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.

Agenda

- 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

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

- 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 $\boldsymbol{\mu} = E(\boldsymbol{X})$ and $\Sigma = \text{Var}(\boldsymbol{X}) = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T]$, where

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

and

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

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.

Notation

- MVN is generalization of univariate normal.
- For the MVN, we write $\boldsymbol{X} \sim \mathcal{MVN}(\boldsymbol{\mu}, \Sigma)$.
- The $(i,j)^{\text{th}}$ component of Σ is the covariance between X_i and X_j (so the diagonal of Σ gives the component variances).

Example: $Cov(X_1, X_2)$ is just one element of the matrix Σ.

Multivariate Normal

Just as the probability density of a scalar normal is

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

the probability density of the multivariate normal is

$$p(\vec{x}) = (2\pi)^{-p/2}(\det\Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}. \tag{2}$$

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

$$Z_1, \ldots, Z_n \overset{iid}{\sim} MVN(0, I)$$

$$\begin{aligned} f_z(z) &= \prod_{i=1}^{n} \frac{1}{2\pi} e^{-z_i^2/2} &(3) \\ &= (2\pi)^{-n} e^{z^T z/2} &(4) \end{aligned}$$

- E[Z] = 0
- Var[Z] = I

Conjugate to MVN

Suppose that

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

Let

$$\pi(\boldsymbol{\theta}) \sim MVN(\boldsymbol{\mu}, \Omega).$$

What is the full conditional distribution of $\boldsymbol{\theta} \mid \boldsymbol{X}, \Sigma$?

Prior

$$
\begin{aligned}
\pi(\boldsymbol{\theta}) &= (2\pi)^{-p/2} \det \Omega^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \Omega^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\} && (5)\\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \Omega^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\} && (6)\\
&\propto \exp -\frac{1}{2}\left\{\boldsymbol{\theta}^T \Omega^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T \Omega^{-1}\mu + \mu^T \Omega^{-1}\mu\right\} && (7)\\
&\propto \exp -\frac{1}{2}\left\{\boldsymbol{\theta}^T \Omega^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T \Omega^{-1}\mu\right\} && (8)\\
&= \exp -\frac{1}{2}\left\{\boldsymbol{\theta}^T A_o \boldsymbol{\theta} - 2\boldsymbol{\theta}^T b_o\right\} && (9)
\end{aligned}
$$

$\pi(\boldsymbol{\theta}) \sim MVN(\boldsymbol{\mu}, \Omega)$ implies that $A_o = \Omega^{-1}$ and $b_o = \Omega^{-1}\mu$.

Likelihood

$$p(\mathbf{X} \mid \boldsymbol{\theta}, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} \det \Sigma^{-n/2} \exp\left\{-\frac{1}{2}(x_i - \boldsymbol{\theta})^T \Sigma^{-1} (x_i - \boldsymbol{\theta})\right\} \tag{10}$$

$$\propto \exp -\frac{1}{2}\left\{\sum_i x_i^T \Sigma^{-1} x_i - 2\sum_i \boldsymbol{\theta}^T \Sigma^{-1} x_i + \sum_i \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\right\} \tag{11}$$

$$\exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T \Sigma^{-1} n\bar{x} + n\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\right\} \tag{12}$$

$$\exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T b_1 + \boldsymbol{\theta}^T A_1 \boldsymbol{\theta}\right\}, \tag{13}$$

where

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

and

$$\bar{x} := (\frac{1}{n}\sum_i x_{i1}, \ldots, \frac{1}{n}\sum_i x_{ip})^T.$$

Full conditional

$$
\begin{aligned}
p(\boldsymbol{\theta} \mid \mathbf{X}, \Sigma) &\propto p(\mathbf{X} \mid \boldsymbol{\theta}, \Sigma) \times p(\boldsymbol{\theta}) && (14)\\
&\propto \exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T b_1 + \boldsymbol{\theta}^T A_1 \boldsymbol{\theta}\right\} && (15)\\
&\times \exp -\frac{1}{2}\left\{\boldsymbol{\theta}^T A_o \boldsymbol{\theta} - 2\boldsymbol{\theta}^T b_o\right\} && (16)\\
&\propto \exp\{\boldsymbol{\theta}^T b_1 - \frac{1}{2}\boldsymbol{\theta}^T A_1 \boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T A_o \boldsymbol{\theta} + \boldsymbol{\theta}^T b_o\} && (17)\\
&\propto \exp\{\boldsymbol{\theta}^T (b_o + b_1) - \frac{1}{2}\theta^T (A_o + A_1)\theta\} && (18)
\end{aligned}
$$

Then

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

and

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

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

Interpretations

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

$$\begin{aligned} \mu_n = A_n^{-1} b_n &= [\Omega^{-1} + n\Sigma^{-1}]^{-1}(b_o + b_1) && (19) \\ &= [\Omega^{-1} + n\Sigma^{-1}]^{-1}(\Omega^{-1}\mu + \Sigma^{-1} n\bar{x}) && (20) \end{aligned}$$

$$\Sigma_n = A_n^{-1} = [\Omega^{-1} + n\Sigma^{-1}]^{-1} \qquad (21)$$

inverse Wishart distribution

Suppose $\Sigma \sim \text{inverseWishart}(\nu_o, S_o^{-1})$ where ν_o is a scalar and S_o^{-1} is a matrix.

Then

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

For the full distribution, see Hoff, Chapter 7 (p. 110).

inverse Wishart distribution

- The inverse Wishart distribution is the multivariate version of the Gamma distribution.
- The full hierarchy we're interested in is

$$\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma \sim MVN(\boldsymbol{\theta}, \Sigma).$$

$$\boldsymbol{\theta} \sim MVN(\mu, \Omega)$$

$$\Sigma \sim \text{inverseWishart}(\nu_o, S_o^{-1}).$$

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

$$\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma \sim MVN(\boldsymbol{\theta}, \Sigma).$$

$$\Sigma \sim \text{inverseWishart}(\nu_o, S_o^{-1}).$$

Continued

What about $p(\Sigma \mid \boldsymbol{X}, \boldsymbol{\theta}) \propto p(\Sigma) \times p(\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma)$. Let's first look at

$$
\begin{aligned}
& p(\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma) && (22)\\
& \propto \det(\Sigma)^{-n/2} \exp\{-\sum_i (\mathbf{X}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{X}_i - \boldsymbol{\theta})/2\} && (23)\\
& \propto \det(\Sigma)^{-n/2} \exp\{-tr(\sum_i (\mathbf{X}_i - \boldsymbol{\theta})(\mathbf{X}_i - \boldsymbol{\theta})^T \Sigma^{-1}/2)\} && (24)\\
& \propto \det(\Sigma)^{-n/2} \exp\{-\mathrm{tr}(S_\theta \Sigma^{-1}/2)\} && (25)
\end{aligned}
$$

where $S_\theta = \sum_i (\mathbf{X}_i - \boldsymbol{\theta})(\mathbf{X}_i - \boldsymbol{\theta})^T$.

Fact:

$$\sum_k b_k^T A b_k = tr(B^T BA),$$

where B is the matrix whose kth row is b_k.

Continued

Now we can calculate $p(\Sigma \mid \boldsymbol{X}, \boldsymbol{\theta})$

$$
\begin{aligned}
& p(\Sigma \mid \boldsymbol{X}, \boldsymbol{\theta}) && (26)\\
& \quad = p(\Sigma) \times p(\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma) && (27)\\
& \quad \propto \det(\Sigma)^{-(\nu_o+p+1)/2} \times \exp\{-\mathrm{tr}(S_o \Sigma^{-1})/2\} && (28)\\
& \qquad \times \det(\Sigma)^{-n/2} \exp\{-\mathrm{tr}(S_\theta \Sigma^{-1})/2\} && (29)\\
& \quad \propto \det(\Sigma)^{-(\nu_o+n+p+1)/2} \exp\{-\mathrm{tr}((S_o + S_\theta)\Sigma^{-1})/2\} && (30)
\end{aligned}
$$

This implies that

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

Continued

Suppose that we wish now to take

$$\boldsymbol{\theta} \mid \boldsymbol{X}, \Sigma \sim MVN(\mu_n, \Sigma_n)$$

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

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

There is no closed form expression for this posterior. Solution?

Gibbs sampler

Suppose the Gibbs sampler is at iteration s.

1. Sample $\boldsymbol{\theta}^{(s+1)}$ from it's full conditional:
   a) Compute μ_n and Σ_n from $\boldsymbol{X}$ and $\Sigma^{(s)}$
   b) Sample $\boldsymbol{\theta}^{(s+1)} \sim MVN(\mu_n, \Sigma_n)$
2. Sample $\Sigma^{(s+1)}$ from its full conditional:
   a) Compute S_n from $\boldsymbol{X}$ and $\theta^{(s)}$
   b) Sample $\Sigma^{(s+1)} \sim \text{inverseWishart}(\nu_n, S_n^{-1})$

Working with Multivariate Normal Distribution

The R package, mvtnorm, contains functions for evaluating and simulating from a multivariate normal density.

```
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))
```

```
##             [,1]        [,2]
## [1,] -0.4956353 -0.02587638
```

Evaluation

Evaluate the multivariate normal density at a single value using the `dmvnorm` function.

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

```
## [1] 0.1591549
```

Working with the Multivariate Normal

- Now let's simulate many multivariate normals.
- Each row is a different sample from this multivariate normal distribution.

```
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
```

Evaluation

We can evaluate the multivariate normal density at several values using the `dmvnorm` function.

```
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]
```

```
##           [,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.

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

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