Multivariate Distributions: Dirichlet-Multinomial

Rebecca C. Steorts
Bayesian Methods and Modern Statistics: STA 360/601

Module 10

Dirichlet-Multinomial

- $\theta = (\theta_1, \ldots, \theta_m)$,
- $X_i \in \{1, \ldots, m\}$,
- $\sum_i \theta_i = 1$.

Assume that

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

or

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

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

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

What is the density of the Dirichlet?

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

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

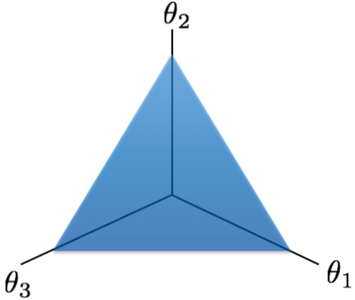

Figure 1: 3 dimensional support of the θ space. Called the simplex!

Likelihood

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

$$p(D \mid \theta) = \prod_{i=1}^{n} P(X_i = x_i \mid \theta) \tag{1}$$

$$= \prod_{i=1}^{n} \theta_{x_i} \tag{2}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \theta_j^{I(x_i = j)} \tag{3}$$

$$= \prod_{j=1}^{m} \theta_j^{\sum_i I(x_i = j)} \tag{4}$$

$$= \prod_{j=1}^{m} \theta_j^{c_j} \tag{5}$$

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

Likelihood, Prior, and Posterior

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

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

Then

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

This implies

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

Takeaways

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

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

[1] The word Categorical seems to be used in CS and ML. The word Multinomial seems to be used in Statistics and Mathematics. I have no idea what is used in other sciences.

Exercises

1. Derive the mean and variance of the Dirichlet distribution.
2. Suppose we have a new data point x. That is calculate $p(x \mid D)$. Derive this on your own. (You need to do 1 to complete 2).