Probability Distributions: Discrete

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SEPTEMBER 27, 2016
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:
    \[
    P(X = 0) = 0.5 \\
    P(X = 1) = 0.5
    \]

  ◦ \( X \) denotes whether or not a TV is defective:
    \[
    P(X = 0) = 0.995 \\
    P(X = 1) = 0.005
    \]
Bernoulli distribution

- Do we need to write out both probabilities?

\[
P(X = 0) = 0.995
\]
\[
P(X = 1) = 0.005
\]

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

• Do we need to write out both probabilities?

\[ P(X = 0) = 0.995 \]
\[ P(X = 1) = 0.005 \]

• What if I only told you \( P(X = 1) \)? Or \( P(X = 0) \)?

\[ P(X = 0) = 1 - P(X = 1) \]
\[ P(X = 1) = 1 - P(X = 0) \]

• 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 $\theta$ denote the probability of success ($0 \leq \theta \leq 1$).
  
  \[ P(X = 0) = 1 - \theta \]
  
  \[ P(X = 1) = \theta \]
  
- An even more compact way to write this:
  
  \[ P(X = x) = \theta^x (1 - \theta)^{1-x} \]
  
  ○ This is called a *probability mass function*. 
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, $\theta$ 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