3: Introduction to multiparameter models

09/11/19

Univariate Normal with unknown mean and variance

Likelihood

$$p(y \mid \mu, \sigma^2) = (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]$$

Prior

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu|\sigma^2) = \text{inv-gamma}(\nu_0/2, \sigma_0^2\nu_0/2)\text{N}(\mu_0, \sigma^2/\kappa_0)$$

Posterior

$$\begin{aligned} p(\mu, \sigma^2|y) = p(\sigma^2|y)p(\mu|\sigma^2, y) &= \text{inv-gamma}(\nu_n/2, \sigma_n^2\nu_n/2)\text{N}(\mu_n, \sigma^2/\kappa_n) \\ &= \text{inverse-}\chi^2(\nu_n, \sigma_n^2)\text{N}(\mu_n, \sigma^2/\kappa_n) \end{aligned}$$

$$p(\mu|y) = \text{t}_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$$

In-class discussion

(Wednesday Sep 18)

1. Investigate the role of priors in the univariate normal likelihood model using simulation.
2. Investigate $p(\mu, \sigma^2 | y_{new}, y_{old})$.

Multivariate Normal Observations

Let each observation y_i follow a multivariate normal distribution. The likelihood $p(y_1, \ldots, y_n \mid \mu, \Sigma)$ is usefully written with a few properties of the trace operator:

$$
\begin{aligned}
&\propto \det(\Sigma)^{-n/2} \exp\left(-\frac{1}{2}\sum_i (y_i - \mu)'\Sigma^{-1}(y_i - \mu)\right) \\
&= \det(\Sigma)^{-n/2} \exp\left[-\frac{1}{2}\sum_i \mathrm{tr}\{(y_i - \mu)'\Sigma^{-1}(y_i - \mu)\}\right] \\
&= \det(\Sigma)^{-n/2} \exp\left[-\frac{1}{2}\sum_i \mathrm{tr}\{\Sigma^{-1}(y_i - \mu)(y_i - \mu)'\}\right] \\
&= \det(\Sigma)^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\overbrace{\sum_i (y_i - \mu)(y_i - \mu)'}^{S_0}\right\}\right]
\end{aligned}
$$

Multivariate Normal Observations with known covariance matrix

A conjugate prior for $p(y \mid \mu) \propto \det(\Sigma)^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\overbrace{\Sigma^{-1}}^{\text{known}} S_0\right\}\right]$ is

$$p(\mu \mid \mu_0, \Lambda_0) = \det(\Lambda_0)^{-1/2} \exp\left[(\mu - \mu_0)'\Lambda_0^{-1}(\mu - \mu_0)\right]$$

This makes the posterior distribution (homework question exercise 3.13) normal with mean and precision

$$\mu_n = (\Lambda_0 + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y})$$

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}.$$

Multivariate Normal Observations with unknown covariance matrix

When all of the elements of Σ are unknown, we need a prior for that as well. This prior must put zero mass on matrices that aren't positive definite or aren't symmetric.

A popular option is the **inverse Wishart** distribution, which is analagous to the inverse-Gamma distribution. It has a degrees of freedom parameter: ν_0. And it has a scale matrix parameter Λ_0.

Multivariate Normal Observations with unknown covariance matrix

When all of the elements of Σ are unknown, we need a prior for that as well. This prior must put zero mass on matrices that aren't positive definite or aren't symmetric.

A popular option is the **inverse Wishart** distribution, which is analagous to the inverse-Gamma distribution. It has a degrees of freedom parameter: ν_0. And it has a scale matrix parameter Λ_0.

If $\Sigma \in \mathbb{R}^{d \times d}$, we will write

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})$$

and we can write (something proportional to) the density as

$$p(\Sigma) \propto \det(\Sigma)^{-(\nu_0+d+1)/2} \exp\left(-\frac{1}{2}\text{tr}\left[\Lambda_0 \Sigma^{-1}\right]\right)$$

Multivariate Normal Observations with unknown covariance matrix

The following is a conjugate prior

$$
\begin{aligned}
p(\mu \mid \Sigma)p(\Sigma) &= \mathrm{N}(\mu_0, \Sigma/\kappa_0)\text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}) \\
&\propto \left[\det(\Sigma)^{-1/2} \exp\left(-\frac{\kappa_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right)\right] \times \\
&\quad \left[\det(\Sigma)^{-(\nu_0+d+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\Lambda_0\Sigma^{-1}\right]\right)\right]
\end{aligned}
$$

Multivariate Normal Observations with unknown covariance matrix

Here's the posterior:

$$
\begin{aligned}
p(\mu, \Sigma \mid y) &\propto p(y \mid \mu, \Sigma) p(\mu \mid \Sigma) p(\Sigma) \\
&\propto \det(\Sigma)^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\sum_i (\mu - y_i)(\mu - y_i)'\right\}\right] \times \\
&\quad \det(\Sigma)^{-1/2} \exp\left(-\frac{\kappa_0}{2}(\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)\right) \times \\
&\quad \det(\Sigma)^{-(\nu_0+d+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\Lambda_0 \Sigma^{-1}\right]\right) \\
&= \cdots
\end{aligned}
$$

Multivariate Normal Observations with unknown covariance matrix

It helps to recognize $p(\mu \mid \Sigma, y)$ first, and then $p(\Sigma \mid y)$. Here is negative twice the log of the exponent:

$$
\begin{aligned}
&\operatorname{tr}\left\{\Sigma^{-1}\sum_i(\mu - y_i)(\mu - y_i)' + \kappa_0(\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)\right\} + c_1 \\
&= \operatorname{tr}\left\{\Sigma^{-1}n(\mu - \bar{y})(\mu - \bar{y})^T + \kappa_0(\mu - \mu_0)(\mu - \mu_0)^T\right\} \\
&+ \operatorname{tr}\left\{\Sigma^{-1}\sum_i(y_i - \bar{y})(y_i - \bar{y})^T\right\} + c_1 \\
&= \operatorname{tr}\left\{B(\mu - \mu_n)(\mu - \mu_n)^T\right\} + \operatorname{tr}\left\{\Sigma^{-1}\frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T\right\} + c_2 \\
&= (\mu - \mu_n)'B(\mu - \mu_n) + c_3
\end{aligned}
$$

where $B = (\Sigma/n)^{-1} + (\Sigma/\kappa_0)^{-1}$ and $\mu_n = B^{-1}\left[(\Sigma/n)^{-1}\bar{y} + (\Sigma/\kappa_0)^{-1}\mu_0\right]$

Multivariate Normal Observations with unknown covariance matrix

Clearly

$$B = (\Sigma/n)^{-1} + (\Sigma/\kappa_0)^{-1} = \Sigma^{-1}(n + \kappa_0)$$

and

$$\mu_n = B^{-1}\left[(\Sigma/n)^{-1}\bar{y} + (\Sigma/\kappa_0)^{-1}\mu_0\right] = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}$$

Multivariate Normal Observations with unknown covariance matrix

Now $p(\Sigma|y)$. Denote $S = \sum_i (y_i - \bar{y})(y_i - \bar{y})^T$

$$p(\Sigma|y) \propto \det(\Sigma)^{-(\nu_0+n+d+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\Lambda_0\Sigma^{-1} + S\Sigma^{-1} + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T\Sigma^{-1}\right]\right)$$

$p(\Sigma|y) \sim \text{Inv-Wishart}_{\nu_n}(\Lambda_n^{-1})$
$\nu_n = \nu_0 + n$
$\Lambda_n = \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T$