Lecture 2: More Condition Probability, Distributions

Statistics 104

Colin Rundel

January 17, 2012

Example - Competing Slots

Imagine two slot machines that have a probability of winning of 50% and 66.6% respectively, but you do not know which machine is which. If we play one machine twice what is the probability that we are playing the good machine given the possible outcomes?

Let W_n be the event of winning on the nth pull and G be the event we played the good machine and B the bad then:

Generalizing Conditional Probability

Rules for conditional probability can be naturally extended to more than one event as follows.

For three events:

$$P(A \cap B \cap C) = P(A \cap B)\ P(C|A, B) = P(A)\ P(B|A)\ P(C|A, B)$$

For n events:

$$P(\cap A_i) = P(A_1)\ P(A_2|A_1)\ P(A_3|A_1, A_2)\ \cdots\ P(A_n|A_1, \ldots, A_{n-1})$$

Return to the Birthday Problem

What we saw last time that the birthday problem,

$$P(\text{no match in } n) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\cdots\left(\frac{365-(n-1)}{365}\right)$$
$$= \frac{365!}{(365-n)!\;365^n}$$

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

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

Bayesian Statistics, in brief

Last time we discussed different definitions of probability, one option was belief. Whereby a probability is defined based on a consistent belief (expert knowledge) that is updated based on observed data.

Bayesian Statistics, in brief

Last time we discussed different definitions of probability, one option was belief. Whereby a probability is defined based on a consistent belief (expert knowledge) that is updated based on observed data.

Imagine there is some event A which we believe has a probability $P(A)$, if we then collect data X then we can update our belief using definitions of conditional probability

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

Example - (Bayesian) Competing Slots

Once again, imagine two slot machines that have a probability of winning of 50% and 66.6% respectively, but you do not know which machine is which.

If we win on the first play how should that change our calculations for the second play?

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 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.

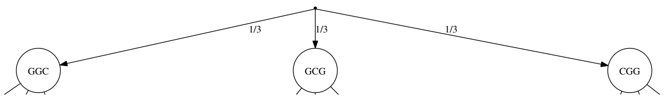
- vos Savant's solution claimed that the contestant should always switch
- About 10,000 (1,000 from Ph.D.s) letters contesting the solution
- vos Savant was right, easy to show with simulation

Moral of the story: trust the math not your intuition

A slightly more entertaining variant of Monty Hall ...

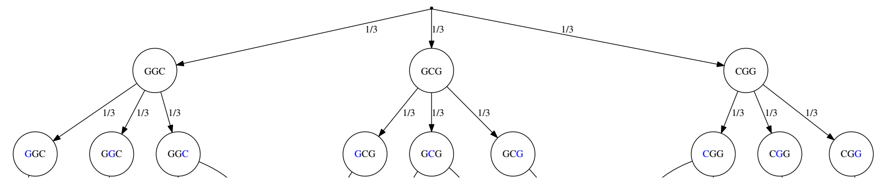
`http://www.youtube.com/watch?feature=player_detailpage&v=vDjm4VLfG_g#t=1722`

Monty Hall - The hard way



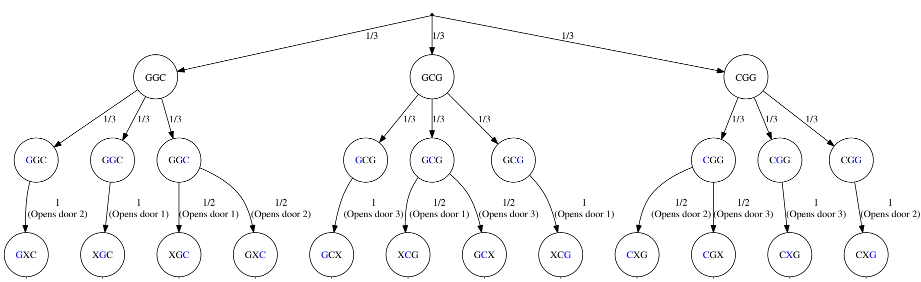
More Conditional Probability

Monty Hall - The hard way



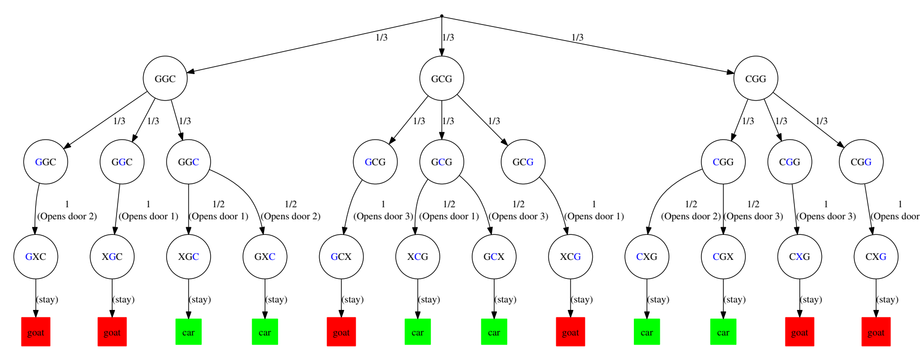
More Conditional Probability

Monty Hall - The hard way

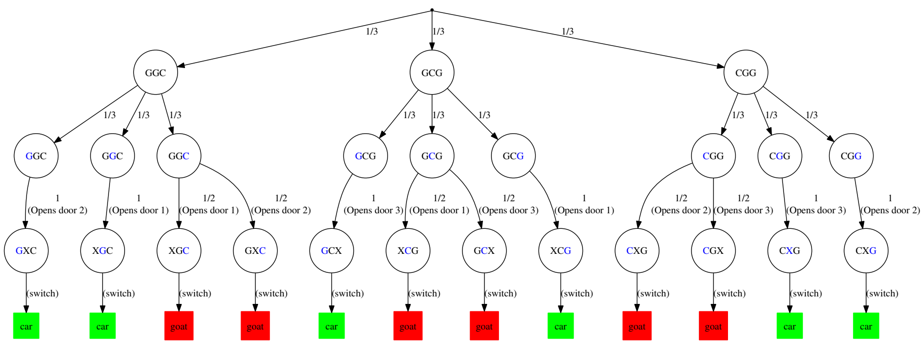


More Conditional Probability

Monty Hall - The hard way - Stay



Monty Hall - The hard way - Switch



Approximating the Birthday Problem

We've already seen the birthday problem where D_n is the event that there are no matches with n people

$$P(D_n) = \frac{365}{365}\frac{364}{365}\cdots\frac{365-(n-1)}{365} = \quad \text{for } n \le 365$$

we can rewrite this as

$$\log P(D_n) = \log 1 + \log\left(1 - \frac{1}{365}\right) + \log\left(1 - \frac{2}{365}\right) + \cdots + \log\left(1 - \frac{(n-1)}{365}\right)$$

A common and useful approximation is $\log(1+x) \approx x$ for small values of x

$$\begin{aligned}
\log P(D_n) &\approx -\frac{1}{365} - \frac{2}{365} - \cdots - \frac{n-1}{365} \\
&\approx -\frac{1}{365}(1+2+\cdots+(n-1)) \\
&\approx -\frac{1}{365} \times \frac{1}{2} n*(n-1) \\
P(D_n) &\approx e^{\frac{-n(n-1)}{2\times 365}}
\end{aligned}$$

Another Approximation to the Birthday Problem

Another common and useful approximation is known as Stirling's approximation

$$k! \approx \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

we can then rewrite the probability as

$$\begin{aligned}
P(D_n) &= \frac{365!}{(365-n)!\,365^n} \\
&\approx \frac{\sqrt{2\pi}}{\sqrt{2\pi}} \frac{\sqrt{365}}{\sqrt{(365-n)}} \frac{365^{365}}{(365-n)^{365-n}} \frac{e^{-365}}{e^{-(365-n)}} \frac{1}{365^n} \\
&\approx \left(\frac{365}{365-n}\right)^{(365.5-n)} e^{-n}
\end{aligned}$$

Approximation Results

n	log approx	Sterling approx	exact
10	0.88401	0.88306	0.88305
20	0.59419	0.58857	0.58856
30	0.30368	0.29369	0.29368
40	0.11801	0.10877	0.10877
50	0.03487	0.02963	0.02963
60	0.00783	0.00588	0.00588
70	0.00133	0.00084	0.00084

Why is Sterling's approximation better?

1 More Conditional Probability

2 Probability Distributions

Probability Distributions

Description of the probability for all outcomes (values) of an outcome space (random variable).

We have to distinguish between the discrete and continuous case:

- Discrete - easy to assign probability to each event (even if there are infinitely many)
  - Value of the role of a six sided die
  - Number of coin flips until the first head
- Continuous - probability defined based on an interval
  - Probability a student's height is *exactly* 5'9"
  - Probability a students height is between 5'9" and 5'10"

Distributions Functions

If X takes only finitely-many or countably-many values, then we can list them and just report

$$f(x) = P(X = x)$$

- When X is discrete - Probability *mass* function

$$P(X \in A) = \sum_{x \in A} f(x)$$

- When X is continuous - Probability *density* function

$$P(X \in A) = \int_A f(x)dx$$

Probability mass function

Let X be the number of aces in two draws *without replacement* from a 52-card deck.

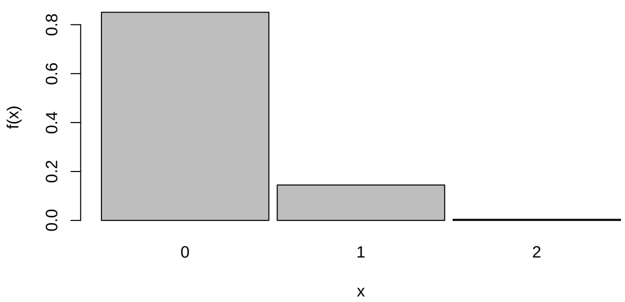
The PMF can be presented in a variety of ways:

	x	f(x)
No aces	0	$(48/52)(47/51) =$ 0.8507
One ace	1	$2(4/52)(48/51) =$ 0.1448
Two aces	2	$(4/52)(3/51) =$ 0.0045

Probability mass function

Let X be the number of aces in two draws *without replacement* from a 52-card deck.

The PMF can be presented in a variety of ways:



Probability mass function

Let X be the number of aces in two draws *without replacement* from a 52-card deck.

The PMF can be presented in a variety of ways:

For $x \in \{0, 1, 2\}$,

$$P(X = x) = \frac{\binom{4}{x}\binom{48}{2-x}}{\binom{52}{2}} = \frac{4!}{x!(4-x)!}\,\frac{48!}{(2-x)!(46+x)!}\,\frac{2!\,50!}{52!}$$

$$\approx \frac{2.25 \times 10^{59}}{x!(4-x)!(2-x)!(46+x)!}$$

Geometric Distribution

Bernoulli trial

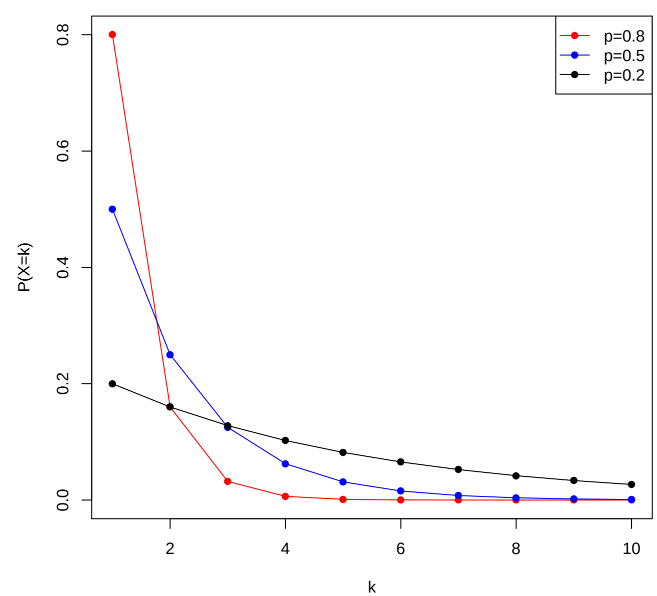
Random trial where there are two possible outcomes: "Success", "Failure", where the probability of a success is defined as p and the probability of a failure is $q = 1 - p$.

Geometric distribution describes the probability of the first success not occuring until the k^{th} Bernoulli trial.

Let X be a random variable of the number of Bernoulli trials with probability p needed to get one success then

$$P(X = k) = (1-p)^{k-1}p$$

Geometric Distribution, cont.



Probability density function

While we haven't seen one yet (we'll see many examples later) some random variables can take on infinitely many values.

Imagine a "spinner" that gives a random value "X" between zero and one.

Probability Distributions

Probability density function

While we haven't seen one yet (we'll see many examples later) some random variables can take on infinitely many values.

Imagine a "spinner" that gives a random value "X" between zero and one.

The PDF is represented as:

$$f(x) = \begin{cases} 1 & \text{for } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

$$P(X = x) = 0 \text{ for all } x$$

$$P(a \le x \le b) = \int_a^b f(x)\, dx = b - a \quad \text{for } 0 \le a \le b \le 1$$

This is a special case of the Uniform Distribution, $U \sim \mathcal{U}(0,1)$.

Uniform Distribution

If a random variable X has constant probability over a range $[a,b]$ then X has a uniform distribution, $X \sim \mathcal{U}(a,b)$.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$