Mathematical Statistics II

An Introduction

Jesse Wheeler

Course Introduction

Course Overview

- The larger focus of last semester (Math 4450) was probability.
- Though we continue where we left off, this semester (Math 4451) will have a much stronger focus on statistics. A complete description of planned course topics can be found at the course website: https://jeswheel.github.io/4451_s26/#planned-topics-spring-2026.

Course Overview II

- Both probability and statistics are, fundamentally, the study of with randomness... so, what's the difference?

Statistical science was the peculiar aspect of human progress which gave to the twentieth century its special character... it is to the statistician that the present age turns for what is most essential in all its more important activities. – Fisher (1954)

Probability and Statistics

What is “Statistics”?

- First, what is statistics?

What is "Statistics"? II

"The science of collecting, displaying, and analysing data."
– Oxford Dictionary (2008)

"The discipline that concerns the collection, organization, analysis, interpretation, and presentation of data."
– Wikipedia contributors (2025)

Something like: "The study of extracting useful information from data in a rigorous way." – Me (it's hard to define an entire discipline).

Probability vs Statistics

- Any of the above definitions (accurately) suggests that probability is a key part of statistics. So where do we draw the line? Does it matter?
- Pawitan (2001) dichotomizes the difference in terms of *deductive* vs *inductive* reasoning.
- Roughly speaking, *deductive* arguments moves from general principles (assumptions) to make specific conclusions. In *inductive* reasoning, we use specific observations (data) to make broader generalizations.

Probability vs Statistics II

Traffic Accidents

Suppose we are interested in the random quantity X_i, the number of accidents during week i at a particular intersection. From last semester, a common model for this situation is a Poisson-process.

- *Probability (deductive)*: If X_i follows a Poisson(λ) distribution (general principle), then what is the expected number of accidents per week (specific conclusion)? What is the probability that we observe more than 10 accidents?

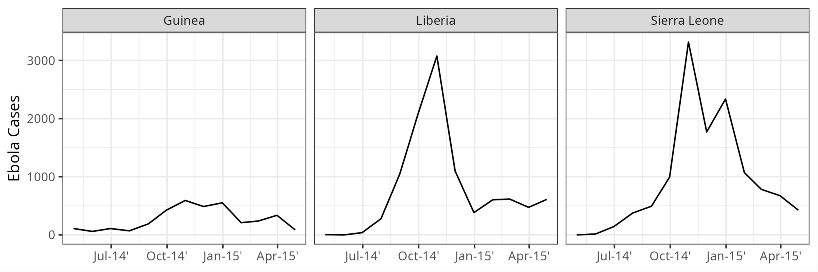
Probability vs Statistics III

- *Statistics (inductive)*: Suppose we count the number of accidents over a 6 week period, observing: $3, 4, 2, 7, 3, 3$ accidents (specific observations). What value λ might describe the Poisson-process that generated the data (broader generalization)? Is the Poisson assumption reasonable given the data?

- From the example above, we can see both ideas used in conjunction for making informed decisions.
- Many statistics problems rely on deductive reasoning in probability, geometry, topology, analysis, etc. to build theory for ways of performing inductive reasoning with specific observations (data).

Probability vs Statistics IV

- Another example related to my own research in population modeling...



- (Statistics) Given the data (specific example), what can we learn about the dynamic system / generative process (generalization)?

ODE, SDE, POMP -

Probability vs Statistics V

- (Probability) Under our assumed process / model (general principle), what is our prediction for the Ebola burden over the next year (specific conclusion)?

Statistics of Math 4451

Statistics and Math 4451

- Pawitan (2001) further categorizes statistics in terms of five key 'statistical activities' in the preface of his book:

  pre 1950s →

  - *Planning*: making decisions about the study design or sampling protocol, what measurements to take, stratification, sample size, etc.
  - *Describing*: summarizing the bulk of data in few quantities, finding or revealing meaningful patterns or trends, etc.
  - *Modeling*: developing mathematical models with few parameters to represent the patterns, or to explain the variability in terms of relationship between variables.
  - *Inference*: assessing whether we are seeing a real or spurious pattern or relationship, which typically involves an evaluation of the uncertainty in the parameter estimates.

  - *Model Checking*: assessing whether the model is sensible for the data.

- A lot of early statistics were focused on the first two activities: *planning* and *describing*. We will not spend much time this semester discussing methods related to these two activities.

Statistics and Uncertainty

- Regardless of the type of "statistical activity" we are doing, a recurring theme in statistics in uncertainty.
- Loosely speaking, we can characterize two distinct forms of uncertainty:
    - *Stocahstic uncertainty*: uncertainty due to inherent randomness in data used to make inference, resulting in different models and estimates. For example, what if we picked different weeks to observe traffic patterns at the intersection? We would likely get different data, resulting in different parameter estimates.

Statistics and Uncertainty II

- *Inductive uncertainty*: uncertainty due to the inductive nature of our estimates. This is a result of having incomplete information (e.g., due traffic accidents at the intersection truly follow a Poisson-process?) We generally can't control this type of uncertainty, or even quantify it.

- The first type can be thought of as the uncertainty present, conditioned on a specific model. The second is related to uncertainty involved in the model selection itself.

Traffic Deaths

Traffic Deaths and Cell-Phone usage

Suppose instead of measuring total accidents at an intersection, what if we model the total number of deaths? We might be interested in estimating how cell-phone usage might be correlated with number of traffic deaths.

- Ideally, we would like to do a randomized experiment, placing drivers into a control (no cell-phone), or treatment group (cell-phone). *Recall from Math 3350, this is the best way to control for confounding.*

Traffic Deaths II

- This isn't possible, so instead we rely on what is called a natural experiment. The "control" group will be the population of drivers the year before the invention of cell-phones, and the "treatment" group is the population of drivers the year cell-phones were invented.
- Thought experiment: What if the number of deaths increase from 170 in one year, to 190 the next? Is this enough to claim cell-phones increase accidents? What about 170 to 300? 170 to 174? What's the cutoff?

Traffic Deaths III

- Now suppose the deaths increased from 170 to 300, but in the second year, a few major accident involved many cars, in which over 50 people died. Alternatively, what if the second year had a much longer winter, or there were more teen drivers? What other factors might change how we think about this?

- Smoking: Leo Breiman

- Alcohol Consumption

Wine: Causal Inference.

Bayesian vs Frequentist Statistics

Frequentist vs Bayesian

- When we finally settle down on a model, we can deal with the inherent stochastic uncertainty in the data in a rigorous way.
- Typically, this is done using probability.
- Probability is a surprisingly abstract topic, however, and how to connect real-world outcomes to probability is non-trivial.
- An at times frustrating (and possibly unique) feature of Statistics as a discipline is that experts cannot agree on the fundamental nature of their subject.
- There are two groups of thought: Bayesian and Frequentist perspective. We will discuss approaches from both in this class.

Frequentist vs Bayesian II

In the extremes, interpretation of probability falls into two main categories:

- Bayesian perspective: probabilities correspond to a (subjective) degree of belief about an event.
- Frequentist perspective: probabilities are interpreted only in long-run frequency of events.

- Note in the Bayesian perspective admits both perspectives, and many modern Bayesian approaches often consider the frequentist properties of their methods.
- In *my experience*, most statisticians actually fall somewhere in the middle.

Frequentist vs Bayesian III

Tossing a coin

Consider tossing a fair coin. We have some sense of uncertainty about the outcome of this experiment, and say that the probability of heads is 0.5.

- What does this mean in the purely frequentist sense? What about the unvertainty related to the *next specific* coin toss?
- What does this mean in the purely Bayesian paradigm?

Frequentist vs Bayesian IV

- We will cover both approaches, and discuss strengths and weaknesses of each when relevant. However, the focus will be on frequentist statistics, as the vast majority of statistical methods in use in the 20th and 21st centuries are based on frequentism.
- There's also a "third way", sometimes called Fisherian, or likelihoodist. To some degree this is a compromise between the two more popular approaches, as it rejects the use of prior distributions, but still interprets likelihood as a way to measure belief about a parameter estimate (Pawitan, 2001).

Frequentist vs Bayesian V

Probability of events

Try to use both Bayesian and frequentist interpretations of probability to describe what is intended by the following statements.

- My weather app states that there is a 40% of rain tomorrow.
- In a sporting event, ESPN says that the probability that team A will win is 57%.
- A particular blood test given by a doctor is said to have a 5% probability of a false-positive result, and you just received a positive result.

Introduction to Estimation

- The primary focus of this class will be parameter estimation. That is, we collect data, pick a model to describe the data generating process, and use the data to "fit" the model to data.
- Models can be complex: machine learning frameworks like random forest, neural networks, gradient boosting machines, etc. These are very good at prediction, but are often referred to as a *black-box*.
- In this class, we focus more on general principles of data analysis, which is most easily demonstrated using simple models.

Flipping coins

A friend gives you a coin from another country. You want to estimate the probability of heads for this particular coin, flip the coin N times, observe $0 \leq n \leq N$ number of heads.

Frequentist Solution:

$\text{Bin}(N, p)$ $\quad X_i \sim \text{Ber}(p)$

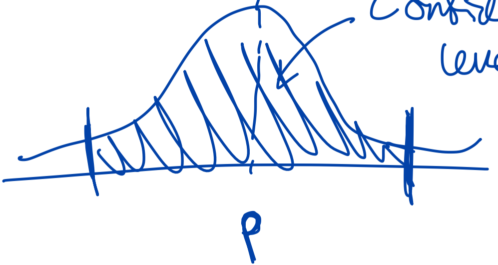
- $\hat{p} = \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$ — point estimate

- CLT: $\hat{p} \approx N\left(p, \frac{p(1-p)}{N}\right)$

$$\hat{p} \approx N\left(p, \; p(1-p)/N\right)$$

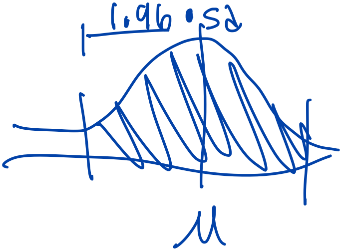
$$\hat{p} \xrightarrow{a.s.} p$$

$$\hat{p} \approx N\left(p, \; \frac{\hat{p}(1-\hat{p})}{N}\right)$$



Confidence level.

95%



$$\hat{p} \approx N\left(p, \frac{\hat{p}(1-\hat{p})}{N}\right)$$

95%

Confidence

point estimate

$$\hat{p} = \frac{n}{N}$$

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

I_p

Following the same procedure many times, the long-run frequency of times that $p \in I_p$ is $\approx 95\%$.

Example suppose $N = 10$

$n = 8$ heads

$$\hat{p} = \frac{n}{N} = \frac{8}{10} = \underline{80\%}$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} = \sqrt{\frac{(0.8)(0.2)}{10}} = 0.126$$

$$0.8 \pm 1.96(0.126)$$

$$= (0.55, 1.05)$$

↑

$N(\mu, \sigma^2)$, $\mu \geq 0$

$(-1.2, -0.3)$ $\mu > 0$

95%

Bayesian Solution:

Bayes Theorem:

$\rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$

- B = observing n Heads, N trials
- A = P = p

p = ~~random~~ probability variable, n is random variable

$$f_{P|X=n}(p|n) = \frac{f_{X|P}(n|p)\, f_P(p)}{\int f_{X|P}(n|p)\, f_P(p)\, dp}$$

→ Assume we know nothing!

→ $f_P(p) = 1 \quad [0 \le p \le 1]$

$$f_{X|P}(x|p) = \binom{N}{x} p^n (1-p)^{N-x}$$

$$f_{P|X=n}(p;n) = \frac{\binom{N}{n} p^n (1-p)^{N-n}}{\int_0^1 \binom{N}{n} \gamma^n (1-\gamma)^{N-n} d\gamma} \cdot \mathbb{I}[0 \le p \le 1]$$

$$= \frac{p^n (1-p)^{N-n}}{\int_0^1 \gamma^n (1-\gamma)^{N-n} d\gamma} \mathbb{I}[0 \le p \le 1]$$

$$B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)} = \int_0^1 p^{z_1-1}(1-p)^{z_2-1} dp$$

$$f_{p|X=n}(p|n) = \frac{\Gamma(N+2)}{\Gamma(n+1)\,\Gamma(N-n+1)} p^n (1-p)^{N-n} \mathbb{1}[0 \le p \le 1]$$

$$p|X=n \sim \text{Beta}(n+1, N-n+1)$$

$n = 8$

$N = 10$

$$E[p|X=n] = \frac{n+1}{N+2} = \frac{9}{12} = 0.75$$

$$\hat{p} = \frac{n}{N} = 0.8$$

0.025 and 0.975 percentiles

of Beta(n+1, N-n+1)

R code

- The R code for finding a credible interval:

```
n <- 8
N <- 10
lower_bound <- qbeta(0.025, n+1, N-n+1)  # lower bound
upper_bound <- qbeta(0.975, n+1, N-n+1)  # upper bound
```

r norm   q Pois

```
95% Credible Interval for p: (0.48, 0.94)
```

r <dist> → random sample from dist.

d <dist> → density evaluate $f(x;\theta)$

q <dist> → quantile gives quantiles

p <dist> → cdf $F(x;\theta)$

Final Thoughts

- The differences between Bayesian vs Frequentist thinking matters, and some get very passionate about the debate.
- In my opinion, most people fall somewhere in the middle.
- Often practitioners choose a Bayes vs Frequentist methodologies not because of their personal interpretation of probability, but for convenience, existing standards, or it is the only way to solve a given problem.
- "All models are wrong but some are useful" – Box (1979), a famous 20th century statistician.

Final Thoughts II

- One example is the following Bayesian approach to image restoration for images with static noise introduced (Chapter 15.6 Keener, 2010), adapted from the seminal paper by Geman and Geman (1984).

$$X_{i,j} \sim N(\theta_{i,j}, \sigma^2)$$

$$\theta \sim N_{n \times m}(\mu_\theta, \Sigma_\theta)$$

$$\theta_{i,j} \mid X_* \sim N$$

Final Thoughts III

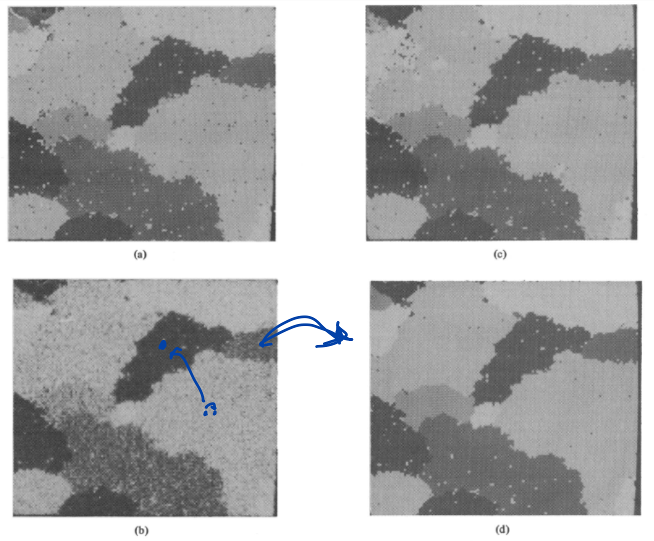


Figure 1: Bayesian image restoration, credit Geman and Geman (1984)

References and Acknowledgements

Box GE (1979). “Robustness in the Strategy of Scientific Model Building.” In *Robustness in Statistics*, pp. 201–236. Elsevier.

Fisher R (1954). “The expansion of statistics.” *American Scientist*, **42**(2), 275–293.

Geman S, Geman D (1984). “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**(6), 721–741. doi: 10.1109/TPAMI.1984.4767596.

Keener RW (2010). *Theoretical statistics: Topics for a core course*. Springer Science & Business Media.

References and Acknowledgements II

Oxford Dictionary (2008). "Statistics." doi: 10.1093/acref/9780199541454.013.1566. URL https://www.oxfordreference.com/view/10.1093/acref/9780199541454.001.0001/acref-9780199541454-e-1566.

Pawitan Y (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Wikipedia contributors (2025). "Statistics — Wikipedia, The Free Encyclopedia." [Online; accessed 9-January-2026], URL https://en.wikipedia.org/w/index.php?title=Statistics&oldid=1328458961.

References and Acknowledgements III

- Compiled on January 12, 2026 using R version 4.5.2.
-
- We acknowledge students and instructors for previous versions of this course / slides.