Multivariate Conjugate Distributions

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Dirichlet

Let K be a number of categories and $\theta_i \in (0, 1), i = 1, \ldots, K$ be R.V. Then, the θ_i have a Dirichlet $(\alpha_i), i = 1, \ldots, K$ distribution with pdf

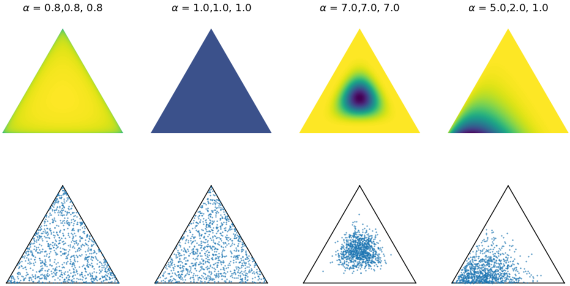
$$f(\boldsymbol{\theta}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1}$$

$$E[\theta_i] = \frac{\alpha_i}{\sum_{i=1}^{K} \alpha_i}$$

Let $\alpha_0 = \sum_{i=1}^{K} \alpha_i$ then

$$Var[\theta_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}$$

Dirichlet Examples



Dirichlet-Multinomial

- $\boldsymbol{x} = (x_1, \ldots, x_k)$ has a multinomial distribution with parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ and N trials, $\sum_i^k x_i = N$

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{i=1}^{k} \theta_i^{x_i}$$

- Prior

$$h(\boldsymbol{\theta}) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}$$

- Posterior

$$h(\boldsymbol{\theta}|\boldsymbol{x}) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i + x_i - 1}$$

Multivariate Gaussian with Unknown Mean & Known Variance-Covariance Matrix

- Mean vector: $\boldsymbol{M}$ with k dimensions
- Precision matrix: $\boldsymbol{\tau}$ with $k \times k$ dimensions
- Likelihood with N trials, $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$

$$f(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{\tau}) \propto exp\left(-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{m})^T\boldsymbol{\tau}(\boldsymbol{x}_i - \boldsymbol{m})\right)$$

- Prior with parameters $\boldsymbol{\mu}_0, \boldsymbol{\tau}_0$

$$f(\mu) = \frac{1}{2}exp\left((\boldsymbol{m} - \boldsymbol{\mu}_0)^T\tau_0(\boldsymbol{m} - \boldsymbol{\mu}_0)\right)$$

- Posterior with $\boldsymbol{\mu}^* = (\boldsymbol{\tau}_0 + N\boldsymbol{\tau})^{-1}(\boldsymbol{\tau}_0\boldsymbol{\mu}_0 + N\boldsymbol{\tau}\bar{\boldsymbol{x}})$

$$f(\boldsymbol{m}|\boldsymbol{x}) = \frac{1}{2}exp\left((\boldsymbol{\mu} - \boldsymbol{\mu}^*)^T(\boldsymbol{\tau}_0 + N\boldsymbol{\tau})(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\right)$$

Multivariate Gaussian with Unknown Mean & Variance-Covariance Matrix

- Let $\boldsymbol{S} = \sum_{i=1}^{N} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$
- $f(\boldsymbol{m}|\boldsymbol{w}) \sim N(\boldsymbol{\mu}_0, v\boldsymbol{w}), v > 0$
- $f(\boldsymbol{w})$ is Wishart with α degrees of freedom and precision matrix $\boldsymbol{r}$, with $\alpha > k - 1$
- Likelihood: $N(\boldsymbol{M}, \boldsymbol{W})$
- Posterior: $f(\boldsymbol{m}|\boldsymbol{x}, \boldsymbol{w}) \sim N\left(\boldsymbol{\mu}^*, (v + N)\boldsymbol{w}\right)$ where

$$\boldsymbol{\mu}^* = \frac{v\boldsymbol{\mu}_0 + N\bar{\boldsymbol{x}}}{v + N}$$

- Posterior: f($\boldsymbol{w}|\boldsymbol{x}$) is Wishart with $\alpha + N$ degrees of freedom & precision matrix $\boldsymbol{r}^*$ where

$$\boldsymbol{r}^* = \boldsymbol{r} + \boldsymbol{S} + \frac{vN}{v + N}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{x}})(\boldsymbol{\mu}_0 - \bar{\boldsymbol{x}})^T$$

Marginal Distribution of the Mean

- Prior:

$$f(\boldsymbol{m},\boldsymbol{w}) \propto |\boldsymbol{w}|^{(\alpha-k)/2} exp[-\frac{1}{2}Tr((\boldsymbol{r}+v(\boldsymbol{m}-\boldsymbol{\mu_0})(\boldsymbol{m}-\boldsymbol{\mu_0})^T)\boldsymbol{w})]$$

- Integrate over $\boldsymbol{w}$ to obtain the marginal prior for $\boldsymbol{m}$

$$\begin{aligned} f(\boldsymbol{m}) &\propto |\boldsymbol{r}+v(\boldsymbol{m}-\boldsymbol{\mu_0})(\boldsymbol{m}-\boldsymbol{\mu_0})^T|^{-(\alpha+1)/2} \\ &\propto [1+v(\boldsymbol{m}-\boldsymbol{\mu_0})\boldsymbol{r}^{-1}(\boldsymbol{m}-\boldsymbol{\mu_0})^T]^{-(\alpha+1)/2} \end{aligned}$$

- So $f(\boldsymbol{m})$ is a multivariate t distribution with $\alpha - k + 1$ d.o.f., location parameter $\boldsymbol{\mu}_0$, and precision $v(\alpha - k + 1)\boldsymbol{r}^{-1}$
- Posterior, $f(\boldsymbol{m}|\boldsymbol{x})$ is a multivariate t distribution with $\alpha + N - k + 1$ d.o.f. and location parameter, $\boldsymbol{\mu}^*$ and precision $(v + N)(\alpha + N - k + 1)(\boldsymbol{r}^*)^{-1}$