



Probability Distributions: Discrete

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SEPTEMBER 13, 2016

Multinomial distribution

- Recall: the binomial distribution is the number of successes from multiple Bernoulli success/fail events
- The **multinomial** distribution is the number of different outcomes from multiple *categorical* events
 - It is a generalization of the binomial distribution to more than two possible outcomes
 - As with the binomial distribution, each categorical event is assumed to be independent
 - **Bernoulli : binomial :: categorical : multinomial**
- Examples:
 - The number of times each face of a die turned up after 50 rolls
 - The number of times each suit is drawn from a deck of cards after 10 draws

Multinomial distribution

- Notation: let \vec{X} be a vector of length K , where X_k is a random variable that describes the number of times that the k th value was the outcome out of N categorical trials.
 - The possible values of each X_k are integers from 0 to N
 - All X_k values must sum to N : $\sum_{k=1}^K X_k = N$
- Example: if we roll a die 10 times, suppose it comes up with the following values:
 $\vec{X} = \langle 1, 0, 3, 2, 1, 3 \rangle$
 - $X_1 = 1$
 - $X_2 = 0$
 - $X_3 = 3$
 - $X_4 = 2$
 - $X_5 = 1$
 - $X_6 = 3$
- The multinomial distribution is a *joint* distribution over multiple random variables: $P(X_1, X_2, \dots, X_K)$

Multinomial distribution

- Suppose we roll a die 3 times. There are 216 (6^3) possible outcomes:

$$P(111) = P(1)P(1)P(1) = 0.00463$$

$$P(112) = P(1)P(1)P(2) = 0.00463$$

$$P(113) = P(1)P(1)P(3) = 0.00463$$

$$P(114) = P(1)P(1)P(4) = 0.00463$$

$$P(115) = P(1)P(1)P(5) = 0.00463$$

$$P(116) = P(1)P(1)P(6) = 0.00463$$

...

...

...

$$P(665) = P(6)P(6)P(5) = 0.00463$$

$$P(666) = P(6)P(6)P(6) = 0.00463$$

- What is the probability of a particular vector of counts after 3 rolls?

Multinomial distribution

- What is the probability of a particular vector of counts after 3 rolls?
- Example 1: $\vec{X} = \langle 0, 1, 0, 0, 2, 0 \rangle$

Multinomial distribution

- What is the probability of a particular vector of counts after 3 rolls?
- Example 1: $\vec{X} = \langle 0, 1, 0, 0, 2, 0 \rangle$
 - $P(\vec{X}) = P(255) + P(525) + P(552) = 0.01389$

Multinomial distribution

- What is the probability of a particular vector of counts after 3 rolls?
- Example 1: $\vec{X} = \langle 0, 1, 0, 0, 2, 0 \rangle$
 - $P(\vec{X}) = P(255) + P(525) + P(552) = 0.01389$
- Example 2: $\vec{X} = \langle 0, 0, 1, 1, 1, 0 \rangle$

Multinomial distribution

- What is the probability of a particular vector of counts after 3 rolls?
- Example 1: $\vec{X} = \langle 0, 1, 0, 0, 2, 0 \rangle$
 - $P(\vec{X}) = P(255) + P(525) + P(552) = 0.01389$
- Example 2: $\vec{X} = \langle 0, 0, 1, 1, 1, 0 \rangle$
 - $P(\vec{X}) = P(345) + P(354) + P(435) + P(453) + P(534) + P(543) = 0.02778$

Multinomial distribution

- The probability mass function for the multinomial distribution is:

$$f(\vec{x}) = \frac{N!}{\underbrace{\prod_{k=1}^K x_k!}_{\text{Generalization of binomial coefficient}}} \prod_{k=1}^K \theta_k^{x_k}$$

- Like categorical distribution, multinomial has a K -length parameter vector $\vec{\theta}$ encoding the probability of each outcome.
- Like binomial, the multinomial distribution has an additional parameter N , which is the number of events.

Multinomial distribution: summary

- Categorical distribution is multinomial when $N = 1$.
- Sampling from a multinomial: same code repeated N times.
 - Remember that each categorical trial is independent.
 - Question: Does this mean the count values (i.e., each X_1 , X_2 , etc.) are independent?

Multinomial distribution: summary

- Categorical distribution is multinomial when $N = 1$.
- Sampling from a multinomial: same code repeated N times.
 - Remember that each categorical trial is independent.
 - Question: Does this mean the count values (i.e., each X_1 , X_2 , etc.) are independent?
 - No! If $N = 3$ and $X_1 = 2$, then X_2 can be no larger than 1 (must sum to N).

Multinomial distribution: summary

- Categorical distribution is multinomial when $N = 1$.
- Sampling from a multinomial: same code repeated N times.
 - Remember that each categorical trial is independent.
 - Question: Does this mean the count values (i.e., each X_1 , X_2 , etc.) are independent?
 - No! If $N = 3$ and $X_1 = 2$, then X_2 can be no larger than 1 (must sum to N).
- Remember this analogy:
 - **Bernoulli : binomial :: categorical : multinomial**