Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Probability Distributions: Discrete

Introduction to Data Science Algorithms

Dirk Grunwald

Overview

- Refresher on random variables, permutations & combinations
- Bernoilli distribute, binomial distribute
- Categorical distribution, Multinomial distribution

Refresher: Random variables

- Random variables take on values in a *sample space*.
- We first focus on *discrete* random variables:
  - Coin flip: $\{H, T\}$
  - Number of times a coin lands heads after N flips: $\{0, 1, 2, \ldots, N\}$
  - Number of words in a document: Positive integers $\{1, 2, \ldots\}$
- 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

$$
\begin{aligned}
P(X=H) &= 0.5 \\
P(X=T) &= 0.5
\end{aligned}
$$

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

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

Permutation

- There are n^k ways to select things *with* replacement (each draw is from n). e.g. AAA, AAB, AAC, ABA, ..., CCC – 3^3 draws

- There are n^k ways to select things *with* replacement (each draw is from n). e.g. AAA, AAB, AAC, ABA, ..., CCC – 3^3 draws
- Recall that $n! = n \times (n-1) \times \ldots \times 2 \times 1$. There are $n!$ orderings of n things (e.g. ABC, ACB, BAC, BCA, CAB, CBA)
- P(n,k) is the number of **permutations** (orderings) of n elements taken k at a time *w/o replacement*.

$$
\begin{aligned}
P(n,k) &= n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) \\
&= \frac{n!}{(n-k)!}
\end{aligned}
$$

- e.g. AB, BA, AC, CA, BC, CB or $P(3,2)$ or $6/1 = 6$ pemutations.

Combinations

- A **combination** is an order-insensitive collection of objects.
- I.e. ABC and CBA are different pemutations but the same combination.

- A **combination** is an order-insensitive collection of objects.
- I.e. ABC and CBA are different pemutations but the same combination.
- The number of combinations of k things from n is $P(n,k)/k!$
  - Start with the $P(n,k)$ unique permutations of lenth k
  - The k things can be ordered $k!$ ways
- AB, AC, BC are the $C(3,2)$ combinations of $\{A,B,C\}$ - $6!/(2!\times 1!) = 3$

Mathematical Conventions

$0!$

If $n! = n \cdot (n-1)!$ then $0! = 1$ if definition holds for $n > 0$.

n^0

Example for 3:

$$3^2 = 9 \tag{1}$$

$$3^1 = 3 \tag{2}$$

$$3^{-1} = \frac{1}{3} \tag{3}$$

Mathematical Conventions

0!

If $n! = n \cdot (n-1)!$ then $0! = 1$ if definition holds for $n > 0$.

n^0

Example for 3:

$$3^2 = 9 \tag{1}$$
$$3^1 = 3 \tag{2}$$
$$3^0 = 1 \tag{3}$$
$$3^{-1} = \frac{1}{3} \tag{4}$$

Today: Types of discrete distributions

- There are many different types of discrete distributions, with different definitions.
- 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 tools for regression, classification, and clustering