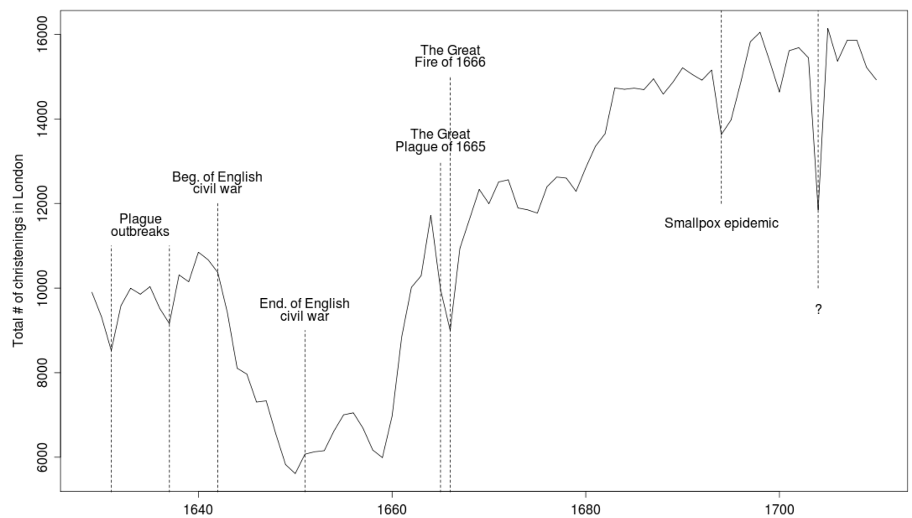
Lecture - More Conditional Probability

Sta102 / BME102

Colin Rundel

September 3, 2014

Lab 1 - Extra Credit



Basic Probability Review

$$0 \le P(E) \le 1$$

$$P(E^c) = 1 - P(E)$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

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

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

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

Disjoint and Independent

$$\text{Independent} \not\Rightarrow \text{Disjoint}$$
$$\text{Disjoint} \not\Rightarrow \text{Independent}^\dagger$$

Your intuition can easily be wrong on this type of problem, always check using the definitions:

Independent:

$$P(E \cap F) = P(E) \times P(F) \text{ or } P(E|F) = P(E)$$

Disjoint:

$$P(E \cap F) = 0$$

Other Important Terms

Joint Distribution - $P(E \cap F)$

Marginal Distribution - $P(E)$

Conditional Distribution - $P(E|F)$

Example - *House*

If you've ever watched the TV show *House* on Fox, you know that Dr. House regularly states, "It's never lupus."

Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease.

The test for lupus is very accurate if the person actually has lupus, however is very inaccurate if the person does not. More specifically, the test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease.

Is Dr. House correct even if someone tests positive for Lupus?

Sampling

Imagine an urn filled with white and black marbles
. . . or a deck of cards
. . . or a bingo cage
. . . or a hat full of raffle tickets

Two common options + one extra for completeness:

- Sampling without replacement
- Sampling with replacement
- Pólya's urn

What is the probability of being dealt two aces?

What if you replace the first card and reshuffle before showing the second?

What is the probability of being dealt a royal flush in poker?

And now a brief magic trick ...

If you have ever shuffled a deck of cards you have done something no one else has ever done before or will ever do again ...

Birthday Problem

Ignoring leap years, and assuming birthdays are equally likely to be any day of the year, what is the probability that there is at least one shared birthday among the students in this class?

As of this morning there are 50 people enrolled in this course,

$$P(\text{at least one match}) = 1 - P(\text{no match}) =$$

Let A_i be the event that student i does not match any of the preceding students then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1, \ldots, A_{n-1})$$

Birthday Problem, cont.

Calculation:

$$P(A_1) = 365/365$$

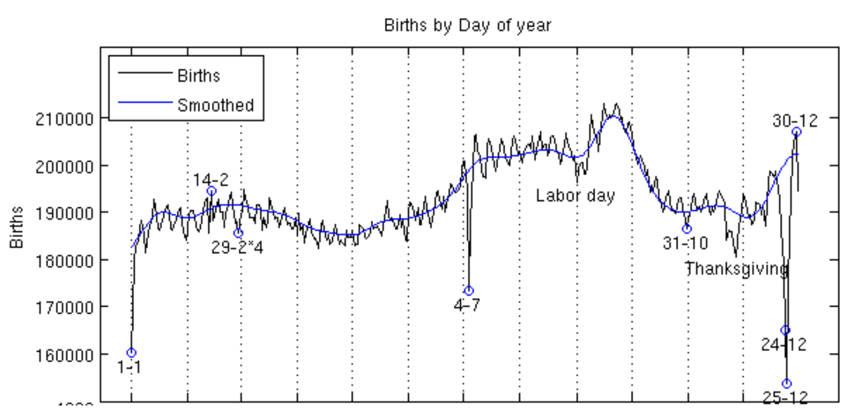
$$\begin{aligned} P(A_1, A_2) &= P(A_1)P(A_2|A_1) \\ &= (365/365) \times (364/365) \end{aligned}$$

$$\begin{aligned} P(A_1, A_2, A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \\ &= (365/365) \times (364/365) \times (363/365) \end{aligned}$$

$$\vdots$$

$$\begin{aligned} P(A_1, \ldots, A_n) &= \frac{365}{365}\frac{364}{365} \cdots \frac{365-(n-1)}{365} \\ &= \frac{365!}{(365-n)!\, 365^n} \end{aligned}$$

Birthday Problem, cont.



The two armed bandit ...

If you know that one of the machines is good (pays out 50% of the time) while the other is bad (pays out 20% of the time) but not which machine is which, what should you do?

Picking a machine

Playing the left machine

What do we know?

- If it is the good machine:

$$P(W|G) = 0.5 \text{ and } P(L|G) = 0.5$$

- If it is the bad machine:

$$P(W|B) = 0.2 \text{ and } P(L|B) = 0.8$$

- We have no idea if it is the good or the bad machine.

$$P(G) = 0.5 \text{ and } P(B) = 0.5$$

What happens if we win?

You put a quarter in the machine on the left, spin, and you win! What does this tell us about the machine on the left? In particular do we now know more about whether it is the good or bad machine?

What happens if we lose?

After winning on on your first play you decide to put another quarter in the machine on the left again, and you lose. Now what do we know about the machine you played?

In case you wanted to see the formulas

$$
\begin{aligned}
P(G|W_1, L_2) &= \frac{P(W_1, L_2|G)}{P(W_1, L_2)}\; P(G) \\
&= \frac{P(W_1|G)P(L_2|G)}{P(W_1)P(L_2|W_1)}\; P(G) \\
&= \frac{P(L_2|G)}{P(L_2|W_1)} \left[\frac{P(W_1|G)}{P(W_1)}\; P(G)\right] \\
&= \frac{P(L_2|G)}{P(L_2|W_1)}\; P(G|W_1) \\
&= \frac{P(L_2|G)}{P(L_2|G, W_1)P(G|W_1) + P(L_2|B, W_1)P(B|W_1)}\; P(G|W_1) \\
&= \frac{P(L_2|G)}{P(L_2|G)P(G|W_1) + P(L_2|B)P(B|W_1)}\; P(G|W_1) \\
&= \frac{1/2}{1/2 \times 5/7 + 4/5 \times 2/7}\; \frac{5}{7} = 25/41 = 0.61
\end{aligned}
$$

Why do we care?

The two-armed (multi-armed) bandit is a very useful model when it comes to clinical trials.

We are trying one or more treatments against a control and we want to know the efficacy of those treatments. This is much more complex in practice because not only do we not know which is better ($P(G)$ in our slot example) we also don't know how much better they are (also need to estimate $P(W|G)$).

Complex optimization problem where we must allocation a limited number of subjects to properly balance:

- Exploration - estimate the payoff of each treatment
- Exploitation - get the best outcome for the most patients

Back to House and Lupus

Last time we worked through a problem on the probability of a patient having lupus given they test positive. We were given

- $P(L) = 0.02$
- $P(+|L) = 0.99$
- $P(-|L^c) = 0.74$

From which we calculated that

$$P(L|+) = \frac{P(L \cap +)}{P(+)} = \frac{P(+|L)P(L)}{P(+|L)P(L) + P(+|L^c)P(L^c)} = \frac{0.02 \times 0.99}{0.02 \times 0.99 + 0.98 \times 0.26} = 0.072$$

If the patient gets a second test, how should our belief in the probability of having lupus, $P(L)$, change?

Let's Make a Deal. . .

Monty Hall Problem

You are offered a choice of three doors, there is a car behind one of the doors and there are goats behind the other two.

Monty Hall, Let's Make a Deal's original host, lets you choose one of the three doors.

Monty then opens one of the other two doors to reveal one of the goats.

You are then allowed to stay with your original choice or switch to the other door.

Which option should you choose?

(a) stay (b) switch (c) it does not matter

A Little History

First known formulation comes from a 1975 letter by Steve Selvin to the American Statistician.

Popularized in 1990 by Marilyn vos Savant in her "Ask Marilyn" column in Parade magazine.

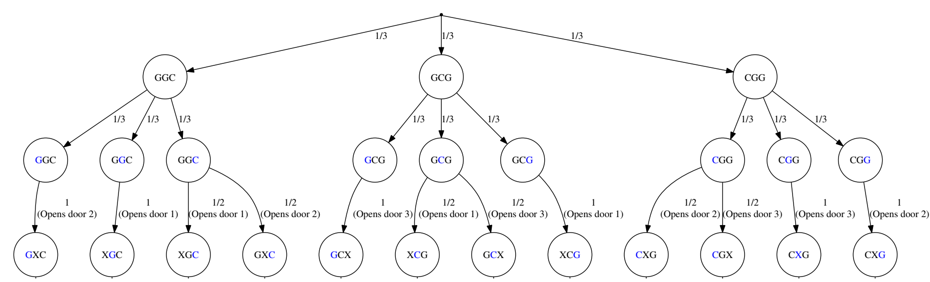
- vos Savant's solution claimed that the contestant should always switch
- About 10,000 (1,000 from Ph.D.s) letters contesting the solution
- vos Savant was right, easy to show with simulation

Moral of the story: trust the math not your intuition

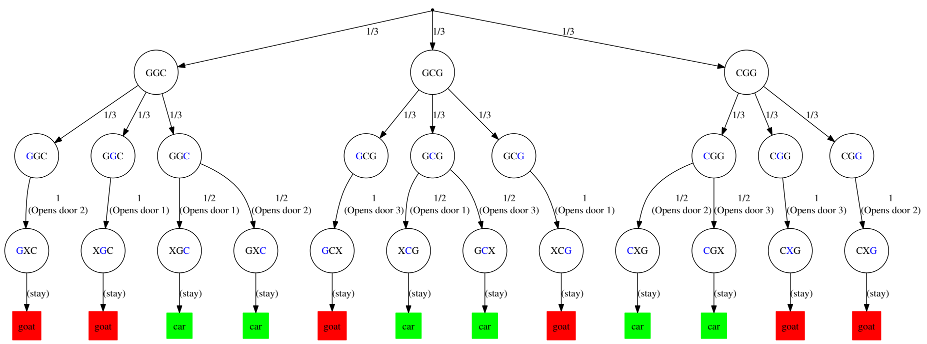
A slightly more entertaining variant of Monty Hall ...

`https://www.youtube.com/watch?v=tvODuUMLLgM&t=3m24s`

Monty Hall - The hard way



Monty Hall - The hard way - Stay



Monty Hall - The hard way - Switch