Probability

J. Alexander Branham

Fall 2016

What is probability?

- Frequency with which an event occurs

Probability

- Frequency with which an event occurs
- Typically:

$$Pr(A) = P(A) = \pi(A) = \frac{\text{Number of ways an event can occur}}{\text{Total number of possible outcomes}}$$

Probability

- Frequency with which an event occurs
- Typically:

$$Pr(A) = P(A) = \pi(A) = \frac{\text{Number of ways an event can occur}}{\text{Total number of possible outcomes}}$$

- There are other ways of thinking about probability, but we'll stick with this one

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
  - where F is the event space

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
  - where F is the event space
  - Probabilities must be non-negative

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
  - where F is the event space
  - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
  - where F is the event space
  - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$
  - Where Ω is the sample space

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
    - where F is the event space
    - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$
    - Where Ω is the sample space
    - Something has to happen

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
    - where F is the event space
    - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$
    - Where Ω is the sample space
    - Something has to happen
    - Probabilities sum/integrate to 1

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
  - where F is the event space
  - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$
  - Where Ω is the sample space
  - Something has to happen
  - Probabilities sum/integrate to 1
- Third: $Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} Pr(E_i)$

Kolmogorov's Axioms

- First: $Pr(E) \in \mathbb{R}, Pr(E) \geq 0 \qquad \forall E \in F$
    - where F is the event space
    - Probabilities must be non-negative
- Second: $Pr(\Omega) = 1$
    - Where Ω is the sample space
    - Something has to happen
    - Probabilities sum/integrate to 1
- Third: $Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} Pr(E_i)$
    - The probability of disjoint (mutually exclusive) sets is equal to their sums

Probability Distributions

Discrete

Discrete probabilities

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"
- We can think of a 3 as a "success"

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"
- We can think of a 3 as a "success"
- And $Y \sim Bernoulli\left(\frac{1}{6}\right)$

Discrete probabilities

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"
- We can think of a 3 as a "success"
- And $Y \sim Bernoulli\left(\frac{1}{6}\right)$
- Fair coins are $\sim Bernoulli(.5)$ for example

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"
- We can think of a 3 as a "success"
- And $Y \sim Bernoulli\left(\frac{1}{6}\right)$
- Fair coins are $\sim Bernoulli(.5)$ for example
- More generally $Bernoulli(\pi)$

- What's the probability that we'll roll a 3 on one die roll:

$$Pr(y = 3) = \frac{1}{6}$$

- If one roll of the die is an "experiment"
- We can think of a 3 as a "success"
- And $Y \sim Bernoulli\left(\frac{1}{6}\right)$
- Fair coins are $\sim Bernoulli(.5)$ for example
- More generally $Bernoulli(\pi)$
    - π represents the probability of success

Binomial distributions

- Before we ran the experiment just once

- Before we ran the experiment just once
- What happens if we run it multiple times?

$$Pr(y_1 = 3, y_2 = 3) = \frac{1}{6} * \frac{1}{6}$$

Binomial distributions

- Before we ran the experiment just once
- What happens if we run it multiple times?

$$Pr(y_1 = 3, y_2 = 3) = \frac{1}{6} * \frac{1}{6}$$

- Now $Y \sim Binomial\left(2, \frac{1}{.6}\right)$

- Before we ran the experiment just once
- What happens if we run it multiple times?

$$Pr(y_1 = 3, y_2 = 3) = \frac{1}{6} * \frac{1}{6}$$

- Now $Y \sim Binomial\left(2, \frac{1}{.6}\right)$
  - Generally, $Y \sim Binomial(n, p)$

- Before we ran the experiment just once
- What happens if we run it multiple times?

$$Pr(y_1 = 3, y_2 = 3) = \frac{1}{6} * \frac{1}{6}$$

- Now $Y \sim Binomial\left(2, \frac{1}{.6}\right)$
  - Generally, $Y \sim Binomial(n, p)$
  - n = number of trials, p = probability of success

- Before we ran the experiment just once
- What happens if we run it multiple times?

$$Pr(y_1 = 3, y_2 = 3) = \frac{1}{6} * \frac{1}{6}$$

- Now $Y \sim Binomial\left(2, \frac{1}{.6}\right)$
  - Generally, $Y \sim Binomial(n, p)$
  - n = number of trials, p = probability of success
- PMF:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

You try!

1. What’s the probability of getting dealt the ace of spades?

You try!

1. What's the probability of getting dealt the ace of spades?
2. What's the probability of rolling six five's in a row?

You try (answers)

1. $$\frac{1}{52}$$

1. $$\frac{1}{52}$$

2. $$\binom{6}{6}\frac{1}{6}^{6}\left(1-\frac{1}{6}\right)^{6-6}\approx 0.00002$$

Continuous distributions

- What happens when our outcome is continous

- What happens when our outcome is continous
- Much harder to think about...

- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes

The basics

- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes
- Makes the denominator of our fraction difficult to work with

The basics

- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes
- Makes the denominator of our fraction difficult to work with
- What's the probability that you make 15,293.19/year?

The basics

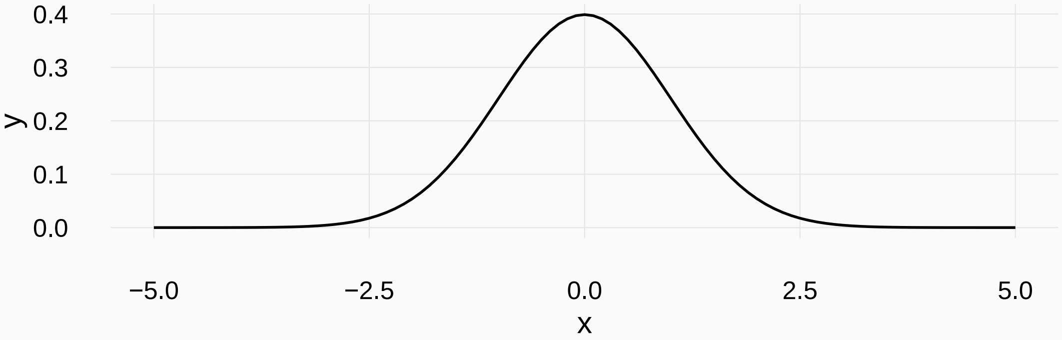
- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes
- Makes the denominator of our fraction difficult to work with
- What's the probability that you make 15,293.19/year?
- Probability of the whole space must equal 1

The basics

- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes
- Makes the denominator of our fraction difficult to work with
- What's the probability that you make 15,293.19/year?
- Probability of the whole space must equal 1
- Even if all events equally likely, $\frac{1}{\infty} = 0$

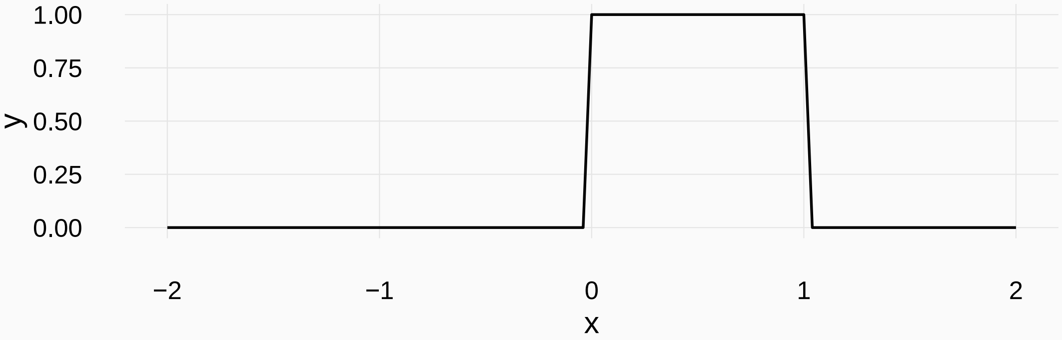
- What happens when our outcome is continous
- Much harder to think about...
- There are infinity possible outcomes
- Makes the denominator of our fraction difficult to work with
- What's the probability that you make 15,293.19/year?
- Probability of the whole space must equal 1
- Even if all events equally likely, $\frac{1}{\infty} = 0$
    - Kinda...

```r
ggplot(data.frame(x = c(-5, 5)), aes(x)) +
  stat_function(fun = dnorm)
```



Continuous distributions - Uniform

```r
ggplot(data.frame(x = c(-2, 2)), aes(x)) +
  stat_function(fun = dunif)
```



- $Pr(y = c) = 0$

Probabilies and continuous distributions

- $Pr(y = c) = 0$
- $Pr(0 < y < .5)$

$$= \int_0^{.5} f(y)dy$$

- $Pr(y = c) = 0$
- $Pr(0 < y < .5)$

$$= \int_0^{.5} f(y)dy$$

- For uniform from previous slide, $Pr(0 < y < .5) = 0.5$

CDF

- CDF = Cumulative Distribution Function

CDF

- CDF = Cumulative Distribution Function
- $F_X(x) = Pr(X \leq x)$

Discrete

- $Y \sim Binom(10, .5)$

- $Y \sim Binom(10, .5)$
- What's the probability that $y \leq 5$?

Discrete

- $Y \sim Binom(10, .5)$
- What's the probability that $y \leq 5$?
- $$\sum_{i=1}^{5} Pr(Y = y_i)$$

$Pr(y \leq 5)$

$$\binom{10}{1}.5^1(1-.5)^{10-1}+$$
$$\binom{10}{2}.5^2(1-.5)^{10-2}+$$
$$\binom{10}{3}.5^3(1-.5)^{10-3}+$$
$$\binom{10}{4}.5^4(1-.5)^{10-4}+$$
$$\binom{10}{5}.5^5(1-.5)^{10-5}$$

Continous

- We can't sum probabilities for continuous distributions (remember the 0 problem)

Continous

- We can't sum probabilities for continuous distributions (remember the 0 problem)
- Solution: integration

Continous

- We can't sum probabilities for continuous distributions (remember the 0 problem)
- Solution: integration
- $F_Y(y) = \int_{-\infty}^{y} f(y)dy$

Marginal probability

- The *Pr* of an event occurring:

Marginal probability

- The *Pr* of an event occurring:
  - $Pr(A)$

Marginal probability

- The *Pr* of an event occurring:
  - $Pr(A)$
- Unconditional probability

Marginal probability

- The *Pr* of an event occurring:
  - $Pr(A)$
- Unconditional probability
- Probability of drawing a red card from a standard deck:

$$Pr(red) = \frac{1}{2}$$

Joint probability

- The probability of both A **and** B

Joint probability

- The probability of both A and B
  - $Pr(A, B)$

Joint probability

- The probability of both A and B
    - $Pr(A, B)$
- This is the intersection of the two sets

- The probability of both A and B
    - $Pr(A, B)$
- This is the intersection of the two sets
- $Pr(A \cap B)$

- The probability of both A and B
  - $Pr(A, B)$
- This is the intersection of the two sets
- $Pr(A \cap B)$
  - Probability of drawing a red card and a 4:

Joint probability

- The probability of both A **and** B
  - $Pr(A, B)$
- This is the intersection of the two sets
- $Pr(A \cap B)$
  - Probability of drawing a red card and a 4:
  - $Pr(red, 4) = \frac{2}{52} = \frac{1}{26}$

Conditional probability

Conditional probability

- The probability of some event happening **given** some other event having occurred

Conditional probability

- The probability of some event happening **given** some other event having occurred
  - $Pr(A|B)$

- The probability of some event happening **given** some other event having occurred
    - $Pr(A|B)$
- You’ve drawn a red card. What’s the probability it’s a four?

- The probability of some event happening **given** some other event having occurred
  - $Pr(A|B)$
- You've drawn a red card. What's the probability it's a four?
  - $Pr(4|red) = \frac{2}{26} = \frac{1}{13}$

$$Pr(A|B) = \frac{Pr(A, B)}{Pr(B)}$$

The conditional probability is equal to the joint probability over the probability of the condition

Bayes' Law (Theorem)

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} = \frac{Pr(B|A)Pr(A)}{Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A)}$$

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers
- Person stopped and fails breathalyzer test

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers
- Person stopped and fails breathalyzer test
    - Test gives positive 98 percent of the time when person is drunk

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers
- Person stopped and fails breathalyzer test
    - Test gives positive 98 percent of the time when person is drunk
    - Also gives positive 5 percent of the time when the person is actually sober

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers
- Person stopped and fails breathalyzer test
    - Test gives positive 98 percent of the time when person is drunk
    - Also gives positive 5 percent of the time when the person is actually sober
    - We think that 1 percent of people going through the checkpoint are drunk

Bayes Example

- Police set up roadblock to randomly screen for potentially drunk drivers
- Person stopped and fails breathalyzer test
    - Test gives positive 98 percent of the time when person is drunk
    - Also gives positive 5 percent of the time when the person is actually sober
    - We think that 1 percent of people going through the checkpoint are drunk
- What is the probability that the man was actually drunk?

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
    - $Pr(drunk) = 0.01$

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
    - $Pr(drunk) = 0.01$
- Test accuracy:

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
  - $Pr(drunk) = 0.01$
- Test accuracy:
  - $Pr(positive|drunk) = 0.98$

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
  - $Pr(drunk) = 0.01$
- Test accuracy:
  - $Pr(positive|drunk) = 0.98$
  - $Pr(positive|\neg drunk) = 0.05$

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
  - $Pr(drunk) = 0.01$
- Test accuracy:
  - $Pr(positive|drunk) = 0.98$
  - $Pr(positive|\neg drunk) = 0.05$
- Want to know:

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
    - $Pr(drunk) = 0.01$
- Test accuracy:
    - $Pr(positive|drunk) = 0.98$
    - $Pr(positive|\neg drunk) = 0.05$
- Want to know:
    - $Pr(drunk|positive) = \frac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$

DWI Example

- Since he's randomly stopped, we think there is a 1 percent change he's drunk
    - $Pr(drunk) = 0.01$
- Test accuracy:
    - $Pr(positive|drunk) = 0.98$
    - $Pr(positive|\neg drunk) = 0.05$
- Want to know:
    - $Pr(drunk|positive) = \dfrac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$
- Everything known, except $Pr(positive)$

$Pr(positive)$

$$
\begin{aligned}
Pr(positive) &= Pr(positive|drunk)Pr(drunk) \\
&\quad + Pr(positive|\neg drunk)Pr(\neg drunk) \\
&= .98(.01) + .05(.99) \\
&\approx .0593
\end{aligned}
$$

$Pr(drunk|positive)$

- $Pr(drunk|positive) = \dfrac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$

$Pr(drunk|positive)$

- $Pr(drunk|positive) = \frac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$
- $Pr(drunk|positive) = \frac{0.01(.98)}{.0593}$

$Pr(drunk|positive)$

- $Pr(drunk|positive) = \dfrac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$
- $Pr(drunk|positive) = \dfrac{0.01(.98)}{.0593}$
- .165

$Pr(drunk|positive)$

- $Pr(drunk|positive) = \frac{Pr(drunk)Pr(positive|drunk)}{Pr(positive)}$
- $Pr(drunk|positive) = \frac{0.01(.98)}{.0593}$
- .165
- So there is a 16.5 percent chance that the man is drunk given that he tested positive

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats
- Once contestants chose a door, Monty would open one of the doors they *didn't* choose that had a goat behind it

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats
- Once contestants chose a door, Monty would open one of the doors they *didn't* choose that had a goat behind it
- The contestant could then either stick with their original choice or switch to the unopened door

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats
- Once contestants chose a door, Monty would open one of the doors they *didn't* choose that had a goat behind it
- The contestant could then either stick with their original choice or switch to the unopened door
- Does switching doors increase your probability of winning?

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats
- Once contestants chose a door, Monty would open one of the doors they *didn't* choose that had a goat behind it
- The contestant could then either stick with their original choice or switch to the unopened door
- Does switching doors increase your probability of winning?
    - YES!

Example, take two: Monty Hall

- *Let's Make a Deal* was a game show where contestants were asked to choose one of three doors (A, B, and C). Behind one is a car and behind the other two are goats
- Once contestants chose a door, Monty would open one of the doors they *didn't* choose that had a goat behind it
- The contestant could then either stick with their original choice or switch to the unopened door
- Does switching doors increase your probability of winning?
    - YES!
    - Nearly 10,000 people wrote in refusing to believe this (including 1,000 PhDs!)

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A
- WLOG assume Monty opens door B

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A
- WLOG assume Monty opens door B
- Then: $Pr(B_{Monty}|A) = \frac{1}{2}$ and $Pr(B_{Monty}|B = 0)$ and $Pr(B_{Monty}|C) = 1$

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A
- WLOG assume Monty opens door B
- Then: $Pr(B_{Monty}|A) = \frac{1}{2}$ and $Pr(B_{Monty}|B = 0)$ and $Pr(B_{Monty}|C) = 1$

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A
- WLOG assume Monty opens door B
- Then: $Pr(B_{Monty}|A) = \frac{1}{2}$ and $Pr(B_{Monty}|B = 0)$ and $Pr(B_{Monty}|C) = 1$

$$Pr(B_{Monty}, A) = Pr(B_{Monty}|A)Pr(A) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$
$$Pr(B_{Monty}, B) = Pr(B_{Monty}|B)Pr(B) = 0 * \frac{1}{3} = 0$$
$$Pr(B_{Monty}, C) = Pr(B_{Monty}|C)Pr(C) = 1 * \frac{1}{3} = \frac{1}{3}$$

Bayes's solution to Monty Hall

- $Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}$ (prior belief about where the car is)
- WLOG assume contestant picks door A
- WLOG assume Monty opens door B
- Then: $Pr(B_{Monty}|A) = \frac{1}{2}$ and $Pr(B_{Monty}|B = 0)$ and $Pr(B_{Monty}|C) = 1$

$$Pr(B_{Monty}, A) = Pr(B_{Monty}|A)Pr(A) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$

$$Pr(B_{Monty}, B) = Pr(B_{Monty}|B)Pr(B) = 0 * \frac{1}{3} = 0$$

$$Pr(B_{Monty}, C) = Pr(B_{Monty}|C)Pr(C) = 1 * \frac{1}{3} = \frac{1}{3}$$

- Note that $Pr(B_{Monty}) = Pr(B_{Monty}, A) + Pr(B_{Monty}, B) + Pr(B_{Monty}, C) = \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}$

$Pr(A|B_{Monty})$ AND $Pr(C|B_{Monty})$

- $Pr(A|B_{Monty}) = \frac{Pr(A)Pr(B_{Monty}|A)}{Pr(B_{Monty})} = \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}$

$Pr(A|B_{Monty})$ AND $Pr(C|B_{Monty})$

- $Pr(A|B_{Monty}) = \dfrac{Pr(A)Pr(B_{Monty}|A)}{Pr(B_{Monty})} = \dfrac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{2}} = \dfrac{1}{3}$
- $Pr(C|B_{Monty}) = \dfrac{Pr(C)Pr(B_{Monty}|C)}{Pr(B_{Monty})} = \dfrac{\frac{1}{3} * 1}{\frac{1}{2}} = \dfrac{2}{3}$

$Pr(A|B_{Monty})$ AND $Pr(C|B_{Monty})$

- $Pr(A|B_{Monty}) = \frac{Pr(A)Pr(B_{Monty}|A)}{Pr(B_{Monty})} = \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}$
- $Pr(C|B_{Monty}) = \frac{Pr(C)Pr(B_{Monty}|C)}{Pr(B_{Monty})} = \frac{\frac{1}{3} * 1}{\frac{1}{2}} = \frac{2}{3}$
- So switch!