



Probability Distributions: Discrete

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SEPTEMBER 13, 2016

Administrivia

- New grader: Aditya Thyagarajan
- HW1 grading nearly done
- Will appear in Moodle (waiting for late days)

Refresher: Random variables

- Random variables take on values in a *sample space*.
- This week we will focus on *discrete* random variables:
 - Coin flip: $\{H, T\}$
 - Number of times a coin lands heads after N flips: $\{0, 1, 2, \dots, N\}$
 - Number of words in a document: Positive integers $\{1, 2, \dots\}$
- Reminder: we denote the random variable with a capital letter; denote a outcome with a lower case letter.
 - E.g., X is a coin flip, x is the value (H or T) of that coin flip.

Refresher: Discrete distributions

- A discrete distribution assigns a probability to every possible outcome in the sample space
- For example, if X is a coin flip, then

$$P(X = H) = 0.5$$

$$P(X = T) = 0.5$$

- Probabilities have to be greater than or equal to 0 and probabilities over the entire sample space must sum to one

$$\sum_x P(X = x) = 1$$

Today: Types of discrete distributions

- There are many different types of discrete distributions, with different definitions.
- Today we'll look at the most common discrete distributions.
 - And we'll introduce the concept of *parameters*.
- These discrete distributions (along with the continuous distributions next) are fundamental
- Regression, classification, and clustering