STAT 400

Applied Statistics and Probability I

Elizabeth Qiu

Prof. Paul J. Smith • Summer 2022 • University of Maryland

Last Revision: July 15, 2022

Contents

1 Monday, July 11, 2022

This class is STAT400: Applied Probability and Statistics I. Topics covered: random variables, standard distributions, moments, law of large numbers and central limit theorem, sampling methods, estimation of parameters, testing of hypotheses.

Logistics

- Textbook: Devore (2018), Probability and Statistics for Engineering and the Sciences (9 ed). Cengage Learning.
- All lectures are recorded and posted on Panopto.
- Frequent homework assignments and possibly pop quizzes. Assignments on ELMS.
- Assignments on ELMS.
- The grade breakdown is 10% for Homeworks, 10% for Quizzes, 25% for each Midterm (there are 2), and 40% for the Final Exam.

Probability, Statistics and Data

In STAT400, we will visualizing and summarize data. At the end we will develop mathematical methods for data analysis.

- **Statistics** is the science of recognizing, collecting, analyzing, interpreting data.
- **Probability** is a real-valued function of all events in a class of events such that unions, intersections and complements are also events. It is the mathematical theory of randomness and most of this course will be devoted to modeling random phenomena.
- **Randomness** almost always arises in real world data.

Example 1: We went over the ”Newcomb’s 1882 Measurements” example, where 66 measurements of the passage time of light were recorded by Newcomb in 1882 (values divided by 1000 plus 24 give the time in microseconds for light to traverse a known distance). We had a dataset with a summary of a minimum, 1st quartile value, median, mean, 3rd quartile value, maximum value, and standard deviation.

What do the summaries tell us?

- The **mean** and **median** might both be considered as ”typical” values of the data.
- The mean is
$$\bar{X} = \frac{1}{66}\sum_{i=1}^{66} X_i,$$
the *sample average* of the data.
- The median $\bar{X}$ is the value such that half of the observations are larger than and half are smaller.
- We need to assess how much variation or *spread* is contained in the data. The sample range (max-min) or interquartile range might be used, but a better choice is the standard deviation, defined by
$$s^2 = \frac{1}{66-1}\sum_{i=1}^{66}(X_i - \bar{X})^2$$
.

Once we have the central values of the data, how far do we expect the values to be? **Standard deviation** is a value that tells you how far away (absolute value) from the typical observation we see data from the mean.

To visualize data, a **Histogram** is a very helpful shape of the graph that tells you typical distributions. **Outliers** are points outside of the bulk of data.

Interpreting histograms:

- Values that are clearly separated from others are outliers.
- The histogram roughly follows the "bell curve", where there are few values at the low ends of scale, rising to a maximum near the mean or median, and then a decline from the maximum. Bell shapes are not easily recognized without visual aid the graph.
- Dropping outliers will only slightly change the mean and median, but may greatly affect the standard deviation, because outliers affect the spread.
- Histograms give good visualizations of shapes and distributions, which allow for comparing samples graphically.

Example 2: We also went over a Cereal example, using boxplots to compare calories from different Cereal brands.

Another way to simplify the visualization of a dataset is with a **Boxplot**. It's made up of the minimum, 1st quartile, median, 3rd quartile, and maximum. The measure spread is the **interquartile range**.

The boxplot of a single sample is not really that useful, especially where calculations can be made instantly with a couple lines of code. But it's just as easy to create a histogram. And histograms are not a reliable indicator when you have small samples of data, like we have in the Newcomb's 1882 Measurements example.

And don't worry too much about Count Data; most data we will encounter in this course is Measurement Data.

We've seen unusual features of the previous two data sets which may be random or due to some systematic differences. Both systematic differences and unusual observations should be investigated. The summaries and graphs suggest features of data that may be interpreted rigorously by employing the mathematical theory of probability. We will see how to make reliable inferences about data sets like these, once we understand probability theory. That will be our next topic.

Recommended: We will use the **R** statistical package in this course. R is a general purpose system which has sophisticate mathematical and statistical functions as well as data handling routines. Learn R. It's also free! Book: Dalgard, P. (2008), *Introductory Statistics with R, Springer.*

Sample Spaces, Outcomes, and Events

Most of this course will be based in probability. Probability theory is based on the concept of a **random experiment**. The outcome of the experiment is random, but the set of all possible outcomes is known.

We are usually not interested in single outcomes, but rather sets of outcomes, known as **events**. Events can be described using "and" (∩), $“or”(\cup), and“not”(').Weassignprobabilitiestotheseevents.$

The set of all outcomes is the **sample space**, denoted S, and the individual outcomes are denoted s. The sample space can also be continuous.

Example 1: tossing a coin n times, whether fair or unfair, will result in 2^n outcomes. The outcomes are H and T. For one coin toss, $S = \{$H, T$\}$. For three consecutive coin tosses, $S = \{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$.

Probabilities are numerical values assigned to events. For example, in the experiment of tossing three coins, events of interest might be $A = \{\text{More heads than tails}\}$ or $B = \{\text{First toss is T}\}$. Each event describes a set of outcomes. For example, $A = \{\text{HHH, HHT, HTH, THH}\}$, $B = \{\text{THH, THT, TTH, TTT}\}$.

To combine two events A and B, we write $A \cap B$ for "A and B" and $A \cup B$ for "A or B". Note that $A \cup B$ means that A occurs or B occurs or both occur. The complement of A, or "not A", is denoted $A^{'}$.

Relative Frequency and Probability

- Probability is a numerical value attached to an event which may or may not occur in a random experiment.
- Baseball is a random experiment: a batter may or may not get a hit. A batter has hundreds of trials of this experiment. A given at-bat is unpredictable, but over a long season the batter's *proportion* of hits becomes stable. This is his batting average. We assume the experiment was conducted more or less identically, and that the player is consistent.
- Consider a long sequence of trials of a random experiment, performed under identical conditions. After N trials, the event A occurs *N(A)* times. The relative frequency of A, $f(A) = \frac{N(A)}{N}$, tends toward a constant value *P(A)*.

Properties of relative frequency:

- $0 \leq \frac{N(A)}{N} \leq 1$
- $\frac{N(S)}{N} = 1$. This means that S is an event. So is $S^{'} = \varnothing$.
- If A and B are **exclusive** or **disjoint**, meaning they can't occur at the same time, then $\frac{N(A \cup B)}{N} = \frac{N(A)}{N} + \frac{N(B)}{N}$.

These properties are obvious and should also hold for probabilities, which are limiting relative frequencies. As for implications of relative frequency for random samples, there is a shared characteristic that after an initial ragged, irregular, or unpredictable behavior over the first few trials, it tends to stabilize towards a value *P(A)*. The randomness never goes away, but it does get averaged out.

Later, we can use data and real world random experiment observations to create mathematical models.

Axioms of Probability

Since probability is a mathematical concept, it must follow some mathematical rules. They must satisfy the following axioms:

1. For any event A, $O \leq P(A) \leq 1$.
2. The sample space S satisfies $P(S) = 1$, where S resembles a finite or infinite sequence of outcomes.
3. Let $A_1, A_2, \ldots, A_n$ be any sequence of disjoint events. Then
$$A_1, A_2, \ldots = \sum_{n=1}^{\infty} P(A_n)$$
.

Axioms 1 and 2 resemble properties of relative frequency, but we regard them as defining a mathematical model.

Why we need Axiom 3: Suppose we have a long sequence of independent trials and cannot describe it the union of finitely many events. Therefore, we assume Axiom 3 to describe its probability; it's extended to handle infinite spaces.

2 Tuesday, July 12, 2022

Probability Calculations

Events can be combined "algebraically" using unions, intersections and complements. There must be corresponding arithmetic rules for calculating their probabilities. These rules are all derived from the axioms. We need to assign probabilities for any theoretical or real-world situation.

For example, we might define the success probability on any trial by p and the failure probability by $1-p$, regardless of what happens on other trials. Then a sequence {SFSS}, where S is a success and F is a fail, is assigned probability $p^3(1-p)$.

Basic Formulas

- As we have seen, the sample space S is an event with probability 1. Its complement is the empty set $\varnothing$. Therefore, S is the certain event and $S^{'} = \varnothing$ is the impossible event. If A and B are disjoint, then $A \cap B = \varnothing$.
- Here are some basic formulas, proved in the text:

$$P(A) = 1 - P(A^{'})$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(A' \cap B) \\ &= P(A) + [P(B) - P(A \cap B)] \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- "In general, the probability of a union of k events is obtained by summing individual event probabilities, subtracting double intersection probabilities, adding triple intersection probabilities, subtracting quadruple intersection probabilities, and so on" (page 63, [**?**]).
- If A occurs whenever B occurs, then A is a **subset** of B, written $A \subseteq B$. In words, A *implies* B. Other interpretation include: "is contained in". In this case, $P(A) \leq P(B)$.
- If S is a certain event that will occur, then the complement of S is an impossible event. Two disjoint events' intersection is an empty set (because it will be impossible that those two events will occur at the same time).
- Formulas like these are suggested by Venn diagrams but are proved using the axioms.

Assigning Probabilities

- One must know the sample space and the form of the outcomes.
- Let $S = S_1, S_2, \ldots, S_m$. Choose numbers $p_j, j = 1, \ldots, m$ such that $0 \leq p_j \leq 1$ (you choose the probabilities for each p_j) and $\sum_{j=1}^{m} p_j = 1$. In this setup, any event has finitely many outcomes. Its probability is the sum

$$P(A) = \sum_{s_j \in A} p_j,$$

where each outcome belongs to A, with a probability p_j attached to it. This rule is used in gambling games with equally likely outcomes.

- Thinking concept for probability in terms of size of a geometric region: if S is a geometric region and A is a subregion, we can define $P(A) = \frac{size\,of\,A}{size\,of\,S}$. "Size" might mean length (of a line segment), area (of a plane region), or volume (of a solid). This only makes sense if S is bounded.

Counting

Probability theory began in the 17th century with the analysis of gambling games, such as cards, coin tossing, dice, or roulette. Often a gamble might be repeated several times. Gambles usually have finitely many outcomes. If the gamble is fair, each outcome should have the same probability.

To calculate probabilities, one must calculate the number of outcomes in S, say n, and the number of outcomes in some event A, say n_A. Then, $P(A) = \frac{n_A}{n}$.

Small sample spaces can be counted explicitly. Then $P(A)$ can be calculated for any A. For larger and more realistic setups, it may be easier to imagine the experiment as conducted in steps.

Example 1: A committee of 2 members is chosen from an organization of 10 members. First a committee chair is selected; there are 10 ways to do this. Then, a secretary is chosen from the remaining 9 members. Therefore, there are $90 = 10 \times 9$ ways to select the committee.

Example 2: Two dice, one red and one green, are tossed. There are 6 outcomes for the red die and 6 for the green die. Therefore, there are $36 = 6 \times 6$ outcomes of the experiment.

The previous examples illustrate a general principle for a two-step process (assuming that each of the outcomes are equally likely). Suppose there are m outcomes in the first step, followed by n outcomes in the second step based on the outcome in the 1st step from m. Then, there are mn total outcomes for the two step process. This principle can be displayed using a **tree diagram**, where each path through the tree is one way to perform the two-step process. It generalizes to k step processes for any k.

Permutations and Combinations

Suppose there are n objects (persons, households, etc.) and we want to select k objects *in order*. That is a **permutation**, or ordered arrangement. But how many permutations are there? We create the permutation in k steps and use the basic principle:

1. Choose the first object: there are n ways.
2. Choose the second object: there are $n-1$ ways.

$\vdots$

k. Choose the k-th object: there are $n-(k-1)$ ways.

Hence there are $n(n-1)(n-2)\dots(n-(k-2))(n-(k-1)) = \frac{n!}{(n-k)!}$ permutations ("permute $n\,k$").

A key assumption about permutations is that the objects are distinguishable. This is obviously true is one is permuting persons, but we can imagine that objects are distinguishable, as in **Example 2**.

But often, order of selection is not important. For example, in a political poll, order of selection is irrelevant. The only event of interest in a poll might be how many sampled persons will vote Democrat or Republican. A **collection** is an unordered set of k objects chosen from a collection of n objects. Imagine that the objects are names drawn simultaneously from a hat.

To count the number of combinations of k objects chosen from a set of $n, C_{k,n}$, we imagine a two-step process:

1. Choose a combination: $C_{k,n}$ ways.
2. Arrange the elements of the combination in order: $k!$ ways.

The result is a permutation of k (size subset) out of n objects (to choose from). Therefore ("choose $n\,k$"):

$$P_{k,n} = \frac{n!}{(n-k)!} = C_{k,n} \times k!$$

$$C_{k,n} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

The $\binom{n}{k}$ ("n choose k") notation is more common, and $\binom{n}{k}$ is called a **binomial coefficient**, because it comes from the binomial theorem (to be discussed later). Note that $\binom{n}{n} = \binom{n}{0} = 1$, and $\binom{n}{1} = n$.

Basic Counting Principle; Sampling and Replacement

Polls and surveys select combinations of persons from a population. This is also true of many card games, such as poker. By contrast, gambling games like dice or roulette repeated trials where the outcomes are the same at each step.

The **basic counting principle** says that if each step in a multi-step experiment has n outcomes, then there are n^k outcomes in the combined experiment. Drawing without replacement results in $\frac{n!}{(n-k)!}$ outcomes, and drawing with replacement results in n^k outcomes. Examples include lotteries, raffles, powerballs, etc.

1. **Example 3**: Urn: Imagine an urn with balls numbered $1, 2, \ldots, n$. If we draw k balls without replacement there are $\frac{n!}{(n-k)!}$ outcomes. If we draw a ball and replace it, there are n^k outcomes.
2. **Example 4**: Daily lotteries: A player bets on a three or four-digit number. There are 10^3 or 10^4 possible outcomes.
3. **Example 5**: Raffles: Suppose 10,000 tickets are sold and there is a first prize, second prize, and third prize. Then, the winning tickets are a permutation of three tickets out of 10,000.
4. **Example 6**: Powerball: A player bets on a permutation of 5 out of 69 numbers along with a powerball number, chosen from $\{1, 2, \ldots, 26\}$. There are $(69 \times 68 \times 67 \times 66 \times 65) \times 26$ outcomes.

Sampling without Replacement, and **Ordered sampling with Replacement**

Example 7: Imagine an urn contains r red balls and b blue balls. If k balls are drawn without replacement, find $P(A) = P(\text{exactly 4 red balls})$ (assuming equally likely outcomes).

The sample space contains $\binom{r+b}{k}$ outcomes. If A occurs, 4 of the r red balls were chosen and $k-4$ of the blue balls were chosen. The red balls are a combination of 4 out of r red balls. Similarly, the blue balls are a combination of $k-4$ out of b blue balls.

Using the basic principle, the formula for combinations and the equally likely assumption,

$$P(A) = \frac{\binom{r}{4}\binom{b}{k-4}}{\binom{r+b}{k}}.$$

Example 8: Now assume k balls are drawn from the urn *with* replacement. This form of sampling automatically identifies a first, second, etc. ball. Now the sample space has $(r+b)^k$ possible outcomes. We want to find $P(B) = P(\text{first red ball on the fifth draw})$.

If B occurs, the outcome must have been of the form $bbbbrxxx\ldots x$, where the sequence begins as shown, b's are blue, r's are red, and the x's are colored arbitrarily. There are $b^4(r+b)^{k-5}$ outcomes in B. Therefore,

$$P(B) = \frac{b^4 r(r+b)^{k-5}}{(r+b)^k}(\text{the number of outcomes of } B \text{ divided by the sample space}) = \frac{b^4 r}{(r+b)^5}.$$

If the sampling were ordered without replacement, then

$$P(B) = \frac{[\frac{b!}{(b-4)!}]r}{[\frac{(r+b)!}{(r+b-4)!}]}.$$

A good application of ordered sampling without replacement is quality testing.

3 Wednesday, July 13, 2022

Due to the power outages caused by the storm yesterday, there is no class today. Continue working on the assigned homework problems.

4 Thursday, July 14, 2022

Due to the power outages caused by the storm on July 12, the professor's video and screensharing Zoom options were not available in class today. Instead, we went over a few questions students had.

Continue working on the assigned homework problems. Missing two days of instruction results in a missed week. We will probably cut off a small portion of chapter 7 content.

5 Friday, July 15, 2022

Conditional Probability

Example 1: Political poll: A political poll is conducted in Maryland. Respondents are classified by resident (urban, suburban, rural) and party preference (Democrat, Independent, Republican). The results are recorded and treated as probabilities. Thus $P(\text{Democrat \& Rural}) = 0.03$, $P(\text{Suburban}) = 0.40$, $P(\text{Independent}) = 0.30$, etc.

A politician may be interested in the proportion of suburbanites $P(\text{Suburban}) = 0.40$ who are Republicans $P(\text{Republican \& Suburban}) = 0.03$. The proportion is $\frac{0.03}{0.40} = 0.075$.

The calculation $\frac{P(\text{Republican \& Suburban})}{P(\text{Suburban})} = 0.075$ is the **conditional probability** that a person is a Republican, given that they are a suburbanite. In symbols, $P(\text{Republican}|\text{Suburban}) = \frac{P(\text{Republican \& Suburban})}{P(\text{Suburban})}$. We are assuming that S has occurred, and we see that the occurrence of S changes the probability of being a Republican. This is because of the sample space changes from the whole state to suburbanites only.

The general definition is that for any two events A and B such that $P(B) > 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

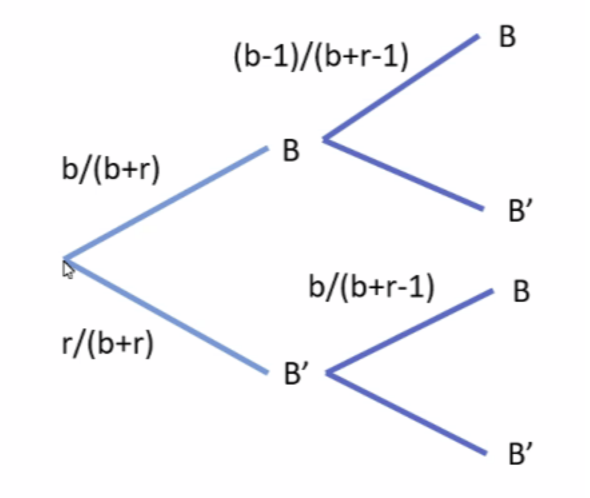
This definition yields the **multiplication rule**:

$$P(A \cap B) = P(A|B) \times P(B).$$

There is no one simple way to compute $P(A|B)$. In some cases we can calculate $P(A \cap B)$ and $P(B)$ directly (as in **Example 1**) and then we can easily evaluate $P(A|B)$ by division. In other problems, we think of B as a new reduced sample space. If probabilities on B are easy to compute, we evaluate $P(A|B)$ directly.

Example 2: Urn problem, revisited: As before the urn contains r red and b blue balls. Balls are selected one at a time without replacement. We want $P(B_2) = P(\text{blue ball on second draw})$. The first ball results in either B_1 or $B_1' = R_1$, where $B_1 = \{\text{blue ball on first draw}\}$.

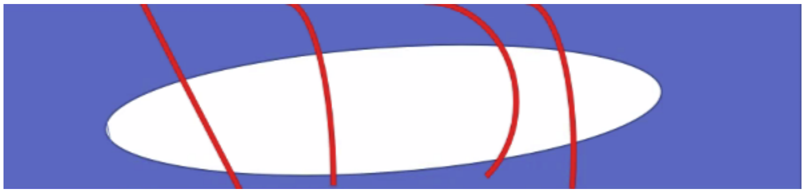
From the problem, we can conclude that $P(B_1) = \frac{b}{b+r}$ and $P(B_1') = \frac{r}{b+r}$. Now consider $P(B_2|B_1)$. Think of B_1 as a reduced sample space with r red and $b+r-1$ blue balls. Then, we see that $P(B_2|B_1) = \frac{b-1}{b-1+r}$. Similarly, $P(B_2|B_1') = \frac{b}{b-1+r}$.



From Axiom 3, we see that $B_2 = (B_1 \cap B_2) \cup (B_1^{'} \cap B_2)$ which is a union of disjoint sets. So $P(B_2) = P(B_1 \cap B_2) + P(B_1^{'} \cap B_2)$. The multiplication rule says that $P(B_1 \cap B_2) = P(B_1) \times P(B_2|B_1)$ and $P(B_1 \cap B_2) = P(B_1) \times P(B_2|B_1)$. We find that $P(B_2) = \frac{b}{b+r}$, which is identical to $P(B_1)$. Drawing a tree to illustrate this scenario would also be helpful.

Total Probability

Suppose S is partitioned into subsets $A_1, A_2, \ldots, A_k$, which are disjoint. This means that $S = A_1 \cup A_2 \cup \cdots \cup A_k$ and $A_i \cap A_j = \varnothing$. In the venn diagram, $k = 5$.



The white oval is the event B and the rectangle is sample space S. The red lines indicate the partitioning sets $A_1, A_2, \ldots, A_k$. Note that B is also partitioned into disjoint subsets $(A_1 \cap B), \ldots, (A_k \cap B)$.

Now Axiom 3 yields the Total Probability Formula:

$$P(B) = P(A1 \cap B) + \cdots + (A_k \cap B) = \sum_{j=1} kP(A_j) \times P(B|A_j).$$

Example 3: Political poll, continued: In the example of the political poll, we can see that $P(\text{Democrat}) = P(\text{Democrat} \cap \text{Republican}) + P(\text{Suburban} \cap \text{Democrat}) + P(\text{Urban} \cap \text{Democrat}) = 0.03 + 0.25 + 0.32$.

Example 4: Friend mail: You ask a friend to mail a letter. They will forget to mail it (event $M^{'}$) with probability 0.1. If it is mailed, the Post Office will fail to deliver the letter (event $D^{'}$) with probability 0.1. With is $P(D^{'})$? From the Total Probability Formula, $P(D^{'}) = P(D^{'} \cap M) + P(D^{'} \cap M^{'}) = (0.1)(0.9) + (1)(0.1) = 0.19$.

Bayes' Formula

In the setup of the Total Probability Formula, let A_j be possible causes and B be a result. We know $P(A_j)$ and $P(B|A_j)$ for $j = 1, \ldots, k$.

If we observe that B occurs, we might ask, "what is the chance that B was caused by A_j?" In other words, what is $P(A_j|B)$?

The answer is **Bayes' Formula**:

$$P(A_j|B) = \frac{P(Aj \cap B)}{P(B)} = \frac{P(A_j) \times P(B|A_j)}{\sum_i P(A_i) \times P(B|A_i)}$$

Bayes' formula is an application of the Total Probability Formula.

Example 5: COVID-19 testing: A diagnostic test for disease, such as COVID-19, is characterized by two parameters: Sensitivity $= SE = P(+|D)$, and Specificity $= SP = P(-|D^{'})$. Let $+ =$ testing positive, $- =$ testing negative, and $D =$ the patient has the disease.

However, the performance of the test also depends on the prevalence of disease in the population, i.e., $P(D)$.

Suppose $P(D) = 0.05$, $SE = 0.90$, and $SP = 0.95$. The real question is "if a patient tests positive, do they really have the disease?" We use Bayes' Theorem to calculate $P(D|+)$:

$$P(D|+) = \frac{P(D) \times P(+|D)}{P(+)} = \frac{(0.05)(0.90)}{(0.05)(0.90) + (0.95)(0.05)} = 0.4865$$

Interpretation: This means that a positive test result is correct less than half the time. Application: That is why we must conduct more than one test.

References

[1] Devore, J. L. (2016). Probability and Statistics for Engineering and the Sciences (9th edition). Cengage.