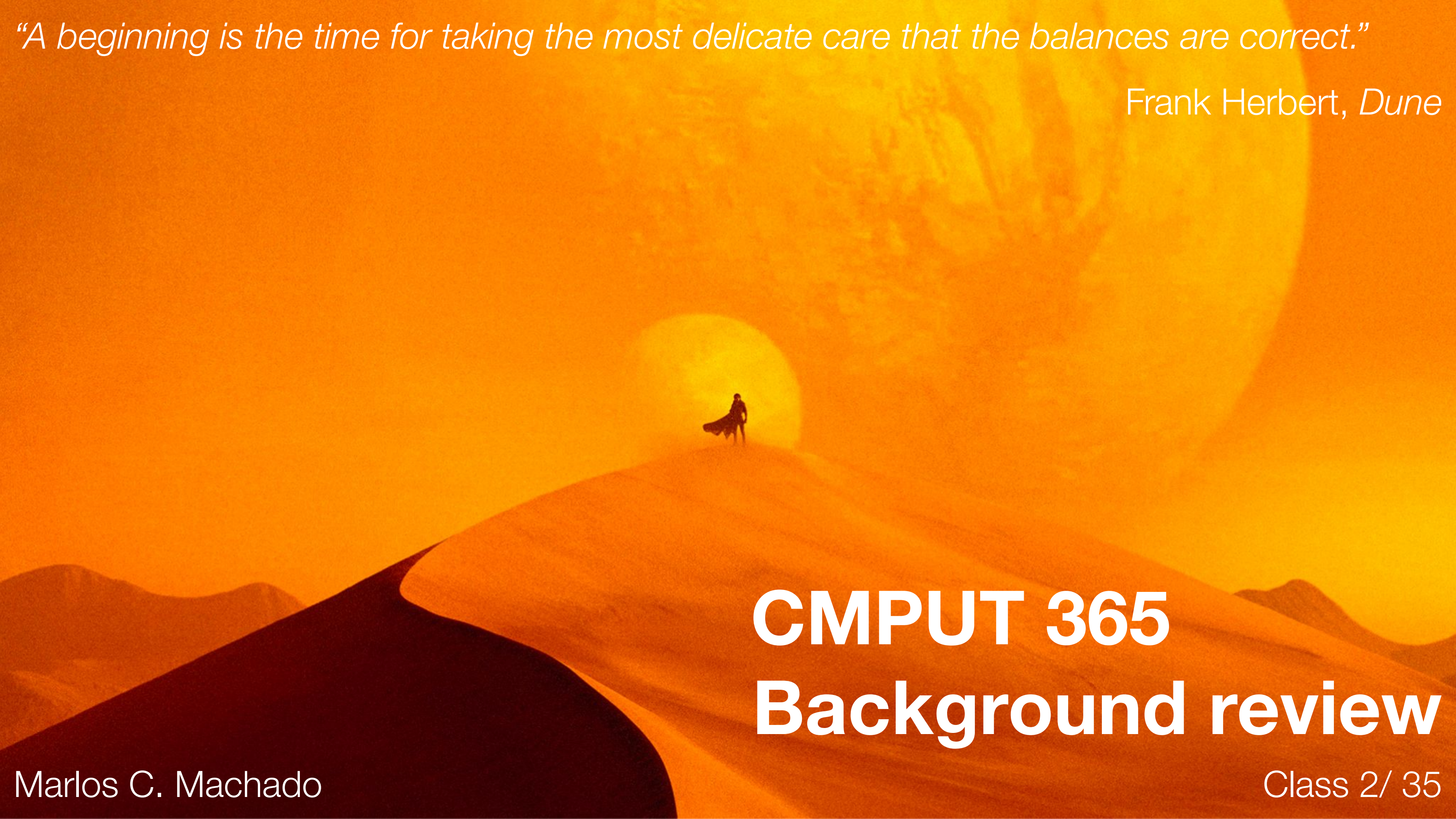


“A beginning is the time for taking the most delicate care that the balances are correct.”

Frank Herbert, *Dune*

CMPUT 365
Background review

Marlos C. Machado

Class 2/ 35

Notification

An audio recording of this class may be made for the purpose of facilitating a student’s approved academic accommodation.

Office hours

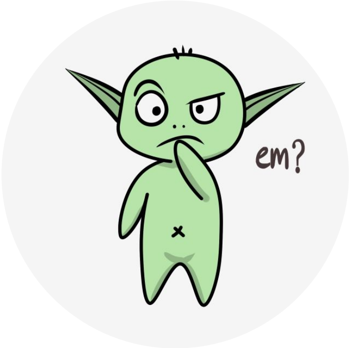
- Slack and eClass: Asynchronous
- Marlos: Thursday 15:00 - 16:45 in ATH 3-08 (Athabasca Hall, 3-08)
- Anna: Monday 12:00 - 14:00 in CAB 3-13
- Bryan: Wednesday 14:00 - 16:00 in CAB 3-13
- David: Tuesday 13:00 - 15:00 in CSC 3-50
- Gabor: Wednesday 9:15-11:15 in CAB 3-13
- Marcos: Friday 10:00 - 12:00 in CAB 3-13

Syllabus [eClass, Slack, website, Google Drive]

Plan

- Probability & statistics
- Linear algebra
- Calculus

Please, interrupt me at any time!



https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

Probability and statistics

Definitions

- **Probability** is about predicting the likelihood of future events.
- **Statistics** is about estimating a model (rule) from past events.

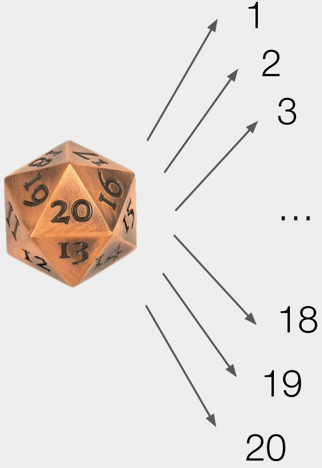
We'll need to understand probability to do statistics.

Probability – The basics

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

Example

Dungeons & Dragons!



Pr(rolling 20) = 1/20 = 5%

Pr(rolling 19 **or** 20) =
Pr(rolling 19) + **Pr**(rolling 20) =
1/20 + 1/20 = 10%

Pr(rolling 20 **and** 20) =
Pr(rolling 20) x **Pr**(rolling 20) =
1/20 x 1/20 = 1/400 = 0.25%

https://i.etsystatic.com/8800676/r/il/6e938b/3226455710/il_794xN.3226455710_64zr.jp
https://cdn.britannica.com/71/241671-050-B5227EBD/Game-of-Dungeons-and-Dragons.jp

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

Probability – Somewhat more formally

*A probability is a function that associates a number between 0 and 1 to an **event**, with this number being a measure of the likelihood of that **set of outcomes**.*

- A **set** is collection of disjoint elements.
- A **sample space** is the set of all possible outcomes of an experiment.
- An **event** is any subset of the sample space.

Example

Sample space. {1, 2, …, 20}

Event. Rolling higher than 16:
{17, 18, 19, 20}

Probability – Somewhat more formally

*A probability is a **function** that **associates** a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.*

- A **set** is collection of disjoint elements.
- A **sample space** is the set of all possible outcomes of an experiment.
- An **event** is any subset of the sample space.
- A **function**, $f: A \rightarrow B$, is a map, a rule, that maps every element of the set A to a unique element in the set B. We call A the *domain*, and B the *codomain*, or the *range*, of the function. Given x ∈ A, the element it is associated with in the set B is called its *image* under f.

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a ***measure of the likelihood*** *of that set of outcomes.*

- A probability distribution is defines how the probability is distributed among the outcomes.

Example

For an unbiased dice, each number is equally likely (i.e., uniform probability distribution). Thus, for each outcome $e \in S$, **Pr**(e) = 1/|S|.

Probability – Somewhat more formally

A probability is a function that associates a number between 0 and 1 to an event, with this number being a ***measure of the likelihood*** *of that set of outcomes.*

- A probability distribution is defines how the probability is distributed among the outcomes.
- A way of calculating the probability of a specific event is a matter of identifying the sample space (set of all possible outcomes) and the probability distribution.

Example 1

For an unbiased dice, the probability of rolling a 20 is **Pr**(rolling 20) = 1/20.

Example 2

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.

Probability – Properties

A probability is a function that associates a number between 0 and 1 to an event, with this number being a measure of the likelihood of that set of outcomes.

- Nonnegativity: $\mathbf{Pr}(A) \geq 0$.
- Normalization: $\sum_{e \in S} \mathbf{Pr}(e) = 1$.
- Additivity: $\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B)$; $A \cap B = \{\ \}$.

Example

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

- The number of all possible subsets of the sample space.
- The power set of the sample space S, denoted 2^S.

Example

$2^S = \{S, \{\}, \{1\}, \{2\}, \{3\}, \ldots, \{20\}, \{1, 2\}, \{1, 3\}, \ldots \{18, 19, 20\}, \ldots \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}, \ldots, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19\}, \ldots\}$

Number of elements in 2^S:
$2^{20} = 1{,}048{,}576$

Probability – Considering all possible events

How many distinct events are possible in a dice rolling experiment?

The number of all possible subsets of the sample space.

The power set of the sample space S, denoted 2^S.

Pr(S) = 1.

Pr({}) = 0.

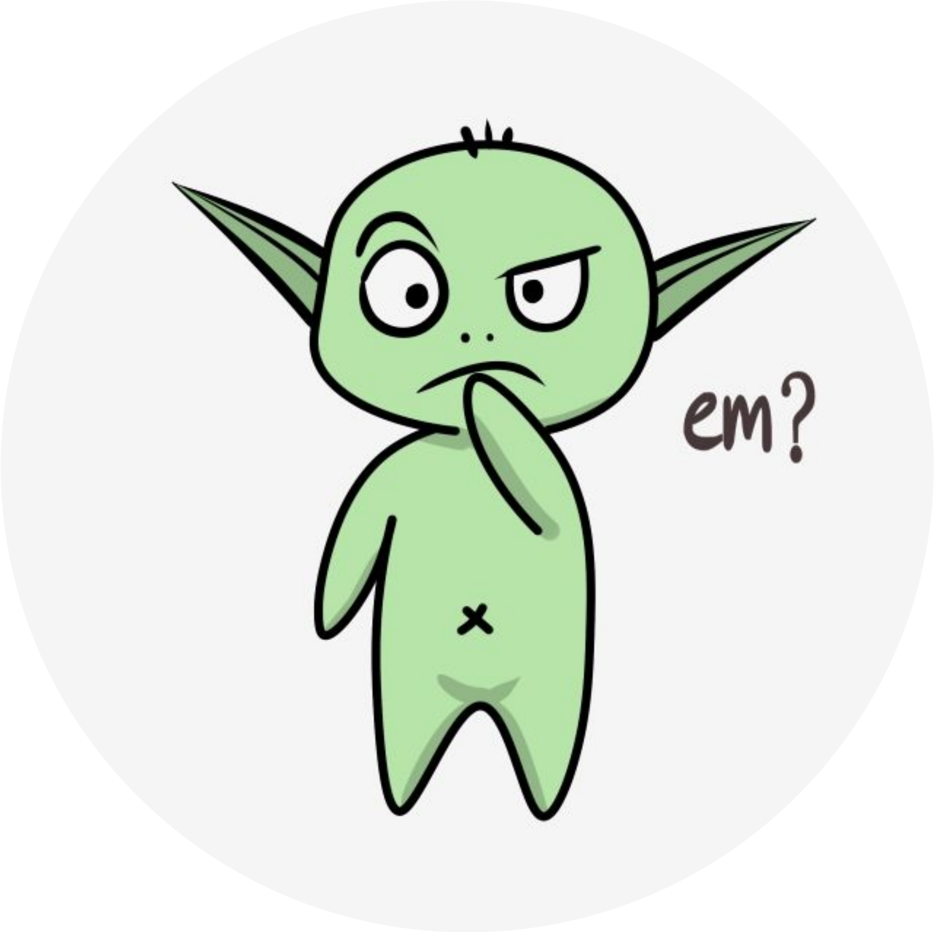
Formally, **Pr**: $2^S \rightarrow [0, 1]$.

Example 1

For an unbiased dice, the probability of rolling a 20 is **Pr**(rolling 20) = 1/20.

Example 2

For an unbiased dice, the probability of rolling higher than 18 is **Pr**(rolling 19 or 20) = 1/20 + 1/20 = 1/10.



https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

Random variables and expectations

Random variables

Random variables are ways to map outcomes of random processes to real numbers.

They are not a traditional variable, nor random 😂

Capital letter!

Example 1

$$X = \begin{cases} 1 \text{ if roll } 1 \\ 2 \text{ if roll } 2 \\ \dots \\ 19 \text{ if roll } 19 \\ 20 \text{ if roll } 20 \end{cases}$$

Example 2

$$Y = \begin{cases} 1 \text{ if heads} \\ 0 \text{ if tails} \end{cases}$$

Example 3

$$Z = \{\text{sum of 2 dice}\}$$

We can write **Pr**(X = 20) to represent **Pr**(rolling 20).

We can write **Pr**(X ≥ 19) to represent **Pr**(rolling 19 or 20).

Examples

When rolling a d20 dice, let *X* be the random variable denoting the outcome of the roll.

Pr$(1 \leq X \leq 20) = 1$

Pr$(X = 15) = 1/20$

Pr$(X = 0) = 0$

Pr$(2X = 1) = 0$

Pr$(X < 19)$?

Pr$([X = 19] \cup [X = 20] \cup [1 \leq X \leq 18]) = 1$
Pr$([X = 19] \cup [X = 20] \cup [X < 19]) = 1$
Pr$(X = 19)$ + **Pr**$(X = 20)$ + **Pr**$(X < 19) = 1$
Pr$(X < 19) = 1 -$ **Pr**$(X = 19) -$ **Pr**$(X = 20)$
Pr$(X < 19) = 1 - 1/20 - 1/20$
Pr$(X < 19) = 18/20$
Pr$(X < 19) = 9/10$

Conditional probabilities

Chain rule:
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A, B) = \mathbf{Pr}(A \mid B)\,\mathbf{Pr}(B)$

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)}\ .$$

In a classroom with 100 students, out of those 100, 20 students play tabletop RPG, and 30 students have read *The Lord of the Rings* books. There are 15 students who play tabletop RPG who have read LOTR. What is the probability that a student has read LOTR given that the student plays tabletop RPG?

Let X be the random variable denoting the probability that a student plays tabletop RPG, and let Y be the random variable denoting the probability that a student has read LOTR.

$\mathbf{Pr}(X) = 0.2$ $\mathbf{Pr}(Y) = 0.3$ $\mathbf{Pr}(X \cap Y) = 0.15$

$\mathbf{Pr}(Y \mid X) = 0.15/0.2 = 0.75$

Conditional probabilities

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)} .$$

When playing D&D, Tristan needs to roll 17 or higher on a d20 to successfully hit the troll. Tristan gets a critical hit when they roll a 20. Knowing that Tristan has successfully hit the target, what's the likelihood that Tristan got a critical hit?

Let X be the random variable denoting the number Tristan rolled on a d20, and Y a binary random variable denoting whether Tristan rolled a 20 (Y=1) or not (Y=0).

$\mathbf{Pr}(X \geq 17) = 1/5$ $\mathbf{Pr}(Y = 1 \cap X \geq 17) = 1/20$

$$\frac{\mathbf{Pr}(Y = 1 \cap X \geq 17)}{\mathbf{Pr}(X \geq 17)} = \frac{1/20}{1/5} = \frac{5}{20} = 25\%$$

Independence

Two events are independent when the likelihood of an event does not change after knowing the other event. A is independent of B if and only if

$$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(A).$$

$$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(A \cap B) \,/\, \mathbf{Pr}(B)$$
$$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A \mid B)$$
$$\mathbf{Pr}(B)$$
$$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A)\, \mathbf{Pr}(B)$$
$$\mathbf{Pr}(B \mid A) = \mathbf{Pr}(B \cap A) \,/\, \mathbf{Pr}(A)$$
$$= \mathbf{Pr}(B)\, \mathbf{Pr}(A) \,/\, \mathbf{Pr}(A)$$
$$= \mathbf{Pr}(B)$$

Example

Tristan now rolls two d20 dice. Given that they rolled a 1 on the first dice, what's the likelihood of them running a 20 on the second dice?

Let X be the random variable denoting the roll on the first dice, and Y be the equivalent for the second dice.

$\mathbf{Pr}(X = 1) = 1/20$ $\quad \mathbf{Pr}(Y = 20) = 1/20$ $\quad \mathbf{Pr}(X = 1 \cap Y = 20) = 1/400$

$\mathbf{Pr}(Y = 20 \mid X = 1) = (1/400)/(1/20) = 1/20$

Conditional probabilities with more than 2 variables

The probability of an event A given another event B is defined as:

$$\mathbf{Pr}(A \mid B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)}.$$

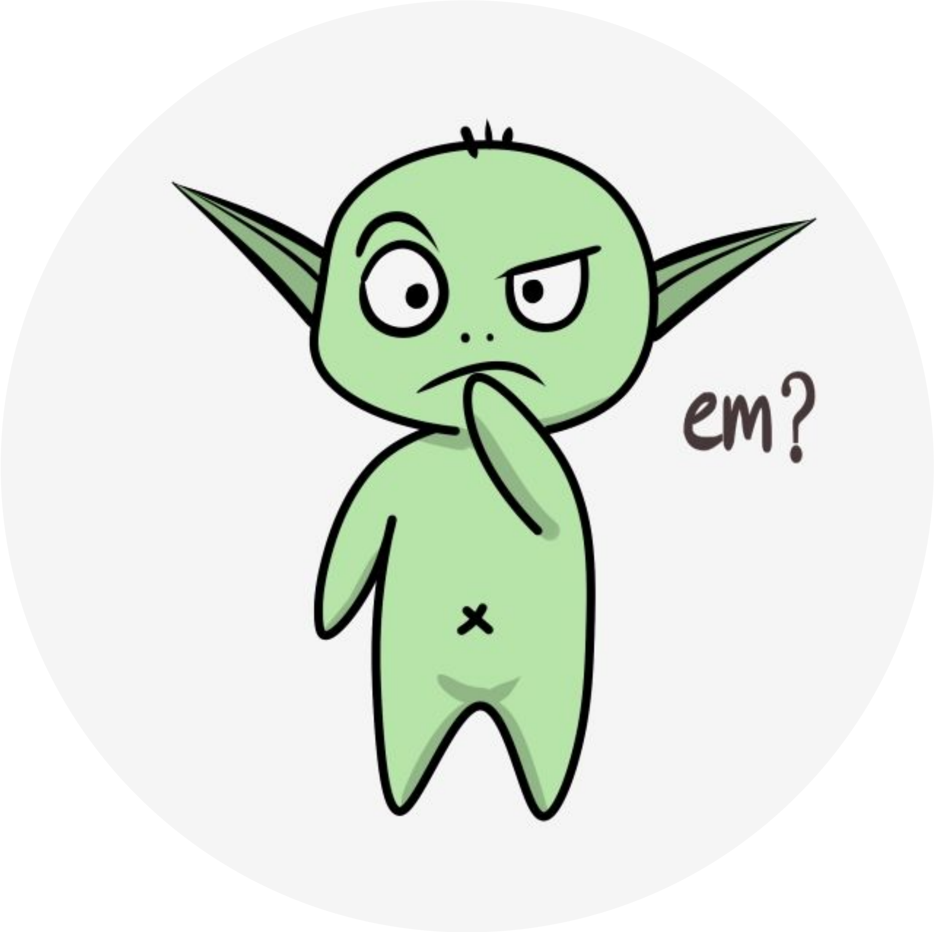
Chain rule:
$\mathbf{Pr}(A \cap B) = \mathbf{Pr}(A, B) = \mathbf{Pr}(A \mid B)\ \mathbf{Pr}(B)$

What's $\mathbf{Pr}(A, B \mid C)$?

Let $D = A \cap B$. Then, $\mathbf{Pr}(D \mid C) = \mathbf{Pr}(D, C) / \mathbf{Pr}(C)$. Thus $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A, B, C) / \mathbf{Pr}(C)$.

Now, let $E = B \cap C$, and recall, by the chain rule, that $\mathbf{Pr}(A, E) = \mathbf{Pr}(A \mid E)\ \mathbf{Pr}(E)$.
We then have $\mathbf{Pr}(A, B, C) = \mathbf{Pr}(A \mid B, C)\ \mathbf{Pr}(B, C) = \mathbf{Pr}(A \mid B, C)\ \mathbf{Pr}(B \mid C)\ \mathbf{Pr}\ (C)$.

Putting these two together: $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A \mid B, C)\ \mathbf{Pr}(B \mid C)\ \mathbf{Pr}\ (C) / \mathbf{Pr}(C)$.
Assuming $\mathbf{Pr}(C) \neq 0$, $\mathbf{Pr}(A, B \mid C) = \mathbf{Pr}(A \mid B, C)\ \mathbf{Pr}(B \mid C)$.



https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

Next class

- What **I** plan to do: Fundamentals of RL: An introduction to sequential decision-making (Bandits)
    - Finish these slides first.
    - Discuss, more in depth, things related to bandits (Chapter 2 of the textbook).

- What I recommend **YOU** to do for next class:
    - Watch videos of Week 1 of Coursera's Fundamentals of RL (Module 1): M1W1.
    - Finish the recommended reading for Coursera's M1W2.
    - Watch videos of Coursera's M1W2.
    - Start collecting (and post) questions in eClass/Slack about Chapter 2 of the textbook.
    - Submit practice quiz for Coursera's Fundamentals of RL: Sequential decision-making (M1 W2).
    - At least start programming assignment for M1 W2.