STA 111: Probability & Statistical Inference

Lecture One – Introduction to Probability

Instructor: Olanrewaju Michael Akande

Department of Statistical Science, Duke University

Outline

- Welcome!
- Interpretations of Probability
- Set theory
- Definition of probability
- Recap

Welcome!

- Syllabus
- Course Website
- Questions?

Learning Objectives

By the end of class, you would

- Be familiar with the common definitions and interpretations of probability,
- Be able to define events based on simple experiments,
- Know how to verify axioms of set theory, and
- Be able to calculate probabilities of events based on simple experiments.

Probability vs. Statistical Inference

– We will spend the first half of this course talking about probability.

– Once we have a good understanding of probability, we will switch to statistical inference.

– Loosely speaking, probability and statistical inference can be thought of as inverses to each other.

– As a probabilist, you wish to make statements about data/samples/subpopulations, given what you know about the overall population.

– As a statistician, you do the opposite. Given the data you have seen, what can you say about the overall population?

Interpretations of Probability

– Probability quantifies the uncertainty about the occurrence of an event.

– If I toss a fair coin once, how likely am I to see one head? What if I toss it twice? What if I use a biased coin? We will revisit these questions once we have a better understanding of how to assign probabilities.

– In the simplest terms, probability lets us assign numbers (which need to satisfy some properties we haven't talked about) to events such that we can make statements about how likely the events are.

– How should we develop a formal definition of probability? First, lets review three frequently used interpretations of probability:

- The frequency interpretation
- The classical interpretation
- The subjective interpretation

The Frequency Interpretation of Probability

– Suppose you are interested in an event A. Then the probability of A can be thought of as the proportion of times A occurs in an infinite sequence (or very long run) of separate, independent and identical tries. That is,

$$P(A) = \lim_{n\to\infty} \frac{\text{\# times A happens}}{n}$$

– Thus for a fair coin for example, if A is the probability of heads/tails and we toss the coin n times with n large enough (say, $\geq$ 10,000), we should expect

$$P(A) = \lim_{n\to\infty} \frac{\text{\# heads}}{n} \approx 0.5$$

– How feasible is this?

The Classical Interpretation of Probability

– The classical interpretation on the other hand is based on the concept of equally likely outcomes. If the outcomes of an experiment must be one of k different outcomes and all k outcomes are equally likely to occur, then the probability of each outcome is simply $1/k$.

– The textbook discusses the two basic difficulties of developing a formal definition of probability from this interpretation:

- First, the concept of equally likely outcomes is essentially based on the concept of probability, which is exactly what we are trying to define. If two outcomes are equally likely, they must have the same probability.
- Second, there is no provision for defining probability for outcomes that are not equally likely.

The Subjective Interpretation of Probability

– The subjective interpretation relies on the notion that the probability of any event should reflect a person's belief about the likelihood of occurrence.

– Let X denote the average age of freshmen at Duke. Pick a number x_1 such that you consider these two outcomes to be equally likely: A $= X \leq x_1$ and B $= X > x_1$. Now pick numbers x_2 and x_3 such that the following outcome are equally likely: A $= X \leq x_2$, B $= x_2 < X < x_3$ and C $= X \geq x_3$. Repeating this experiment for all (or most of) the possible values X can take, defines the probability of X through your beliefs.

– Discuss plausible values of x_2 and x_3 with the person seated closest to you. Can both of you agree on the same values? (1 min).

– Note: we will not spend any more time discussing the pros and cons of each interpretation, but rather we will see as we move on, how each one can be useful depending on the experiment of interest. Interestingly, the theory of probability doesn't actually depend on any particular interpretation.

Set Theory

– An experiment is any process such that the collection of every possible outcome of the process can be described and "perhaps" even enumerated. An event is any subset of the sample space – the collection of all possible outcomes of an experiment.

– Now for some relations:

- Ω denotes the sample space.
- $x \in A$ denotes that an element x is a member of A,
- $A \subset B$ denotes that a set A is a subset of set B, so that A and B are equal if and only if $A \subset B$ and $B \subset A$.
- If $A \subset B$ and $B \subset C$, then $A \subset C$.
- $\emptyset$ denotes a set that contains no element; the empty set. Thus, $\emptyset \subset A$ for any A.

Set Operations

Let $A, B \subset \Omega$

- Complement: $x \in A \Rightarrow x \notin A^c$.
- Union: $x \in A \cup B \Rightarrow x \in A$ or $x \in B$ or both
- Intersection: $x \in A \cap B \Rightarrow x \in A$ and $x \in B$. Thus, $(A \cap B) \subset (A \cup B)$. A and B are disjoint if $(A \cap B) = \emptyset$.
- Set Difference: $x \in A \backslash B \Rightarrow x \in A$ and $x \notin B$. That is, $A \backslash B = A \cap B^c$.
- Symmetric difference: $x \in A \triangle B \Rightarrow x \in (A \backslash B) \cup (B \backslash A)$.
- De Morgan's Laws: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.
- Cardinality: $|A|$ is the number of elements in A.

Examples

Example 1: Toss a fair coin twice and let A be the event that you observe only one head and C be the event that you observe at least one head. Then,

- The sample space $\Omega = \{HH, HT, TH, TT\}$
- $A = \{HT, TH\}$
- $|A| = 2$
- $A^c = \{HH, TT\}$
- $\{HT\}$ and $\{TH\}$ are subsets of A.
- $A \cup A^c = \{HH, HT, TH, TT\} = \Omega$
- $C = \{HH, HT, TH\}$. Then $A \subset C$, $A \cup C = C$, $A \cap C = A$, $A \backslash C = \emptyset$, $C \backslash A = \{HH\} = A \triangle C$.

Definition of Probability

Let $\mathcal{F}$ be a well-defined collection of subsets of Ω. A probability measure (or simply a probability) P is a function $P : \mathcal{F} \to [0, 1]$ such that

- For every event $A \in \mathcal{F}$, $P(A) \geq 0$.
- $P(\Omega) = 1$
- For any sequence of disjoint events $A_1, A_2, \ldots \in \mathcal{F}$,

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Some Properties of Probability

- $P(\emptyset) = 0$
- $P(A) = 1 - P(A^c)$
- If $A \subset B$, then $P(A) \leq P(B)$
- For any event A, $0 \leq P(A) \leq 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- For any sequence of events,

$$P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i) \quad \text{and} \quad P\left(\bigcap_i A_i\right) \geq 1 - \sum_i P(A_i^c)$$

Examples

Example 1 again: Since this is a fair coin, all outcomes are equally likely.

- $P(\Omega) = P\{HH\} + P\{HT\} + P\{TH\} + P\{TT\} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$
- $P(A) = P\{HT\} + P\{TH\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
- $P(A^c) = 1 - P(A) = \frac{1}{2} = P\{HH\} + P\{TT\}$
- $P(C) = \frac{3}{4}$
- $P(A \cup C) = \frac{3}{4}$
- $P(A \cap C) = \frac{1}{2}$
- $P(A \backslash C) = 0$
- $P(C \backslash A) = P\{HH\} = \frac{1}{4} = P(A \triangle C)$

Examples

Example 1 yet again: Suppose the coin is biased, with $P(H) = \frac{3}{5}$, can you recalculate the same probabilities?

- $P\{HH\} = \frac{9}{25}$ $\quad P\{HT\} =$ $\quad P\{TH\} =$ $\quad P\{TT\} = \frac{4}{25}$
- $P(A) =$
- $P(A^c) =$
- $P(C) =$
- $P(A \cup C) =$
- $P(A \cap C) =$
- $P(A \backslash C) =$
- $P(C \backslash A) =$
- $P(A \triangle C) =$

In-class Exercise

To be done in-class with one teammate: Suppose you roll a fair die twice.

- What is $|\Omega|$?
- What is the probability of observing an even number for each roll?
- What is the probability of observing an odd number for each roll?
- What is the probability of observing the same number for both rolls?
- What is the probability of observing different numbers for both rolls?
- What is the probability of observing numbers whose sum is at most 7?
- What is the probability of observing numbers whose sum is at least 4?

Recap

Today, we talked about

- Common interpretations of probability.
- Set theory and basic set operations.
- How to calculate probability in simple experiments.