Statistics for Life Sciences

Probability

Nathan Harmston

January 14, 2022

Review from Tuesday

not A	$P(A^C)$	1 - P(A)
A or B	$P(A \cup B)$	P(A) + P(B) - $P(A \cap B)$
A and B	$P(A \cap B)$	$P(A\|B)P(B)$

de Mere problem

Chevalier de Méré was a gambler who played two different dice games and was confused by the outcome,

- bet on the chance of at least one six on four rolls of a six-sided die
- bet on the chance of at least one double-six on 24 rolls of two dice

Hey Pascal, why do I win money with the first bet, but I lose my money with the second bet?

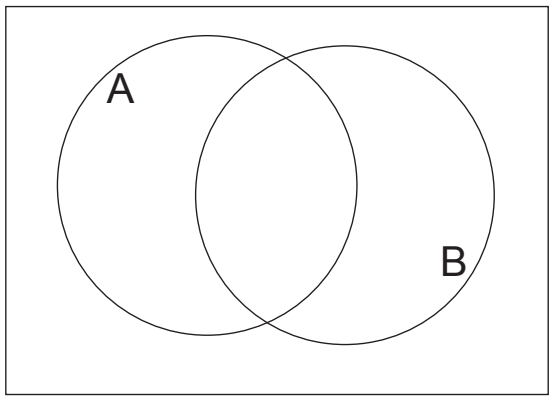
because P(Roll a 6 on throw 1 $\cup$ Roll 6 on throw 2) =1/6 + 1/6 = 1/3
... so therefore 4* 1/6 = 2/3

because P(Roll a double 6) =1/36 = ... so therefore 24* 1/36 = 2/3

de Meres problem

Conditional probability

$P(A|B)$ - probability that event A will occur given that event B has occurred

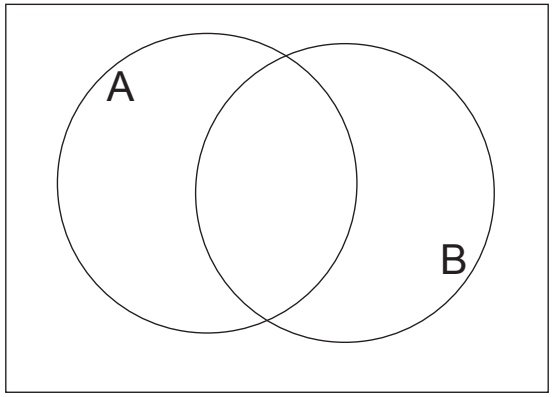


$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(B) > 0$

Conditional probability

$P(B|A)$ - probability that event B will occur given that event A has occurred



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

$P(A) > 0$

https://setosa.io/conditional/

Conditional probability

$P(A|B)$ - probability that event A will occur given that event B has occurred

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

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

$P(B) > 0$

https://setosa.io/conditional/

Conditional probability

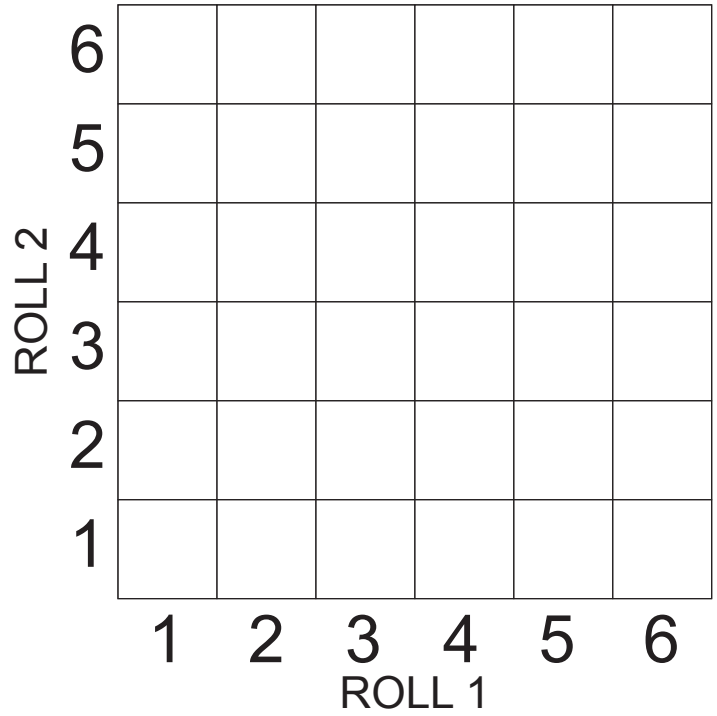
so P(A|B) is not the same as P(B|A) ....

- the probability that a statistics book is boring is not the same as the probability that a boring book is about statistics
- the probability that someone with abdominal pain has appendicitis is not the same as the probability that someone diagnosed with appendicitis will have had abdominal pain

Motulsky CHAPTER 2

Sometimes probability is just simply just counting

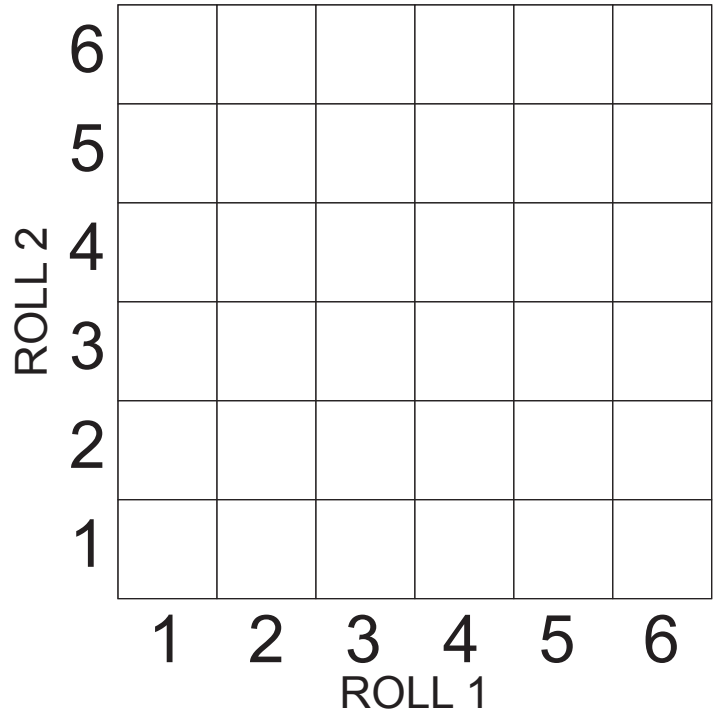
Sample space



$$B = \min(X,Y) = 2$$
$$M = \max(X,Y)$$
$$P(M = 1 \mid B)?$$

Sometimes probability is just simply just counting

Sample space



$$B = \min(X,Y) = 2$$
$$M = \max(X,Y)$$
$$P(M = 2 \mid B)?$$

Linda is thirty-one years old, single, outspoken and bright. In college, she majored in philosophy. While a student she was deeply concerned with discrimination and social justice and participated in antinuclear demonstrations

Linda is thirty-one years old, single, outspoken and bright. In college, she majored in philosophy. While a student she was deeply concerned with discrimination and social justice and participated in antinuclear demonstrations

Rank these statements from 1 to 8 according to their probability* ...

- Linda is active in the feminist movement
- Linda is a psychiatric social worker
- Linda works in a bookstore and takes yoga classes
- Linda is a bank teller and is active in the feminist movement
- Linda is a teacher in elementary school
- Linda is a member of the League of Women voters
- Linda is a bank teller
- Linda is an insurance salesperson

1 is most likely, 8 is least likely

So tell me about Linda?

Linda is thirty-one years old, single, outspoken and bright. In college, she majored in philosophy. While a student she was deeply concerned with discrimination and social justice and participated in antinuclear demonstrations

Linda is active in the feminist movement	2.1
Linda is a psychiatric social worker	3.1
Linda works in a bookstore and takes yoga classes	3.3
Linda is a bank teller and is active in the feminist movement	4.1
Linda is a teacher in elementary school	5.2
Linda is a member of the League of Women voters	5.4
Linda is a bank teller	6.2
Linda is an insurance salesperson	6.4

1 is most likely, 8 is least likely

Linda is thirty-one years old, single, outspoken and bright. In college, she majored in philosophy. While a student she was deeply concerned with discrimination and social justice and participated in antinuclear demonstrations

Linda is active in the feminist movement	2.1
Linda is a psychiatric social worker	3.1
Linda works in a bookstore and takes yoga classes	3.3
Linda is a bank teller and is active in the feminist movement	4.1
Linda is a teacher in elementary school	5.2
Linda is a member of the League of Women voters	5.4
Linda is a bank teller	6.2
Linda is an insurance salesperson	6.4

whats wrong here?

1 is most likely, 8 is least likely

Judgment Under Uncertainty: Heuristics and Biases - Kahneman, Slovic, Tversky

So did you get this right?

How can P(*Linda is a bank teller and is active in the feminist movement*) be greater than P(*Linda is a bank teller*)?

So did you get this right?

How can P(*Linda is a bank teller and is active in the feminist movement*) be greater than P(*Linda is a bank teller*)?

the probability that two events will both occur can never be greater than the probability that each will occur individually
sometimes peoples intuition leads them to make mistakes

Human beings can be really bad at understanding probability, and understanding risk*

i.e. *risk of vaccine side-effects vs. risk from covid

Bayes' theorem

Inverse problems

given P(A|B) what is P(B|A)?

reverse conditional probabilities

Baye's rule

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

We don't do Bayesian calculations intuitively

Question

Only a tiny fraction (0.1%) of the people have a disease D. A test for this disease is highly accurate but not quite perfect. It correctly identifies 99% of patients with the disease but also incorrectly concludes that 1% of the noninfected samples have the disease. When this test identifies a blood sample as having HIV present, **if you have a positive result what is the chance that you have the disease**?

Bayes' theorem

- P(D) - probability of having the disease -
- P(T) - probability of having a positive test
  - $P(b) = \Sigma P(a_i) x P(b|a_i)$ - law of total probability
- P(D | T) = what we want to find - probability of disease given a positive test
- P(T | D) = probability of a positive test given you have the disease

Bayes' theorem

$$P(D|T) = 0.09$$

the interpretation of the result depends on the fraction of the population that has the disease

what happens if we test again and it comes back positive? (replication is always good)

Bayes' theorem

- statistical inference
- spam filters
- Bayes theory is to theory of probability what the Pythagorean theorem is to geometry - *some smart statistician*
- allows you to invert probabilities
- *Bayesian statistics*

Now .... back to flipping a coin

$P(H) = 0.5$, $P(T)=0.5$

what is the probability of $P(H|HHH)$?

Now .... back to flipping a coin

$P(H) = 0.5$, $P(T)=0.5$

what is the probability of $P(H|HHH)$?

sometimes people think that probability has a memory - **Gambler's fallacy**

Independence

independence of two events?

$$P(A|B) = P(A)$$

$$P(A \cap B) = P(A)P(B)$$

seeing B provides no information about the occurrence of A (and its symmetric)

statistical independence refers to the case where one cannot predict anything about one variable from the value of another variable

Independence

a married person has a 1/50 chance of getting divorce each year

a police officer has 1/5000 chance of getting killed on the job

what is the chance that a married police officer will be killed in the same year?

Independence

is preferring a specific show independent from sex?

	Male	Female	
The Wire	220	180	400
My little pony	200	50	250
Game of Thrones	100	250	350
	460	540	1000

is preferring a specific show independent from sex?

	Male	Female	
The Wire	220	180	400
My little pony	200	50	250
Game of Thrones	100	250	350
	460	540	1000

	Male	Female	
The Wire	0.22	0.18	0.4
My little pony	0.20	0.05	0.25
Game of Thrones	0.1	0.25	0.35
	0.46	0.54	1

Independence

	Male	Female	
The Wire	0.22	0.18	0.4
My little pony	0.20	0.05	0.25
Game of Thrones	0.1	0.25	0.35
	0.46	0.54	1

Independence

	Male	Female	
The Wire	0.22	0.18	0.4
My little pony	0.20	0.05	0.25
Game of Thrones	0.1	0.25	0.35
	0.46	0.54	1

Independence

	Male	Female	
The Wire	0.22	0.18	0.4
My little pony	0.20	0.05	0.25
Game of Thrones	0.1	0.25	0.35
	0.46	0.54	1

we'll discuss testing for independence of two variables later on

The King's sibling problem

The King comes from a family of two children .... what is the probability that his sibling is female?

what are my assumptions?

Odds

some fields prefer to use odds (odds ratios are pretty important in epidemiology / case-control studies) instead of probabilities - it is the probability that an event will occur divided by the probability that the event will not occur

$$odds = \frac{p}{1-p} \tag{1}$$

$$p = \frac{odds}{1+odds} \tag{2}$$

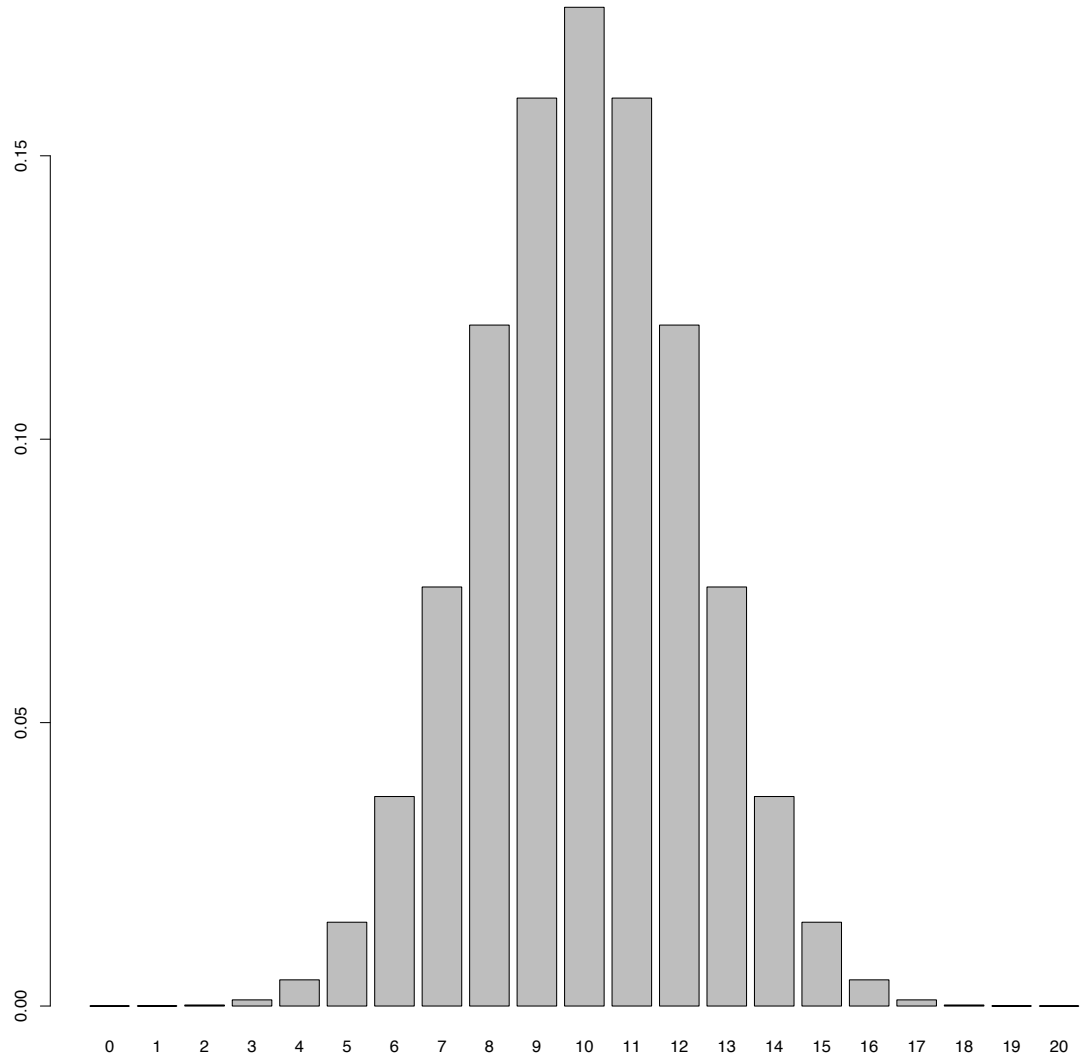
p=0.5 is the same as odds of 1.0, whiles p=0.75, odds 3 to 1

Probability distributions

gives the probabilities of occurrence of different events for an experiment

Binomial distribution - coin flipping

discrete probability distribution - sequence of Bernoulli trials



Binomial distribution

Requirements

- two outcomes per trial
- probability of success is the same for each trial (θ)
- number of trials is fixed (N)
- independence of trials

$$P(X) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$X \sim Binomial(X|\theta, N)$$

?dbinom

Binomial distribution - left handedness

10% of people are left-handed, a random sample of 20 people are taken

whats the probability

- that all 20 people are left-handed?
- no-one is left-handed?
- exactly 5 people are left-handed?
- at least 2 people are left-handed?

Binomial distribution - left handedness

all 20 people are left-handed?

Binomial distribution - left handedness

no-one people are left-handed?

Binomial distribution - left handedness

5 people are left-handed?

Binomial distribution - left handedness

at least 2 people are left-handed?

Distribution functions in R

- **dbinom** - probability **d**ensity/mass function at each point
- **pbinom** - cumulative **p**robability of an event
- **qbinom** - takes a probability and gives the number whose cumulative value matches the probability value - **q**uantile of
- **rbinom** - generates a number of **r**andom values of given probability from a given sample

Questions?

- Next week - handling and visualising data - probability distributions
- Reading - Motulsky - Chapters 7,8,9, 10
- Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm - https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.10
- First problem set