Mathematical Statistics I

Chapter 1: Probability

Jesse Wheeler

Contents

1 Course Overview

Course Overview

- This course is the first part of a two semester introductory course on Mathematical Statistics.
- Our goal is to cover Chapters 1-10 of "Mathematical Statistics and Data Analysis", by John A. Rice.
- Topics include: Probability, Random Variables, Discrete and Continuous distributions, Order Statistics, Limit Theorems, Point and Interval Estimation, Uniformly most powerful tests, likelihood ratio tests, chi-squre and F tests, and nonparameteric tests.
- Roughly speaking, 4450 and 4451 can be broken into two parts:
  - Math 4450: Probability (mathematics of randomness)
  - Math 4451: Statistics (procedures for analyzing data)

1.1 Logistics

Course Logistics

- About Me (TODO)

- Course Website: https://jeswheel.github.io/4450_f25/.
- Canvas: Canvas will be used to submit assignments, view grades, and for course announcments.
- Course Syllabus

1.2 Chapter I Overview

Probability: Chapter I Overview

- Probability has been around for a long time.
- Probability theory originated in the study of games of chance (i.e., dice, cards, etc.). These provide some nice introductory examples.
- More modern examples of probability in practice include:
  - Modeling mutations in genetics, playing a central role in bioinformatics.
  - Designing and analyzing computer operating systems.
  - Modeling atmospheric turbulence.
  - Probability theory is a cornerstone of the theory of finance, machine learning, and artificial intelligence.
  - Much more...

This semester will focus on the theory of probability as a mathematical model for chance phenomena. This will be essential for building statistical theory in 4451.

2 Sample Spaces

2.1 Experiments

Sample Spaces
Probability theory is concerned with situations in which the outcomes occur randomly. We call these situations *experiments*. The set of all possible outcomes is called the *sample space*.

Flipping a coin
Flipping a coin is an *experiment*, with possible outcomes $\{H, T\}$, which defines the *sample space* of the experiment.

An arbitrary sample space is typically denoted Ω, and an element of Ω is denoted ω.

- In the example above, $\Omega = \{H, T\}$.

Sample Space Examples

Example: Stoplights
Driving to work, a commuter passes through a sequence of three intersections with traffic lights. At each light, they either stop (s), or continues (c). The sample space Ω is the set of all possible outcomes:

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

where csc denotes the outcome that the commuter continues at the first light, stops at the second, and continues through the third.

Example: Printing

The number of jobs in a print queue of a printing machine may be modeled as random. Here the sample space is all non-negative integers:

$$\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}$$

Example: Earthquakes
We may want to model the *time* between successive earthquakes in a particular region. In this case, the experiment is the length of time between earthquakes, and our sample space is (uncountably) infinite:

$$\Omega = \{t \in \mathbb{R} | t \geq 0\}.$$

2.2 Events

Events
As we saw, sample spaces can be comprised of many different possible outcomes. In probability, we are often interested in *subsets* of specific outcomes that we call *events*. For example, let A be the event that the commuter stops at the first of three lights:

$$A = \{sss, ssc, scc, scs\}.$$

Note that *events* are sets of outcomes; any algebra you know about sets can be applied to events. That is, suppose that $B \subset \Omega$, where Ω is the sample space in the three stoplight example. We can then consider some event C that is the *union* or *intersection* of events A and B. E.g., $C = B \cup A$.

Common set operations
For the below definitions, we assume that Ω is the sample space, and $A, B, C \subset \Omega$ are events.

Intersection
The event (set) $A \cap B$ is the set of all outcomes ω that are in both events A and B. That is, $\omega \in A \cap B$ if and only if $\omega \in A$ and $\omega \in B$.

Union
The event (set) $A \cup B$ is the set of all outcomes ω that are in A or B; That is, $\omega \in A \cup B$ if and only if $\omega \in A$ or $\omega \in B$. Note that this definition does NOT use exclusive-or. In other words, $\omega \in A$ and $\omega \in B$ still implies that $\omega \in A \cup B$.

Compliment
The event (set) A^c (sometimes written A') is the set of all outcomes ω that are NOT in A; That is, $\omega \in A^c$ if and only if $\omega \in \Omega$, and $\omega \notin A$.

Empty Set
The *empty set* ($\emptyset$) is the set with no elements, or the event with no outcomes. If $A \cap B = \emptyset$, then A and B are *disjoint*.

Properties

- Commutative:

$$A \cup B = B \cup A,$$
$$A \cap B = B \cap A.$$

- Associative:

$$(A \cup B) \cup C = A \cup (B \cup C),$$
$$(A \cap B) \cap C = A \cap (B \cap C).$$

- Distributive:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

3 Probability Measures

Probability Measures
Measure theory is a branch of mathematics that allows us to rigorously talk about the "size" or "length" of sets in interesting ways. Measure theory is the basis of probability (and thereofore statistics), but a complete treatment of measure theory is outside the scope of this course.
Instead, we will focus on a specific type of measure, called a *probability measure*. Roughly speaking, we can think of a probability measure P on a sample space Ω as a function that assigns real-valued weights (or probabilities) to subsets of Ω.

Definition
A probability measure P on a sample space Ω is a function that assigns real values to subsets of Ω, and must satisfy the following conditions:

Probability Measure

1. $P(\Omega) = 1$.
2. If $A \subset \Omega$, then $P(A) \geq 0$.
3. If A_1 and A_2 are disjoint, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

More generally, if $A_1, A_2, \ldots$ are a set of mutually disjoint sets, then

$$P\big(\cup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i).$$

Properties

Property A (Theorm 1.1)
$P(A^c) = 1 - P(A)$.

Proof.

Properties II

Property B (Corrolary 1.1)
$P(\emptyset) = 0$.

Proof.

Properties II

Property C (Theorem 1.2)
If $A \subset B$, the $P(A) \leq P(B)$.

Proof.

Proprties III

Property D (Theorem 1.3)
"Addition Law": $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof as an exercise. Often, it can be helpful to draw Venn diagrams. See section 1.3 of our textbook for help.

Assigning Probabilities
Recall that the sample space Ω can consist of many types of outcomes. A probability measure can assign probabilities to subsets of Ω in a variety of ways, as long as it satisfies the axioms provided above.
The most simple examples are finite sample spaces. Suppose that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. A probability measure for this sample space defines a mapping such that for all $\omega_i \in \Omega$, $P\big(\{\omega_i\}\big) = p_i$, with $\sum_i p_i = 1$. That is, each element $\omega_i \in \Omega$ gets assigned a probability value p_i.

Tossing a coin
Consider the simple experiment of tossing a coin once. The sample space is $\Omega = \{h, t\}$. There are many possible probability measures that can be assigned to this space:

Fair coin
One probability measure P assumes that the coin is fair, or that each outcome is equally as likely. Thus, $P(\{h\}) = \frac{\text{Total Number of Outcomes of} h}{\text{Total Number of Outcomes}} = 1/2$.

Unfair coin
Alternatively, we could define a probability measure P' that represents an unfair coin. In this case, we could assign $P'(\{h\}) = p_h$, and $P'(\{t\}) = 1 - p_h$. p_h can be any value such that $0 \leq p_h \leq 1$, and this proposed measure P' will satisfy the axioms.

Computing probabilities
Formally, probability measures are defined on subsets (events) of Ω, not elements of Ω, which is why we write $P(\{h\})$ rather than $P(h)$, though often we use the later notation for convenience.
For finite spaces, we can think of each outcome (ω) as it's own event ($\{\omega\}$). Then, more more interesting events, the probability is computed by summing the disjoint events that make up the more interesting subset.

Flipping two fair coins
Consider flipping a fair coin twice. The sample space is:

$$\Omega = \{hh, ht, th, tt\}.$$

Each event is equally likely, that is, for all $\omega \in \Omega$, $P(\{\omega\}) = 1/4$. Now consider the event A, that there is at least one tails in the two coin tosses. Then,

$$P(A) = P\big(\{ht, th, tt\}\big) = P\big(\{ht\}\big) + P\big(\{th\}\big) + P\big(\{tt\}\big) = \frac{3}{4}.$$

4 Computing Probabilities

Counting Methods
In the above example, all outcomes were equally likely. This type of situation is very common, which leads us to the first general method for computing probabilities in situations where it is not so easy to write down all possibilities.

Finite, equal probabilities

Suppose Ω has N elements, and the probability measure P assigns equal weight to all outcomes. Then, for any event $A \subset \Omega$ that can occur in n possible ways, then:

$$P(A) = \frac{\text{Number of ways } A \text{ can occur}}{\text{Total number of outcomes}} = \frac{n}{N}.$$

Because of this, we now will focus on some counting strategies.

Multiplication Principle
Consider an experiment that consists of two smaller experiments with sample spaces Ω_1 and Ω_2, such that $|\Omega_1| = n$, and $|\Omega_2| = m$. Then the total number of outcomes is $n \times m$.

Proof. Write the outcomes of the first experiment as $(a_1, \ldots, a_n)$, and the outcomes of the second experiment as $(b_1, \ldots, b_m)$. The outcome of the complete experiment can be expressed as pairs (a_i, b_j). We can then write the complete set of experiments as an $n \times m$ matrix, with entries the unique combinations of (a_i, b_j). This matrix has $n \times m$ elements. $\square$

Student Government
In a class, a teacher would like to randomly select 1 boy and 1 girl to serve as representatives to the student government. If there are 12 boys and 18 girls, then the total number of ways she can pick students (outcomes) is $12 \times 18 = 216$.

The multiplication principle also extends to the case where there are many experiments.

Binary Numbers
An 8-bit binary number contains a sequence of 8 digits, each being 0 or 1. How many different 8-bit words are there? For each bit, there are two choices. Thus, using the multiplication principal, there are:

$$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^8 = 256$$

digits.

Permutations
A *permutation* is an ordered arrangement of objects. For instance, consider a set $A = \{a_1, a_2, \ldots, a_n\}$, and suppose that we want to choose r elements from this set and list them in order. How many ways can we do this?
Answer: Depends on if we sample *with* or *without* replacement.

With Replacement
Suppose there are n labeled marbles of in a bag, and I want to perform the following experiment: draw out a marble, record it's label, put it back in the bag, and repeat this experiment r times.
By the multiplication principle, we will treat these as r experiments. Each experiment has n possible outcomes, so there are $n \times n \times \ldots \times n = n^r$ possible outcomes.

Without Replacement
Suppose there are n labeled marbles of in a bag, and I want to perform the following experiment: draw out a marble, record it's label, and then repeat this experiment r times without putting marbles back in the bag.
We can still use the multiplication principle, but now the experiments change. The first time, there are n possible outcomes. After taking a marble out, there are only $n-1$ outcomes for the second experiment, and so on. Thus, the total number of outcomes is:

$$n \times (n-1) \times (n-2) \ldots \times (n-r+1).$$

The previous examples can be used as a basic "proof sketch" to the following proposition:

Proposition 1.1: Ordering objects
For a set of size n, and a sample size of r, there are n^r different ordered samples with replacement and

$$\begin{aligned} n(n-1)(n-2)\dots(n-r+1) &= \\ n(n-1)(n-2)\dots(n-r+1)\frac{(n-r)(n-r-1)\dots1}{(n-r)(n-r-1)\dots1} &= \\ \frac{n!}{(n-r)!} &:= nPr(n,r) \end{aligned}$$

without replacement.

Corollary 1.1.1
The number of orderings of n elements is $n(n-1)(n-2)\dots1 = n!$

License plates
In some states, license plates have six characters: three letters followed by three numbers. How many unique plates are possible?
This is an example of sampling with replacement (as each license plate can have duplicated numbers or letters). Thus, there are 26^3 different ways to choose the letters, and 10^3 different ways to choose the numbers. Using the multiplication principle, there are $26^3 \times 10^3 = 17,576,000$ possible unique license plate numbers (using this rule).

Birthday probabilities
Suppose that a room contains n people. Assuming birthdays are uniformly distributed for 365 days, what is the probability that at least two of them have a common birthday?

Solution: See examples E and F from the textbook.

Combinations
So far, we have considered the case where we care about the order. What if the order doesn't matter? Instead, when just want to know what makes up a sample and don't care about the order in which they are obtained.
Question: If r objects are taken from a set of n objects without replacement (and disregarding order), how many different samples are possible?
Probably a few ways to think about this, but let's use the theorems / corollaries / propositions that we have already derived. First, consider how many unique ordered samples there are (without replacement). Then, how many times are we counting the same sample? That is, how many unique ways can we arrange the sample?

Combinations formula
First, there are $nPr(n,r)$ ordered samples, which is equal to the number of unique samples (the thing we want, let's call it $\binom{n}{r}$), times the number of ways to order each unique sample. From Corollary 1.1.1, the latter value is $r!$. Thus: $\binom{n}{r} = nPr(n,r)/r! = \frac{n!}{(n-r)!r!}$.

Proposition 1.2: Binomial Coefficient
The number of unordered samples of r objects selected from n objects without replacement is:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

We call this "*n choose r*", or sometimes the binomial coefficients.

The term *binomial coefficient* comes from the Binomial Thereom, which states:

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}.$$

In particular, this implies that $2^n = \sum_{k=0}^{n} \binom{n}{k}$, which can be interpreted as the number of subsets of a set of n objects; that is, it is the sum of the number of subsets of size 0 (which, is taken to be 1), the number of subsets of size 1, the number of subsets of size 2, etc. The set of all possible subsets is known as the *power set*, and for finite sets of size n, this set has size 2^n based on this calculation.

Lottery
Suppose to win a jackpot for a given lottery, you must correctly choose 6 numbers from 1 to 53, where the order doesn't matter. How many possible combinations of numbers are there, and if you play once, what is the probability that you win (these numbers come from California lottery in the 90's)
Answer: There are $\binom{53}{6} = 22,957,480$ possible combinations. If you play once, your probability of winning is $1/22,957,480 \approx 0.00000004$.

Quality Control
Suppose you are tasked with quality control of a manufacturing process, and that there are n total items, and k defective items. If you randomly sample r items, what is the probability that you find exactly m defective items in your sample (this question is relevant because it can be used to design effective sample schemes).
Solution:

Extending the binomial coefficient

Proposition 1.3: Multinomials
The number of ways that n objects can be grouped into r classes of size n_i, $i = 1, \ldots, r$ is

$$\binom{n}{n_1\, n_2\, \cdots\, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!},$$

where $n = \sum_{i=1}^{r} n_i$.

Proof Sketch: There are $\binom{n}{n_1}$ ways to choose the objects for the first class. Having done that, there are $\binom{n-n_1}{n_2}$ ways of choosing the objects for the second class, etc. Thus:

$$\binom{n}{n_1\, n_2\, \cdots\, n_r} = \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \frac{(n-n_1-n_2)!}{(n-n_1-n_2-n_3) n_3!} \cdots$$

Canceling out factors, we get the desired result.

Multinomial Theorem
Like the binomial theorem, the numbers $\binom{n}{n_1\, n_2\, \cdots\, n_r}$ are called *multinomial coefficients*. They appear in the expansion:

$$(x_1 + x_2 + \ldots + x_r)^k = \sum \binom{n}{n_1\, n_2\, \cdots\, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r},$$

where the sum is over all nonnegative integers $n_1, n_2, \ldots, n_r$ that satisfy $n_1 + n_2 + \ldots + n_r = n$.

Multinomial Examples

Subcommittees
How many ways can a committee of seven members be divided into three subcommittees of sizes three, two, and two, respectively?

$$\binom{7}{3\,2\,2} = \frac{7!}{3!2!2!} = 210.$$

Genomics

In how many ways can the set of nucleotides $\{A, A, G, G, G, G, C, C, C\}$ be arranged in a sequence of nine letters?
To answer, let's re-frame the question. How many ways can nine positions be divided into subroups of sizes two, four, and three (i.e., the locations of the letters A, G, and C?)

$$\binom{9}{2\,4\,3} = \frac{9!}{2!4!3!} = 1260$$

5 Conditional Probability

test cond prob
Start of Conditional Prob.

6 Independence

test 1
Start of independence. (Bretó, 2014)

Acknowledgments

- Compiled on August 11, 2025 using R version 4.5.1.

- 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.

References

Bretó C (2014). "On idiosyncratic stochasticity of financial leverage effects." *Statistics & Probability Letters*, **91**, 20–26. doi: 10.1016/j.spl.2014.04.003. 29