Binomial and Discrete Distributions

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

Binomial distribution

- Bernoulli: distribution over two values (success or failure) from a single event
- **binomial**: number of successes from multiple 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
- Important: each Bernoulli event is assumed to be independent
- Notation: let X be a random variable that describes the number of successes out of N trials.
  - The possible values of X are integers from 0 to N: $\{0, 1, 2, \ldots, N\}$

Binomial distribution

- Suppose we flip a coin 3 times. There are 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

- The probability mass function for the binomial distribution is:

$$f(x) = \underbrace{\binom{N}{x}}_{\text{“}N \text{ choose } x\text{”}} \theta^x (1-\theta)^{N-x}$$

- Like the Bernoulli, the binomial parameter θ is the probability of success from one event.
- Binomial has second parameter N: number of trials.
- The PMF important: difficult to figure out the entire distribution by hand.

Aside: Binomial coefficients

- The expression $\binom{n}{k}$ is called a binomial coefficient.
    - Also called a combination in combinatorics.
- $\binom{n}{k}$ is the number of ways to choose k elements from a set of n elements.
- For example, the number of ways to choose 2 heads from 3 coin flips:
  HHT, HTH, THH
  $\binom{3}{2} = 3$
- Formula:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1

Pascal's triangle depicts the values of $\binom{n}{k}$.

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?

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

Categorical distribution

- Recall: the Bernoulli distribution is a distribution over two values (success or failure)
- **categorical** distribution generalizes Bernoulli distribution over any number of values
  - Rolling a die
  - Selecting a card from a deck
- AKA discrete distribution.
  - Most general type of discrete distribution
  - specify all (but one) of the probabilities in the distribution
  - rather than the probabilities being determined by the probability mass function.

Categorical distribution

- If the categorical distribution is over K possible outcomes, then the distribution has K parameters.
- We will denote the parameters with a K-dimensional vector $\vec{\theta}$.
- The probability mass function can be written as:

$$f(x) = \prod_{k=1}^{K} \theta_k^{[x=k]}$$

  where the expression $[x = k]$ evaluates to 1 if the statement is true and 0 otherwise.
    - All this really says is that the probability of outcome x is equal to θ_x.
- The number of free parameters is $K-1$, since if you know $K-1$ of the parameters, the Kth parameter is constrained to sum to 1.

Categorical distribution

- Example: the roll of a (unweighted) die

$$
\begin{aligned}
P(X=1) &= \frac{1}{6} \\
P(X=2) &= \frac{1}{6} \\
P(X=3) &= \frac{1}{6} \\
P(X=4) &= \frac{1}{6} \\
P(X=5) &= \frac{1}{6} \\
P(X=6) &= \frac{1}{6}
\end{aligned}
$$

- If all outcomes have equal probability, this is called the uniform distribution.
- General notation: $P(X = x) = \theta_x$

Sampling from a categorical distribution

- How to randomly select a value distributed according to a categorical distribution?
- The idea is similar to randomly selected a Bernoulli-distributed value.
- Algorithm:
  1. Randomly generate a number between 0 and 1
     r = random(0, 1)
  2. For $k = 1, \dots, K$:
     - Return smallest r s.t. $r < \sum_{i=1}^{k} \theta_k$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.452383$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}$$

Random number in $(0,1)$:
$r = 0.452383$

$r < \theta_1$?

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.452383$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.452383$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.452383$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?

- Return $X = 3$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.117544$

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.117544$

$r < \theta_1$?

Sampling from a categorical distribution

- Example: simulating the roll of a die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.166667 \\
P(X=2) &= \theta_2 &= 0.166667 \\
P(X=3) &= \theta_3 &= 0.166667 \\
P(X=4) &= \theta_4 &= 0.166667 \\
P(X=5) &= \theta_5 &= 0.166667 \\
P(X=6) &= \theta_6 &= 0.166667
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.117544$

$r < \theta_1$?

- Return $X = 1$

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$\begin{aligned} P(X=1) &= \theta_1 &= 0.01 \\ P(X=2) &= \theta_2 &= 0.01 \\ P(X=3) &= \theta_3 &= 0.01 \\ P(X=4) &= \theta_4 &= 0.01 \\ P(X=5) &= \theta_5 &= 0.01 \\ P(X=6) &= \theta_6 &= 0.95 \end{aligned}$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6$?

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$:
$r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6$?

- Return $X = 6$

Sampling from a categorical distribution

- Example 2: rolling a biased die

$$
\begin{aligned}
P(X=1) &= \theta_1 &= 0.01 \\
P(X=2) &= \theta_2 &= 0.01 \\
P(X=3) &= \theta_3 &= 0.01 \\
P(X=4) &= \theta_4 &= 0.01 \\
P(X=5) &= \theta_5 &= 0.01 \\
P(X=6) &= \theta_6 &= 0.95
\end{aligned}
$$

Random number in $(0,1)$: $r = 0.209581$

$r < \theta_1$?
$r < \theta_1 + \theta_2$?
$r < \theta_1 + \theta_2 + \theta_3$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5$?
$r < \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6$?

- Return $X = 6$

- We will always return $X = 6$ unless our random number $r < 0.05$.
  - 6 is the most probable outcome