Multinomial Dirichlet Conjugacy

Rebecca C. Steorts

Agenda

- Dirichlet distribution
- The Dirichlet-Multinomial

Dirichlet

A Dirichlet distribution[1] is a distribution of the K-dimensional probability simplex[2]

$$\triangle_K = \{(\pi_1, \ldots, \pi_k) : \pi_k \geq 0, \sum_k \pi_k = 1\}.$$

We say that $(\pi_1, \ldots, \pi_k)$ is Dirichlet distributed:

[1] This is the multivariate version of the Beta distribution.

[2] In geometry, a simplex is a generalization of the notion of a triangle or tetrahedron to arbitrary dimensions.

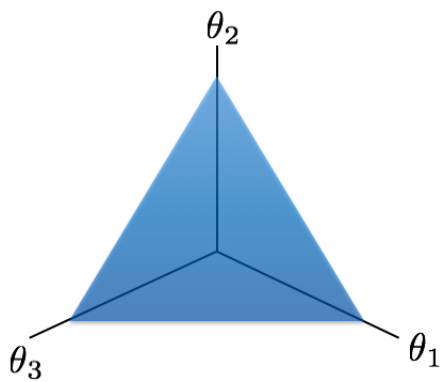
Dirichlet distribution

Let

where the probability density function is

$$p(\theta \mid \alpha) \propto \prod_{k=1}^{m} \theta_k^{\alpha_k - 1},$$

where $\sum_k \theta_k = 1, \theta_i \geq 0$ for all i



Dirichlet distribution

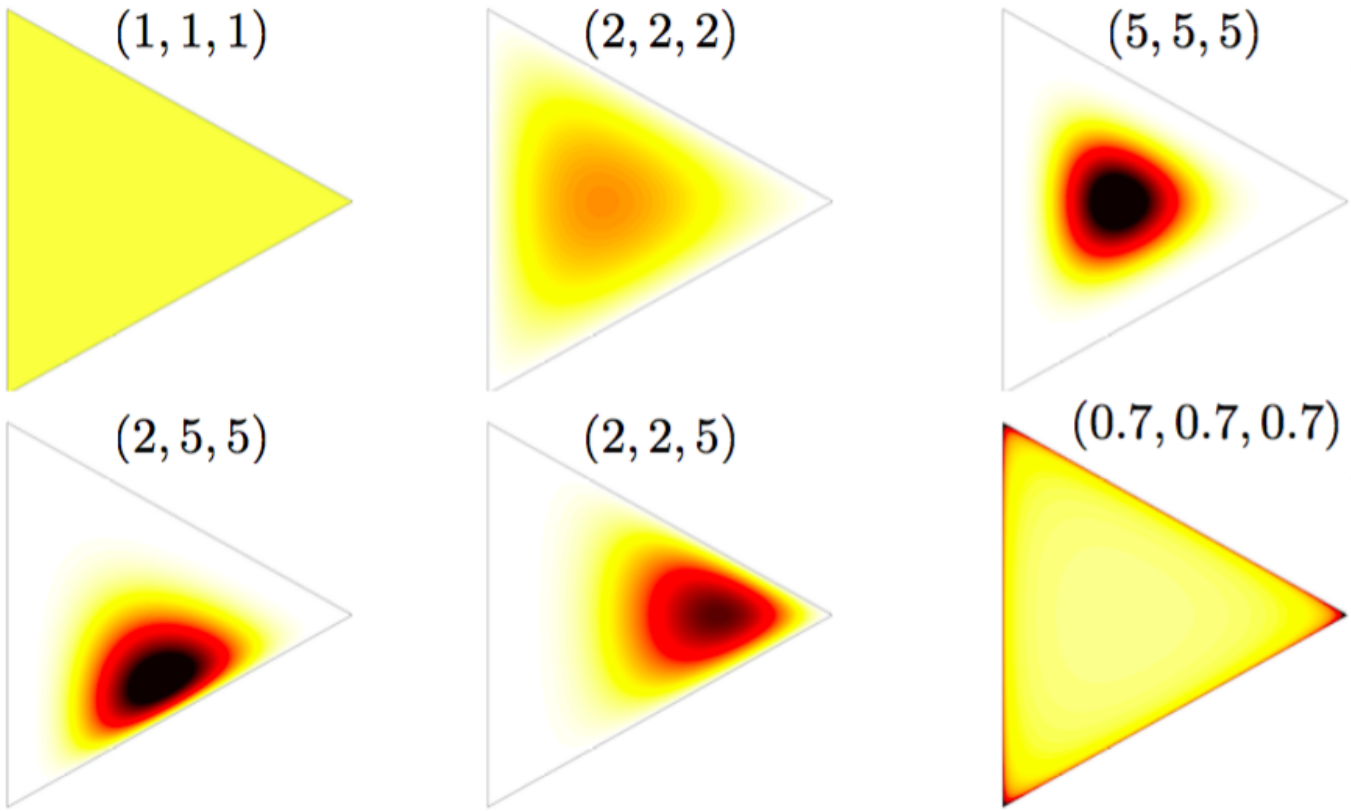


Figure 1: Far left: We get a uniform prior on the simplex. Moving to the right we get things unimodal. On the bottom, we get distributions that are multimodal at the corners.

Multinomial-Dirichlet

In this exercise, we'll learn about the Multinomial or Categorical distribution.[3]

[3] This is the multivariate generalization of of the Binomial distribution.

Multinomial or Categorical distribution

Conjugate prior (Dirichlet)

$$\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$$

Recall the density of the Dirichlet is the following:

$$p(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) \propto \prod_{j=1}^{m} \theta_j^{\alpha_j - 1},$$

where $\sum_j \theta_j = 1, \theta_i \geq 0$ for all i

Likelihood

Likelihood, Prior, and Posterior

Takeaways

1. Dirichlet is conjugate for Categorical or Multinomial.[4]
2. Useful formula:

$$\prod_i \text{Multinomial}(x_i \mid \theta) \times \text{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) \propto \text{Dir}(\boldsymbol{\theta} \mid \boldsymbol{c} + \boldsymbol{\alpha}).$$

[4] The word Categorical seems to be used in CS and ML. The word Multinomial seems to be used in Statistics and Mathematics.