Week 8: Probability Rules and Conditional Probabilities

Data Analysis for Psychology in R 1

Alex Doumas

Department of Psychology
The University of Edinburgh

Course Overview

Exploratory Data Analysis	Research design and data
	Describing categorical data
	Describing continuous data
	Describing relationships
	Functions
Probability	Probability theory
	Probability rules
	Random variables (discrete)
	Random variables (continuous)
	Sampling

Foundations of inference	Confidence intervals
	Hypothesis testing (p-values)
	Hypothesis testing (critical values)
	Hypothesis testing and confidence intervals
	Errors, power, effect size, assumptions
Common hypothesis tests	One sample t-test
	Independent samples t-test
	Paired samples t-test
	Chi-square tests
	Correlation

Learning objectives

- Understand the use of probability rules
- Understand the basics of and how to use Bayes' equation
- Understand use of probability to test independence
- Apply probability rules and compute probability using R

Recap

Probabilities of multiple events

- Probability of A **and** B = $P(A \bigcap B)$
  - Joint event or *intersection* of A and B
  - Both A and B have occurred
- Probability of A **or** B = $P(A \bigcup B)$
  - *Union* of A and B
  - A, B, or both have occurred
- Probability of **not** A = $P(A^c)$
  - We describe the event not A as the *complement* of A
  - Elements in U that are not A

Relations between events

- **Mutually exclusive** events: If A occurs, B can not occur
- **Independent** events: The occurence of event A does not impact event B
- **Dependent** events: The occurence of event A **does** impact event B
    - "Impact" means that whether A occurs changes the probability of B

Rules of Probability

Rules of probability

Rule	Formula
1. Range	$0 \leq P(A) \leq 1$
2. Sum of Outcomes	$P(a_1) + P(a_2) + \ldots P(a_i) = 1$
3. Complement Rule	$P(A^c) = 1 - P(A)$
4a. Simple Addition Rule	$P(A \bigcup B) = P(A) + P(B)$
4b. General Addition Rule	$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
5a. Simple Multiplication Rule	$P(A \bigcap B) = P(A)P(B)$
5b. General Multiplication Rule	$P(A \bigcap B) = P(A)P(B\|A)$
6. Conditional Probability	$P(B \mid A) = \frac{P(A \bigcap B)}{P(A)}$

- As we discuss these rules in more depth, imagine that our random experiment is *drawing a single card from a fair deck*

Rules of probability

1. The probability of any event will fall between 0 and 1

$$0 \leq P(A) \leq 1$$

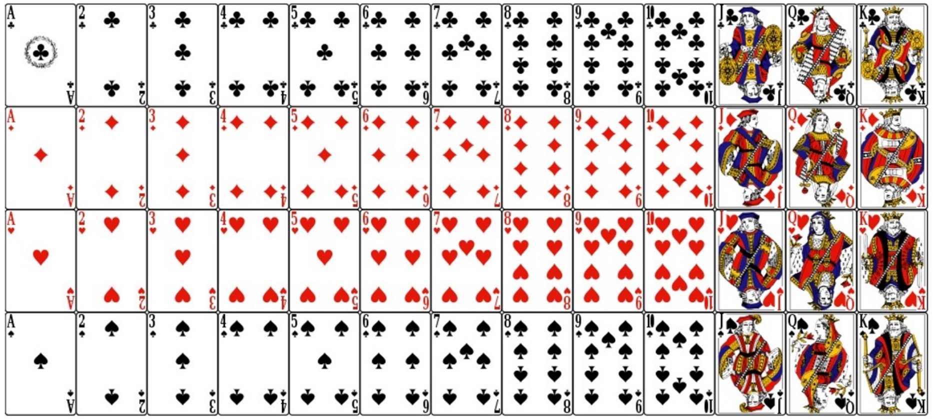
- As $P(A)$ approaches 1, A becomes more likely to occur
- $P(A) = 0$ indicates that event A is impossible
- $P(A) = 1$ indicates that event A will absolutely occur

Rules of probability

2. The sum of the probabilities of all possible outcomes = 1

$$P(a_1) + P(a_2) + \ldots P(a_i) = 1$$

- Because the sample space includes *all possible* outcomes of an experiment, one of these outcomes *must occur*
- $P = 1$ indicates an event will certainly occur
- Therefore, the probability of all events within the sample space will sum to 1



$$P(black) + P(red) = 1$$

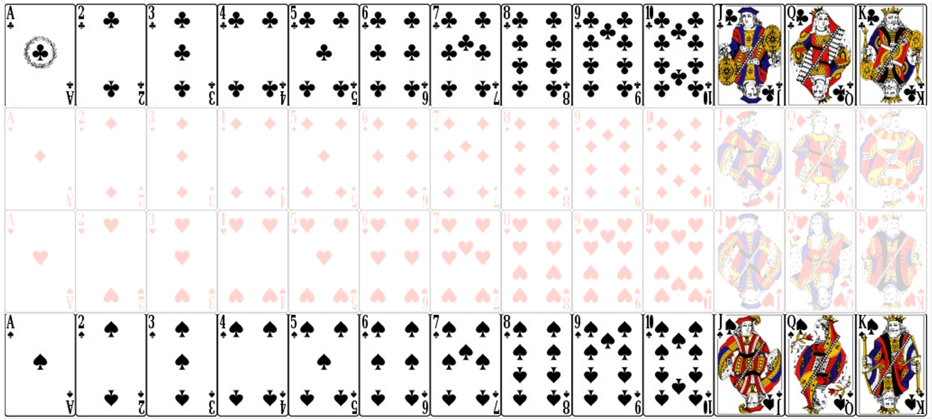
$$\frac{26}{52} + \frac{26}{52} = 1$$

Rules of probability

3. Complement Rule

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

- The probability of A^c (Not A) is equal to 1 - the probability of A
- Notice how this is connected to Rule 2 (the probabilities of all possible outcomes sum to 1)



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

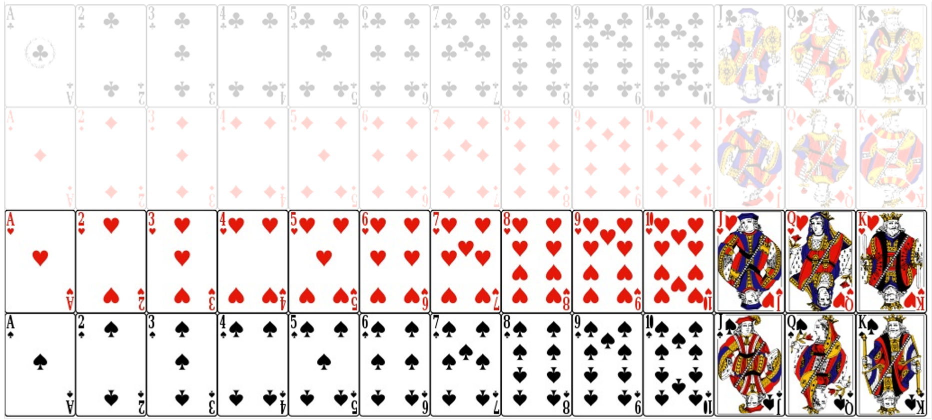
$$P(red^c) = 1 - \frac{26}{52}$$

Rules of probability

4a. Simple Addition Rule

$$P(A \bigcup B) = P(A) + P(B)$$

- The probability that one or both events occur
- This rule can only be used for mutually exclusive events



$$P(spade \bigcup heart) = \frac{13}{52} + \frac{13}{52}$$

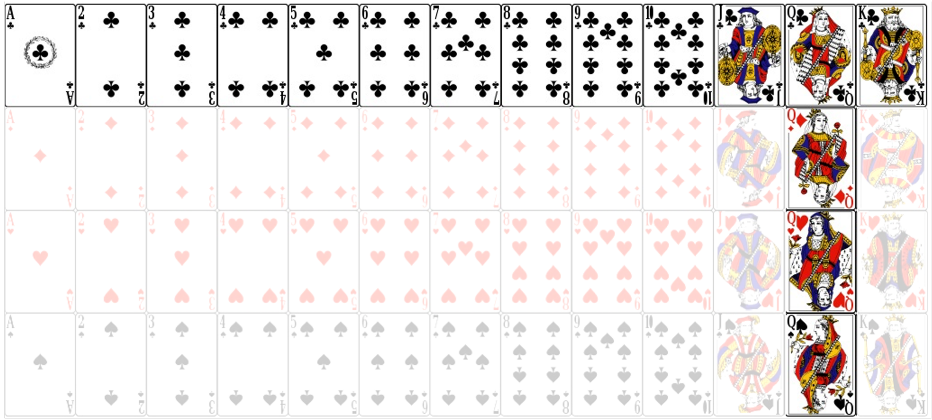
$$P(spade \bigcup heart) = \frac{1}{2}$$

Rules of probability

4b. General Addition Rule

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

- The probability that one or both events occur when events *are not necessarily* mutually exclusive
- However, it could also be used when events are mutually exclusive, as $P(A \bigcap B) = 0$



$$P(club \bigcup queen) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$P(club \bigcup queen) = \frac{16}{52} = \frac{4}{13}$$

Rules of probability

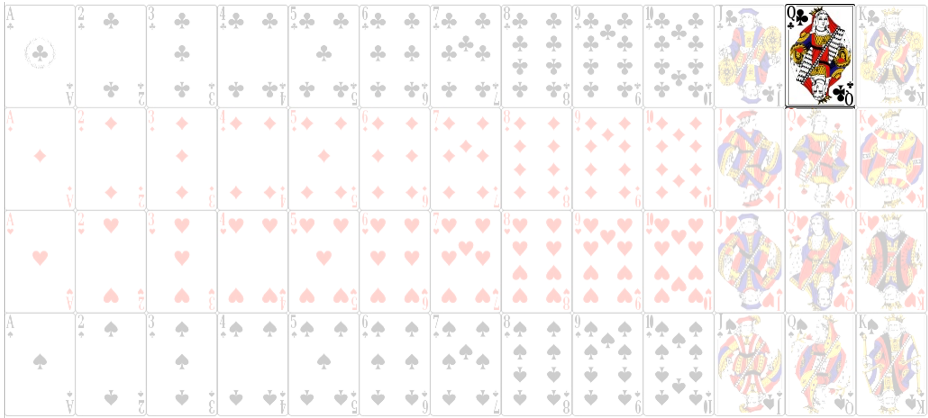
5a. Simple Multiplication Rule

$$P(A \bigcap B) = P(A) \times P(B)$$

also written as:

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

- The probability of the co-occurrence of A and B *if* they are independent events
- The probability of A and B will always be less than or equal to the probability of either single event



$$P(club \bigcap queen) = \frac{13}{52} \times \frac{4}{52}$$

$$P(club \bigcap queen) = \frac{1}{52}$$

Question

- Why is the probability of A and B always less than or equal to the probability of either single event?

- Because the probability of A and B is reflective of the intersection of these events, and this intersection will never be larger than the probability of the least likely event

- Suppose we have two events
  - Event A occurs $\frac{1}{4}$ of the time
  - Event B occurs $\frac{1}{2}$ of the time

- If the events are independent (the probability of one does not impact the probability of the other), then event B will happen with equal probability when both A and not A

  - So, of the total number of times A occurs, B will occur half the time
  - $P(A \bigcap B) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$

Rules of probability

5b. General multiplication rule

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

- The probability of the co-occurrence of A and B when they are not necessarily independent events
- Note that *and* is commutative (i.e., $A \bigcap B = B \bigcap A$).
  - Probability of rain and Tuesday can't be different from the probability of Tuesday and rain
- But what is $P(B|A)$?

Conditional probability

- $P(B|A)$: Probability of B **given** A
- This is referred to as **conditional probability**
  - When events are dependent, the likelihood of one event changes based on the outcome of the other
- Note that when $P(A)$ and $P(B)$ are independent, then the $P(B|A) = P(B)$, hence the simple multiplication rule for independent events
- We can calculate conditional probability as:

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

- or the inverse:

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

Conditional probability: an example

- Let's imagine we have data on the handedness of children who like and dislike comic books in a classroom
- Just by looking at our data, the likelihood of being left-handed changes depending on whether the student is a boy or a girl
- Specifically, if a student is a boy, they are more likely to be left-handed than if they are a girl
- **If we select one comic liker from the class, what is the probability that they are left handed?**

	Left	Right	Marginal
Dislike	25	24	49
Like	10	41	51
Marginal	35	65	100

Conditional probability: an example

$$P(Left|Like) = \frac{P(Left \bigcap Like)}{P(Like)}$$

$$P(Left|Like) = \frac{\frac{10}{100}}{\frac{51}{100}}$$

$$P(Left|Like) = \frac{0.10}{0.51} = .196$$

	Left	Right	Marginal
Dislike	25	24	49
Like	**10**	41	**51**
Marginal	35	65	100

> An important note: while $P(A \bigcap B)$ is equal to $P(B \bigcap A)$, $P(A|B)$ is not equal to $P(B|A)$
>
> - $P(Left|Like) = \frac{.10}{.51} = .196$
> - $P(Like|Left) = \frac{.10}{.35} = .286$
> - Another way to think about it: You draw a single card from a fair deck. Is the probability it's red given that it's a diamond equal to the probability that it's a diamond given that it's red?

Questions?

Bayes' Equation

Bayes' Rule

- Bayes' rule follows from the idea of conditional probability
- In a basic sense, it allows you to update your assessment of probability as you gather more evidence

We know that $P(B|A) = \frac{P(A \bigcap B)}{P(A)}$ and $P(A \bigcap B) = P(B \bigcap A)$

and $P(B \bigcap A) = P(B)P(A|B)$

So by substitution we get:

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

- Rearrange this to get Bayes' rule:

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

Who cares?

- Turns out to be a big deal in psychology and statistics
  - Many psychological models based on Bayes' rule
  - Bayesian statistics are becoming more prominent
- Can be really useful for calculating conditional probabilities when you don't know things like $P(A \bigcap B)$
- Can also be useful if you know $P(B|A)$ but not $P(A|B)$, or vice versa
  - E.g., $P(clouds|rain)$ is much easier to calculate than $P(rain|clouds)$
- Let's apply the equation to our previous handedness example

Conditional probability: Bayes' equation

$$P(Left|Like) = \frac{P(Like|Left)}{P(Like)} P(Left)$$

	Left	Right	Marginal
Dislike	25	24	49
Like	**10**	41	**51**
Marginal	**35**	65	100

$$P(Like|Left) = \frac{10}{35} = .2857$$

$$P(Left|Like) = \frac{0.2857}{.51} \times .35 = .196$$

Questions?

Assessing Independence of Events

Are events independent?

- The multiplication rule, and the use of contingency tables, provides a way to assess if events are independent
- **Example**: Consider we have a sample of 200 students and faculty members at the university
    - 110 students and 90 faculty members
    - Suppose we ask whether they live within one mile of campus. 140 say yes, 60 say no
- We can tabulate the proportions and think about the probabilities

Are events independent?

- Notice that we're not using raw frequencies as in the handedness example, but relative frequencies
- E.g. 110 students out of our sample of 200 = 55%
- 70% of our sample live within a mile of campus (`Yes`), while 30% live further away (`No`)

	Yes	No	Marginal
Students	$P(Student, Yes)$	$P(Student, No)$	.55
Faculty	$P(Faculty, Yes)$	$P(Faculty, No)$	.45
Marginal	.70	.30	1.00

Are events independent?

- **If role and distance are independent,** then the probability of living near campus should be 70% whether someone is a student or a faculty member

- With independent events, we can use the *multiplication rule* to compute the probability of each outcome:
  - $P(yes \bigcap student) = P(yes)P(student) = .70 \times .55 = .385$

	Yes	No	Marginal
Students	0.385	0.165	0.55
Faculty	0.315	0.135	0.45
Marginal	0.70	0.30	1.00

- The values within each cell are the proportions we would expect *if* role and distance are absolutely independent
  - E.g. we would expect ~32% of our sample to be faculty members who live within a mile of campus

Are events independent?

	Yes	No	Marginal
Students	0.45	0.10	0.55
Faculty	0.25	0.20	0.45
Marginal	0.70	0.30	1.00

- But what if, when we looked within our data, we actually observed the proportions shown here?
- Given this outcome, our conditional probabilities are actually:
  - $P(yes|student) = \frac{.45}{.55} = .82$
  - $P(yes|faculty) = \frac{.25}{.45} = .56$
- Using Bayes:
  - $\frac{P(student|yes)}{P(student)} P(yes) = \frac{\frac{.45}{.70}}{.55} \times .70 = .82$
  - $P(yes|faculty) = \frac{P(faculty|yes)}{P(faculty)} P(yes) = \frac{\frac{.25}{.70}}{.45} \times .70 = .56$

Are events independent?

Expected (if independent)

	Yes	No	Marginal
Students	0.385	0.165	0.55
Faculty	0.315	0.135	0.45
Marginal	0.70	0.30	1.00

$$P(yes|student) = \frac{.385}{.55} = .70$$

$$P(yes|faculty) = \frac{.315}{.45} = .70$$

$$P(no|student) = \frac{.165}{.55} = .30$$

$$P(no|faculty) = \frac{.135}{.45} = .30$$

Observed

	Yes	No	Marginal
Students	0.45	0.10	0.55
Faculty	0.25	0.20	0.45
Marginal	0.70	0.30	1.00

$$P(yes|student) = \frac{.45}{.55} = .82$$

$$P(yes|faculty) = \frac{.25}{.45} = .56$$

$$P(no|student) = \frac{.10}{.55} = .18$$

$$P(no|faculty) = \frac{.20}{.45} = .44$$

- Because the probabilities we observe within our data are not at all similar to the probabilities we would expect if role and distance are independent, we might conclude that these events actually *are* dependent

Our first statistical test

- What we have just done (more or less) is do all the background work to understand a χ^2 test of independence
- This tests the independence of two nominal category variables
- To use this in practice, we have a couple of extra steps covered later in DAPR1, but the above is the fundamental principle

Questions?

Summary of today

1. Review and extension of the rules of probability
2. Bayes' rule can be used to calculate conditional probabilities
3. How to do a simple test for independence

- Tomorrow, I'll present a live R session focused on computing probability based on these rules

This week

Tasks

- Attend both lectures
- Attend your lab and work together on the lab tasks
- Complete the weekly quiz
    - Opened Monday at 9am
    - Closes Sunday at 5pm

Support

- **Office hours**: for one-to-one support on course materials or assessments
  (see LEARN > Course information > Course contacts)
- **Piazza**: help each other on this peer-to-peer discussion forum
- **Student Adviser**: for general support while you are at university
  (find your student adviser on MyEd/Euclid)