# Statistics Refresher

CMPUT 397
Fall 2019

# Reminders: Sept 9, 2019

- Schedule updated on github pages (https://marthawhite.github.io/rlcourse/schedule.html)
- Practice Quiz and Discussion Question **due tomorrow** (participation marks)
- Graded Notebook due on Friday
- **Usually**: Practice Quiz and Discussion Question **due on Sunday** and Graded Notebook **due on Friday**:
  - Exception this week to make up for class starting on Wednesday
  - Exception for Graded Quizzes and Graded Peer Review

# Practice Quiz participation marks

- 10 marks in total, with 11 weeks of Practice Quizzes
- You have to complete the quiz and submit a discussion question to get the 1%
- You get 1 mulligan (i.e., you can miss one and still get the full 10%)

# Modeling random outcomes

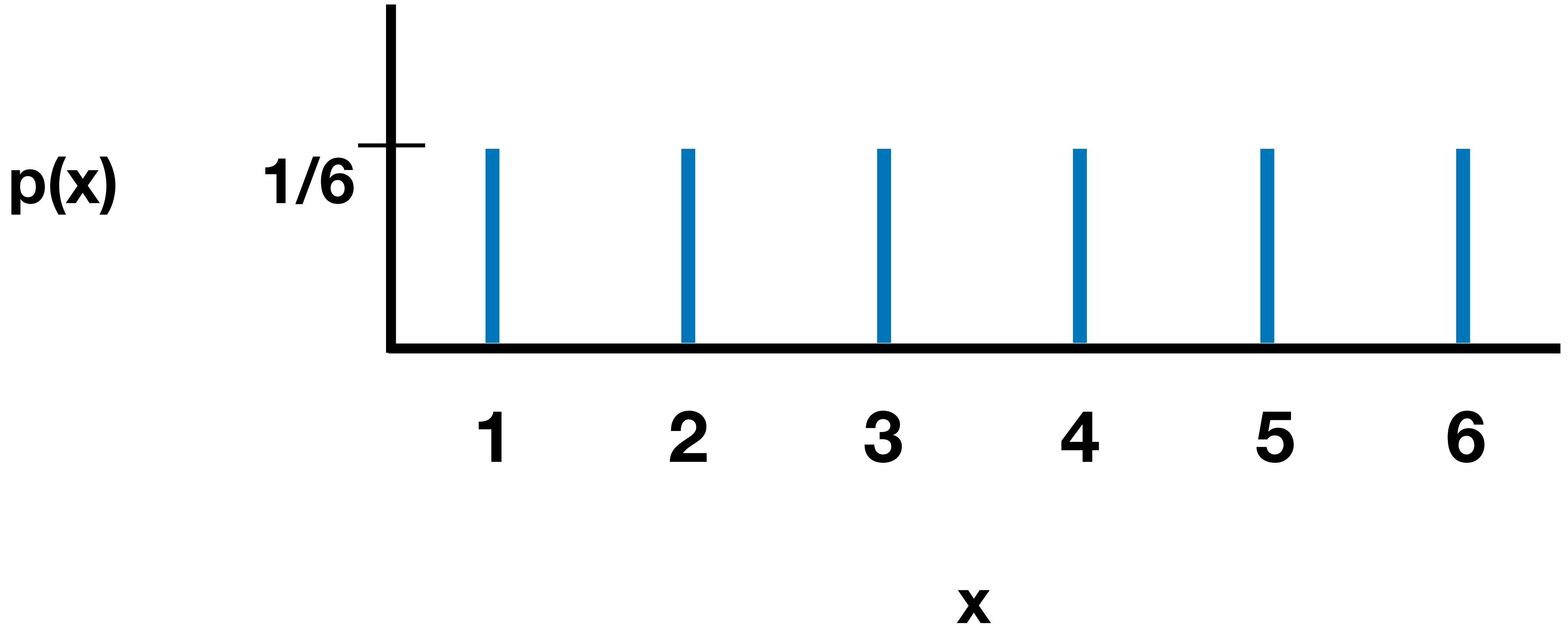
- Imagine you would like to model stochastic outcomes (or dynamics), such as the outcome of a dice role
- We need to define:
  - the possible set of outcomes (e.g., 1, 2, 3, 4, 5, 6)
  - the probability of each outcome (e.g., 1/6 for each)

# Random Variable

- (Informally) A random variable X is a variable with stochastic values
- Depends on underlying random phenomena
- Examples:
  - X = the outcome of a dice roll
  - X = the temperature tomorrow

# Probability Mass Function

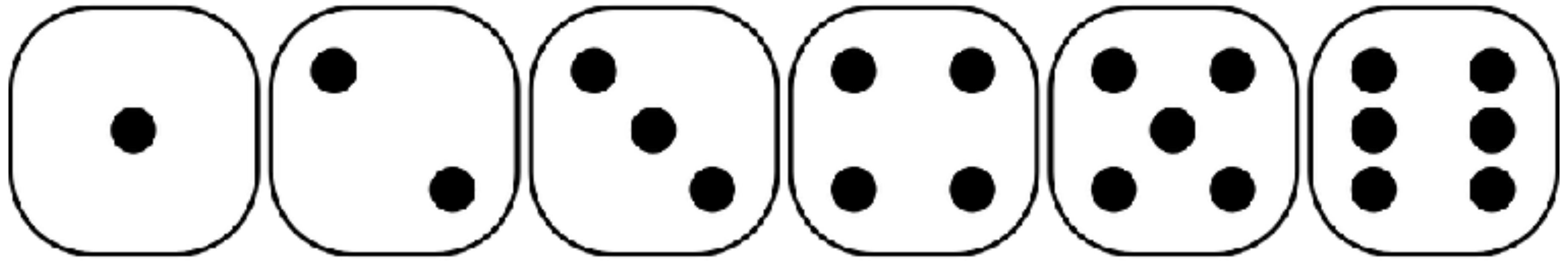
- Outcome space (or sample space) is {1, 2, 3, 4, 5, 6}
- p(x) = 1/6 for all x in {1, 2, 3, 4, 5, 6}
- Sometimes we write: P(X = x) = 1/6




**Random Variable**: A variable that can take one of many possible values.

$X =$ 

**Sample space**: The set of possible values for a random variable.



# Notation

Probability!

$$P(X = x)$$

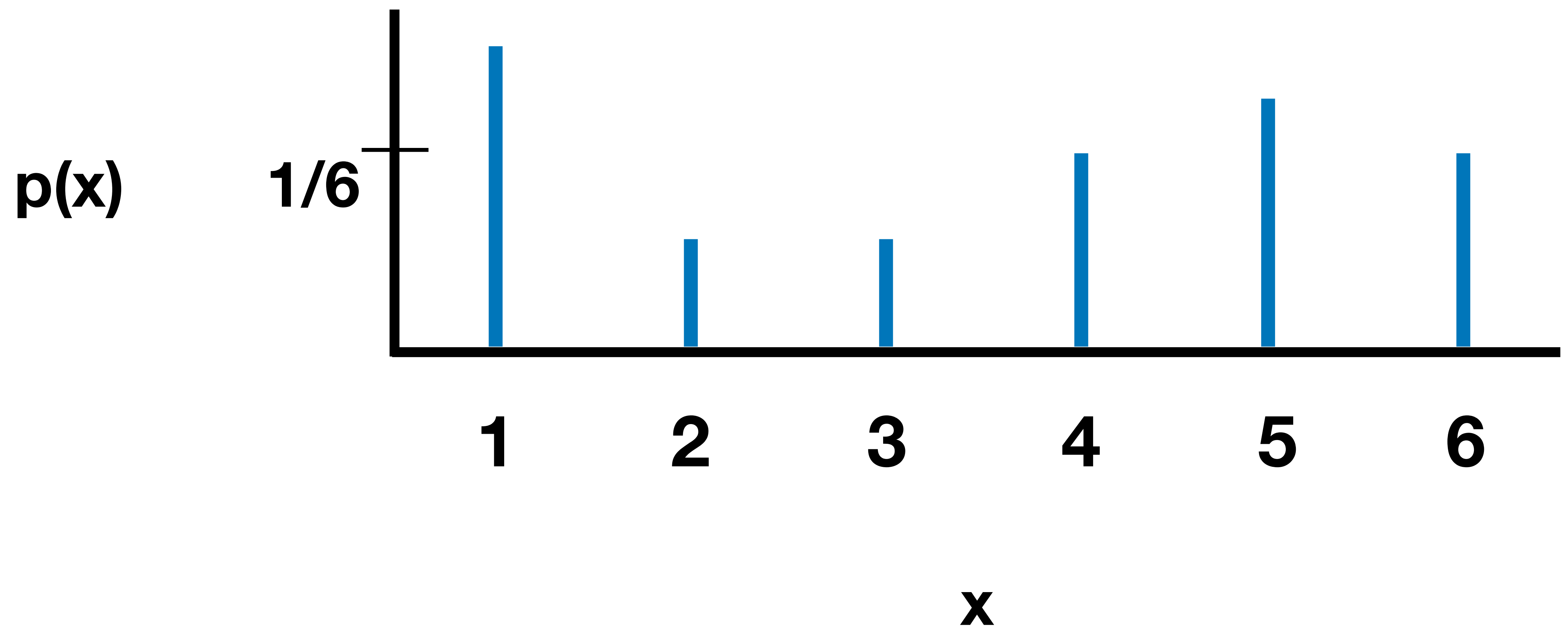
The name of the
random variable

One of the elements in
the sample space for X

**"The probability that the random variable 'X' takes the value 'x'"**

# Another PMF

- Outcome space is {1, 2, 3, 4, 5, 6}
- What does this PMF say?



# Properties of the PMF

1. $p : \mathcal{X} \to [0, 1]$

   i.e., $0 \leq p(x) \leq 1$

2. $\displaystyle\sum_{x \in \mathcal{X}} p(x) = 1$

e.g., $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$

# Expectation and Variance

- The Expected Value is defined as $$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} p(x)x$$

- The Variance is defined as $$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \sum_{x \in \mathcal{X}} p(x)(x - \mathbb{E}[X])^2\end{aligned}$$

# Back to the dice example

- The expected value of a dice roll with 1/6 probability for each outcome is 3.5
- The variance is $\sum_{x=1}^{6} p(x)(x - \mathbb{E}[X])^2 = \frac{1}{6}\sum_{x=1}^{6}(x - 3.5)^2$

$$= \frac{17.5}{6} \approx 2.92$$

- **Exercise**: Imagine you can manufacture the dice to give biased probabilities.
  Then what probability values would give a lower variance outcome?
  How about a higher variance outcome?

# Another example

- Imagine that you are a medical doctor, prescribing treatments to patients
- You give the treatment for many patients and note the outcome (X = treatment outcome). You find that you have the following probabilities:

$$p(x) = \begin{cases} 0.2 & \text{if } x \text{ is Bad} \\ 0.5 & \text{if } x \text{ is Neutral} \\ 0.3 & \text{if } x \text{ is Good.} \end{cases}$$

- But that's pretty stochastic…

# How can we make this less stochastic?

- What if we knew something about the patient?
- What information could help?

# Conditional Probabilities

- X = outcome of treatment
- Y = age of the patient
- P(X = Good | Y = 22) is a different value than P(X = Good | Y = 78)

$$P(X = x \mid Y = y)$$

**"The probability that the random variable 'X' takes the value 'x', given that the random variable 'Y' has taken the value 'y'"**

The name of another random variable

One of the elements in the sample space for Y

# Another example of conditional probabilities

- Often the value of a random variable is dependent on or correlated with another random variable

- Example: A store restocks on Wednesday, so the probability that they have pencils in stock depends on the day of the week

| | | |
|---|---|---|
| 1 | Wed | Pencil Delivery! |
| 2 | Thu | |
| 3 | Fri | |
| 4 | Sat | |
| 5 | Sun | |
| 6 | Mon | |
| 7 | Tues | |

$P(\text{Pencils in stock} \mid \text{Tuesday}) = 0.2 \qquad P(\text{Pencils in stock} \mid \text{Wednesday}) = 1.0$

**"if" statement of probability**

# Same rules for conditional probabilities as for unconditioned probabilities

1. $p(\cdot|Y = y) : \mathcal{X} \to [0, 1]$

2. $\sum_{x \in \mathcal{X}} p(x|Y = y) = 1$

**e.g.** If $P(\text{Pencils in stock} \mid \text{Tuesday}) = 0.2$, then $P(\text{Pencils NOT in stock} \mid \text{Tuesday}) = 0.8$

# Independence and Conditional Independence

- X and Y are **independent** if and only if $p(x, y) = p(x)\, p(y)$
  - equivalently, if $p(x \mid y) = p(x)$
  - recall: $p(x, y) = p(x \mid y)\, p(y)$
- X and Y are **conditionally independent**, given Z = z, if and only if
  $p(x, y \mid z) = p(x \mid z)\, p(y \mid z)$
- Exercise: Show that this definition means $E[X \mid Y = y] = E[X]$ if X and Y are independent

# Exercise 1

- Find $P(X = 3 \mid X = 3)$ and $P(X = 9 \mid X = 4)$.
- You flip a coin and get two heads in a row. What is the probability that your third coin flip also results in heads?
- You roll two standard dice. One of the die shows a 3. What is the probability that the sum of the dice is greater than 7?

# Exercise 2

- Find $P(X = 3 \mid X = 3)$ and $P(X = 9 \mid X = 4)$.

Known

Known

Must be true

Impossible

# Exercise 2: Answer

- Find $P(X = 3 \mid X = 3)$ and $P(X = 9 \mid X = 4)$.

Known

Known

Must be true

Impossible

$$P(X = 3 \mid X = 3) = 1 \qquad P(X = 9 \mid X = 4) = 0$$

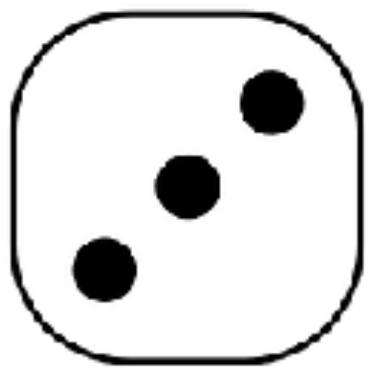
# Exercise 3

- You flip a coin and get two heads in a row. What is the probability that your third coin flip also results in heads?

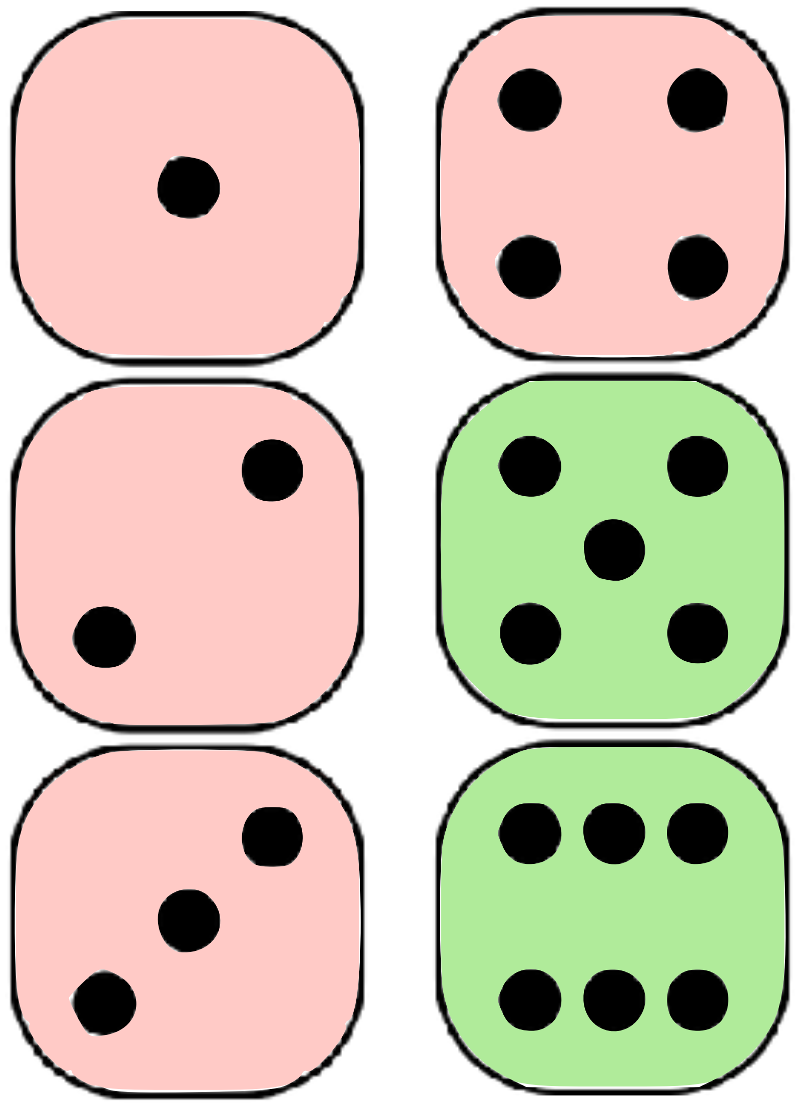
# Exercise 4

- You roll two standard dice. One of the die shows a 3. What is the probability that the sum of the dice is greater than 7?

**Roll 1**



**Roll 2**



# Worksheet Exercise

- Adam and Martha propose a simple dice game to you. You can throw a die up to two times, and they will reward you with the amount equivalent to the face value of the die. If you throw a die once and 3 comes up, you can choose to take \$3 or throw again. If you choose to throw again and 2 comes up, you earn only \$2. **The amount you earn is not additive and you only earn the amount of your last roll.**

  - Suppose in your first roll, the die comes up as a 1. What is the expected amount you would earn in your second roll?

  - For what values in your first roll should you re-roll the die?

  - What is the expected amount you would earn in this game, if you play optimally ?