Probability Distributions: Discrete

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SEPTEMBER 13, 2016

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

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

- 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 $k = 1, \ldots, K$:
   - 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 $(0,1)$:
$r = 0.209581$

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

- 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