

probability

Adam Okulicz-Kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Wednesday 6th October, 2021 05:28

outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

your own research

- how is it going?
- again, get data asap and start looking at it
 - “get your hands dirty with data”!
- don't wait—it takes time to get the data
 - AND you may design some great framework, only to find there are no data

outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

ps2 comments fa2021

- always do pay attention to results, be critical, does it make sense?
- apparently scatterplots in excel came out wrong for many folks, eg camden is 70 and 0, not 70 and 4!
- #2 is about proportions or percents; not really about mean, median, mode; correlation is fine, yet just 3 data points so not super useful

ps2

- always show your work, eg formula for mean, and plugin numbers
- always interpret, say you calculated corr as -0.95 , do say it is very strong and so strongly supports the hypothesis that...
- same with scatterplot and everything you calculate, always interpret!!
- for compliments/complains the key of course is to calculate proportions or percentages; interpretations could be quite different
- remember that corr is not causation!

- few of you confused pos with neg correlation!
- as a custom put independent/explanatory variable (counseling) on X axis; and outcome or explained (dropout) on Y
- do not connect points on scatterplot! if anything software must fit the line that goes kind of in the middle (it is NOT simple line connecting points)
- always label all axes in graphs

outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

evolutionary and counterintuitive!

- evolution made us to survive in an environment that is long time gone!
- so is our cognitive function and probability calculation off!
- and so we need statistics to help us think better!
- eg: overestimate prob of memorable/flashy events
- terrorist attack and airplane crash are similar to mistaking stick for snake etc—better be extra careful and see even if it's not there
- we underestimate, on the other hand
- much more deadly effects of sugar and fat (which were always rare and desirable)

making right decisions

- people and orgs make mistakes bc miscalculate prob
 - eg gambling and lotteries
- smoking (hundreds or thousands of % increased risks!)
- flying v driving, etc
 - 9/11 killed additional thousands
 - because people chose to drive (Wheelan, 2013, p.72-3)
- think about probabilities when making a decision
- the easiest (but already informative and helpful):
 - $\frac{\text{occurrences}}{\text{total}}$; eg : $\frac{\text{cancers}}{\text{smokers}}$, $\frac{\text{crashes}}{\text{miles}}$, $\frac{\text{crashes}}{\text{hours travelled}}$
 - so depends how you measure ! car v flight: hours v miles !

why is prob relevant to MPA student?

- probability may be confusing !
- but probability improves thinking/decision making
 - it is everywhere !
- usually don't realize it, but we calculate prob all the time!
 - and it helps our lives enormously if can do it better!

important for organizations!

- eg identify teachers who cheat or doctors who overcharge
- (just identify outliers, and unlikely events (nontypical))
- (or overall changes/trends (for everyone) that are suspect (collusion?))
- public policy makers determining dangerous items
eg swimming pools kill more kids than guns
 - (just see how many there are and how many people die from them!—simple des stat)
- for more interesting examples see Levit's Freakonomics
<http://www.freakonomics.com/>

i used to have a policy about undocumented

emergencies

- and i do not have it anymore! thank god for probability!
- there were 7 students and 4 of them had their grandmas die
- the reported probability was too high to be plausible!
- so what's the prob of grandma dying during this semester?
- say avg grandma is expected to last at least (if not more) about 10 years, or 40 three-month periods in a year (about semester long)
- so about $1/40$, so for 10 students class: every 4 semesters one grandma dead

outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

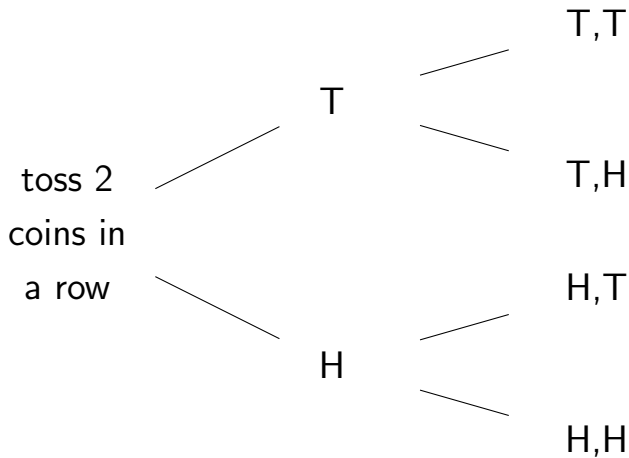
what is it?

- the numerical measure of the likelihood that the event will occur (the proportion of times the outcome would occur)
- it ranges from 0 to 1
- 0 means impossible
 - never 0, almost nothing is impossible eg $\frac{1 \text{ (crashes)}}{10m \text{ (flights)}}$
- 1 means certain
 - also almost never 1; almost nothing is certain

concepts

- event (A) , (B) , etc: a set of outcomes, eg $A = \{T, T\}$
- sample space (S) : the finest grain, mutually exclusive, collectively exhaustive listing of all possible outcomes
- mutually exclusive (disjoint):
 - 2 outcomes can't occur at the same time, eg T and H
- collectively exhaustive:
 - 1 outcome must occur: eg T or H
- eg toss a coin twice
- $S = \{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}$
- $A = \{T, T\}$

tree



table

HH	HT
TH	TT

- 1st row H in first flip
- 2nd row T in first flip
- 1st column H in second flip
- 2nd column T in second flip

exercises...

- what is the probability of getting 2 T in 2 flips ?
- we just showed with tree and table that there are 4 possible events, and only one outcome with 2 T, so $P = \frac{1}{4}$
- how about at least 1 T ?
- $P = \frac{3}{4}$
- how about exactly 1 T ?
- $P = \frac{2}{4}$

some useful properties and useful language

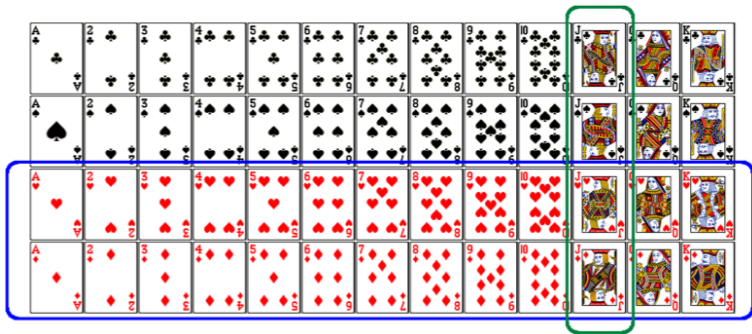
- intersection (and) \cap (both have to happen)
 - multiply prob ($= < 1$), so less likely to have both (frownie)
- union (or) \cup (at least one has to happen)
 - add prob, so more likely to have both (smiley)
- eg $A = H \ B = H$
 - $A \cap B = \{H, H\}$
 - $A \cup B = \{H, H\}, \{H, T\}, \{T, H\}$
- if mutually exclusive: $P(A \cup B) = P(A) + P(B)$
- if mutually exclusive and collectively exhaustive:
$$P(A) + P(B) = 1$$
 - eg $A = \text{tail} \ B = \text{head}; A = \text{male} \ B = \text{female}$

cards examples

- $P(\text{heart}) = \frac{1}{4}$
- mutually exclusive
$$P(A \cup B) = P(A) + P(B)$$
- $P(\text{ace or king}) = P(\text{ace}) + P(\text{king}) = \frac{1}{13} + \frac{1}{13} = 2/13$
- not mutually exclusive
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
- $P(\text{ace or black}) = P(\text{ace}) + P(\text{black}) - P(\text{ace and black}) = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = 7/13$
- $P(\text{heart}|\text{red}) = 1/2$ because sample space is reduced to 26 red cards (will get back to it at the end!)

Union of non-disjoint events

What is the probability of drawing a jack or a red card from a well shuffled full deck?



$$P(\text{jack or red}) = P(\text{jack}) + P(\text{red}) - P(\text{jack and red})$$

$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

What is the probability that a randomly sampled student thinks marijuana should be legalized or they agree with their parents' political views?

<i>Legalize MJ</i>	<i>Share Parents' Politics</i>		<i>Total</i>
	<i>No</i>	<i>Yes</i>	
No	11	40	51
Yes	36	78	114
Total	47	118	165

(a) $(40 + 36 - 78) / 165$

(b) $(114 + 118 - 78) / 165$

(c) $78 / 165$

(d) $78 / 188$

(e) $11 / 47$

Product rule for independent events

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$\text{Or more generally, } P(A_1 \text{ and } \dots \text{ and } A_k) = P(A_1) \times \dots \times P(A_k)$$

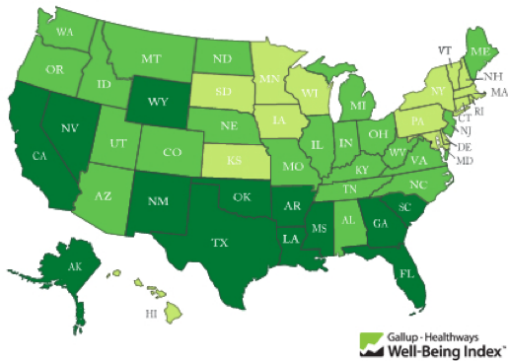
You toss a coin twice, what is the probability of getting two tails in a row?

$$\begin{aligned} &P(\text{T on the first toss}) \times P(\text{T on the second toss}) \\ &= (1 / 2) \times (1 / 2) = 1 / 4 \end{aligned}$$

A recent Gallup poll suggests that 25.5% of Texans do not have health insurance as of June 2012. Assuming that the uninsured rate stayed constant, what is the probability that two randomly selected Texans are both uninsured?

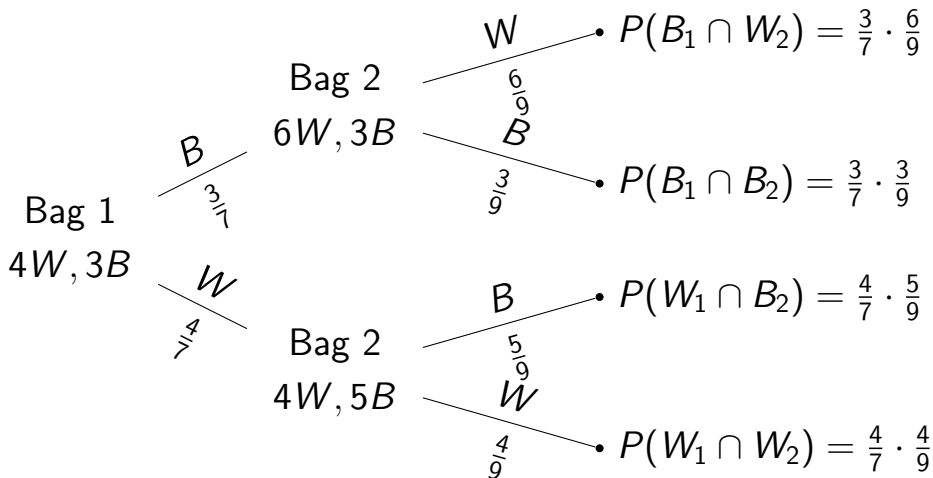
% Uninsured, January-June 2012

■ Higher range ■ Midrange ■ Lower range



- (a) 25.5^2
- (b) 0.255^2
- (c) 0.255×2
- (d) $(1 - 0.255)^2$

tree example

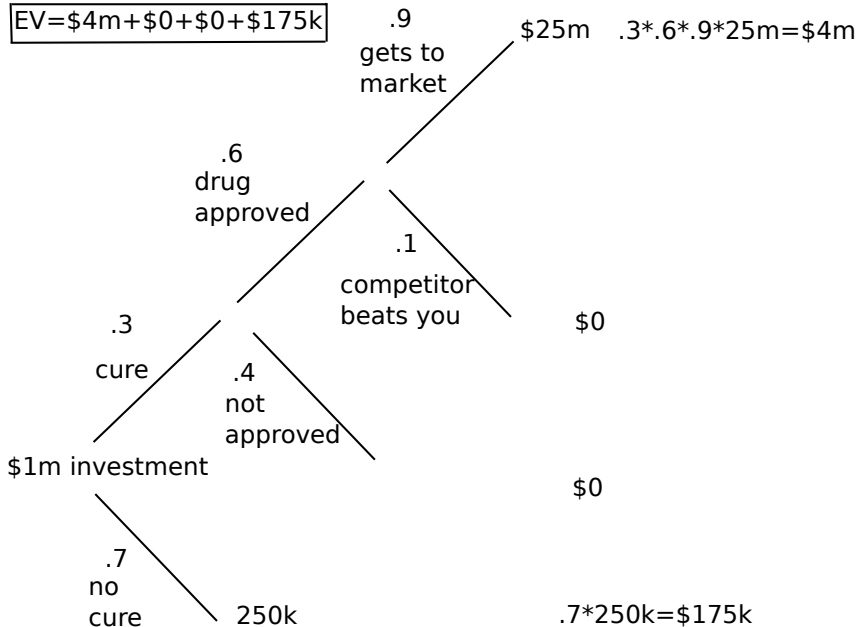


http://www.onemathematicalcat.org/Math/Algebra_II_obj/prob_tree_diagrams.htm

Expected Value (Wheelan, 2013, p83)

- just multiply value (\$ amount) by associated probability(ies) AND add them up
- and this is how much you are expected to get on average

$$EV = \$4m + \$0 + \$0 + \$175k$$



outline

ps1, ps1-1

general ps2 comments

intuition and why bother?

computing probability

conditional probability and independence

conditional probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- you have $P(A \cap B)$ in numerator because both A and (\cap) B need to happen to be conditional on B, if A happens but not B, then it cannot be conditional on B

table practice (all numbers in the body are "n")

Category	Type of Policy (%)			Total %
	Fire	Auto	Other	
Fraudulent	6	1	3	10
Nonfraudulent	14	29	47	90
Total	20	30	50	100

- $P(\text{fire}) = \frac{20}{100} = .2$
- $P(F|\text{fire})$ is $6/20$ or $.06/.2$
- $P(\text{fire}|F)$ is $.06/.1$

strategy

- yes, probability can be confusing...but
- probability is not a rocket science
- if you think about it you can figure it out
- formulas may be more confusing than revealing
- you can just use tables or possibly best to use trees...
- i don't care what method you use
- we'll practice more next week!

LEVITT, S. D. AND S. J. DUBNER (2010): Freakonomics, vol. 61, Sperling & Kupfer.

WHEELAN, C. (2013): Naked statistics: stripping the dread from the data, WW Norton & Company.