# Introduction to Probability

Myra O'Regan

October 13, 2015

# What is probability?

- The study of probability is the study of randomness
- Started in 1600 with Blaise Pascal and Pierre de Fermat
- How they could use mathematics to describe random events
- To increase their chances of winning in games of chance

# What is probability?

- Probability is a number between 0 and 1
- Probability of 1 - event is certain to happen
- Probability of 0 - event is impossible, definitely will not happen
- The higher the number the more likely it is to happen
- Can be calculated either mathematically or by carrying out an experiment

# Chance Experiment

- A chance experiment is any activity or situation in which there is uncertainty about which of two or more possible outcomes will result
- Flipping a coin
- Rolling a dice
- Drawing a card from a deck of cards

# Sample Space

- A sample space is the set of all possible outcomes of a chance experiment
- Flipping a coin S={Heads, Tails} 2 outcomes
- Rolling a dice S={1,2,3,4,5,6 } 6 outcomes
- Draw a card S={ 1♣, 2♣....$K$♠} 52 outcomes

# An event

- An event is a subset of outcomes from a sample space
- Heads when we flip a coin
- 5 on dice
- Hearts on a card draw
- How do we calculate probability of an event E?
- If all outcomes in a sample space are equally likely
- $p(E) = \frac{\text{number of outcomes in E}}{\text{number of outcomes in S}}$

## Some simple examples

- What is the probability of tails on a coin flip
- E={tails } S= { heads, tails }
- $p = \frac{1}{2}$
- What is the probability of a spade on a card draw?
- E= $\{A\heartsuit,\ 2\heartsuit,\ 3\heartsuit ... K\heartsuit\}$ 13 outcomes
- S= {deck of cards} 52 outcomes
- p (heart) $= \frac{13}{52}$

# Laws of probability

- We calculated the p(E)
- Define p(E') - probability of the event not happening
- $p(E) + p(E') = 1$
- $p(E) = 1 - p(E')$
- Sometime it easier to calculate the p(E')

# Two events A and B

- p(A and B) sometimes written $A \cap B$
- Event that both A and B occur
- p(A or B) sometimes written as $A \cup B$
- Event that A or B or both occur
- Two events are mutually exclusive if both cannot occur together

# Two events A and B - example

- Example 30 marbles in a bag 10 red, 15 blue and 5 green
- p(red marble)$= \frac{10}{30}$
- p(green marble) $= \frac{5}{30}$
- p(red or green marble ) $\frac{10}{30} + \frac{5}{30}$
- Two mutually exclusive events; no ball can be red or green simultaneously

# Another example

- Select a card at random from a deck
- What is the probability that it is a diamond (A) or a 7(B)?
- These are not mutually exclusive events
- You can be both a diamond and a 7
- p(A or B)= p(A) +p(B)- p(A and B)
- p(diamond or a 7) $= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$

# Conditional probability

- Another bag of marbles with 6 red and 12 white
- Two marbles are chosen without replacement
- A: event first marble is red
- B: event second marble is red
- $p(A) = \frac{6}{18}$
- p(B) depends on whether A happens or not

## Two events A and B

- If A happens there are only 5 red balls left
- p(B) in that case is $\frac{5}{17}$
- Conditional probability of an event B given A $=$ $p(B \mid A)$
- We have a new sample space which is the event A
- For B to occur some of its members must also be in A
- $p(B \mid A) = \frac{\text{number of outcomes in A and B}}{\text{number of outcomes in A}}$
- Divide numerator and denominator by total number of outcomes
- We get $p(B \mid A) = \frac{p(\text{A and B})}{p(A)}$

# Probability of two events

- Rewrite above formula
- p(A and B) $= p(B \mid A)$ p(A) or $p(A \mid B)p(B)$

# Another useful finding

- To calculate p(B): probability that the second ball is red
- p(B) = p(A and B) + p(A' and B)
- Event(A and B) and Event(A' and B) are two mutually exclusive events
- p(B)= $p(B \mid A)$p(A) + $p(B \mid A')$p(A')
- Makes it easy to use information given

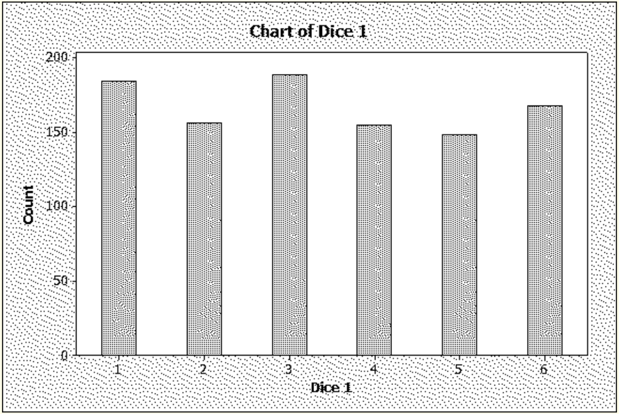
# Independent events

- A and B are independent events if the occurrence of A does not effect or is not affected by another event B
- In this case the $p(A \mid B)$= p(A) and $p(B \mid A)$=p(B)
- p(A and B) = p(A)p(B)

# Independent events

- Toss two dice and calculate the probability of getting each of the totals
- Going to look at this two ways mathematically and by doing experiment
- Two fair 6 sided dice
- Toss both of them and compute sum
- Write out possible outcomes - easy there is not that many

# Results for one dice



# Independent events

Table : sample space for sum of two dice

| | Dice 1 | | | | | |
|---|---|---|---|---|---|---|
| Dice 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

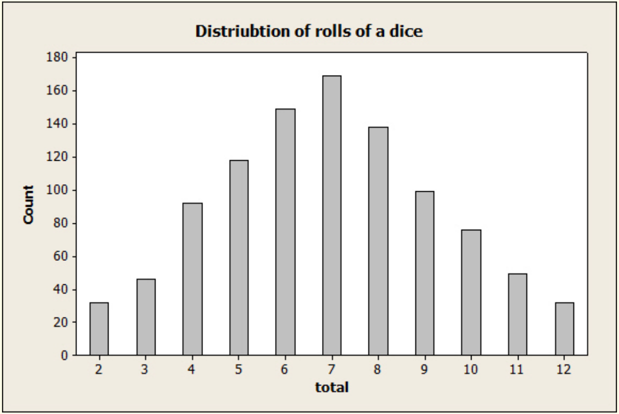
# Some comments

- 11 outcomes
- Not all equally likely
- p(Dice 1=1 and Dice 2 =1) $= \frac{1}{6} * \frac{1}{6}$
- p(sum=7)= $\frac{7}{36}$

# Experiment Results

| Sum | Count | Percent |
|---|---|---|
| 2 | 32 | 3.2 |
| 3 | 46 | 4.6 |
| 4 | 92 | 9.2 |
| 5 | 118 | 11.8 |
| 6 | 149 | 14.9 |
| 7 | 169 | 16.9 |
| 8 | 138 | 13.8 |
| 9 | 99 | 9.9 |
| 10 | 76 | 7.6 |
| 11 | 49 | 4.9 |
| 12 | 32 | 3.2 |

# Results for Total



# Prosecutor's fallacy

- A crime has been committed and a criminal has left some evidence such as blood
- One in every 1000 people has the matching type
- Person caught with matching type
- Prosecutor states that probability that an innocent person has the matching blood is 0.001
- Therefore the probability that they are innocent is 0.001
- Is this right?

# Have to look at this in a different way

- Have to consider the number of people
- 100,000 people in area
- Expect 100 to have matching blood type
- Probability that Fred is innocent is 0.99
- There is a mix up in conditional probabilities here

# And some more....

- Event A Fred is innocent
- Event B Fred has matching blood type
- We really want to know $p(A \mid B)$
- What we actually know is $p(B \mid A)$
- $p(A \mid B) = 0.99$
- $p(B \mid A) = 0.001$

# Bayes Theorem

- Rev Bayes (1701-1761) English statistician philosopher and Presbyterian minister
- A firm buys computer chips from two companies X and Z. They have in store 400 chips from Company X and 200 from company Z - 600 chips in all.
- We know that rate of defects for Company X is 10% and company Z is 20%
- We sample one at random
- What is the probability it comes from Company Z?
- $p(Z) = \frac{200}{600} = .33$

# Bayes Theorem

- What if we were now told that the chip was defective?
- Does this change the probability?
- We are given extra information
- How do we use this to update probability?
- We use Bayes Theorem

# Bayes Theorem

- Two events A and B
- $p(A \mid B) = \frac{p(B|A)p(A)}{p(B)}$
- $p(A \mid B) = \frac{p(B|A)p(A)}{p(B|A)p(A)+p(B|A')p(A')}$

# Bayes Theorem

- Event A : chip comes from Z; Event A' :chip comes from X
- Event B: chip is defective; Event B': chip is not defective
- p(A)=.33; p(A')=0.67;
- $p(B \mid A)$=0.20;p$(B \mid A') = 0.10$;
- Bayes theorem tells us
- $p(A \mid B) = \frac{0.20*0.33}{0.20*0.33+0.10*0.67} = 0.496$

# Using some Mathematics for Monty Hall

https://www.khanacademy.org/math/precalculus/prob_comb/dependent_events_precalc/v/monty-hall-problem

- Let the 3 doors be A,B,C
- Let your selection be A
- p(car)behind door x $=\frac{1}{3}$
- p(MH opens door B $|$ prize were behind A) $= \frac{1}{2}$
- p(MH opens door B $|$ prize were behind B) $= 0$
- p(MH opens door B $|$ prize were behind C) $=1$

# Some Notation

- p(A) = prob. prize behind A
- p(B) =prob. prize behind B
- p(C) = prob. prize behind C

# Using some Mathematics for Monty Hall

- p(MH opens B) $=$

  p(A)*p(M opens B $\mid A$)$+$p(B)*p(M opens B $\mid B$)$+$ p(C)*p(M opens B $\mid C$)
- $= \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}$
- Using Bayes theorem
- p(A $\mid$ MH opens B )$= \frac{p(A)*p(MH\ \ opens\ \ B|A)}{p(MH\ \ opens\ \ B)} = \frac{1/6}{1/2} = \frac{1}{3}$

- p(C $\mid$ MH opens B )$= \frac{p(C)*p(MH\ \ opens\ \ B|C)}{p(MH\ \ opens\ \ B)} = \frac{1/3}{1/2} = \frac{2}{3}$

# What does this mean?

- If you chose A and Monty opens door B to reveal a goat
- Probability of car behind C is $\frac{2}{3}$
- Probability of car behind A is $\frac{1}{3}$
- So you should always switch

# Another example

- There is a disease BigBad that has a prevalence of 1% in population
- There is a special test to detect it
- Test has a 98% accuracy for positive results
- 97% accuracy for negative results
- You took the test and it comes back positive
- What are the chances that you have the disease?

# What info do we have in probability terms

- Let D be event that you are sick
- p(D) = 0.01; p(D')=0.99
- Let T= event test is right ; T' test is wrong
- $p(T \mid D) = 0.98$; and $p(T \mid D') = 0.97$
- We want to know $p(D \mid T)$?

# Using Bayes theorem we have

- $p(D \mid T) = \frac{p(T|D)p(D)}{p(T|D)p(D)+p(T|D')p(D')}$
- $p(D \mid T) = \frac{0.01*0.98}{0.01*0.98+0.99*0.97} = 0.18$

# So what!!!

- We have looked at two ways of calculating probabilities
- Using formulae or by carrying out experiments or simulations on computer
- Relative frequency distribution - uniform distribution