STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 2

Outline

Outline

Conditional Probability

- Recall, from week 1, that a sample space Ω lists the possible outcomes of an experiment.
- Most of the examples we have seen so far have been for one run of an experiment.
- In practice, scientists run an experiment multiple times to obtain an idea of the long-run probabilities of events (recall our gambler example).
- The results of these multiple runs of an experiment can be represented in tables.
- When the elements of the sample space consist of multiple pieces (e.g. tossing a coin twice), we can present the results of these multiple experiments in a **contingency table**.

Example: Tossing a coin

Our experiment is to toss a coin twice and record the outcome. The sample space is

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

Suppose we want to determine if our coin is fair. We will run this experiment (e.g. toss the coin twice) 100 times, and record the result of the first and the second flip. We can do this in a contingency table as below:

	First flip		
Second flip	H	T	Total
H	28	24	52
T	21	27	48
Total	49	51	100

Contingency Tables

- Contingency tables represent the results of experiments involving two or more components.
- For example, if we choose to toss our coin 3 times, instead of twice, we would need a $2 \times 2 \times 2$ table to represent the results of our repeated experiment.
- Such tables can also be used to present results from experiments in which there are more than two outcomes of each trial.
  - suppose an experiment is to roll a dice and then flip a coin.
  - there are 6 possible outcomes of the dice roll and 2 outcomes to the coin flip.
  - we would then present results in a 6×2 contingency table
- Contingency tables are especially useful for thinking about conditional probabilities.

Conditional Probability - Motivating Example

- Conditional probability reflects the probability of some event that depends on whether some other event has already occurred.
- For example, a certain therapy is often beneficial to patients with a certain condition
  - however serious side effects (toxicity) may occur but are difficult to diagnose.
  - a test can be used to detect such side effects, by measuring the concentration of the therapy in the blood.
- To determine how good the test is, a study is conducted on 135 patients and both the diagnosis of toxicity (D^+ or D^-) and the blood test result (T^+ or T^-) of the patients are collected.
  - in this case, $\Omega = \{D^+T^+, D^+T^-, D^-T^+, D^-T^-\}$

Conditional Probability - Motivating Example

- The results of this study are presented in the contingency table:

	D^+	D^-	Total
T^+	25	14	39
T^-	18	78	96
Total	43	92	135

- Thus, for example,
  - 25 of the 135 patients have both toxicity present and a positive test, i.e. $P(D^+ \cap T^+) = {}^{25}\!/_{135}$.
  - 39 of the 135 patients have a positive blood test, without considering toxicity status, i.e. $P(T^+) = {}^{39}\!/_{135}$
  - 43 of the 135 patients have toxicity, without considering the result of the test, i.e. $P(D^+) = {}^{43}\!/_{135}$

Conditional Probability - Motivating Example

- Now, if a doctor knows the blood test is positive, what is the probability of the patient having toxicity given this knowledge?
- This is a job for conditional probability!
- **Conditional probability** considers the probability of an event, but restricted to the sample space defined by the occurrence of another event.
  - in this case, restricted only to those patients who have a positive test, T^+
  - it makes no sense to use the information from patients with a negative test, because *the doctor already knows the test was positive*.

Conditional Probability - Motivating Example

- We can now restrict our attention only to the first row of the table, namely

	D^+	D^-	Total
T^+	25	14	39

- We are now interested in the probability of toxicity (D^+) out of those patients with a positive test (T^+).
- We can denote this by $P(D^+ \mid T^+)$, which is called the conditional probability of D^+ given T^+.
- Based on our restricted table, we have 25 patients with toxicity out of the 39 with a positive test, therefore

$$P(D^+ \mid T^+) = {}^{25}\!/_{39}$$

Conditional Probability

- Note that we could also calculate this by considering the whole original contingency table:

	D^+	D^-	Total
T^+	25	14	39
T^-	18	78	96
Total	43	92	135

- We had $^{25}/_{135}$ patients with both a positive test and toxicity
- We also had $^{39}/_{135}$ patients with a positive test.
- Therefore

$$P(D^+ \mid T^+) = \frac{^{25}/_{135}}{^{39}/_{135}}$$

- This motivates the mathematical definition of conditional probability.

Conditional Probability

Definition

Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is defined to be

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

- Important: make sure the event behind the "$|$" is the same as the one in the denominator!
- Technically, because we are restricting the sample space, we are defining a new measure.
- Therefore, in order for it to be a valid probability measure, it <u>must satisfy the axioms</u> in week 1, and this can be shown.

Multiplication Law

- In some situations $P(A \cap B)$ cannot be found as easily as $P(B)$ and $P(A \mid B)$.
- The following result can be used in this case:

Multiplication Law

Let A and B be events and assume $P(B) \neq 0$. Then

$$P(A \cap B) = P(A \mid B)P(B).$$

- The multiplication law is just a reorganization of the definition of conditional probability
- It can also be handy in finding probabilities of intersections.

Example: Another Urn Problem

An urn contains three red balls and one blue ball. Two balls are selected without replacement. What is the probability that they are both red?

- Of course, this can be approached using counting methods.
- However we can use the multiplication law quite easily by the way in which we define our events:
  - Let R_1 denote the event that a red ball is drawn on the first trial and R_2 denote the event that a red ball is drawn on the second trial.
  - The multiplication law tells us that

$$P(R_1 \cap R_2) = P(R_1)P(R_2 \mid R_1).$$

  - It's clear that $P(R_1) = 3/4$, and that, if we have drawn a red ball first, $P(R_2 \mid R_1) = 2/3$. Therefore $P(R_1 \cap R_2) = 2/4$.

Exercise - Give it a try!

Suppose that if it is cloudy (B), the probability that it is raining (A) is 0.3, and that the probability that it is cloudy is $P(B) = 0.2$. What is the probability that it is cloudy and raining?

Outline

Law of Total Probability

- We saw that the multiplication law can be used to find probability of intersections.
- Similarly, conditional probability can be used to find the probability of a single event.
- It essentially considers the conditional probability across all possible conditioning events in Ω.

Law of Total Probability (LTP)

Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$ (union of mutually disjoint events), with $P(B_i) > 0$ for all i. Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i).$$

Law of Total Probability (LTP)

Proof

$\square$

Example: Back to our urn

What is the probability that a red ball is selected on the second draw (R_2), when we are sampling without replacement?

- The intuition here is a little muddy...
- But, using LTP, we'll see that this is not true:
  - To get a red ball on the second draw, only two things can happen on the first draw: a blue ball (B_1) or a red one (R_1)
  - these are the two conditioning events.

$$P(R_2) = P(R_1)P(R_2 \mid R_1) + P(B_1)P(R_2 \mid B_1)$$

Example: Back to our urn

What is the probability that a red ball is selected on the second draw (R_2), when we are sampling without replacement?

- The intuition here is a little muddy...
- But, using LTP, we'll see that this is not true:
  - To get a red ball on the second draw, only two things can happen on the first draw: a blue ball (B_1) or a red one (R_1)
  - these are the two conditioning events.

$$
\begin{aligned}
P(R_2) &= P(R_1)P(R_2 \mid R_1) + P(B_1)P(R_2 \mid B_1) \\
&= \frac{3}{4} \times \frac{2}{3} + \frac{1}{4} \times 1
\end{aligned}
$$

Exercise - Give it a try!

Suppose occupations are grouped into upper (U), middle (M), and lower (L) levels. U_1 denotes the event that a father's occupation is upper-level, while U_2 denotes that a son's occupation is upper-level, etc. We have the following occupation statistics:

	U_2	M_2	L_2
U_1	0.45	0.48	0.07
M_1	0.05	0.70	0.25
L_1	0.01	0.50	0.49

The table is read in the following way: If a father is in U, the probability that his son is in U is 0.45. Thus the table gives conditional probabilities, for example $P(U_2 \mid U_1) = 0.45$.

Exercise (cont.)

	U_2	M_2	L_2
U_1	0.45	0.48	0.07
M_1	0.05	0.70	0.25
L_1	0.01	0.50	0.49

Suppose that, of the father's generation, 10% are in U, 40% are in M, and 50% in L. What is the probability that a son from the next generation is in U?

Outline

Motivation - Occupation Example

- The previous exercise asked about probabilities of the form $P(U_2), P(M_2)$ and $P(L_2)$.
- However, the study that collected this data can also be used to answer a slightly different question.
- If we know the occupation level of the son, can we determine the chance that the father had a specific occupation level?
- This can be seen as 'reversing' the conditional probabilities given in the table, e.g. $P(U_1 \mid U_2)$ instead of $P(U_2 \mid U_1)$.
- **Bayes' rule** can let us compute such a quantity.

Bayes' Rule

Bayes' Rule

Let A and $B_1, B_2, \ldots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

Proof

$\square$

Example: Coronary Artery Disease (CAD)

A procedure is used to determine presence of calcification of coronary arteries and thus used to diagnose CAD. Let $T_i, i = 0, 1, 2, 3$ denote the number of arteries in which the test detected calcification. Let D^+ or D^- denote whether the disease is present or absent. We have the following data

i	$P(T_i \mid D^+)$	$P(T_i \mid D^-))$
0	0.42	0.96
1	0.24	0.02
2	0.20	0.02
3	0.15	0.00

For male patients between the ages of 30 and 39, it is known that $P(D^+) \approx 0.05$. What is the probability that such a patient has CAD given that the test identified 1 artery with calcification?

Example (cont.)

- First, identify the probability that we are interested in here: $P(D^+ \mid T_1)$
- So we can restrict our attention to one row of the table:

i	$P(T_i \mid D^+)$	$P(T_i \mid D^-))$
1	0.24	0.02

  and we are told that for our patient, $P(D^+) \approx 0.05$.
- It's helpful to realize that the denominator in Bayes' rule is just LTP, so we can split up our calculation:
  1. LTP:
  $$\begin{aligned} P(T_1) &= P(T_1 \mid D^+)P(D^+) + P(T_1 \mid D^-)P(D^-) \\ &= 0.24 \times 0.05 + 0.02 \times 0.95 = 0.031 \end{aligned}$$
  2. Now Bayes': $P(D^+ \mid T_1) = \frac{P(T_1|D^+)P(D^+))}{P(T_1)} = \frac{0.24 \times 0.05}{0.031} = 0.387$

Exercise - Give it a try!

Polygraph tests: Let $+$ denote the event that the polygraph is positive (lying); let T denote the subject is truthful, L that they are lying. According to studies, $P(+ \mid L) = 0.88$ and $P(- \mid T) = 0.86$. Suppose that $P(T) = 0.99$. A subject produces a positive response on the polygraph. What is the probability that the polygraph is wrong and the subject is telling the truth?

Bayes' Rule: Discussion

- Bayes' rule is the foundation for a different philosophical approach to statistics, called the **Bayesian** approach.
- This course only covers the **Frequentist** approach to statistics.
- The biggest difference between the two approaches is that the Bayesian one let's a statistician incorporate previous beliefs into an analysis, whereas frequentist approach does not.
- One reason this makes intuitive sense is that, if you are analyzing data in a study, it seems reasonable to include information that has been gained from prior studies, rather than 'starting from scratch'.
- But we will continue with the frequentist approach.

Outline

Independence of Events

- Intuitively, we say that two events, A and B, are **independent** if knowing one happened gave us no information about whether the other happened.
- Thinking about this with conditional probability, if the events have no effect on each other, then

$$P(A \mid B) = P(A) \text{ and } P(B \mid A) = P(B).$$

- By considering the definition of conditional probability, we can easily deduce the definition of independent events:

Definition: Independence

A and B are said to be independent events if

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

We can then write that $A \perp\!\!\!\perp B$.

Example: Cards in a Deck

A card is selected randomly from a deck. Let A denote the event that it is an ace, and D the event that it is a diamond. Are these events independent?

- The key here is to realize that knowing if a card is an ace does not give us any information about the suit of that card.
- This can be checked formally using the definition by:
  - $P(A) = {}^{4}/_{52}$ and $P(D) = {}^{1}/_{4}$ by basic counting.
  - We can also find $P(A \cap D)$, the probability that the card is both an ace and a diamond, as ${}^{1}/_{52}$.
  - Now apply the definition:

$$P(A)P(D) = \frac{4}{52} \times \frac{1}{4} = \frac{1}{52} \equiv P(A \cap D)$$

so these events are independent.

Exercise - Give it a try!

A player throws darts at a target. On each trial, independently of the other trials, he hits the bull's-eye with probability 0.05. How many times should he throw so that his probability of hitting the bull's-eye **at least once** is 0.5?

Pairwise Independence

- Things become more complicated when we are dealing with more than 2 events.
- Now we need to distinguish between two types of independent events
- **Pairwise independence** occurs when any two events (pairs) are independent.
- It is important to note that just because two events are independent of each other, it does not necessarily guarantee that all of the events are **mutually independent**.

Example: A fair coin toss

A fair coin is tossed twice. Let A denote the event of heads on the first toss, B the event of heads on the second toss, and C the event that exactly one head is thrown.

- It should be fairly obvious that A and B are independent.
- We can also easily see that $P(A) = P(B) = P(C) = 0.5$.
- We can check that $A \perp\!\!\!\perp C$ (similarily $B \perp\!\!\!\perp C$) by noting that

$$P(C \mid A) = \frac{P(A \cap C)}{P(A)} = \frac{P(\{ht\})}{P(\{hh, ht\})} = 0.5$$

- But

$$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$$

so they are pairwise but not mutually independent.

Mutual Independence

- Mutual independence is thus harder to achieve
- Generally, events are **mutually independent** when each event is independent of every combination of events.
- Formally, this means that a collection of events $A_1, A_2, \ldots, A_n$ are mutually independent if for any subcollection, $A_i, \ldots, A_{i_m}$, $P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$.

Example: HIV transmission

Suppose that transmission in 500 acts of intercourse are mutually independent events and that the probability of transmission in any one act is $1/500$. Under this model, what is the probability of infection?

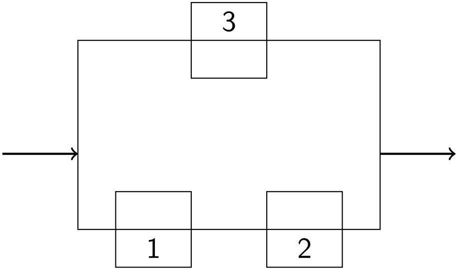
- Infection occurs if at least one act results in virus transmission.
- Since the events are mutually independent, we can solve the problem under the complement (which is easier).
- We define the sequence of events $C_1, \ldots, C_{500}$ denoting no transmission for each of the 500 encounters.
- Now we may compute $P(\text{infection}) = 1 - P(\text{no infection})$ by

$$P(\text{no infection}) = P(C_1 \cap C_2 \cap \cdots \cap C_{500}) = \left(1 - \frac{1}{500}\right)^{500} = 0.37$$

and finally $P(\text{infection}) = 1 - 0.37 = 0.63$

Example: Circuit

Consider a circuit with three relays. Let A_i denote the event that the ith relay works, and assume that $P(A_i) = p$ and that the relays are mutually independent. If F denotes the event that current flows through the circuit, then find the probability of F.



- We need to first define F in terms of the circuits:

$$F = A_3 \cup (A_1 \cap A_2)$$

- Using the addition law and independence, we get:

$$P(F) = (A_3) + P(A_1 \cap A_2) + (A_1 \cap A_2 \cap A_3) = p + p^2 - p^3$$

Exercise - Give it a try!

Suppose that a system consists of components connected in series, so the system fails if any one component fails. If there are n mutually independent components and each fails with probability p, what is the probability that the system will fail?