F_m) = 1 - P(F_1)P(F_2) \cdots P(F_m) \\ &= 1 - \left(\frac{n-1}{n}\right)^m \end{aligned}$$

Hash Tables (again)

- m strings are hashed (unequally) into a hash table with n buckets
- Each string hashed is an independent trial with probability p_i of getting hashed into bucket i
- Event E is that at least one of buckets 1 to k has ≥ 1 string hashed to it.
- What is $P(E)$?

Solution:

- Event F_i is that at no string is hashed into bucket i
- $P(F_i) = (1 - p_i)^m$
- Event $F_1 \cap F_2 \cap \cdots F_k$ is that no strings hashed to buckets 1 to k.

$$\begin{aligned} P(E) &= 1 - P(F_1 \cap F_2 \cap \cdots F_m) = 1 - P(F_1)P(F_2)\cdots P(F_m) \\ &= 1 - (1 - p_1)^m (1 - p_2)^m \cdots (1 - p_k)^m \end{aligned}$$

A Word of Caution

- When we assume events E and F are independent and use the product $P(E)P(F)$ this can be v small.
- Housing example:
  - Suppose $P(\text{one household defaults in mortgage}) = \frac{1}{100}$. If assume independent, then probability that two households default is $\frac{1}{100} \times \frac{1}{100} = \frac{1}{10{,}000}$. And of three households $\frac{1}{1{,}000{,}000}$ etc.
  - But what if this assumption is wrong ? Then probability of joint events might be <u>much</u> higher. E.g. suppose large employers closes down in a small town then prob of > 3 households defaulting might be much greater than 1 in 1M.
- Crypto example:
  - Assume random number generator produces independent samples
  - But what if not true ?

Conditional Independence

Say we rolled two 6-sided dice.

- $S = \{(1,1), (1,2), (1,3), \cdots, (6,1), (6,2), \cdots\}$ (36 possibilities)
- E is the event that the first dice comes up 1
- F is the event that the second dice comes up 6
- So $E \cap F$ is the event that the first dice is 1 and the second 6
- G is the event that the dice sum to 7

Clearly E and F are independent: $P(E \cap F) = P(E)P(F) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

- Now suppose that we have observed event G. What are the probabilities of events E, F and $E \cap F$ now ?
- $S \cap G = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$
- $E \cap G = \{(1,6)\}$, $F \cap G = \{(1,6)\}$
- $P(E|G) = \frac{1}{6}$, $P(F|G) = \frac{1}{6}$
- $P(E \cap F|G) = \frac{1}{6} \neq P(E|G)P(F|G) \rightarrow$ dependent.

Key takeaway: Independent events can become dependent when we condition on additional information. Also dependent events can become independent.

Conditional Independence

- Two events E and F are called **conditionally independent given G** if:

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

  It follows that $P(E|F \cap G) = P(E|G)$ (apply Bayes rule $P(E|F \cap G) = P(E \cap F|G)/P(F|G)$)
- In English, even after observing event G the events E and F still do not depend on one another
- If E and F are independent, does it follow that $P(E \cap F|G) = P(E|G)P(F|G)$? No.

Breaking Dependence

Take the following three events:

- Sample space $S = \{\text{days of week}\}$
- A is that it is not a Monday, $P(A) = \frac{6}{7}$
- B is that it is a Saturday, $P(B) = \frac{1}{7}$
- C is that it is the weekend

Note that A and B are dependent events ($P(A \cap B) = \frac{1}{7} \neq P(A)P(B)$).
What happens when we condition on C ?

- $P(A|C) = 1$, $P(B|C) = \frac{1}{2}$.
- $P(A \cap B|C) = \frac{1}{2} = P(A|C)P(B|C)$
- Dependent events can become independent by conditioning on additional information.