Lecture 7 - Key

Multivariate Normal Distribution

While the normal distribution has two parameters, up to now we have focused on univariate data. Now we will consider multivariate responses from a multivariate normal distribution.

Example.

Multivariate Normal Distribution: The multivariate normal distribution has the following sampling distribution:

$$p(\tilde{y}|\tilde{\theta}, \Sigma) =$$

where

$$\tilde{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \qquad \tilde{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,p} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,p} & \sigma_{2,p} & \cdots & \sigma_p^2 \end{pmatrix}$$

The vector $\tilde{\theta}$ is the mean vector and the matrix Σ is the covariance matrix, where the diagonal elements are the variance terms for observation i and the off diagonal elements are the covariance terms between observation i and j.

Marginally, each $y_i \sim$

Linear Algebra Review

- Let A be a matrix, then $|A|$ is the
  For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $|A| =$.

- Let A be a matrix, then A^{-1} is the **inverse** of A such that

- Let $\tilde{\theta}$ be a vector of dimension $p \times 1$, $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}$, then the transpose of θ, denoted θ^T is a vector of dimension $1 \times p$. Then $\theta^T = (\theta_1, \theta_2, \cdots, \theta_p)$.

- Let A be a $n \times p$ matrix and B be a $p \times m$ matrix, then the matrix product, AB will be a $n \times m$ matrix.

- Let A be a matrix, then the **trace** of A is the sum of the diagonal elements. A useful property of the trace is that

- The `mnormt` package in R includes functions `rmnorm, dmnorm` which are multivariate analogs of functions used for a univariate normal distribution.

Multivariate Normal Exercise

1. Simulate data from two dimensional multivariate normal distributions.
    a. select a variety of mean and covariance structures and visualize your results.
    b. how do the mean and covariance structures impact the visualization?

2. Similar to our earlier exercise using the Gibbs sample on the tri-modal example, create a two-dimensional mixture distribution of multivariate normal distributions and plot your results.

Priors for multivariate normal distribution

In the univariate normal setting, a normal prior for the mean term was *semiconjugate*. Does the same hold for a multivariate setting? Let $p(\tilde{\theta}) \sim N_p(\tilde{\mu}_0, \Lambda_0)$.

$$\begin{aligned}
p(\tilde{\theta}) &= (2\pi)^{-p/2}|\Lambda_0|^{-1/2}\exp\left[-\frac{1}{2}(\tilde{\theta}-\tilde{\mu}_0)^T\Lambda_0^{-1}(\tilde{\theta}-\tilde{\mu}_0)\right] \\
&\propto \\
&\propto
\end{aligned}$$

Now combine this with the sampling model, only retaining the elements that contain θ.

$$\begin{aligned}
p(\tilde{y}_1,\ldots,\tilde{y}_n|\tilde{\theta},\Sigma) &\propto \prod_{i=1}^{n}\exp\left[-\frac{1}{2}(\tilde{y}_i-\tilde{\theta})^T\Sigma^{-1}(\tilde{y}_i-\tilde{\theta})\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\tilde{y}_i-\tilde{\theta})^T\Sigma^{-1}(\tilde{y}_i-\tilde{\theta})\right] \\
&\propto
\end{aligned}$$

Next we find the full conditional distribution for θ, $p(\tilde{\theta}|\Sigma,\tilde{y}_1,\ldots,\tilde{y}_n)$.

$$p(\tilde{\theta}|\Sigma,\tilde{y}_1,\ldots,\tilde{y}_n) \propto$$

We have a similar trick to the univariate normal case, where the variance (matrix) is A^{-1} and the expectation is $A^{-1}B$.

Hence the full conditional follows a multivariate normal distribution with variance $\Lambda_n =$

and expectation= $\tilde{\mu}_n$

Sometimes a uniform prior $p(\tilde{\theta}) \propto \tilde{1}$ is used. In this case the variance and expectation simplify to $V = \Sigma/n$ and $E = \bar{\tilde{y}}$.

Using this semiconjugate prior in a Gibbs sampler we can make draws from the full conditional distribution using `rmnorm(.)` in R. However, we still need to be able to take samples of the covariance matrix Σ to get draws from the joint posterior distribution.

Inverse-Wishart Distribution

A covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,p} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,p} & \sigma_{2,p} & \cdots & \sigma_p^2 \end{pmatrix}$ has the variance terms on the diagonal and covariance terms for off diagonal elements.

Similar to the requirement that σ^2 be positive, a covariance matrix, Σ must be **positive definite**, such that: $\tilde{x}^T \Sigma \tilde{x} > 0$ for all vectors $\tilde{x}$. With a positive definite matrix, the diagonal elements (which correspond the marginal variances σ_j^2) are greater than zero and it also constrains the correlation terms to be between -1 and 1.

A covariance matrix also requires symmetry, so that

The covariance matrix is closely related to the sum of squares matrix with is given by:

where $z_1, ..., z_n$ are $p \times 1$ vectors containing a multivariate response.

Thus $\tilde{z}_i \tilde{z}_i^T$ results in a $p \times p$ matrix, where

$$\tilde{z}_i \tilde{z}_i^T = \begin{pmatrix} z_{i,1}^2 & z_{i,1}z_{i,2} & \cdots & z_{i,1}z_{i,p} \\ z_{i,2}z_{i,1} & z_{i,2}^2 & \cdots & z_{i,2}z_{i,p} \\ \vdots & & & \vdots \\ z_{i,p}z_{i,1} & z_{i,p}z_{i,2} & \cdots & z_{i,p}^2 \end{pmatrix}$$

Now let the $\tilde{z}_i's$ have zero mean (are centered). Recall that the sample variance is computed as $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2/(n-1) = \sum_{i=1}^n z_i^2/(n-1)$. Similarly the matrix $\tilde{z}_i \tilde{z}_i^T/n$ is the contribution of the i^{th} observation to the sample covariance.

In this case:

1. $\frac{1}{n}[Z^T Z]_{j,j} = \frac{1}{n}\sum_{i=1}^n z_{i,j}^2 = s_j^2$

2. $\frac{1}{n}[Z^T Z]_{j,k} = \frac{1}{n}\sum_{i=1}^n z_{i,j} z_{i,k} = s_{j,k}$

If $n > p$ and the $\tilde{z}_i's$ are linearly independent then $Z^T Z$ will be positive definite and symmetric.

Consider the following procedure with a positive integer, ν_0, and a $p \times p$ covariance matrix Φ_0:

- sample

- calculate

then the matrix $Z^T Z$ is a random draw from a

The expectation of $Z^T Z$ is $\nu_0 \Phi_0$. The Wishart distribution can be thought of as a multivariate analogue of the gamma distribution.

Accordingly, the Wishart distribution is semi-conjugate for the precision matrix (Σ^{-1}); whereas, the Inverse-Wishart distribution is semi-conjugate for the covariance matrix.

The density of the inverse-Wishart distribution with parameters S_0^{-1}, a $p \times p$ matrix and ν_0 ($IW(\nu_0, S_0^{-1})$ is:

$$
\begin{aligned}
p(\Sigma) &= \left[2^{\nu_0 p/2} \pi^{\binom{p}{n}/2} |S_0|^{-\nu_0/2} \prod_{j=1}^{P} \Gamma([\nu_0 + 1 - j]/2)\right]^{-1} \times \\
& \quad |\Sigma|^{-(\nu_0+p+1)/2} \times \exp[-tr(S_0 \Sigma^{-1})/2]
\end{aligned}
$$

Inverse Wishart Full Conditional Calculations

$$
\begin{aligned}
p(\Sigma|\tilde{y}_1, \ldots, \tilde{y}_n, \tilde{\theta}) &\propto p(\Sigma) \times p(\tilde{y}_1, \ldots, \tilde{y}_n|\Sigma, \tilde{\theta}) \\
&\propto \left(|\Sigma|^{-(\nu_0+p+1)/2} \exp[-tr(S_0 \Sigma^{-1})/2]\right) \times \left(|\Sigma|^{-n/2} \exp[-\frac{1}{2} \sum_{i=1}^{n} (\tilde{y}_i - \tilde{\theta})^T \Sigma^{-1} (\tilde{y}_i - \tilde{\theta})]\right) \\
& \quad \text{note that } \sum_{i=1}^{n} (\tilde{y}_i - \tilde{\theta})^T \Sigma^{-1} (\tilde{y}_i - \tilde{\theta}) \text{ is a number so we can apply the trace operator} \\
& \quad \text{using properties of the trace, this equals } tr\left(\sum_{i=1}^{n} (\tilde{y}_i - \tilde{\theta})(\tilde{y}_i - \tilde{\theta})^T \Sigma^{-1}\right) = tr(S_\theta \Sigma^{-1}) \\
\text{so } p(\Sigma|\tilde{y}_1, \ldots, \tilde{y}_n, \tilde{\theta}) &\propto \left(|\Sigma|^{-(\nu_0+p+1)/2} \exp[-tr(S_0 \Sigma^{-1})/2]\right) \times \left(|\Sigma|^{-n/2} \exp[-tr(S_\theta \Sigma^{-1}/2)]\right) \\
&\propto \left(|\Sigma|^{-(\nu_0+n+p+1)/2} \exp[-tr([S_0 + S_\theta]\Sigma^{-1})/2]\right) \\
\text{thus} \Sigma|\tilde{y}_1, \ldots, \tilde{y}_n, \tilde{\theta} &\sim IW(\nu_0 + n, [S_0 + S_\theta]^{-1})
\end{aligned}
$$

Thinking about the parameters in the prior distribution, ν_0 is the prior sample size and S_0 is the prior residual sum of squares.

Gibbs Sampling for Σ and $\tilde{\theta}$

We now know that the full conditional distributions follow as:

$$\begin{aligned}
\tilde{\theta}|\Sigma, \tilde{y}_1, \ldots, \tilde{y}_n &\sim MVN(\mu_n, \Lambda_n) \\
\Sigma|\tilde{\theta}, \tilde{y}_1, \ldots, \tilde{y}_n &\sim IW(\nu_n, S_n^{-1}).
\end{aligned}$$

Given these full conditional distributions the Gibbs sampler can be implemented as:

1. Sample $\tilde{\theta}^{(j+1)}$ from the full conditional distribution
   a. compute $\tilde{\mu}_n$ and Λ_n from $\tilde{y}_1, \ldots, \tilde{y}_n$ and $\Sigma^{(j)}$

   b. sample $\tilde{\theta}^{(j+1)} \sim MVN(\tilde{\mu}_n, \Lambda_n)$. This can be done with `rmnorm(.)` in R.

2. Sample $\Sigma^{(j+1)}$ from its full conditional distribution
   a. compute S_n from $\tilde{y}_1, \ldots, \tilde{y}_n$ and $\tilde{\theta}^{(j+1)}$

   b. sample $\Sigma^{(j+1)} \sim IW(\nu_0 + n, S_n^{-1})$

   As $\tilde{\mu}_n$ and Λ_n depend on Σ they must be calculated every iteration. Similarly, S_n depends on $\tilde{\theta}$ and needs to be calculated every iteration as well.

Gibbs Sampler Exercise

Now extend our first Gibbs sampler to accomodate multivariate data. This will require simulating multivariate responses.