Discrete Probability Distributions

Data Science: Jordan Boyd-Graber
University of Maryland
JANUARY 18, 2018

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: $\{0, 1, 2, \ldots, N\}$
  - 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

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

Mathematical Conventions

n^0

Example for 3:

$$3^2 = 9 \quad (1)$$

$$3^1 = 3 \quad (2)$$

$$3^{-1} = \frac{1}{3} \quad (3)$$

$0!$

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

Mathematical Conventions

0!

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

n^0

Example for 3:

$$3^2 = 9 \quad (1)$$

$$3^1 = 3 \quad (2)$$

$$3^0 = 1 \quad (3)$$

$$3^{-1} = \frac{1}{3} \quad (4)$$

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

Bernoulli distribution

- A distribution over a sample space with two values: $\{0, 1\}$
  - Interpretation: 1 is “success”; 0 is “failure”
  - Example: coin flip (we let 1 be “heads” and 0 be “tails”)
- A Bernoulli distribution can be defined with a table of the two probabilities:
  - X denotes the outcome of a coin flip:

$$\begin{aligned} P(X=0) &= 0.5 \\ P(X=1) &= 0.5 \end{aligned}$$

  - X denotes whether or not a TV is defective:

$$\begin{aligned} P(X=0) &= 0.995 \\ P(X=1) &= 0.005 \end{aligned}$$

Bernoulli distribution

- Do we need to write out both probabilities?

$$\begin{aligned} P(X=0) &= 0.995 \\ P(X=1) &= 0.005 \end{aligned}$$

- What if I only told you $P(X=1)$? Or $P(X=0)$?

Bernoulli distribution

- Do we need to write out both probabilities?

$$
\begin{aligned}
P(X=0) &= 0.995 \\
P(X=1) &= 0.005
\end{aligned}
$$

- What if I only told you $P(X=1)$? Or $P(X=0)$?

$$
\begin{aligned}
P(X=0) &= 1-P(X=1) \\
P(X=1) &= 1-P(X=0)
\end{aligned}
$$

- We only need one probability to define a Bernoulli distribution
  - Usually the probability of success, $P(X=1)$.

Bernoulli distribution

Another way of writing the Bernoulli distribution:

- Let θ denote the probability of success ($0 \leq \theta \leq 1$).

$$\begin{aligned} P(X=0) &= 1-\theta \\ P(X=1) &= \theta \end{aligned}$$

- An even more compact way to write this:

$$P(X=x) = \theta^x (1-\theta)^{1-x}$$

  - This is called a probability mass function.

Probability mass functions

- A probability mass function (PMF) is a function that assigns a probability to every outcome of a discrete random variable X.
  - Notation: $f(x) = P(X = x)$
- Compact definition
- Example: PMF for Bernoulli random variable $X \in \{0, 1\}$

$$f(x) = \theta^x (1 - \theta)^{1-x}$$

- In this example, θ is called a parameter.

Parameters

- Define the probability mass function
- Free parameters not constrained by the PMF.
- For example, the Bernoulli PMF could be written with two parameters:

$$f(x) = \theta_1^x \theta_2^{1-x}$$

  But $\theta_2 \equiv 1 - \theta_1$ ...only 1 free parameter.
- The complexity $\approx$ number of free parameters. Simpler models have fewer parameters.

Sampling from a Bernoulli distribution

- How to randomly generate a value distributed according to a Bernoulli distribution?
- Algorithm:
  1. Randomly generate a number between 0 and 1
     $r = \text{random}(0, 1)$
  2. If $r < \theta$, return success
     Else, return failure