First Day of Class

EES 4891-06/5891-01

Bayesian Statistical Methods

Jonathan Gilligan

Class #1: Sunday, January 05 2026

Introduce Yourselves

1. Name
2. What are your major or program, and research interests?
3. What's your year or how long you've been in your graduate program?
4. What previous math or stats have you taken?
5. What do you hope to get from this class?
6. Something interesting about yourself
7. (Optional) Ask me a question about some aspect of probability or statistics (preferably Bayesian)

What Is Statistics?

What Is Statistics?

- Define “statistics”
- What do you use statistics for?
- What kinds of questions do you expect statistics to answer?

What is Bayesian Statistics?

- Why are you taking this class?
- What do you know about Bayesian statistics?
- Any questions for me?

What is the probability the sun will rise tomorrow?

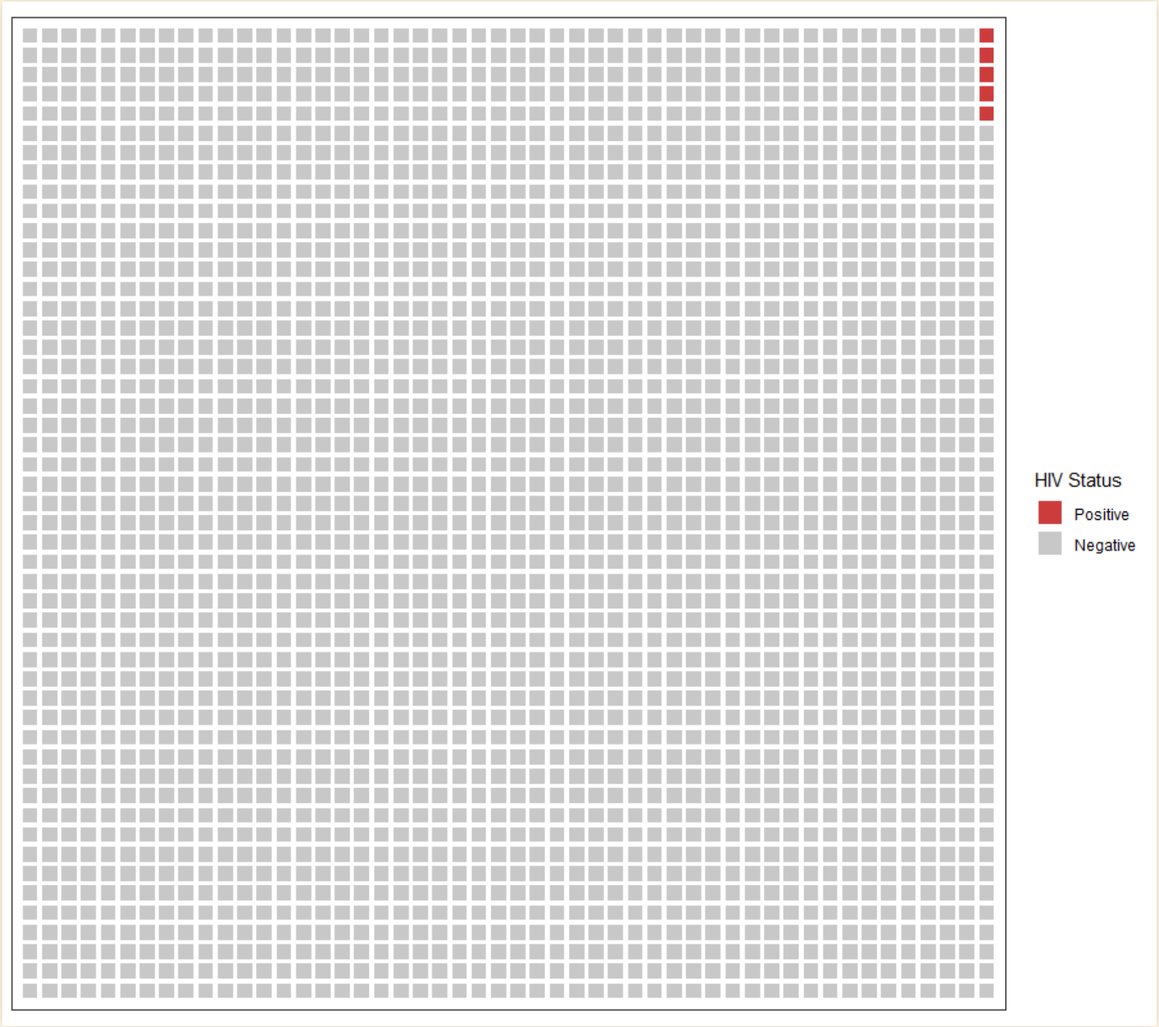
Disease testing

- Roughly 1 in 500 university students (0.2%) are infected with HIV
- Suppose a test has
  - 99.0% sensitivity (if someone has HIV, there's a 99.0% probability the test will correctly say "positive")
  - 99.0% specificity (if someone doesn't have HIV, there's a 99.0% probability the test will correctly say "negative")
- If the test is positive, what is the probability the person really has HIV?
- If the test is negative, what is the probability the person doesn't have HIV?

H.D. Gayle *et al.*. 1990. "Prevalence of Human Immunodeficiency Virus among University Students," *New England J. Medicine* **323**, 1538. doi: 10.1056/NEJM199011293232206

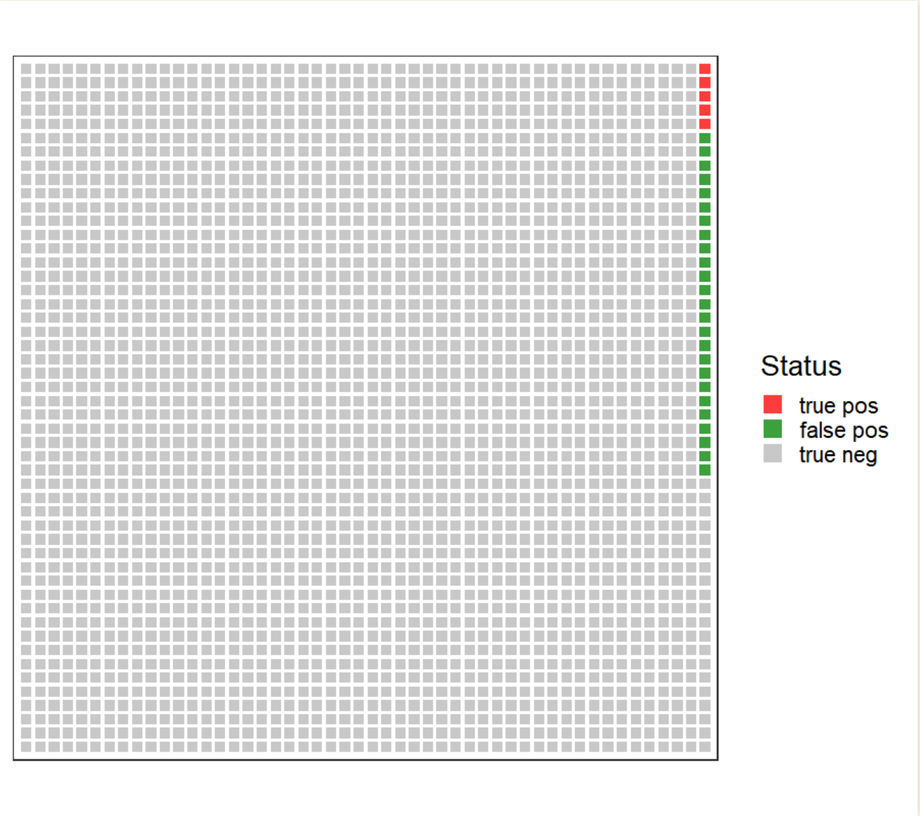
Disease testing:

- 2500 people
- Prevalence of HIV = 0.2%
  - 5 people are HIV positive.
- Sensitivity: 99.0%
- Specificity: 99.0%



Disease testing:

- 2500 people
- Prevalence of HIV = 0.2%
  - 5 people are HIV positive.
- Sensitivity: 99.0%
  - 5 HIV-positive people test *true positive*
  - 0 HIV-positive people test *false negative*
- Specificity: 99.0%
  - 2470 HIV-negative people test *true negative*
  - 25 HIV-negative people test *false positive*
- 30 people test positive.
  - 5 (16.7%) are *true-positive* results, representing people actually have HIV
  - 25 (83.3%) are *false-positive* results, representing people who don't have HIV.
- More exactly, if a university student tests positive, the probability they actually have HIV is 16.6%.
- If a student tests negative, the probability they don't have HIV is 99.998%.



What does this mean?

- When you do a standard statistical significance test, and it tells you $p = 0.01$, what does that mean?
- It means that *if* the null hypothesis were true, there would be a 1% chance that you'd observe data at least this far away from the exact null hypothesis.
  - If you don't have HIV, you have a 1.0% chance of getting a positive test ($p = 0.01$).
- But what we usually want to know is not the probability seeing the data we collected *if* the null hypothesis is true.
  - We want to know the probabilty that the null hypothesis is true *if* we see the data we observed.
  - That's a different question.
- To answer it, we need to know the probability that the null hypothesis was true before we collected the data, and then calculate how much our data changes that probability by giving us more information.

Bayes's Theorem

$$P(H|D, I) = \frac{P(D|H, I) \times P(H|I)}{P(D|I)}$$

- H: Test hypothesis (student has HIV)
- h: Null hypothesis (test hypothesis is false: student does not have HIV)
- D: New data (HIV test results)
- I: Other evidence we had before learning the new data

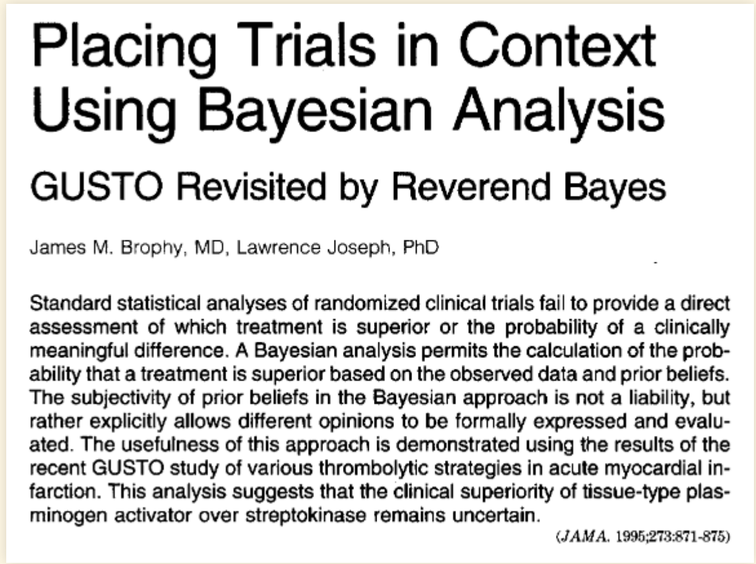
- $P(H|D, I)$ = *posterior:* Probability H is true, after learning D.
- $P(D|H, I)$ = *sensitivity:* likelihood we'd see D if H is true (true positive, if student has HIV).
- $P(H|I)$ = *prior* probability that H is true, before learning D.
- $P(D|I)$ = *evidence*: A measure of the additional information we get from learning D.

$$P(D|I) = P(D|H, I) \times P(H|I) + P(D|h, I) \times P(h|I)$$

- Tricky parts: Figuring out the *prior* and calculating the *evidence*

Applications

- Mid-1990s, two drugs used to treat heart attacks:
  - streptokinase (SK): widely used, inespensive.
  - Tissue plasminogen activator (t-PA): new, expensive, poorly understood.
- GUSTO: big clinical trial with 30,516 patients:
  - 6.3% of patients treated with t-PA died
  - 7.3% of patients treated with SK died.
  - Statistical significance (t-PA better than SK) was $p < 0.001$
- But there had been two previous clinical trials with a total of 47,294 patients found no benefit to t-PA.
- Combine all three trials using Bayesian methods:
  - If GUSTO was twice as credible as the other studies, 44% probability that t-PA was better.
  - If the other studies were just as credible as GUSTO, 17% chance that t-PA was better.



Placing Trials in Context Using Bayesian Analysis

GUSTO Revisited by Reverend Bayes

James M. Brophy, MD, Lawrence Joseph, PhD

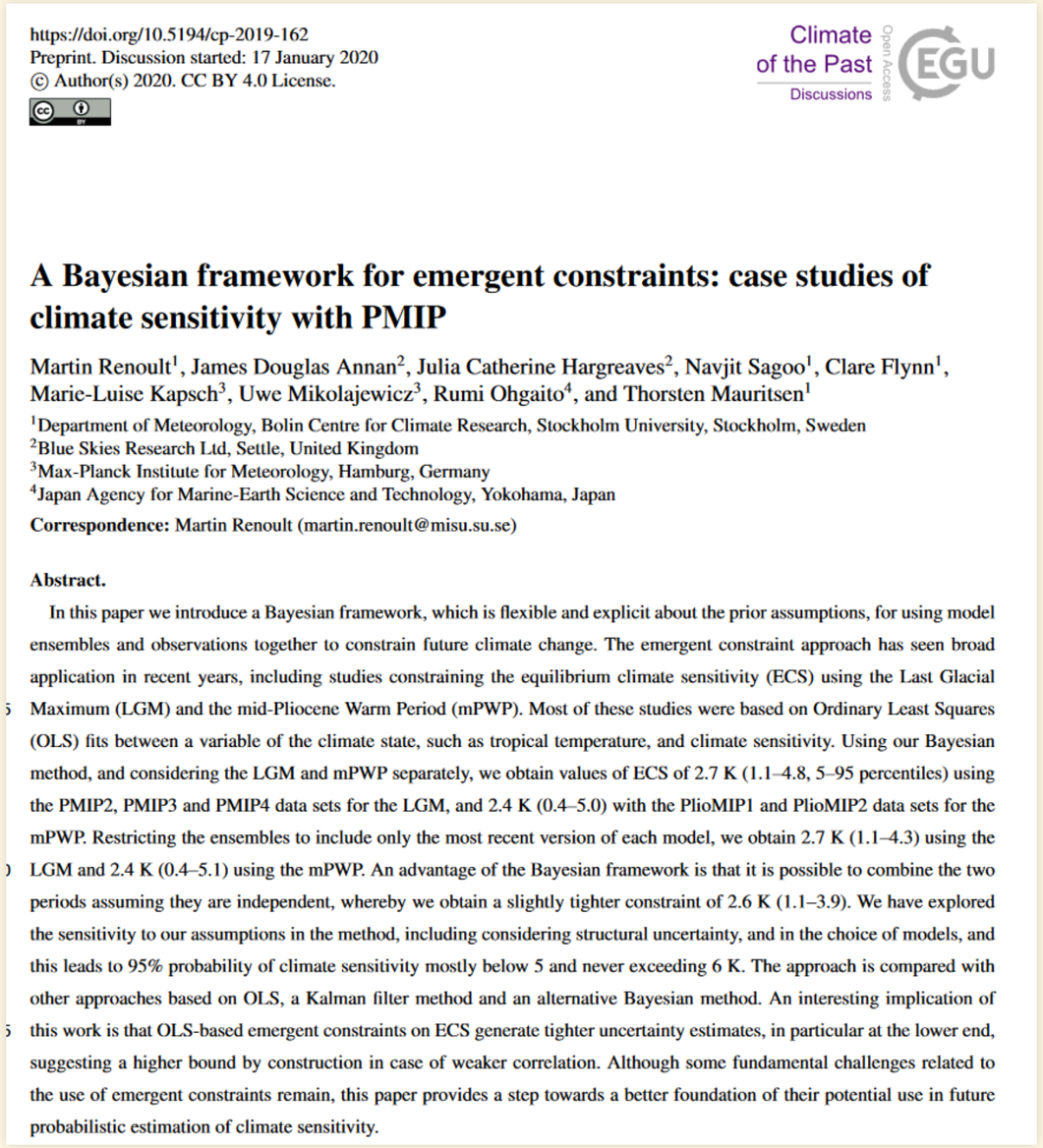
Standard statistical analyses of randomized clinical trials fail to provide a direct assessment of which treatment is superior or the probability of a clinically meaningful difference. A Bayesian analysis permits the calculation of the probability that a treatment is superior based on the observed data and prior beliefs. The subjectivity of prior beliefs in the Bayesian approach is not a liability, but rather explicitly allows different opinions to be formally expressed and evaluated. The usefulness of this approach is demonstrated using the results of the recent GUSTO study of various thrombolytic strategies in acute myocardial infarction. This analysis suggests that the clinical superiority of tissue-type plasminogen activator over streptokinase remains uncertain.

(*JAMA*. 1995;273:871-875)

- With either prior, there is less than 50% probability that t-PA is better than SK.

Sensitivity of the Climate System

- Climate models predict global warming, but we only have a few decades with which to test their accuracy
  - CO_2 has varied about 315–430 ppm (37%) during this time.
- Use Bayesian methods to combine paleoclimate data with current models.
  - More than 800,000 years of Pleistocene data
    - Much greater variations in CO_2 (180–280 ppm, 56%).
- Including paleoclimate data helps set an upper limit on how big climate sensitivity could be.



https://doi.org/10.5194/cp-2019-162
Preprint. Discussion started: 17 January 2020

Climate of the Past Discussions

A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP

Martin Renoult[1], James Douglas Annan[2], Julia Catherine Hargreaves[2], Navjit Sagoo[1], Clare Flynn[1], Marie-Luise Kapsch[3], Uwe Mikolajewicz[3], Rumi Ohgaito[4], and Thorsten Mauritsen[1]

[1]Department of Meteorology, Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[2]Blue Skies Research Ltd, Settle, United Kingdom
[3]Max-Planck Institute for Meteorology, Hamburg, Germany
[4]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

Correspondence: Martin Renoult (martin.renoult@misu.su.se)

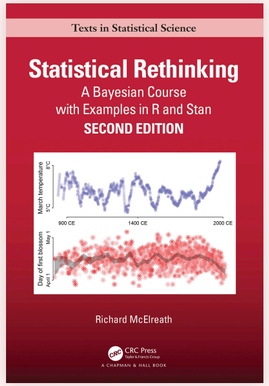
Abstract.

In this paper we introduce a Bayesian framework, which is flexible and explicit about the prior assumptions, for using model ensembles and observations together to constrain future climate change. The emergent constraint approach has seen broad application in recent years, including studies constraining the equilibrium climate sensitivity (ECS) using the Last Glacial Maximum (LGM) and the mid-Pliocene Warm Period (mPWP). Most of these studies were based on Ordinary Least Squares (OLS) fits between a variable of the climate state, such as tropical temperature, and climate sensitivity. Using our Bayesian method, and considering the LGM and mPWP separately, we obtain values of ECS of 2.7 K (1.1–4.8, 5–95 percentiles) using the PMIP2, PMIP3 and PMIP4 data sets for the LGM, and 2.4 K (0.4–5.0) with the PlioMIP1 and PlioMIP2 data sets for the mPWP. Restricting the ensembles to include only the most recent version of each model, we obtain 2.7 K (1.1–4.3) using the LGM and 2.4 K (0.4–5.1) using the mPWP. An advantage of the Bayesian framework is that it is possible to combine the two periods assuming they are independent, whereby we obtain a slightly tighter constraint of 2.6 K (1.1–3.9). We have explored the sensitivity to our assumptions in the method, including considering structural uncertainty, and in the choice of models, and this leads to 95% probability of climate sensitivity mostly below 5 and never exceeding 6 K. The approach is compared with other approaches based on OLS, a Kalman filter method and an alternative Bayesian method. An interesting implication of this work is that OLS-based emergent constraints on ECS generate tighter uncertainty estimates, in particular at the lower end, suggesting a higher bound by construction in case of weaker correlation. Although some fundamental challenges related to the use of emergent constraints remain, this paper provides a step towards a better foundation of their potential use in future probabilistic estimation of climate sensitivity.

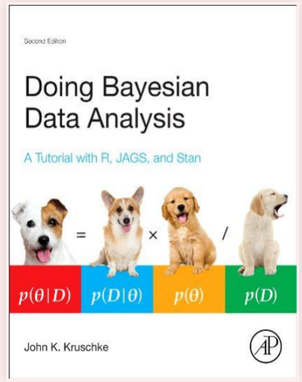
Organization of the Course

- The main class website is at https://ees5891.jgilligan.org
- Copies of the
  - syllabus,
  - reading assignments,
  - homework assignments
  - slides from class (also link from QR code on title slide)
  - instructions for installing necessary software on your computer
- Links to helpful resources.
- Slides:
  - The title slide has QR code with link to online version.
  - PDF versions are also posted to course web site (link on title slide)
  - Slides have two-dimensional navigation (in a browser, hit “?” for help)

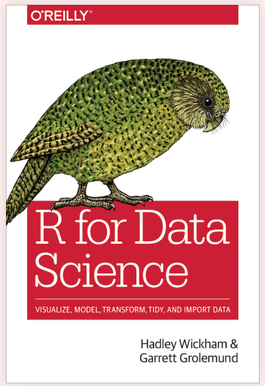
Textbooks



Richard McElreath, *Statistical Rethinking* (2nd Edition)
Required
Online version at VU libraries



John K. Kruschke, *Doing Bayesian Data Analysis* (2nd Edition)
Optional supplementary book, not required
Online version at VU libraries



Hadley Wickham, Mine Çetinkaya-Rundel & Garrett Grolemund, *R for Data Science*
Optional recommended book, not required
Free web version online at `https://r4ds.had.co.nz/`

Course Material

- Main source of material: ees5891.jgilligan.org
  - Syllabus
  - Reading assignments for the semester
    - Do the assigned reading **before** class on the day it's assigned for.
  - Homework is due at the beginning of class on the assigned due date.
- Slides from class
  - Web-based and PDF versions
  - Posted on ees5891.jgilligan.org/schedule/
  - Slides:
    - The title slide has QR code with link to online version.
    - PDF versions are also posted to course web site (link on title slide)
    - Web-based slides have two-dimensional navigation (in a browser, hit "?" for help)

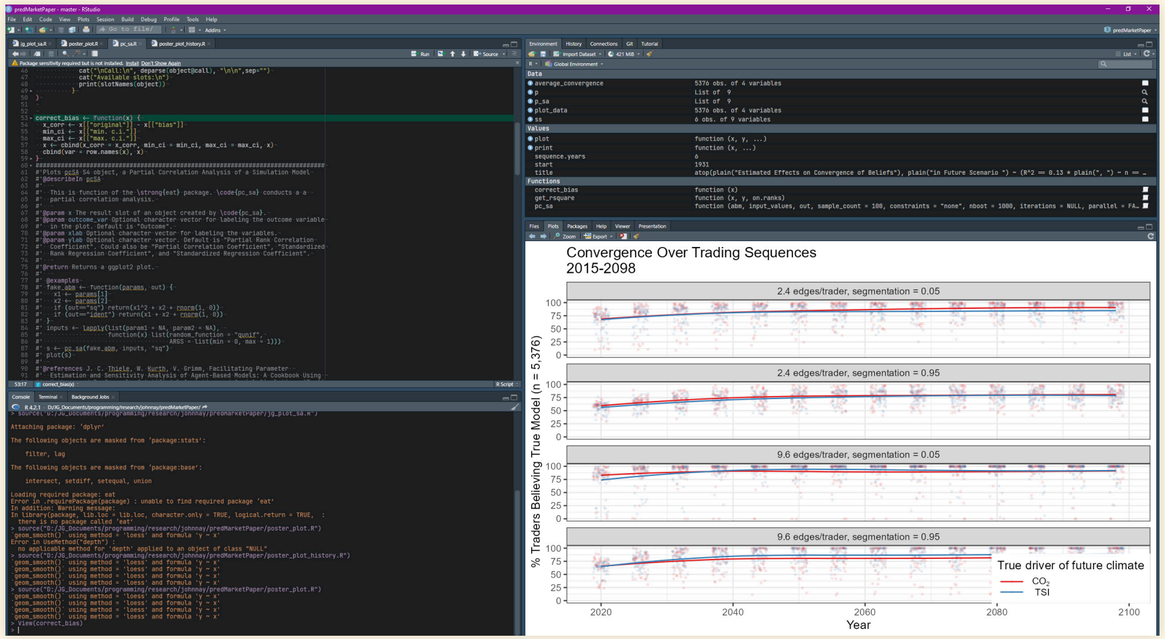
Computational Resources

R

- We will use R for all the statistical analysis in this course
  - Free and open source
  - Widely used and powerful
  - Written specifically for statistical analysis
  - Extensive library of free packages people have written to extend it.
  - You can find instructions for installing R and the other software we will use this semester in the "Tools" page of the class web site and in the homework assignment for Wednesday.
    - We will use the `rethinking` package, which in turn uses the `cmdstanr` package and the `cmdstan` tool for Bayesian Hamiltonian Monte Carlo analysis.
      - `cmdstan` can be difficult to install, but we won't use it until later in the semester.
      - You can install `cmdstanr` and `rethinking` before installing `cmdstan` itself.

RStudio

- We will use the RStudio integrated development environment for working with R.
  - RStudio combines an editor for writing code and an environment for running R scripts and also using R interactively, displaying graphics, etc.
  - Like R, RStudio is free and open-source.



Additional R Resources

- We will also install a lot of packages other people have written to add powerful Bayesian data analysis to R.
- To use these, it will be necessary to install the R development environmment. There are links to detailed instructions in the homework assignment for Tuesday and on the “Tools” page of the class web site.

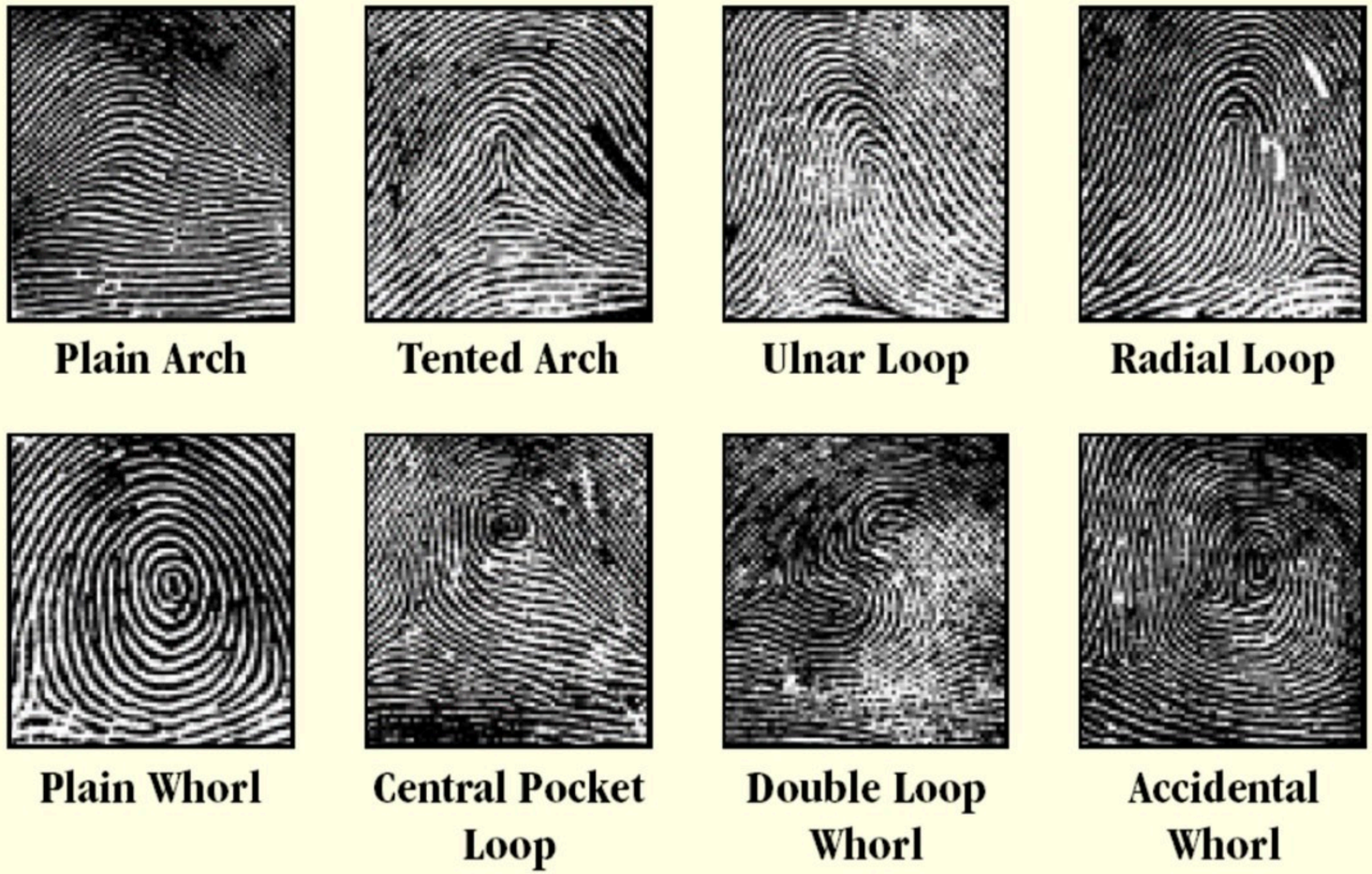
Application to Law and Justice

Weighing Evidence of Guilt

- In a city with 1 million people, someone committed a burglary.
- Police found a partial fingerprint at the crime scene
- A suspect was arrested and his fingerprints match the partial print at the scene

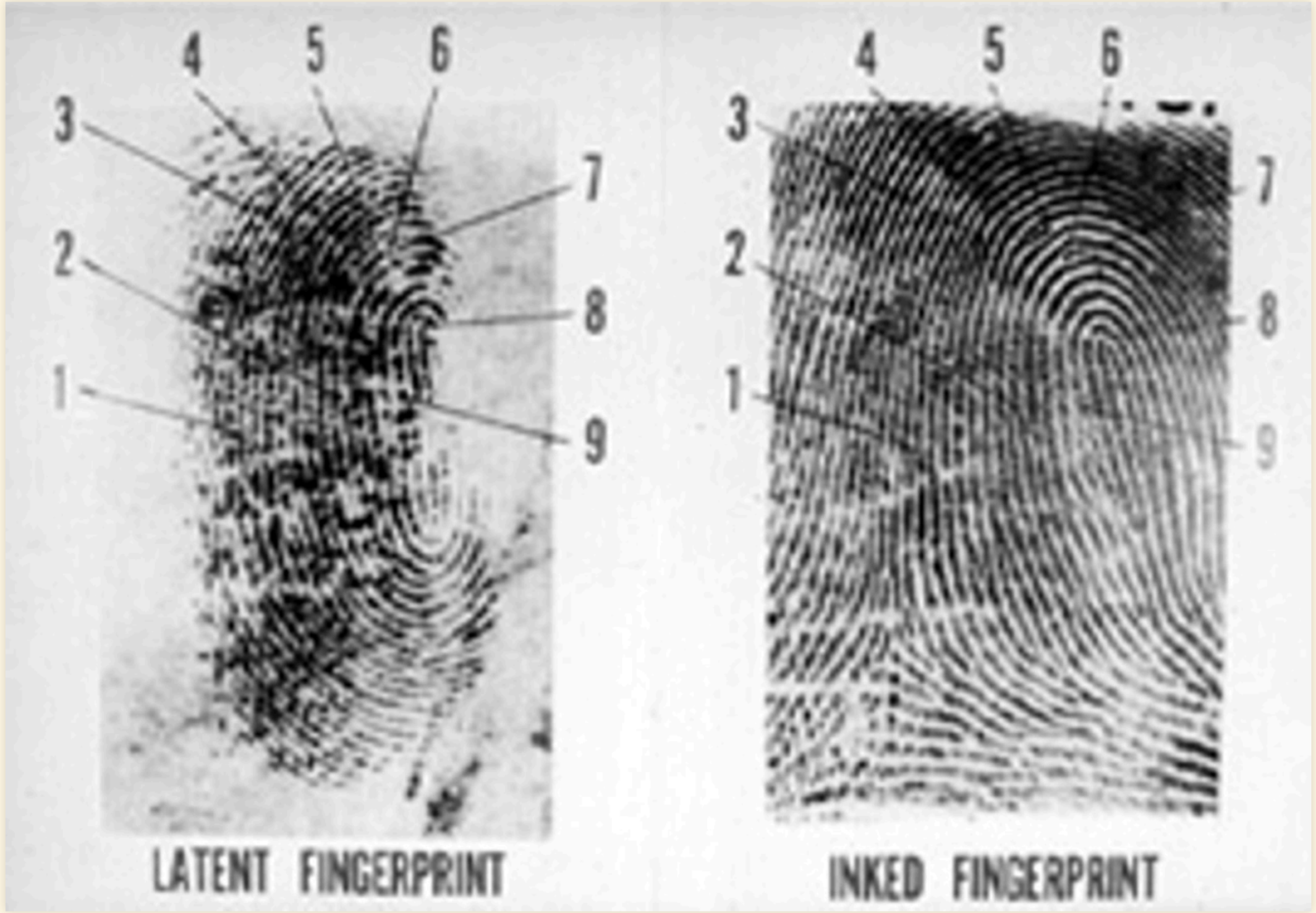
Fingerprint Analysis

- People’s fingerprints have distinctive patterns



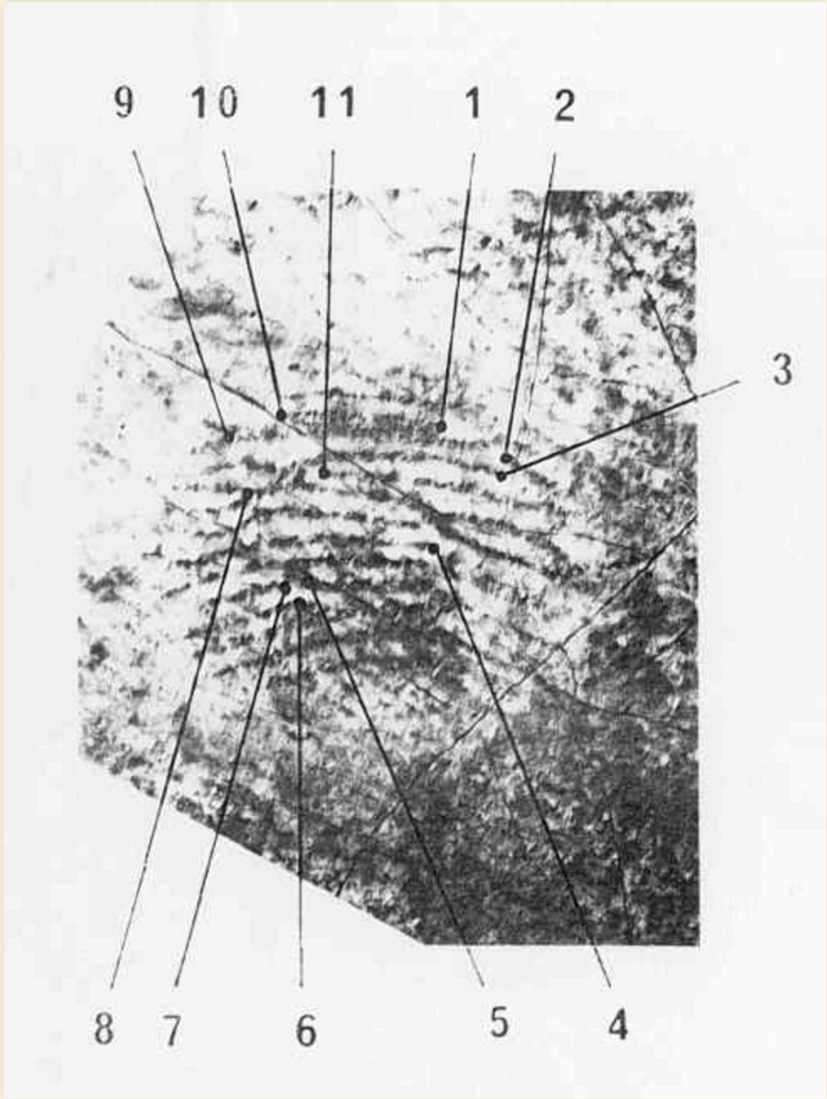
Comparing Fingerprints

Not a match



Comparing Fingerprints

Do these match?



Fingerprint Examinations

Suppose the sensitivity and specificity of fingerprint analysis are as follows:

- Sensitivity = 0.999
  - If the suspect actually left the print, there is a 99.9% probability that the fingerprint experts will declare a match.
- Specificity = 0.99999
  - If someone else left the print, there is only a 1 in 100,000 chance that the experts will declare a match (a *false positive*)

Prosecutor’s Fallacy

- Specificity = 99.999%. If someone else left the print, there is only a 1 in 100,000 chance that the experts would declare a match.
- Therefore, if experts declare a match, we can be 99.999% certain the suspect is guilty.
- This is **not true**.

Bayes's Theorem

$$P(H|D,I) = \frac{P(D|H,I) \times P(H|I)}{P(D|I)}$$

- H: Hypothesis, the suspect left the fingerprint
- h: Null hypothesis, someone else left the fingerprint
- D: New data, experts say the fingerprints match
- d: New data, experts say the fingerprints don't match
- I: Other evidence we had before the fingerprint analysis.

- $P(D|H,I)$ = sensitivity of the fingerprint analysis (99.9%)
- $P(d|h,I)$ = specificity of the fingerprint analysis (99.999%)
- $P(H|I)$ = *prior* probability that the suspect left the fingerprint
- $P(D|I)$ = *evidence*: A measure of the additional information we get from learning D.

$$\begin{aligned} P(D|I) &= P(D|H,I) \times P(H|I) + P(D|h,I) \times P(h|I) \\ &= P(D|H,I) \times P(H|I) + (1 - P(d|h,I)) \times (1 - P(H|I)) \end{aligned}$$

Prior Probability

$$P(H|D,I) = \frac{P(D|H,I) \times P(H|I)}{P(D|I)}$$

- $P(D|H,I)$ = sensitivity (99.9%)
- $P(d|h,I)$ = specificity (99.999%)
- $P(D|I) = P(D|H,I) \times P(H|I) + (1 - P(d|h,I)) \times (1 - P(H|I))$
- We need to know $P(H|I)$, the probability, before we learn the fingerprint evidence, that the suspect left the fingerprint at the scene. (this is called the *prior probability*)

Different values of the prior

$$P(H|D,I) = \frac{P(D|H,I) \times P(H|I)}{P(D|I)}$$

- $P(D|H,I)$ = sensitivity (99.9%)
- $P(d|h,I)$ = specificity (99.999%)
- Suppose the prior $P(H|I)$ is very small. The fingerprint is the only useful evidence we have.
- There are 1 million people in the city, so $P(H|I) = 1 \times 10^{=6}$

$$P(H|D,I) = \frac{0.999 \times 1 \times 10^{-6}}{0.999 \times 1 \times 10^{-6} + (1 - 0.99999) \times (1 - 1 \times 10^{-6})} = 0.091$$

- So even if the fingerprint matches, there is only a 9.1% probability that the suspect was the culprit, and a 90.9% probability that they were mistakenly accused.

Different values of the prior

$$P(H|D,I) = \frac{P(D|H,I) \times P(H|I)}{P(D|I)}$$

- $P(D|H,I)$ = sensitivity (99.9%)
- $P(d|h,I)$ = specificity (99.999%)
- Suppose the suspect was seen near the scene of the crime. This narrows it down so before we get the fingerprint evidence, we estimate that $P(H|I)$ is 1 in 1000 (0.001).

$$P(H|D,I) = \frac{0.999 \times 0.001}{0.999 \times 0.001 + (1 - 0.99999) \times (1 - 0.001)} = 0.99$$

- So if the fingerprint matches, there is now a 99% probability that the suspect is the true culprit, and only a 1% probability that they were mistakenly accused.

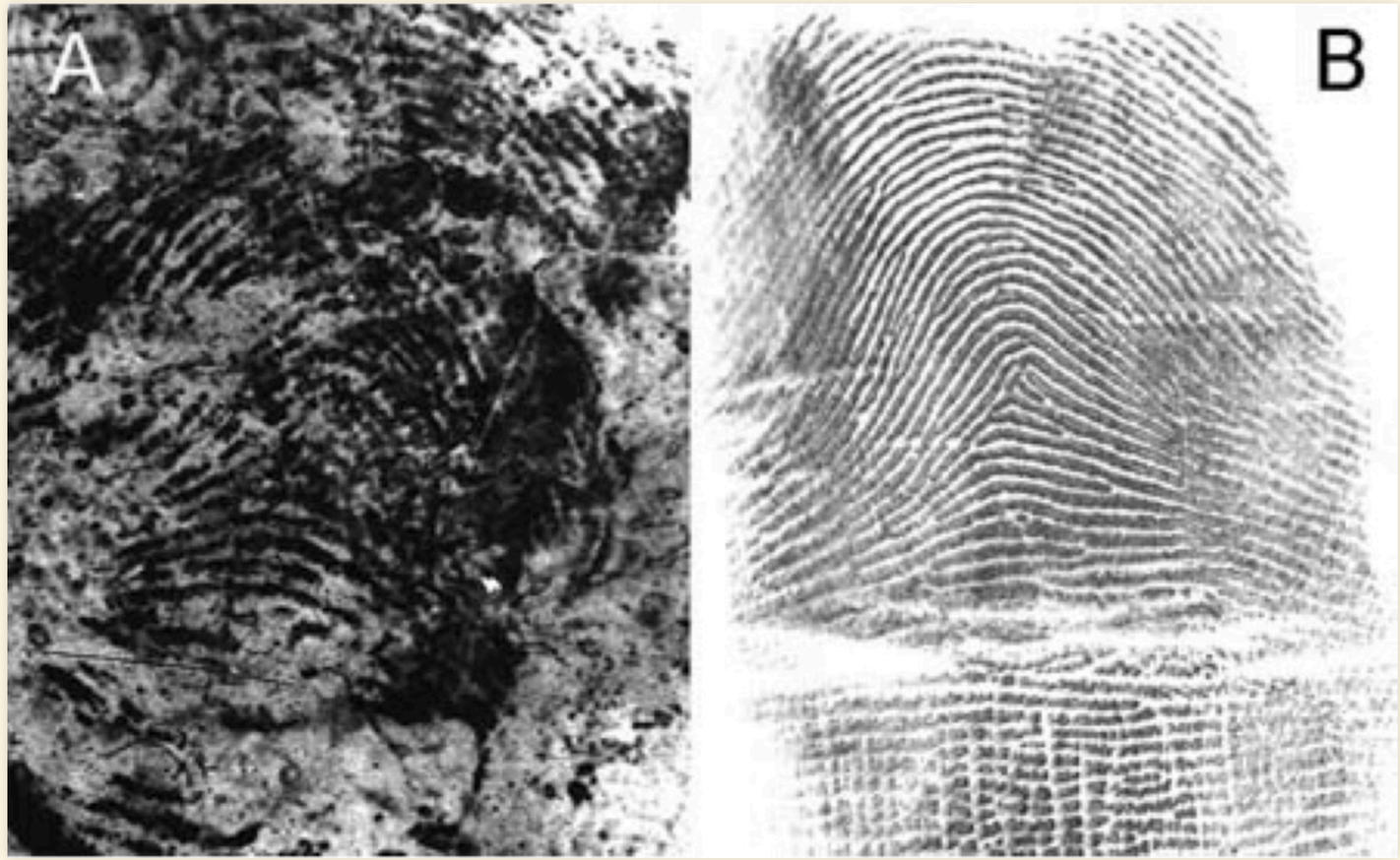
Real-World Example

2004 Madrid Train Bombing

- March 11, 2004: Terrorists set off ten bombs in commuter trains headed for Madrid
- 193 people were killed
  - The deadliest terrorist incident in Spanish history

Fingerprints

- Investigators recovered a plastic bag containing detonating devices.
- Latent fingerprints from the bag were circulated worldwide
- FBI investigators found 20 possible matches in FBI databases
  - FBI announced a definitive “100%” match of the latent fingerprint



Suspect

- May 6, 2004:
  - The FBI arrested Brandon Mayfield, a Portland OR lawyer and U.S. military veteran who had converted to Islam in 1986.
  - FBI experts claimed Mayfield’s fingerprint was “a 100% match” to the Madrid fingerprint, with no possibility of error
  - Spanish authorities disagreed and doubted that Mayfield had anything to do with the bombing

Exoneration

- May 19: Spanish authorities matched the fingerprint to Daoud Ouhnane, a known Algerian terrorist.
  - Ouhnane subsequently joined insurgent forces in Iraq and was killed in combat in 2006.
- May 20: Mayfield was released from custody.
- An FBI review of the investigation concluded that prejudice against Mayfield's religion led investigators to ignore evidence of his innocence
- In 2006, the FBI issued an apology and paid Mayfield $2 million to settle a wrongful arrest lawsuit.

- Bad statistics associated with forensic scientific examinations are a leading factor in false-convictions of people later proved to be innocent.