USEFUL PROPERTIES OF THE MULTIVARIATE NORMAL*

3.1. Conditionals and marginals

For Bayesian analysis it is very useful to understand how to write joint, marginal, and conditional distributions for the multivariate normal.

Given a vector $x \in \mathbb{R}^p$ the multivariate normal density is

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Now split the vector into two parts

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \text{of size } \begin{bmatrix} q \times 1 \\ (p-q) \times 1 \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \text{of size } \begin{bmatrix} q \times q & q \times (p-q) \\ (p-q) \times q & (p-q) \times (p-q) \end{bmatrix}.$$

We now state the joint and marginal distributions

$$x_1 \sim \mathrm{N}(\mu_1, \Sigma_{11}), \quad x_2 \sim \mathrm{N}(\mu_2, \Sigma_{22}), \quad x \sim \mathrm{N}(\mu, \Sigma),$$

and the conditional density

$$x_1 \mid x_2 \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

The same idea holds for other sizes of partitions.

3.2. Conjugate priors

3.2.1. Univariate normals

3.2.1.1. *Fixed variance, random mean.* We consider the parameter σ^2 fixed so we are interested in the conjugate prior for μ:

$$\pi(\mu \mid \mu_0, \sigma^2) \propto \frac{1}{\sigma_0} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right),$$

where μ_0 and σ^2 are hyper-parameters for the prior distribution (when we don't have informative prior knowledge we typically consider $\mu_0 = 0$ and σ^2 large).

The posterior distribution for $x_1, ..., x_n$ with a univariate normal likelihood and the above prior will be

$$\text{Post}(\mu \mid x_1, ..., x_n) \sim \text{N}\left(\frac{\sigma_0^2}{\frac{\sigma^2}{n} + \sigma_0^2} \bar{x} + \frac{\sigma^2}{\frac{\sigma^2}{n} + \sigma_0^2} \mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1} \right).$$

3.2.1.2. *Fixed mean, random variance.* We will formulate this setting with two parameterizations of the scale parameter: (1) the variance σ^2, (2) the precision $\tau = \frac{1}{\sigma^2}$.

The two conjugate distributions are the Gamma and the inverse Gamma (really they are the same distribution, just reparameterized)

$$\text{IG}(\alpha, \beta) : f(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp(-\beta(\sigma^2)^{-1}), \quad \text{Ga}(\alpha, \beta) : f(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\beta\tau).$$

The posterior distribution of σ^2 is

$$\sigma^2 \mid x_1, ..., x_n \sim \text{IG}\Big(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (x_i - \mu)^2 \Big).$$

The posterior distribution of τ is not surprisingly

$$\tau \mid x_1, ..., x_n \sim \text{Ga}\Big(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (x_i - \mu)^2 \Big).$$

3.2.1.3. *Random mean, random variance.* We now put the previous priors together in what is called a Bayesian hierarchical model:

$$\begin{aligned} x_i \mid \mu, \tau &\overset{iid}{\sim} \text{N}(\mu, (\tau)^{-1}) \\ \mu \mid \tau &\sim \text{N}(\mu_0, (\kappa_0 \tau)^{-1}) \\ \tau &\sim \text{Ga}(\alpha, \beta). \end{aligned}$$

For the above likelihood and priors the posterior distribution for the mean and precision is

$$\begin{aligned} \mu \mid \tau, x_1, ..., x_n &\sim \text{N}\Big(\frac{\mu_0 \kappa_0 + n\bar{x}}{n + \kappa_0}, (\tau(n + \kappa_0))^{-1} \Big) \\ \tau \mid x_1, ..., x_n &\sim \text{Ga}\Big(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (x_i - \bar{x})^2 + \frac{n}{n+1} \frac{(\bar{x} - x_i)^2}{2} \Big). \end{aligned}$$

3.2.2. Multivariate normal

Given a vector $x \in \mathbb{R}^p$ the multivariate normal density is

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right).$$

We will work with the precision matrix instead of the covariance and we will consider the following Bayesian hierarchical model:

$$\begin{aligned} x_i \mid \mu, \Lambda &\overset{iid}{\sim} \text{N}(\mu, (\Lambda)^{-1}) \\ \mu \mid \Lambda &\sim \text{N}(\mu_0, (\kappa_0 \Lambda)^{-1}) \\ \Lambda &\sim \text{Wi}(\Lambda_0, n_0), \end{aligned}$$

the precision matrix is modeled using the Wishart distribution

$$f(\Lambda; V, n) = \frac{|\Lambda|^{(n-d-1)/2} \exp(-.5\text{tr}(\Lambda V^{-1}))}{2^{nd/2} |V|^{n/2} \Gamma_d(n/2)}.$$

For the above likelihood and priors the posterior distribution for the mean and precision is

$$\begin{aligned} \mu \mid \Lambda, x_1, ..., x_n &\sim \mathrm{N}\Big(\frac{\mu_0\kappa_0 + n\bar{x}}{n+\kappa_0}, (\Lambda(n+\kappa_0))^{-1}\Big) \\ \Lambda \mid x_1, ..., x_n &\sim \mathrm{Wi}\left(n_0 + \frac{n}{2}, \Lambda_0 + \frac{1}{2}\left[\bar{\Sigma} + \frac{\kappa_0}{\kappa_0 + n}(\bar{x} - \mu_0)(\bar{x} - \mu_0)^T\right]\right). \end{aligned}$$