Mathematical Statistics II

Chapter 1: Probability

Jesse Wheeler

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

Course Overview 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?

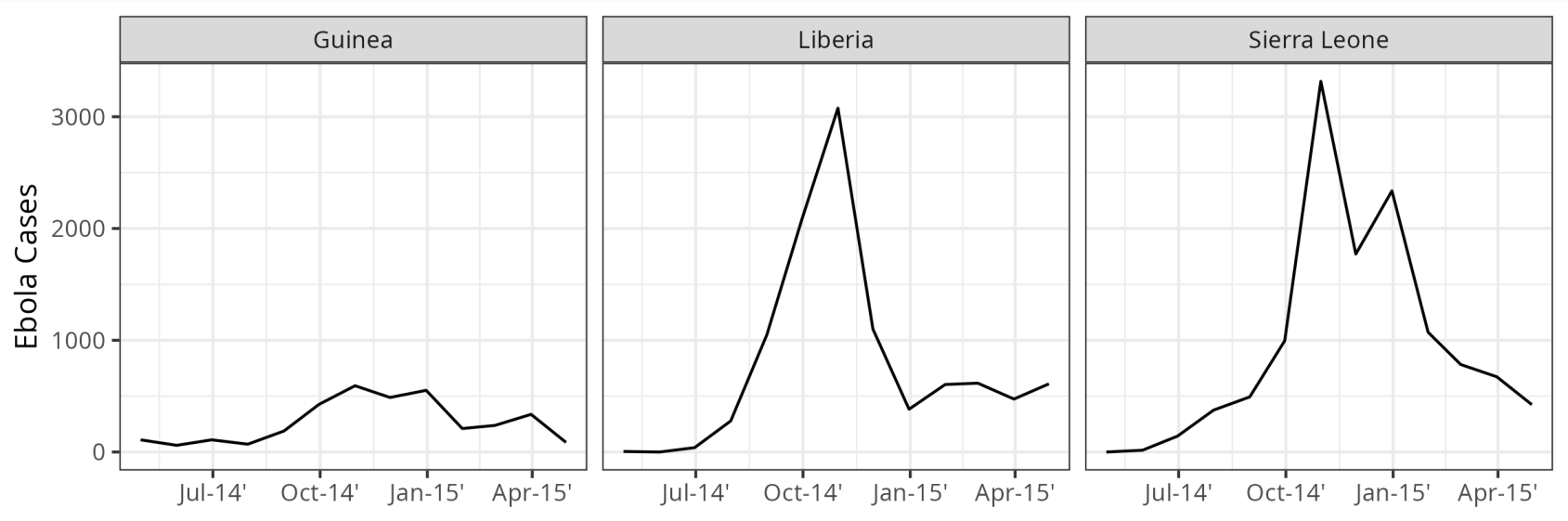
Probability vs Statistics III

- *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)?

Probability vs Statistics V

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

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

References and Acknowledgements

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 II

- Compiled on January 9, 2026 using R version 4.5.2.
- Licensed under the Creative Commons Attribution-NonCommercial license. Please share and remix non-commercially, mentioning its origin.
- We acknowledge students and instructors for previous versions of this course / slides.