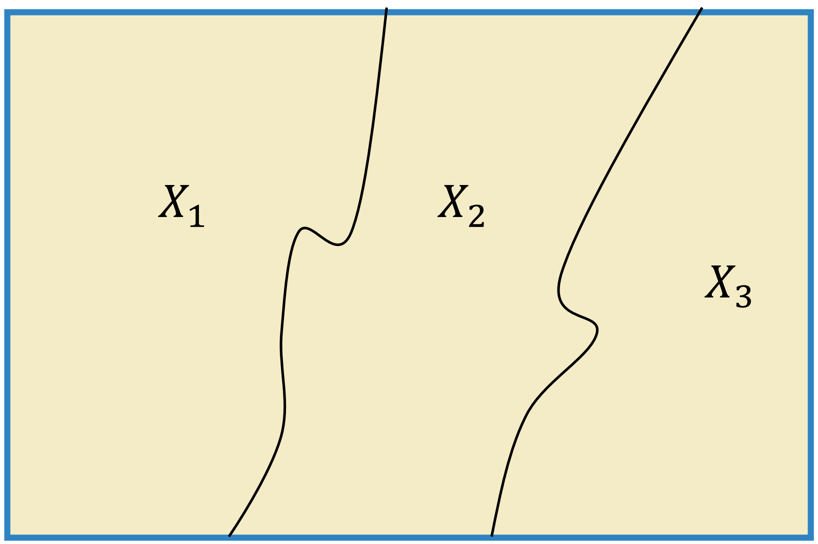
ITCS 121 Statistics

Lecture 5

More Probability and Random Variables

Partitions (reminder)

- A **partition** of a set X is a division of X into two or more mutually exclusive parts $X = X_1 \cup X_2 \ldots \cup X_n$.
- E.g.



- I.e. events $X_1, \ldots, X_n$ partition the sample space if they are **mutually exclusive** and **exhaustive**.

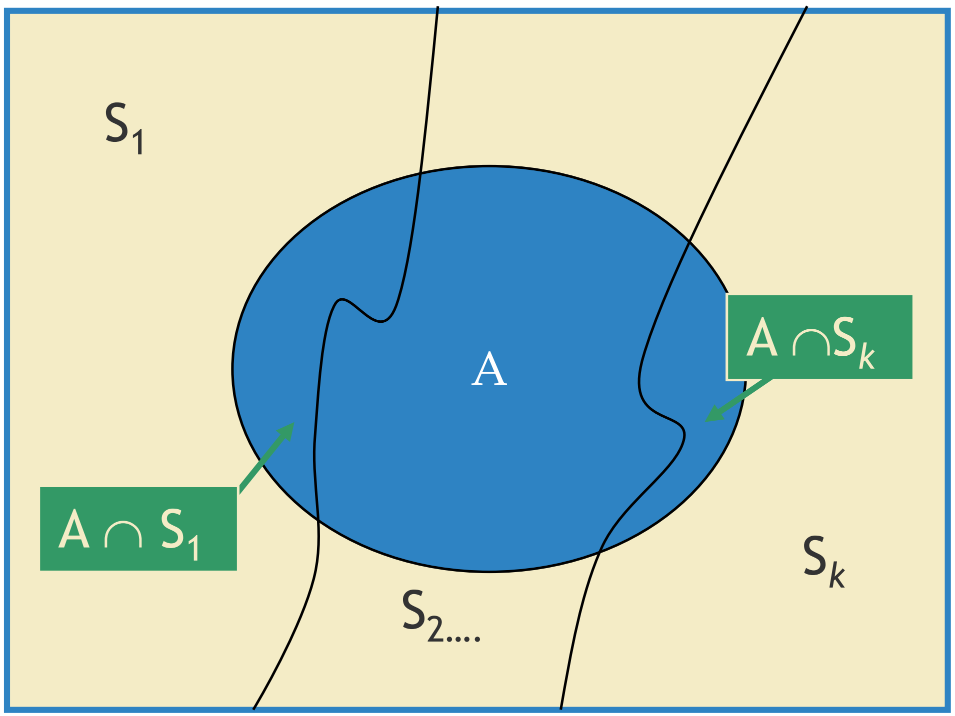
The Law of Total Probability

- Suppose events $S_1, \dots, S_n$ partition the sample space.
- We know that we must have $P(S_1 \cup \cdots \cup S_n) = P(S_1) + \cdots + P(S_n) = 1.$
- Let A be another event.
- Then

$$P(A) = P(A \cap S_1) + \cdots + P(A \cap S_n)$$

$$= P(A|S_1)P(S_1) + \cdots + P(A|S_n)P(S_n)$$

Visualizing the Total Probability Law



$$P(A) = P(A \cap S_1) + \cdots + P(A \cap S_n)$$

Class activity 1

A survey of people in a given region showed that 20% were smokers. The probability of death due to lung cancer, given that a person smoked, was roughly 10 times the probability of death due to lung cancer, given that a person did not smoke. If the probability of death due to lung cancer in the region is 0.006, what is the probability of death due to lung cancer given that a person is a smoker?

Bayes' Theorem

- Let A and B be events.
- Then $P(A \cap B) = P(A|B)P(B)$.
- And $P(A \cap B) = P(B|A)P(A)$.
- (Remember that $P(A|B)$ and $P(B|A)$ are not usually the same!)
- So, $P(A|B)P(B) = P(B|A)P(A)$.
- Rearranging this gives Bayes' Formula:

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

The importance of Bayes' formula

- Bayes' formula is very important to Bayesian statisticians, because it gives them their system for updating beliefs when they get new information.
- E.g. My prior belief is that statement A is true with probability $P(A)$. If I observe event B has occurred then I can update my beliefs about the truth of A using $P(A|B)=P(B|A)P(A)/P(B)$ (assuming I have values for the other probabilities).
- Other people can use the formula without accepting the Bayesian philosophy, because it is useful for calculating conditional probabilities.

Example – heart disease

- A test for a heart disease has 95% accuracy.
- 1% of people have this heart disease.
- If you test positive for the disease, what is the probability you have the disease?

$$D = have\ disease \qquad T = test\ positive \qquad P(D|T) = ?$$

$$P(D) = 0.01 \qquad P(T|D) = 0.95$$

$$\begin{aligned} P(T) &= P(T|D)P(D) + P(T|D^c)P(D^c) \\ &= (0.95)(0.01) + (0.05)(0.99) \\ &= 0.059 \end{aligned}$$

Law of Total Probability

$$P(D|T) = \frac{(0.95)(0.01)}{0.059} = 0.16$$

Even with positive test result, still a low probability

Class activity 2

The CIA is afraid that an evil alien species is taking over the world by kidnapping politicians and replacing them with clones. It is calculated that 1 in every 1000 politicians has been cloned in this way.

The CIA has developed a special ‘clone detector’, which is known to be accurate 90% of the time. A CIA agent scans Donald Trump and the detector says he is an alien clone. What is the probability that he really is a clone?

The Monty Hall Problem

- Imagine a game show where there are three closed doors.
- Behind one door is the prize, behind the other two doors are goats.

- The contestant chooses a door. If the prize is behind the door she chooses then she wins.

The twist

- But, there’s a twist.
- After the contestant has chosen, but before the door has been opened to reveal either the prize or a goat…
- … the host will always open one of the doors not picked and reveal a goat.
- The host then asks the contestant if she wants to change her choice to being the other unopened door.

Monty Hall example

- For example, we have doors A, B and C. The prize is behind door B, and behind doors A and C are goats.

A B C

- The contestant chooses door A.
- The host opens a different door with a goat behind it. In this case it must be door C.
- The host asks the contestant “do you want to stay with door A, or switch to door B?”.

Should you switch?

- If you are the contestant, should you switch?
- In our example, the contestant should switch, because if she switches to door B she will win.
- But we only know that because we have information she does not have.
- The question is, if you are the contestant and you don’t have extra information, is it better to switch or to keep your original choice, or does it not make a difference?

Flawed intuition

- Most people seem to think it shouldn't make a difference.
- Maybe because, since there are two doors left, and one has the goat and the other has the prize, the chance is 50/50 either way.
- But most people are wrong about this.
- You are actually twice as likely to win if you switch.
- Why???

Applying Bayes

- I want to find $P(b.d.I.c|s.m.a.g)$.
- According to Bayes' theorem

$$P(b.d.I.c|s.m.a.g) = \frac{P(s.m.a.g|b.d.I.c)P(b.d.I.c)}{P(s.m.a.g)}$$

- Now, $P(b.d.I.c)$ is my initial estimate of the probability that the prize is behind the door I picked, so we should have $P(b.d.I.c) = \frac{1}{3}$.
- Similarly, $P(s.m.a.g)$ is the probability that the host shows me a goat. So $P(s.m.a.g) = 1$, as the host will always opens a door with a goat behind it.
- Finally, $P(s.m.a.g|b.d.I.c) = 1$ too, for the same reason. So…

$$P(b.d.I.c|s.m.a.g) = \frac{1 \times \frac{1}{3}}{1} = \frac{1}{3}$$

Drawing a conclusion

- We proved that the probability that the prize is behind the door I chose is 1/3.
- Since the prize is definitely not behind the door opened by the host, there's only one other door.
- The prize must be behind the other door with probability $1 - \frac{1}{3} = \frac{2}{3}$.
- So I double my chance of winning if I switch.
- Fact.

Using intuition

- This conclusion is very disturbing to some people.
- But, if we think about it a little bit more, it’s what we should expect.
- Think about it like this.
- If you chose correctly with your first guess and then switch, you will lose.
- But if you chose incorrectly with your first guess and then switch, you will win.
- Since you are twice as likely to choose incorrectly with your first guess than to choose correctly, switching doubles your chances of winning.

Another Monty

- Think about the following variation of the Monty Hall problem.
- The basic setup with the doors and the goats is the same.
- But now, the host opens one of the doors at random (i.e. he’s not guaranteed to show you a goat).
- Suppose that by chance the door the host opens has a goat behind it.
- Should you switch now?

Bayes again

- It turns out it doesn't make a difference this time!
- Again I want to find $P(b.s.I.c|s.m.a.g)$.
- $P(b.d.I.c) = \frac{1}{3}$ like before.
- But $P(s.m.a.g)$ is not 1 anymore.
- To find $P(s.m.a.g)$ now we need the law of total probability

$$\begin{aligned} P(s.m.a.g) = {} & P(s.m.a.g|b.d.I.c)P(b.d.I.c) \\ & +P(s.m.a.g|not\ b.d.I.c)P(not\ b.d.I.c) \end{aligned}$$

- So $P(s.m.a.g) = \left(1 \times \frac{1}{3}\right) + \left(\frac{1}{2} \times \frac{2}{3}\right) = \frac{2}{3}$.
- And $P(s.m.a.g|b.d.I.c) = 1$.
- Now Bayes gives

$$P(b.d.I.c|s.m.a.g) = \frac{1 \times \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

Confusion

- This is now very confusing.
- If the host is guaranteed to show me a goat, then switching my choice is the best strategy.
- But if the host shows me a goat by accident, then it doesn't make a difference!
- What's going on?

Making sense of it

- I think about it is like this.
- If you made a mistake with your first choice, then switching is good.
- If the host is guaranteed to show you a goat, then him showing you a goat doesn’t give you any extra information about your original choice.
- So your belief that you made the correct choice first time should stay at 1/3, and so switching is better.
- But, if the host shows you a goat by chance, then you *do* get more information.
- Why? Because the host is more likely to show you a goat if you made the correct choice first time!
- So the host showing you a goat by chance gives you evidence that your original choice was correct.
- And if your original choice was correct you shouldn’t switch.

Class activity 3

- Suppose that, in a particular city, airport A handles 50% of all airline traffic, and airports B and C handle 30% and 20%, respectively.

- The detection rates for weapons at the three airports are 0.9, 0.8, and 0.85, respectively. Suppose a person tries to take a weapon through security at one of the airports and that the weapon is detected.

- *What is the probability that the person is using airport A?*

- You should assume the person has no information about the detection rates at each airport, so they picked their airport randomly according to airport traffic rates.

Random Variables

- A **random variable** X is a variable whose value is considered to be random.
- What does random mean here?
- This means the values that X takes are controlled by some probability law.
- We will explain exactly how this works later.
- Random variables can be qualitative or quantitative, but we are mainly interested in quantitative ones here.

Examples of random variables

- The outcome of tossing a coin.
- The outcome of rolling a dice.
- The number we get if we roll 3 dice and take the mean.
- How long it takes a particle to decay.
- How many car accidents in Thailand in a year.
- How many people out of a random sample of 100 people like cheese.
- Etc.

Expected values

- The **expected value** $E(X)$ of a random variable X is a kind of `average', like the mean of a population.

- It tells us what result, on average, we should expect from the variable.

- Exactly what this means is not obvious, but it will become clearer when we make some explicit definitions later.

- Not all random variables have expected values, but the most important ones do.

Linearity of expectation

- If X and Y are random variables, and if a and b are numbers, then $aX + bY$ is also a random variable.
- It turns out that $E(aX + bY) = aE(X) + bE(Y)$.
- In other words, taking expected values is *linear*.
- To see why this is true we would need to explicitly define the expected value of a random variable, which we haven't done yet.

Variance of random variables

- If a random variable X has an expected value, then we can also think about how spread out, on average, the values it takes are around $E(X)$.
- This gives us the **variance** of X, $Var(X)$, just like we get the variance of a population.
- We define $Var(X) = E\left(\left(X - E(X)\right)^2\right).$ ← If this exists!
- This is similar to the formula for the variance of a population.
- Using linearity of expectation, we can also derive an alternative formula

$$Var(X) = E(X^2) - E(X)^2$$

Nonlinearity of variance

- Consider again the composite random variable $aX + bY$.
- In general it is **not true** that $Var(aX + bY) = aVar(X) + bVar(Y)$.
- But, if X and Y are **independent**, i.e. if the values they take do not affect each other at all, then

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$$

Discrete random variables

- Remember a variable is discrete if it only takes whole number values (e.g. …,-2,-1,0,1,2,3,…).
- Discrete random variables are defined by their **probability mass function** (pmf).
- A pmf p is a function from the integers to the real numbers (i.e. all the possible decimals).
- If p is the pmf for X, and if z is an integer, then $p(z)$ tells us the probability that X has value z.
- I.e. $p(z) = P(X = z)$.
- We must have $\sum_{z \in \mathbb{Z}} p(z) = 1$.
  - Why? Because X must take some value, and it can only have one value at a time, so the sum of all the probabilities across all the integers must be 1.

Example – rolling a dice

- The outcome of rolling a dice is a discrete random variable.
- The possible outcomes are 1,2,3,4,5,6.
- Each outcome has the same probability, $\frac{1}{6}$.
- I.e. $P(X=1)=P(X=2)=\cdots=P(X=6)=\frac{1}{6}$.
- If z is not in $\{1,2,3,4,5,6\}$ then $P(X=z)=0$.
- So the pmf for this variable is what we call the **uniform distribution** over $\{1,2,3,4,5,6\}$.
- Check: $\sum_{z\in\mathbb{Z}} P(X=z)=P(X=1)+\cdots+P(X=6)$

$$=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}$$

$$=1$$

Expectation for discrete random variables

- If X is a discrete random variable, then we define its expected value by

$$E(X) = \sum_{z \in \mathbb{Z}} zP(X = z)$$

- E.g. If X is the outcome of rolling a dice, then

$$E(X) = \left(1 \times \frac{1}{6}\right) + \left(2 \times \frac{1}{6}\right) + \left(3 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{6}\right) + \left(5 \times \frac{1}{6}\right) + \left(6 \times \frac{1}{6}\right) = 3.5$$

Variance for discrete random variables

- We can also calculate the variance of a discrete random variable.

- Remember $Var(X) = E\left(\left(X - E(X)\right)^2\right)$.

- $\left(X - E(X)\right)^2$ is not technically a discrete random variable, because it may not take whole number values, but we can still calculate $E\left(\left(X - E(X)\right)^2\right)$ using similar logic, which gives us the formula

$$Var(X) = \sum_{z \in \mathbb{Z}} P(X = z)(z - E(X))^2$$

Variance of rolling a dice

- We calculated $E(X) = 3.5$.

- So

$$Var(X) = \sum_{z=1,2,\ldots,6} \frac{1}{6}(z-3.5)^2 = \frac{35}{12}$$

Strategies in gambling games

- In the last class we imagined your friend asked you to play a game.
- She wanted to roll a 6-sided dice. If the result was 6 you would win, and she said she would pay you \$40.
- If the result was odd she would win, and she wanted you to pay her \$10.
- I said this would be a good game for you to play, so long as it wouldn't be a big problem if you lost. Why?
- We can use expected values for this.
- Let X be the change in your money after playing the game (yes, this is a discrete random variable).

$$E(X) = \left(40 \times \frac{1}{6}\right) + \left(-10 \times \frac{3}{6}\right) = \frac{40 - 30}{6} = \frac{10}{6} = 1.67$$

- $E(X)$ is positive, so on average you expect to win money.

Class activity 4

Suppose your friend offers to play another gambling game with you.

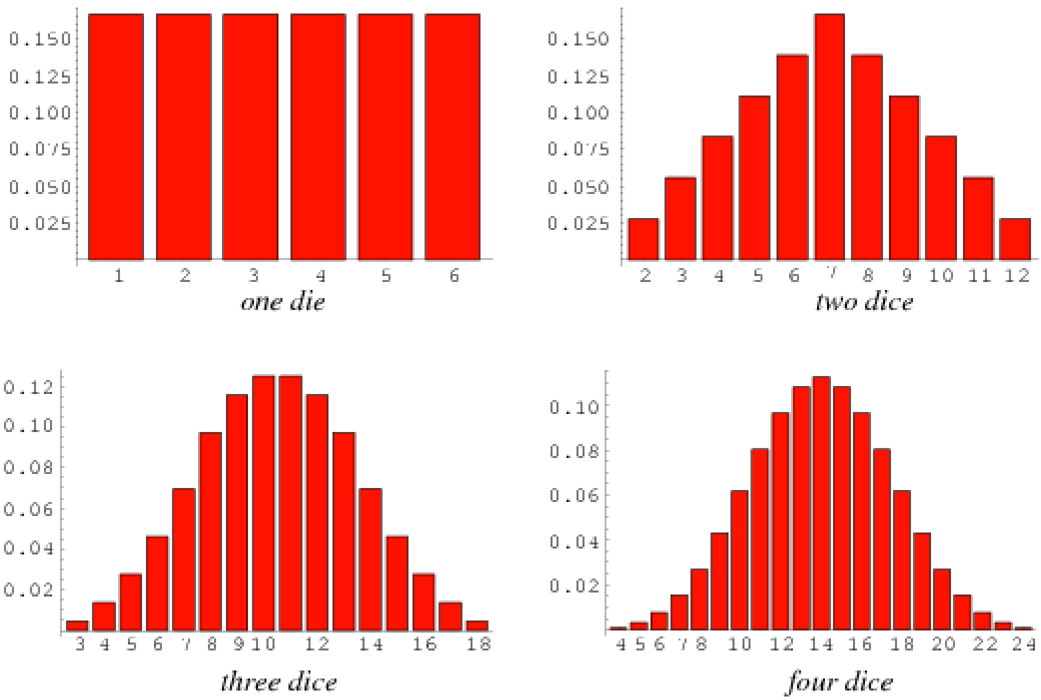
In this game you roll two 6-sided dice. If either but not both of the rolls result in 6 (i.e. if you roll exactly one 6), then she will pay you $12.

But, if both rolls show the same number, you have to pay her $15.

Is this a good bet for you?

Probability distributions

- We can draw probability distributions.
- For discrete random variables, which have pmf functions, these look like histograms.
- The height of the bar for an integer z tells us the probability that the random variable takes value z.



Probability distributions for sums of dice rolls. Image taken from Wolfram mathworld.

Class activity 5

How many times should a coin be tossed to obtain a probability equal to or greater than 0.9 of observing at least one head?