Introduction & Revision of Basic Probability

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- Instructor: Dr. **Sharon** McNicholas.
- Take a look at the course outline.
- Notes will be posted online ahead of classes; assignments and the final report will also be posted online.
- Office hours.
- Learning outcomes.
- We have one block of (almost) three hours per week for class; we will take a 10 minute break somewhere around the middle of each block.

Data Science

- What is data science anyway?
- That is a good question and I will not attempt to answer it today.
- Some words of wisdom:

  "...we must help the student to recognise the computer for what it is — a sophisticated tool, not a substitute for thought." (Barnett, 1999)[a]

[a]Barnett, V. (1999), *Comparative Statistical Inference*, 3rd edition, Chichester: England.

More Words of Wisdom

- From Tukey and Wilk (1966)[a]:

  > Nothing — not the careful logic of mathematics, not statistical models and theories, not the awesome arithmetic power of modern computers — nothing can substitute here for the flexibility of the informed human mind...Accordingly, both analysis approaches and techniques need to be structured so as to facilitate human involvement and intervention.

[a]Tukey, J. and Wilk, M. (1966), 'Data Analysis and Statistics: An Expository Overview', AFIPS, International Workshop on Managing Requirements Knowledge, pp. 695–709.

Your Background

- Everyone takes a different background into this course.
- Before going any further, it is helpful to understand each other's background and experience with real data.
- I will start, stating my background and experience as I would have as a beginning graduate student.
- Now, it is your turn...
- Next, a revision of (very) basic probability.
- Important note: the material in this set of notes is particularly easy and should not be taken as indicative of the difficulty level of this course.

Terminology

- Consider an **experiment**. This can be something as simple as a coin toss or the roll of a die.
- The **sample space** is the set of all possible outcomes. For the roll of a die, this is $\{1, 2, 3, 4, 5, 6\}$.
- An **event** is any subset of the sample space. For the roll of a die, this could be the odd numbers $\{1, 3, 5\}$.

What is Probability?

- Probability is a way to quantify the uncertainty surrounding the outcome of an experiment.

- We talk about the probability of an event. If E is an event then we denote the probability of E occurring by $P(E)$.

- For an experiment with possible outcomes $E_1, E_2, \ldots, E_n$, the probability $P(E_i)$ must obey the following rules:
  a. $0 \leq P(E_i) \leq 1$ for all E_i
  b. $P(E_1) + P(E_2) + \cdots + P(E_n) = 1$
  c. $P(\emptyset) = 0$
  d. The OR law for mutually exclusive events (see later).

Fundamental Probability

- This is the very basic idea of a probability.

- Suppose an experiment has n possible outcomes, r of which satisfy some event E, then
$$P(E) = \frac{r}{n}.$$

- Example 1: A fair die is rolled (the experiment). What is the probability of it showing an even number? In this case, E is the event of an even number, there are $n = 6$ possible outcomes and there are $r = 3$ even numbers, so
$$P(E) = \frac{r}{n} = \frac{3}{6} = \frac{1}{2}.$$

Multiple Events: Mutually Exclusive

- What would happen with more than 2 dice? Would we have to write out all of the possible outcomes?
- What if we rolled a die and chose a card — imagine writing out all of those outcomes ($6 \times 52 = 312$ in all)!
- Thankfully, we do not need to write out all of these outcomes — there are rules we can use... but first a definition.
- Two or more events are said to be **mutually exclusive** if they cannot occur at the same time.

OR Rule

- Consider rolling a die; the event of getting a 4 and the event of getting a 5 are mutually exclusive.
- The **OR Rule for mutually exclusive events**: if E and F are mutually exclusive events then

$$P(E \text{ or } F) = P(E \cup F) = P(E) + P(F).$$

- This makes sense when you consider that E and F are mutually exclusive, i.e., $E \cap F = \emptyset$.

Examples I

- Example 1: A card is selected from a well shuffled pack. What is the probability of it being a jack or a 5? One card cannot be both a jack and a 5 so the events are mutually exclusive. Let $J =$ the event of a jack and $5 =$ the event of a 5, then

$$P(J \cup 5) = P(J) + P(5) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}.$$

- Example 2: A fair die is rolled. What is the probability of it landing on 1 or 6? A die cannot land on both 1 and 6 so the events are mutually exclusive. Let $1 =$ the event of a 1 and $6 =$ the event of a 6, then

$$P(1 \cup 6) = P(1) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}.$$

AND Rule

- A card is selected from a well shuffled pack. What is the probability of it being a jack or a red card? There are two red jacks, so the events are not mutually exclusive.

- Before we can learn how to deal with this question, we must consider the AND Rule.

- The **AND Rule**: if E and F are two events then

$$P(E \text{ and } F) = P(E \cap F) = P(E)P(F \mid E),$$

where $F \mid E$ means the occurrence of an event F given that an event E has already occurred.

Examples II

- Example 3: Two cards are selected from a well shuffled pack. What is the probability that they are both jacks? Let $J1$ be the event that the first card is a jack and let $J2$ be the event that the second card is a jack, then

$$P(J1 \cap J2) = P(J1)P(J2 \mid J1) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}.$$

- Example 4: There are 3 red and 4 green balls in a bag. Two balls are selected consecutively, at random. What is the probability that first ball is red and the second green? Let $R1$ be the event that the first ball is red and let $G2$ be the event that the second ball is green then,

$$P(R1 \ \cap \ G2) = P(R1)P(G2 \mid R1) = \frac{3}{7} \times \frac{4}{6} = \frac{2}{7}.$$

Examples III

- Example 5: There are 3 red, 4 green and 7 blue balls in a bag. Three balls are selected consecutively, at random. What is the probability that the first is a red, the second green and the third blue?
Let $R1$ be the event that the first ball is red, let $G2$ be the event that the second ball is green and let $B3$ be the event that the third ball is blue, then,

$$\begin{aligned} P(R1 \ \cap \ G2 \ \cap \ B3) &= P(R1)P(G2 \mid R1)P(B3 \mid R1, G2) \\ &= \frac{3}{14} \times \frac{4}{13} \times \frac{7}{12} \\ &= \frac{84}{2184} \\ &= \frac{1}{26}. \end{aligned}$$

Statistical Independence

- We can take two interesting points from the last example:
  a. The AND and OR rules can be extended to more than two events.
  b. The notation $P(E, F)$ also means $P(E$ and $F)$.

- There is a famous special case of the AND rule — the case where the events are independent. In fact, statistical independence is often defined in terms of the AND rule.

- Two events E and F are said to be **statistically independent** if

$$P(E \cap F) = P(E)P(F).$$

Examples IV

- Practically, two events are independent if the outcome of one has no effect over the outcome of the other.

- Hereafter, we take *independent* to mean *statistically independent*.

- Example 6: A card is selected from a well shuffled pack and a die is rolled. What is the probability of obtaining a club and a 3? Let C be the event that the card is a club and let 3 be the event that the die shows 3, then

$$P(C \cap 3) = P(C)P(3) = \frac{13}{52} . \frac{1}{6} = \frac{1}{24}.$$

Examples V

- Example 7: A card is selected from a well shuffled pack and a die is rolled. What is the probability of obtaining a red card and an even number? Let R be the event that the card is red and let E be the event that the die shows an even number, then

$$P(R \cap E) = P(R)P(E) = \frac{26}{52}.\frac{3}{6} = \frac{1}{4}.$$

- Example 8: A card is selected from a well shuffled pack and a coin is tossed. What is the probability of obtaining a queen and a tail? Let Q be the event that the card is a queen and let T be the event that the coin shows a tail, then

$$P(Q \cap T) = P(Q)P(T) = \frac{4}{52}.\frac{1}{2} = \frac{1}{26}.$$

A Very Useful Trick

- For any event E, we use $\overline{E}$ to denote '*not* E'.

- For any event E,

$$P(E) = 1 - P(\overline{E}).$$

- Example 9: A coin is flipped 20 times, what is the probability that it shows heads at least once? Let H be the event of a head, then,

$$P(\text{at least one } H) = 1 - P(\text{no } H) = 1 - \left(\frac{1}{2}\right)^{20} = \frac{1048575}{1048576}.$$

General Or Rule

- We looked at the OR rule for mutually exclusive events.

- What if the events are not mutually exclusive?

- For any two events E and F, the probability of E **or** F occurring is given by
$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

- Note that if E and F are mutually exclusive then $P(E \cap F) = 0$ and we have the familiar formula $P(E \cup F) = P(E) + P(F)$.

Examples VI

- Example 10: A card is selected from a well shuffled pack. What is the probability of it being a jack or a red card? Because there are two red jacks, the answer is
$$P(R \cup J) = P(R) + P(J) - P(R \cap J) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13}.$$

- Example 11: A card is selected from a well shuffled pack. What is the probability of it being an ace or a spade? Besause there is one ace of spades, the answer is
$$P(A \cup S) = P(A) + P(S) - P(A \cap S) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

Conditional Probability

- We have seen the **AND Rule**: if E and F are two events then

$$P(E \cap F) = P(E)P(F \mid E),$$

where $F \mid E$ means the occurrence of an event F given that an event E has already occurred.

- Dividing both sides of the above equation by $P(E)$ gives us the definition of a conditional probability.

- The probability that an event F occurs given that an event E has already occurred is given by

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)}.$$

Conditional Probability II

- Using the fact that $P(E \cap F) = P(F \cap E)$ and noting that $P(F \cap E) = P(E \mid F)P(F)$, we can write the expression for $P(F \mid E)$ as follows.

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)} \tag{1}$$

- Now, we can also rewrite the term $P(E)$ in this equation. However, first we will look at an example to help motivate the situation.

Conditional Probability Example

- Example 12: In a region, 31% of people are smokers, 19% of smokers develop lung cancer and 2% of non-smokers develop lung cancer. What is the probability that a person chosen at random will develop lung cancer; i.e. what proportion of this region will suffer from lung cancer?

- Given, $P(S) = 0.31$, so $P(\overline{S}) = 1 - 0.31 = 0.69$.
  Also given, $P(C \mid S) = 0.19$ and $P(C \mid \overline{S}) = 0.02$.
  We deduce $P(C)$ as follows:

$$\begin{aligned} P(C) &= P(C \cap S) + P(C \cap \overline{S}) = P(C \mid S)P(S) + P(C \mid \overline{S})P(\overline{S}) \\ &= (0.19)(0.31) + (0.02)(0.69) = 0.0589 + 0.0138 = 0.0727. \end{aligned}$$

  Therefore, the answer is 7.27%.

STATS 780/CSE 780 Prof. Sharon McNicholas

Why?

- Why is it the case that $P(C) = P(C \cap S) + P(C \cap \overline{S})$?

- Consider the set C (people with lung cancer).

- The set C is **partitioned** by S and $\overline{S}$

- The result follows.

The Partition Theorem

- This approach can be generalized to get a general formula for $P(E)$ in terms of conditional probabilities.

- **The Partition Theorem**: Suppose the outcome of an event E depends on an event F which has possible outcomes $F_1, F_2, \dots, F_n$, then

$$P(E) = \sum_{i=1}^{n} P(E \mid F_i) P(F_i)$$

- In the lung cancer example we used the Partition Theorem with $E = C$, $F_1 = S$ and $F_2 = \overline{S}$.

Bayes' Theorem

- This is a basic introduction to Bayes' theorem.

- The Partition Theorem replaces $P(E)$ in the denominator of (1) to give Bayes' Theorem.

- **Bayes' Theorem**: Suppose the outcome of an event E depends on an event F which has possible outcomes $F_1, F_2, \dots, F_n$, then

$$P(F_j \mid E) = \frac{P(E \mid F_j) P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) P(F_i)},$$

$j = 1, 2, \dots, n$.

Bayes' Theorem Example

- Example 13: Returning to the lung cancer example, work out the probability that someone smokes given that they have lung cancer.

- This can be using Bayes' Theorem.

$$\begin{aligned} P(S \mid C) &= \frac{P(C \mid S)P(S)}{P(C \mid S)P(S) + P(C \mid \overline{S})P(\overline{S})} \\ &= \frac{(0.19)(0.31)}{(0.19)(0.31) + (0.02)(0.69)} \\ &= \frac{0.0589}{0.0589 + 0.0138} \\ &= \frac{0.0589}{0.0727} = 0.810 \end{aligned}$$

Introduction to Counting

- Without understanding the principals of counting, it is not possible to completely answer even some the most rudimentary probability problems.

- We are often required to count different permutations in order to solve problems.

- First, we will look at arranging objects in a line.

Factorial Notation I

- The number of ways of arranging n distinct objects in a line is

$$n! = n(n-1)(n-2)\ldots 3.2.1.$$

- $n!$ is called $\mathbf{n}$ **factorial**.

- Example 14: How many ways can the letters KATE be arranged to give a four letter 'word'? The answer is

$$4! = 4.3.2.1 = 24 \text{ ways.}$$

- Example 15: You know that a person has two children and one is a boy. What is the probability that the other is a girl?

Factorial Notation II

- The number of ways of arranging n objects, of which m are identical is

$$\frac{n!}{m!}.$$

- Example 16: How many ways can the letters CIARA be arranged if all must be used? There are two identical letters, so the answer is

$$\frac{5!}{2!} = 5 \times 4 \times 3 = 60 \text{ ways.}$$

- Example 17: How many ways can the letters ENGINEER be arranged if all must be used? There are two sets of identical letters, so

$$\frac{8!}{3!2!} = 3,360 \text{ ways.}$$

Combinatorics

- The number of ways of choosing r items from n distinct items (in any order) is
$$^nC_r = \frac{n!}{r!(n-r)!}.$$
- Example 18: How many ways can a committee of 6 people be chosen from 10 people? The answer is
$$\begin{aligned} ^{10}C_6 &= \frac{10!}{6!4!} = \frac{10.9.8.7.6.5.4.3.2.1}{6.5.4.3.2.1.4.3.2.1} \\ &= \frac{10.9.8.7}{4.3.2.1} = \frac{5040}{24} = 210 \text{ ways.} \end{aligned}$$
- ^nC_r buttons are common on scientific calculators — but you'll probably be using a computer anyway.

Fundamental Principle of Counting

- If a particular task may be accomplished n ways and then a second task may be accomplished m ways, then the first task followed by the second task may the accomplished in nm different ways.
- Example 9: How many ways can a committee of 4 people be chosen from 6 men and 4 women if there must be 2 people of each gender on the committee? The answer is
$$^6C_2 \ ^4C_2 = (15)(6) = 90 \text{ ways.}$$
- Example 19: In a lottery, 6 numbers are drawn from a drum with 42 numbers. How many different winning lines are possible?
$$^{42}C_6 = 5,245,786 \text{ lines.}$$

Probability & Counting

- We could use the method of Example 19 to work out the probability of winning a 42-choose-6 lottery (if one line is played) as follows.

$$P(\text{Winning the lottery}) = \frac{1}{{}^{42}C_6} = \frac{1}{5245786}.$$

- This solution is much easier, computationally, than multiplying out the conditional probabilities.

- Several probability distributions use ${}^{n}C_r$, three common ones are:
  - the hypergeometric distribution,
  - the binomial distribution, and
  - the negative binomial distribution.

Comments

- You may wonder why we have spent the last few minutes going over some (very) basic ideas in probability.

- Consider a very large binary data set, e.g., a set of transactions.

- Believe it or not, using little more sophistication than illustrated in this lecture, we can carry out effective analyses.

- The approach we will look at uses association rules.