Mathematical Statistics II
Chapter 1: Probability

Jesse Wheeler

Contents

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? **Depends on who you ask!**

    "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).

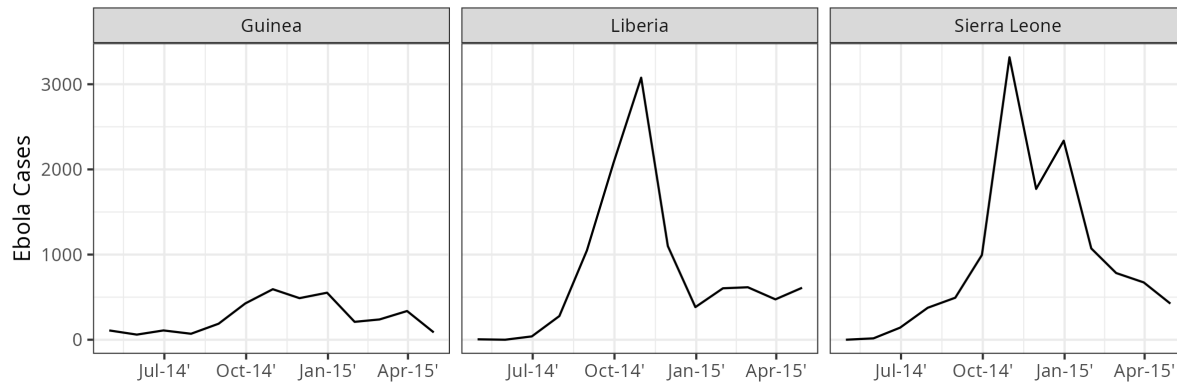**Probability vs Statistics**

- 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.

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?

- *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).

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

Statistics and Math 4451

- 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.

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.
  - – *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.*

- 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?

  - – The question above is related to changes in observed data, not the model. That is, fixed on a model, we can make statements about what is considered a "large enough" change in order to consider cell-phones having an impact.

- 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?

  - – This time, uncertainty arises because of model choice. We can increase model detail by splitting by age, month, etc., but the data then fall into smaller and smaller groups, increasing the stochastic uncertainty. We have to stop adding detail at some point, but where?

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.

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.

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.

- In the extreme frequentist paradigm, this is interpreted to mean that the long-run proportion of heads (many coin flips) is 0.5. However, interpreting the uncertainty of the next *specific* toss is difficult (something like: the next toss belongs to an infinite sequence of tosses of the same coin, which have a limiting frequency of 0.5.)

- In the extreme Bayesian paradigm, a 0.5 probability corresponds to a *belief* about the behavior of the coin; this belief can be entirely subjective. Saying a probability of 0.5 means that even in the next specific coin toss, our sense of uncertainty is 0.5.

- *Important*! Both approaches have their strengths and weaknesses, and some of the smartest scientists / mathematicians in the world can be found in either extreme. Though most statisticians will claim to think one-way versus the other, most fall somewhere in the middle (frequentists' will often be founding making subjective probability statements, whereas Bayesians' often admit to caring about frequentist properties of their methods).

- 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).

Acknowledgments

References

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. 1

Pawitan Y (2001). *In all likelihood: statistical modelling and inference using likelihood.* Oxford University Press. 2, 3, 6

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. 1