Multinomial Dirichlet Conjugacy

Rebecca C. Steorts

Agenda

- Dirichlet distribution
- The Dirichlet-Multinomial

Dirichlet

A Dirichlet distribution[1] is a distribution of the $\overbrace{K\text{-dimensional}}^{\text{triangle}}$ probability simplex[2]

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

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

fixed, known

$$(\pi_1, \dots, \pi_k) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k)$$

$$p(\pi_1, \pi_2, \dots, \pi_k) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$$

[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

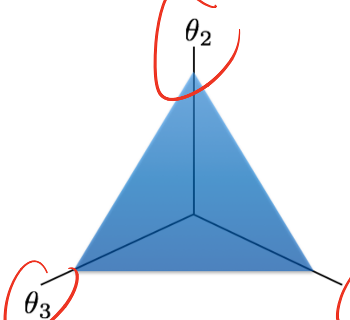
$$\theta \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_k)$$

where the probability density function is

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

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

used to model multivariate weight e.g. gender



Triangle.

Dirichlet distribution

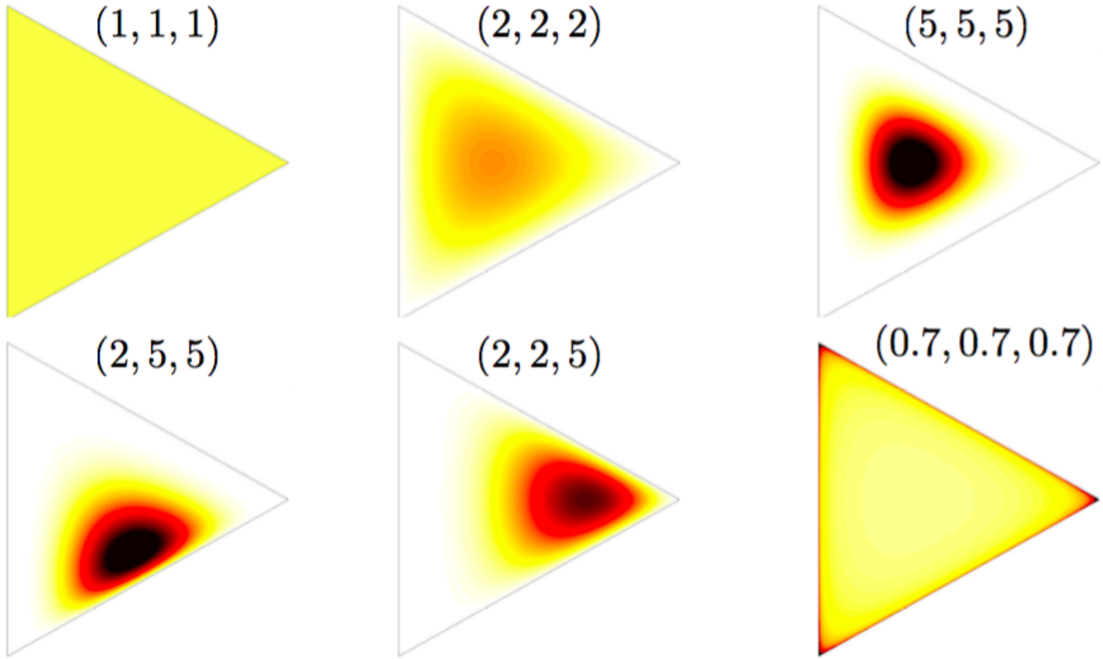


Figure 1: Far left: We get a uniform prior on the simplex. Moving to the right we get things unimodal. On the bottom, we get distributions that are multimodal at the corners.

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

$P(x_i = j \mid \theta)$

$= \theta_j$

Assume $X = (x_1, \ldots, x_n)$

$x_i \in \{1, \ldots, m\}$

$\theta = (\theta_1, \ldots, \theta_m)$, $\sum_{i=1} \theta_i = 1$.

Assume

$\underline{X} \mid \underline{\theta} \overset{ind}{\sim} \text{Multinomial}(\underline{\theta})$

$x \mid \theta \overset{ind}{\sim} \text{Categorical}(\underline{\theta})$

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

$$X = (x_1, \ldots, x_n) \qquad x_i \in \{1, \ldots, m\}.$$

$$p(X \mid \theta) = \prod_{i=1}^{n} p(X_i = x_i \mid \theta) \qquad \text{(defn)}$$

$$= \prod_{i=1}^{n} \theta_{x_i} = \theta_{x_1} \times \theta_{x_2} \cdots \theta_{x_n} \qquad \text{(defn)}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \theta_j^{I(x_i = j)} = \prod_{j=1}^{m} \prod_{i=1}^{n} \theta_j^{I(x_i = j)}$$

swap

$$= \prod_{j=1}^{m} \theta_j^{\underbrace{\sum_i I(x_i = j)}_{c_j}}$$

Likelihood, Prior, and Posterior

$$= \boxed{\prod_{j=1} \theta_j}$$

$$c = (c_1, \ldots, c_m)$$

$$c_j = \#\{i : x_i = j\}$$

$$p(\theta \mid x) \propto \prod_{j=1}^{m} \theta_j^{c_j} \times \prod_{j=1}^{m} \theta_j^{\alpha_j - 1} \times I\left(\sum_j \theta_j = 1;\ \theta_i \geq 0\ \forall i\right)$$

$$= \prod_{j=1}^{m} \theta_j^{c_j + \alpha_j - 1} I\left(\qquad \right)$$

$$\theta \mid x \sim \mathrm{Dir}(\underline{c} + \underline{\alpha}).$$

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.