The Normal Exponential Family with Normal-Inverse-Gamma Prior

Yee Whye Teh ywteh@gatsby.ucl.ac.uk
Gatsby Computational Neuroscience Unit, UCL, 17 Queen Square, London WC1N 3AR, U.K.
http://www.gatsby.ucl.ac.uk/∼ywteh

Data:

n	The number of data items.
$\boldsymbol{x}$	The data items $x_1, \ldots, x_n$.
X	$= \sum_{i=1}^n x_i$.
C	$= \sum_{i=1}^n x_i^2$.

Parameters:

μ	Mean of data.
ρ	Precision of data.

Likelihood of data:

$$p(\boldsymbol{x} \mid \mu, \rho) = (2\pi)^{-n/2} \rho^{n/2} \exp\left(-\frac{1}{2}\left[\rho\left(\sum_{i=1}^n (\mu - x_i)^2\right)\right]\right) \tag{1}$$

$$= (2\pi)^{-n/2} \rho^{n/2} \exp\left(-\frac{1}{2}\left[\rho\left(n\mu^2 - 2\mu X + C\right)\right]\right) \tag{2}$$

Hyperparameters:

r	Relative precision of μ versus data. The precision of μ is $r\rho$.
ν	Degrees of freedom of precision of ρ.
m	Mean of μ is m.
s	Mean of ρ is ν/s.

Prior:

Gamma:
$$p(\rho) = 2^{-\nu/2} \frac{s^{\nu/2}}{\Gamma(\nu/2)} \rho^{\nu/2-1} \exp\left(-\frac{1}{2} s\rho\right) \tag{3}$$

Gaussian:
$$p(\mu \mid \rho) = (2\pi)^{-1/2} (r\rho)^{1/2} \exp\left(-\frac{1}{2}\left[r\rho(\mu - m)^2\right]\right) \tag{4}$$

$$p(\mu, \rho) = \frac{1}{Z(r, \nu, s)} \rho^{(\nu-1)/2} \exp\left(-\frac{1}{2}\left[\rho\left(r(\mu - m)^2 + s\right)\right]\right) \tag{5}$$

$$= \frac{1}{Z(r, \nu, s)} \rho^{(\nu-1)/2} \exp\left(-\frac{1}{2}\left[\rho\left(r\mu^2 - 2r\mu m + rm^2 + s\right)\right]\right) \tag{6}$$

where
$$Z(r, \nu, s) = 2^{(\nu+1)/2} \pi^{1/2} r^{-1/2} s^{-\nu/2} \Gamma\left(\nu/2\right) \tag{7}$$

Joint likelihood:

$$p(\boldsymbol{x}, \mu, \rho) = \frac{(2\pi)^{-n/2}}{Z(r, \nu, s)} \rho^{(\nu+n-1)/2} \exp\left(-\frac{1}{2}\left[\rho\left((r+n)\mu^2 - 2\mu(rm + X) + rm^2 + C + s\right)\right]\right) \tag{8}$$

$$= \frac{(2\pi)^{-n/2}}{Z(r, \nu, s)} \rho^{(\nu+n-1)/2} \exp\left(-\frac{1}{2}\left[\rho\left((r+n)\left(\mu - \tfrac{rm+X}{r+n}\right)^2 - \tfrac{(rm+X)^2}{r+n} + rm^2 + C + s\right)\right]\right) \tag{9}$$

Posterior hyperparameters:

$$\begin{aligned} r' &= r + n \\ \nu' &= \nu + n \\ m' &= \frac{rm + X}{r + n} \\ s' &= s + C + rm^2 - r'm'^2 \end{aligned} \tag{10}$$

Marginal likelihood:

$$p(\boldsymbol{x}) = (2\pi)^{-n/2} \frac{Z(r', \nu', s')}{Z(r, \nu, s)} = \pi^{-n/2} \frac{r'^{-1/2} s'^{-\nu'/2}}{r^{-1/2} s^{-\nu/2}} \frac{\Gamma(\nu'/2)}{\Gamma(\nu/2)} \tag{11}$$

Posterior:

$$p(\mu, \rho \,|\, \boldsymbol{x}) = \frac{1}{Z(r', \nu', s')} \rho^{(\nu'-1)/2} \exp\left(-\frac{1}{2}\left[\rho\left(r'(\mu - m')^2 + s'\right)\right]\right) \tag{12}$$

Predictive:

$$\begin{aligned} p(y|\boldsymbol{x}) = \frac{p(y, \boldsymbol{x})}{p(\boldsymbol{x})} &= (2\pi)^{-1/2} \frac{Z(r'', \nu'', s'')}{Z(r', \nu', s')} r'' && = r + n + 1 \\ \nu'' &= \nu + n + 1 \\ m'' &= \frac{rm + X + y}{r + n + 1} \\ s'' &= s + C + y^2 + rm^2 - r''m''^2 \end{aligned} \tag{13}$$

Derivatives of hyperparameters:

$$\frac{\partial \log p(\boldsymbol{x})}{\partial \log r} = \frac{1}{2} - \frac{r}{2r'} - \frac{\nu' r (m - m')^2}{2s'} \tag{14}$$

$$\frac{\partial \log p(\boldsymbol{x})}{\partial \log \nu} = \frac{\nu}{2}\left(\log s - \log s' + \Psi\left(\frac{\nu'}{2}\right) - \Psi\left(\frac{\nu}{2}\right)\right) \tag{15}$$

$$\frac{\partial \log p(\boldsymbol{x})}{\partial m} = -\frac{\nu' r (m - m')}{s'} \tag{16}$$

$$\frac{\partial \log p(\boldsymbol{x})}{\partial \log s} = \frac{\nu}{2} - \frac{\nu' s}{2s'} \tag{17}$$