# Module 7: Exercise 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:

$$(\pi_1, \ldots, \pi_k) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_k)$$

if

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

[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

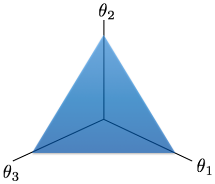
Let

$$\theta \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$$

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

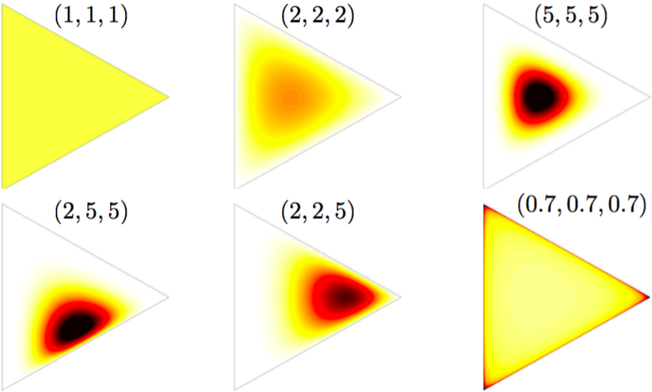



Figure 1: Far left: We get a uniform prior on the simplex. Moving to the right we get things unimodal. On the bottom, we get distributions that are multimodal at the corners.

# Multinomial-Dirichlet

In order to proceed with the lab, we'll need to learn about the Multinomial or Categorical distribution.[3]

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

# Multinomial or Categorical distribution

Assume $X = (x_1, x_2, \ldots x_n)$ where $x_i \in \{1, \ldots, m\}$. Assume $\theta = (\theta_1, \ldots, \theta_m)$, where $\sum_i \theta_i = 1$.

Assume that

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

or

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

$$P(X_i = j \mid \theta) = \theta_j$$

# 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

Define the data as $\boldsymbol{X} = (x_1, \ldots, x_n)$, $x_i \in \{1, \ldots m\}$. Consider

$$
\begin{aligned}
p(\boldsymbol{X} \mid \theta) &= \prod_{i=1}^{n} P(X_i = x_i \mid \theta) && (1)\\
&= \prod_{i=1}^{n} \theta_{x_i} = \theta_{x_1} \times \theta_{x_2} \times \theta_{x_n} && (2)\\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} \theta_j^{I(x_i=j)} = \prod_{j=1}^{m} \prod_{i=1}^{n} \theta_j^{I(x_i=j)} && (3)\\
&= \prod_{j=1}^{m} \theta_j^{\sum_i I(x_i=j)} && (4)\\
&= \prod_{j=1}^{m} \theta_j^{c_j} && (5)
\end{aligned}
$$

where $c = (c_1, \ldots c_m)$, $c_j = \#\{i : x_i = j\}$.

# Likelihood, Prior, and Posterior

$$p(\boldsymbol{X} \mid \boldsymbol{\theta}) = \prod_{j=1}^{m} \theta_j^{c_j}$$

$$P(\boldsymbol{\theta}) \propto \prod_{j=1}^{m} \theta_j^{\alpha_j - 1} I(\sum_j \theta_j = 1, \theta_i \geq 0 \forall i)$$

Then

$$\begin{aligned} P(\boldsymbol{\theta} \mid \boldsymbol{X}) &\propto \prod_{j=1}^{m} \theta_j^{c_j} \times \prod_{j=1}^{m} \theta_j^{\alpha_j - 1} I(\sum_j \theta_j = 1, \theta_i \geq 0 \forall i) \qquad (6) \\ &\propto \prod_{j=1}^{m} \theta_j^{c_j + \alpha_j - 1} I(\sum_j \theta_j = 1, \theta_i \geq 0 \forall i) \qquad (7) \end{aligned}$$

This implies

$$\boldsymbol{\theta} \mid \boldsymbol{X} \sim \text{Dirichlet}(\boldsymbol{c} + \boldsymbol{\alpha}).$$

# 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.