The Dirichlet distribution

The multinomial distribution answers the question: "What is the probability of obtaining a particular set of counts $\boldsymbol{c} = (c_1, ..., c_k) \in \mathbb{N}_0^k$ from repeated draws from a categorical distribution $\boldsymbol{\beta} = (\beta_1, ..., \beta_k) \in [0,1]^k$?" The answer is:

$$\text{Multinomial}(\boldsymbol{c} \mid \boldsymbol{\beta}) = \frac{(\sum_i c_i)!}{\prod_i c_i!} \prod_i \beta_i^{c_i} = \frac{(c_1 + \cdots + c_k)!}{c_1! \cdots c_k!} \beta_1^{c_1} \cdots \beta_k^{c_k}$$

The Dirichlet distribution asks the opposite question: "What is the probability that a particular categorical distribution is responsible for a given set of counts?" The answer is:

$$\text{Dirichlet}(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \beta_i^{\alpha_i - 1} \tag{1}$$

Note that these equations are equal with $\alpha_i = c_i + 1$. The difference between the distributions is in the interpretation, and in that α_i is not assumed to be integral.

Mean

The expectation value of $\boldsymbol{\beta}$ can be computed for each component.

$$\mathbb{E}\{\beta_i\} = \int_0^1 \beta_i \Pr(\beta_i \mid \boldsymbol{\beta}_{[i]}, \boldsymbol{\alpha}) \, \mathrm{d}\beta_i = \int_\Delta \beta_i \, \text{Dirichlet}(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\beta}$$

The integral $\int_\Delta$ is taken over the simplex $\sum_i \beta_i = 1$. Substituting eq. 1 yeilds

$$\mathbb{E}\{\beta_i\} = \frac{\Gamma\left(\sum_j \alpha_j\right)}{\prod_j \Gamma(\alpha_j)} \int_\Delta \beta_i \prod_{j \neq i} \beta_j^{\alpha_j - 1} \, \mathrm{d}\boldsymbol{\beta} \tag{2}$$

where the integrand is now proportional to eq. 1, but with one of the shape factors offset as $\alpha_i \mapsto \alpha_i + 1$ due to the extra β_i. We can circumnavigate this integral by noting that the Dirichlet distribution is normalised to establish

$$\int_\Delta \text{Dirichlet}(\boldsymbol{\beta} \mid \alpha_1, ..., \alpha_i + 1, ..., \alpha_k) \, \mathrm{d}\boldsymbol{\beta} = \frac{\Gamma\left(\sum_j \alpha_j + 1\right)}{\Gamma(\alpha_i + 1) \prod_{j \neq i} \Gamma(\alpha_j)} \int_\Delta \beta_i \prod_{j \neq i} \beta_j^{\alpha_j - 1} \, \mathrm{d}\boldsymbol{\beta} = 1$$

which, substituted into eq. 2, gives

$$\mathbb{E}\{\beta_i\} = \frac{\Gamma\left(\sum_j \alpha_j\right)}{\prod_j \Gamma(\alpha_j)} \frac{\Gamma(\alpha_i + 1) \prod_{j \neq i} \Gamma(\alpha_j)}{\Gamma\left(\sum_j \alpha_j + 1\right)} = \frac{\Gamma(\alpha_i + 1)}{\Gamma(\alpha_i)} \frac{\Gamma\left(\sum_j \alpha_j\right)}{\Gamma\left(\sum_j \alpha_j + 1\right)} = \frac{\alpha_i}{\sum_j \alpha_j} =: \frac{\alpha_i}{\alpha_0}$$

For $\boldsymbol{\beta} \in \text{Dirichlet}(\boldsymbol{\alpha})$, the mean of a component β_i is

$$\mathbb{E}\{\beta_i\} = \frac{\alpha_i}{\alpha_0}$$

where $\alpha_0 = \sum_i \alpha_i$.