# Models

kiraselby

April 2018

## 1 Univariate Normal

The Univariate Normal model consists of a Normal likelihood (in canonical form), and a conjugate Normal-Gamma prior.

### 1.1 Parameters

The canonical Normal is parametrized by the mean $\mu$ and precision (or inverse-variance) $\lambda$.

The Normal-Gamma prior is parametrized by the hyperparameters $\mu_0$, $\nu_0$, $\alpha_0$, and $\beta_0$. $\mu_0$ is the prior mean. $\nu_0$ can be viewed as the prior sample weight - the effective number of existing observations upon which the prior mean has been estimated. $\alpha_0$ and $\beta_0$ are the shape and rate parameters for a gamma distribution respectively.

By default, these hyperparameters are initialized to $\mu_0 = 0$, $\nu_0 = 1e-8$, $\alpha_0 = 2$, and $\beta_0 = 0.5$. This choice of parameters approximates an improper flat prior over the mean (we cannot set $\nu_0 = 0$ directly due to limitations from other packages we are using which do not accept improper distributions). Unfortunately, we cannot use a similarly weak prior over the precision as this will result in an improper posterior (see Prior distributions for variance parameters in hierarchical models by A. Gelman) [CITATION - http://www.stat.columbia.edu/ gelman/research/published/taumain.pdf]. We use some simple values for $\alpha_0$, and $\beta_0$ that seem to perform well, but more work may be necessary to find a more principled choice of defaults for these parameters.

### 1.2 Likelihood

$$F(y|\mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{-1/2} e^{-\frac{\lambda}{2}(y-\mu)^2} \tag{1}$$

## 1.3 Prior

$$G_0(\mu, \lambda | \mu_0, \nu_0, \alpha_0, \beta_0) = N(\mu | \mu_0, \nu_0 \lambda) \cdot Gamma(\lambda | \alpha_0, \beta_0) \tag{2}$$

$$= \left(\frac{\nu_0}{2\pi}\right)^{-\frac{1}{2}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0 - \frac{1}{2}} e^{-\beta_0 \lambda} e^{-\frac{\lambda}{2}(\mu - \mu_0)^2} \tag{3}$$

## 1.4 Posterior

$$\nu_n = \nu_0 + n \tag{4}$$

$$\alpha_n = \alpha_0 + \frac{n}{2} \tag{5}$$

$$\mu_n = \frac{\nu_0}{\nu_n} \mu_0 + \frac{n}{\nu_n} \bar{x} \tag{6}$$

$$\beta_n = \beta_0 + \frac{n}{2} \left( s^2 + \frac{\nu_0}{\nu_n} (\bar{x} - \mu)^2 \right) \tag{7}$$

where $\bar{x} = \frac{1}{n} \sum_x x$ and $s^2 = \frac{1}{n} \sum_x (x - \bar{x})^2$

## 1.5 Marginal Likelihood

$$p(y | \eta) = \int F(y | \mu, \lambda) dG_0(\mu, \lambda | \mu_0, \nu_0, \alpha_0, \beta_0) \tag{8}$$

$$= \frac{1}{2\pi} \frac{\Gamma(\alpha_n)}{\Gamma \alpha_0} \sqrt{\frac{\nu_0}{\nu_n}} \frac{\beta_0^{\alpha_0}}{\beta_n^{\alpha_n}} \tag{9}$$

# 2 Multivariate Normal

The Multivariate Normal model consists of a Multivariate Normal likelihood (in canonical form), and a conjugate Normal-Wishart prior.

## 2.1 Parameters

The canonical Multivariate Normal is parametrized by the mean vector $\mu$ and precision (or inverse-covariance) matrix $\Lambda$.

The Normal-Wishart prior is parametrized by the hyperparameters $\mu_0$, $\kappa_0$, $_0$, and $\nu_0$. $\mu_0$ is the prior mean vector. $\kappa_0$ can be viewed as the prior sample weight - the effective number of existing observations upon which the prior mean has been estimated (similar to $\nu_0$ in the univariate case). $T_0$ can be thought of as the prior covariance matrix, or the scale matrix for a Wishart distribution. $\nu_0$ is the number of degrees of freedom for a Wishart distribution (must be ¿d where d is the number of dimensions).

By default, these hyperparameters are initialized to $\mu_0 = 0$, $\kappa_0 = 1e - 8$, $_0$=id, and $\nu_0 =$ d. Similarly to the univariate case, this choice of parameters approximates an improper flat prior over the mean (we cannot set $\kappa_0 = 0$ directly due to limitations from other packages we are using which do not accept improper distributions). Again, the default choice of parameters for the Wishart is not particularly principled, and requires further development.

Rather than using these default parameters, a much better option is to calculate reasonable hyperparameters from the data. Since the mean of a Wishart distribution is $\nu T_0$, setting $T_0 = \nu^{-1}\Sigma$, where $Sigma$ is the covariance matrix for the data, uses the covariance for the data as the prior mean for the cluster covariances.

## 2.2 Likelihood

$$F(y|\mu, \Lambda) = (2\pi)^{-d/2}|\Lambda|^{1/2}e^{-\frac{1}{2}(y-\mu)^T\Lambda(y-\mu)} \tag{10}$$

## 2.3 Prior

$$\begin{aligned} G_0(\mu, \Lambda|\mu_0, \kappa_0,_0, \nu_0) &= N(\mu|\mu_0, \kappa_0\Lambda) \cdot Wishart(\Lambda|_0, \nu_0) && (11) \\ &= \frac{|\Lambda|^{\frac{(\nu-d)}{2}} e^{-\frac{1}{2}tr(T^{-1}\Lambda)} e^{-\frac{1}{2}(\mu-\mu_0)^T\Lambda(\mu-\mu_0)}}{(2\pi)^{\frac{d}{2}} 2^{\frac{\nu d}{2}} |T|^{\frac{\nu}{2}} \Gamma_d(\frac{\nu}{2})} && (12) \end{aligned}$$

## 2.4 Posterior

$$\kappa_n = \kappa_0 + n \tag{13}$$
$$\nu_n = \nu_0 + n \tag{14}$$
$$\mu_n = \frac{\kappa_0}{\kappa_n}\mu_0 + \frac{n}{\kappa_n}\bar{x} \tag{15}$$
$$T_n^{-1} = T_0^{-1} + nS + \frac{n\kappa_0}{\kappa_n}(\bar{x} - \mu_0)(\bar{x} - \mu_0)^T \tag{16}$$

where $\bar{x} = \frac{1}{n}\sum_x x$ and $S^2 = \frac{1}{n}\sum_x (x - \bar{x})(x - \bar{x})^T$

## 2.5 Marginal Likelihood

$$p(y|\eta) = \int F(y|\mu, \lambda) dG_0(\mu, \Lambda|\mu_0, \kappa_{0,0}, \nu_0) \tag{17}$$
$$= \frac{1}{2\pi}\frac{\Gamma(\alpha_n)}{\Gamma\alpha_0}\sqrt{\frac{\nu_0}{\nu_n}}\frac{\beta_0^{\alpha_0}}{\beta_n^{\alpha_n}} \tag{18}$$