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

Probability Distributions: Discrete

Introduction to Data Science Algorithms
Dirk Grunwald

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, $P(X=1)$.

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 = random(0, 1)
  2. If $r < \theta$, return success
     Else, return failure

Binomial distribution

- Bernoulli: distribution over two values (success or failure) from a single event
- **Binomial**: distribution of successes from *multiple* independent Bernoulli events.
- Examples:
  - The number of times “heads” comes up after flipping a coin 10 times
  - The number of defective TVs in a line of 10,000 TVs

Binomial distribution

- Suppose we flip a coin 3 times. There are $2^3 = 8$ possible outcomes:

$$
\begin{aligned}
P(HHH) &= P(H)P(H)P(H) &= 0.125 \\
P(HHT) &= P(H)P(H)P(T) &= 0.125 \\
P(HTH) &= P(H)P(T)P(H) &= 0.125 \\
P(HTT) &= P(H)P(T)P(T) &= 0.125 \\
P(THH) &= P(T)P(H)P(H) &= 0.125 \\
P(THT) &= P(T)P(H)P(T) &= 0.125 \\
P(TTH) &= P(T)P(T)P(H) &= 0.125 \\
P(TTT) &= P(T)P(T)P(T) &= 0.125
\end{aligned}
$$

- What is the probability of landing heads x times during these 3 flips?

Binomial distribution

- What is the probability of landing heads x times during these 3 flips?
- 0 times:
  - $P(TTT) = 0.125$
- 1 time:
  - $P(HTT) + P(THT) + P(TTH) = 0.375$
- 2 times:
  - $P(HHT) + P(HTH) + P(THH) = 0.375$
- 3 times:
  - $P(HHH) = 0.125$

Binomial distribution

- If the probability of success is p, this pattern

$$\underbrace{\text{SSS.....SSS}}_{x}\;\underbrace{\text{FFF...FFF}}_{n-x}$$

  occurs with probability $p^x(1-p)^{n-x}$. There are $\binom{n}{x}$ possible orderings of precisely x successes in n trials.
- The probability mass function for the binomial distribution is:

$$f(x) = \underbrace{\binom{N}{x}}_{\text{"}N\text{ choose }x\text{"}} p^x(1-p)^{N-x}$$

- Like the Bernoulli, the binomial parameter p is the probability of success from one event.
- Binomial has second parameter N: number of trials.

Bernoulli vs Binomial

- A Bernoulli distribution is a special case of the binomial distribution when $N = 1$.
- For this reason, sometimes the term binomial is used to refer to a Bernoulli random variable.

Example

- Probability that a coin lands heads *at least* once during 3 flips?

Example

- Probability that a coin lands heads *at least* once during 3 flips?

$$P(X \geq 1)$$

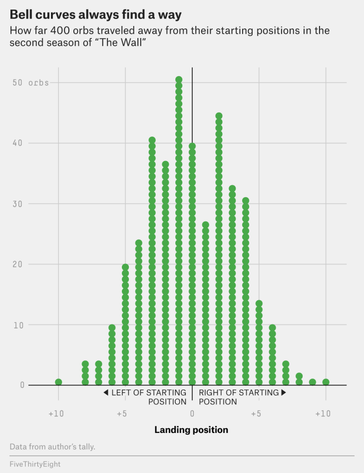
Example

- Probability that a coin lands heads *at least* once during 3 flips?

$$
\begin{aligned}
P(X \geq 1) &= P(X = 1) + P(X = 2) + P(X = 3) \\
&= 0.375 + 0.375 + 0.125 = 0.875
\end{aligned}
$$

Binomial in real life

- Assume you drop balls that hit pegs, making a series of Bernoulli trials, each of which leads to further trials.
- This is a "galton board" - see `https://youtu.be/jiWt77xme64`
- TV game "The Wall" uses a "galton board" to awards wins and losses - see `https://youtu.be/bLhZI_h2yXo`



The Geometric Distribution

- Assume we have a series of Bernoulli trials for Success (p) or Failure ($1-p$).

The Geometric Distribution

- Assume we have a series of Bernoulli trials for Success (p) or Failure ($1-p$).
- Let X be a discrete random variable that describes the number of trials until a success occurs.

The Geometric Distribution

- Assume we have a series of Bernoulli trials for Success (p) or Failure ($1 - p$).
- Let X be a discrete random variable that describes the number of trials until a success occurs.
- Then X has a *geometric distribution* with parameter p.

- Assume we have a series of Bernoulli trials for Success (p) or Failure ($1-p$).
- Let X be a discrete random variable that describes the number of trials until a success occurs.
- Then X has a *geometric distribution* with parameter p.
- The probablity mass function is

$$p_X(k) = P(X = k) = (1-p)^{k-1} p \quad \text{for } k = 1, 2, \ldots$$

- Application: number of events until a failure...