Module 9: The Multivariate Normal Distribution

Rebecca C. Steorts
Rest of semester

- Last day of class is Tuesday, April 18
- Last homework is HW 7 (see Sakai). This goes along with Lab 6.
- For HW 7, your TAs have gone over tasks 1-3 with you and posted solutions are available.
- Code has been posted for Task 4 - 5.
- TAs will help with the remainder of the assignment in lab on April 12. If you miss lab, no reviews will be done.
Final exam

- You may bring a cheat sheet into the exam (front and back). It must be a standard 9 in by 11 inch side piece of paper. You must turn this in with your final exam.
- Final exam is posted on the syllabus (no make ups). May 4, 7–10, Old Chem 116.
- Final exam is cumulative and will be similar to exam I.
- Practice problems have now been posted. Solutions will be posted after April 19th.
- I will do a review class on Tuesday, April 25. This is optional.
Moving from univariate to multivariate distributions.

The multivariate normal (MVN) distribution.

Conjugate for the MVN distribution.

The inverse Wishart distribution.

Conjugate for the MVN distribution (but on the covariance matrix).

Combining the MVN with inverse Wishart.

See Chapter 7 (Hoff) for a review of the standard Normal density.
Notation

Assume a matrix of covariates

\[ X_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ x_{i1} & x_{i2} & \cdots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \]

- A column of x represents a particular covariate we might be interested in, such as age of a person.
- Denote \( x_i \) as the ith row vector of the \( X_{n \times p} \) matrix.

\[ x_i = \begin{pmatrix} x_{i1} \\ x_{ip} \\ \vdots \\ x_{ip} \end{pmatrix} \]
Distribution of MVN

We assume that the population mean is $\mu = E(X)$ and $\Sigma = \text{Var}(X) = E[(X - \mu)(X - \mu)^T]$, where

$$
\mu = \begin{pmatrix}
\mu_1 \\
\mu_2 \\
\vdots \\
\mu_p
\end{pmatrix}
$$

and

$$
\Sigma = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\
\sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{p1} & \sigma_{p2} & \ldots & \sigma_p^2
\end{pmatrix}.
$$
Notation

Suppose matrix $A$ is invertible. The

$$\det(A) = \sum_{i=1}^{j=n} a_{ij}A_{ij}.$$ 

I recommend using the det() commend in R.

Suppose now we have a square matrix $H_{p \times p}$.

$$\text{trace}(H) = \sum_{i} h_{ii},$$

where $h_{ii}$ are the diagonal elements of $H$. 
- MVN is generalization of univariate normal.
- For the MVN, we write $\mathbf{X} \sim \mathcal{MVN}(\mu, \Sigma)$.
- The $(i, j)^{th}$ component of $\Sigma$ is the covariance between $X_i$ and $X_j$ (so the diagonal of $\Sigma$ gives the component variances).

Example: $\text{Cov}(X_1, X_2)$ is just one element of the matrix $\Sigma$. 
Multivariate Normal

Just as the probability density of a scalar normal is

\[ p(x) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right\}, \]

(1)

the probability density of the multivariate normal is

\[ p(\mathbf{x}) = (2\pi)^{-p/2} (\det \Sigma)^{-1/2} \exp \left\{-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\}. \]

(2)

Univariate normal is special case of the multivariate normal with a one-dimensional mean “vector” and a one-by-one variance “matrix.”
Standard Multivariate Normal Distribution

Consider

\[ Z_1, \ldots, Z_n \overset{iid}{\sim} \text{MVN}(0, I) \]

\[ f_z(z) = \prod_{i=1}^{n} \frac{1}{2\pi} e^{-z_i^2/2} \]

\[ = (2\pi)^{-n} e^{z^T z/2} \]  

\begin{itemize}
  \item \( \mathbb{E}[Z] = 0 \)
  \item \( \text{Var}[Z] = I \)
\end{itemize}
Suppose that

\[ X_1 \ldots X_n \overset{iid}{\sim} \text{MVN}(\theta, \Sigma). \]

Let

\[ \pi(\theta) \sim \text{MVN}(\mu, \Omega). \]

What is the full conditional distribution of \( \theta \mid \mathbf{X}, \Sigma \)?
\[
\pi(\theta) = (2\pi)^{-p/2} \det \Omega^{-1/2} \exp \left\{ -\frac{1}{2} (\theta - \mu)^T \Omega^{-1}(\theta - \mu) \right\}
\]

(5)

\[
\propto \exp \left\{ -\frac{1}{2} (\theta - \mu)^T \Omega^{-1}(\theta - \mu) \right\}
\]

(6)

\[
\propto \exp -\frac{1}{2} \left\{ \theta^T \Omega^{-1} \theta - 2\theta^T \Omega^{-1} \mu + \mu^T \Omega^{-1} \mu \right\}
\]

(7)

\[
\propto \exp -\frac{1}{2} \left\{ \theta^T \Omega^{-1} \theta - 2\theta^T \Omega^{-1} \mu \right\}
\]

(8)

\[
= \exp -\frac{1}{2} \left\{ \theta^T A_o \theta - 2\theta^T b_o \right\}
\]

(9)

\[
\pi(\theta) \sim \text{MVN}(\mu, \Omega) \text{ implies that } A_o = \Omega^{-1} \text{ and } b_o = \Omega^{-1} \mu.
\]
Likelihood

\[ p(\mathbf{X} \mid \theta, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} \det \Sigma^{-n/2} \exp \left\{ -\frac{1}{2} (x_i - \theta)^T \Sigma^{-1} (x_i - \theta) \right\} \]

(10)

\[ \propto \exp -\frac{1}{2} \left\{ \sum_i x_i^T \Sigma^{-1} x_i - 2 \sum_i \theta^T \Sigma^{-1} x_i + \sum_i \theta^T \Sigma^{-1} \theta \right\} \]

(11)

\[ \exp -\frac{1}{2} \left\{ -2\theta^T \Sigma^{-1} n\bar{x} + n\theta^T \Sigma^{-1} \theta \right\} \]

(12)

\[ \exp -\frac{1}{2} \left\{ -2\theta^T b_1 + \theta^T A_1 \theta \right\}, \]

(13)

where

\[ b_1 = \Sigma^{-1} n\bar{x}, \quad A_1 = n\Sigma^{-1} \]

and

\[ \bar{x} := \left( \frac{1}{n} \sum_i x_{i1}, \ldots, \frac{1}{n} \sum_i x_{ip} \right)^T. \]
$p(\theta \mid X, \Sigma) \propto p(X \mid \theta, \Sigma) \times p(\theta)$ \hspace{1cm} (14)

$\propto \exp -\frac{1}{2} \left\{ -2\theta^T b_1 + \theta^T A_1 \theta \right\}$ \hspace{1cm} (15)

$\times \exp -\frac{1}{2} \left\{ \theta^T A_o \theta - 2\theta^T b_o \right\}$ \hspace{1cm} (16)

$\propto \exp \left\{ \theta^T b_1 - \frac{1}{2} \theta^T A_1 \theta - \frac{1}{2} \theta^T A_o \theta + \theta^T b_o \right\}$ \hspace{1cm} (17)

$\propto \exp \left\{ \theta^T (b_o + b_1) - \frac{1}{2} \theta^T (A_o + A_1) \theta \right\}$ \hspace{1cm} (18)

Then

$A_n = A_o + A_1 = \Omega^{-1} + n\Sigma^{-1}$

and

$b_n = b_o + b_1 = \Omega^{-1} \mu + \Sigma^{-1} n\bar{x}$

$\theta \mid X, \Sigma \sim MVN(A_n^{-1} b_n, A_n^{-1}) = MVN(\mu_n, \Sigma_n)$
Interpretations

\[ \theta | \mathbf{X}, \Sigma \sim \text{MVN}(A_n^{-1} b_n, A_n^{-1}) = \text{MVN}(\mu_n, \Sigma_n) \]

\[ \mu_n = A_n^{-1} b_n = [\Omega^{-1} + n\Sigma^{-1}]^{-1}(b_o + b_1) \]  
\[ = [\Omega^{-1} + n\Sigma^{-1}]^{-1}(\Omega^{-1} \mu + \Sigma^{-1} n\bar{x}) \]  
\[ \Sigma_n = A_n^{-1} = [\Omega^{-1} + n\Sigma^{-1}]^{-1} \]
Suppose \( \Sigma \sim \text{inverseWishart}(\nu_0, S_o^{-1}) \) where \( \nu_0 \) is a scalar and \( S_o^{-1} \) is a matrix.

Then

\[
p(\Sigma) \propto \det(\Sigma)^{-(\nu_0+p+1)/2} \times \exp\left\{ -\text{tr}(S_o \Sigma^{-1})/2 \right\}
\]

For the full distribution, see Hoff, Chapter 7 (p. 110).
The inverse Wishart distribution is the multivariate version of the Gamma distribution. The full hierarchy we’re interested in is

\[ X \mid \theta, \Sigma \sim MVN(\theta, \Sigma). \]
\[ \theta \sim MVN(\mu, \Omega) \]
\[ \Sigma \sim \text{inverseWishart}(\nu_o, S_o^{-1}). \]

We first consider the conjugacy of the MVN and the inverse Wishart, i.e.

\[ X \mid \theta, \Sigma \sim MVN(\theta, \Sigma). \]
\[ \Sigma \sim \text{inverseWishart}(\nu_o, S_o^{-1}). \]
What about \( p(\Sigma \mid X, \theta) \propto p(\Sigma) \times p(X \mid \theta, \Sigma) \). Let’s first look at

\[
p(X \mid \theta, \Sigma) \propto \det(\Sigma)^{-n/2} \exp\left\{ -\sum_i (X_i - \theta)^T \Sigma^{-1} (X_i - \theta) / 2 \right\}
\]

(22)

\[
\propto \det(\Sigma)^{-n/2} \exp\left\{ -\text{tr}\left( \sum_i (X_i - \theta)(X_i - \theta)^T \Sigma^{-1} / 2 \right) \right\}
\]

(23)

\[
\propto \det(\Sigma)^{-n/2} \exp\left\{ -\text{tr}(S_\theta \Sigma^{-1} / 2) \right\}
\]

(24)

\[
\propto \det(\Sigma)^{-n/2} \exp\left\{ -\text{tr}(S_\theta \Sigma^{-1} / 2) \right\}
\]

(25)

where \( S_\theta = \sum_i (X_i - \theta)(X_i - \theta)^T \).

Fact:

\[
\sum_k b_k^T A b_k = \text{tr}(B^T BA),
\]

where \( B \) is the matrix whose \( k \)th row is \( b_k \).
Now we can calculate $p(\Sigma \mid X, \theta)$

$$p(\Sigma \mid X, \theta) = p(\Sigma) \times p(X \mid \theta, \Sigma)$$

$$\propto \text{det}(\Sigma)^{-(\nu_o+p+1)/2} \times \exp\{-\text{tr} (S_o \Sigma^{-1})/2\}$$

$$\times \text{det}(\Sigma)^{-n/2} \exp\{-\text{tr} (S_\theta \Sigma^{-1})/2\}$$

$$\propto \text{det}(\Sigma)^{-(\nu_o+n+p+1)/2} \exp\{-\text{tr} ((S_o + S_\theta) \Sigma^{-1})/2\}$$

This implies that

$$\Sigma \mid X, \theta \sim \text{inverseWishart}(\nu_o + n, [S_o + S_\theta]^{-1} =: S_n)$$
Continued

Suppose that we wish now to take

\[ \theta \mid X, \Sigma \sim \text{MVN}(\mu_n, \Sigma_n) \]

(which we finished an example on earlier). Now let

\[ \Sigma \mid X, \theta \sim \text{inverseWishart}(\nu_n, S_n^{-1}) \]

There is no closed form expression for this posterior. Solution?
Suppose the Gibbs sampler is at iteration $s$.

1. Sample $\theta^{(s+1)}$ from it’s full conditional:
   a) Compute $\mu_n$ and $\Sigma_n$ from $X$ and $\Sigma^{(s)}$
   b) Sample $\theta^{(s+1)} \sim MVN(\mu_n, \Sigma_n)$

2. Sample $\Sigma^{(s+1)}$ from its full conditional:
   a) Compute $S_n$ from $X$ and $\theta^{(s)}$
   b) Sample $\Sigma^{(s+1)} \sim \text{inverseWishart}(\nu_n, S_n^{-1})$
The R package, `mvtnorm`, contains functions for evaluating and simulating from a multivariate normal density.

```r
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 3.3
```
Simulating Data

Simulate a single multivariate normal random vector using the `rmvnorm` function.

```r
rmvnorm(n = 1, mean = rep(0, 2), sigma = diag(2))
```

```r
## [,1]       [,2]
## [1,] -0.4956353 -0.02587638
```
Evaluate the multivariate normal density at a single value using the `dmvnorm` function.

```r
dmvnorm(rep(0, 2), mean = rep(0, 2), sigma = diag(2))
```

```r
# [1] 0.1591549
```
Now let’s simulate many multivariate normals.
Each row is a different sample from this multivariate normal distribution.

```r
rmvnorm(n = 3, mean = rep(0, 2), sigma = diag(2))
```

```
[,1]       [,2]
[1,] 0.48096760 1.857402
[2,] 0.73460381 -1.579971
[3,] -0.06040684 -1.885633
```
We can evaluate the multivariate normal density at several values using the `dmvnorm` function.

```r
dmvnorm(rbind(rep(0, 2), rep(1, 2), rep(2, 2)),
    mean = rep(0, 2), sigma = diag(2))
```

```
## [1] 0.159154943 0.058549832 0.002915024
```
Work with the Wishart density

- The R package, stats, contains functions for evaluating and simulating from a Wishart density.
- We can simulate a single Wishart distributed matrix using the rWishart function.

```r
nu0 <- 2
Sigma0 <- diag(2)
rWishart(1, df = nu0, Sigma = Sigma0)[, , 1]
```

```r
## [,1] [,2]
## [1,] 1.521133 -1.623125
## [2,] -1.623125 1.758854
```
inverse Wishart simulation

We can simulate a single inverse-Wishart distributed matrix using the `rWishart` function as well.

```r
nu0 <- 2
Sigma0 <- diag(2)
solve(rWishart(1, df = nu0,
               Sigma = solve(Sigma0))[, , 1])
```

```r
# [,1]           [,2]  
# [1,] 157.23769  62.23885
# [2,]  62.23885  24.92614
```