The Multivariate Distributions: Normal and inverse Wishart

Rebecca C. Steorts

Bayesian Methods and Modern Statistics: STA 360/601

Module 10

- 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.

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

Determinant and Trace of Matrices

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 $H_{p \times p}$.

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

where h_{ii} are the diagonal elements of H.

- MVN is generalization of univariate normal.
- For the MVN, we write $\boldsymbol{X} \sim \mathcal{MVN}(\boldsymbol{\mu}, \Sigma)$.
- 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).

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."

Please review the first section for Chapter 7 if you're uncomfortable with the density of the standard normal or MVN.

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

$$\pi(\boldsymbol{\theta}) = (2\pi)^{-p/2} \det \Omega^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \Omega^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\} \quad (3)$$

$$\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T \Omega^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\} \quad (4)$$

$$\propto \exp -\frac{1}{2}\{\boldsymbol{\theta}^T\Omega^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\Omega^{-1}\mu + \mu^T\Omega^{-1}\mu\} \quad (5)$$

$$\propto \exp -\frac{1}{2}\{\boldsymbol{\theta}^T\Omega^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\Omega^{-1}\mu\} \quad (6)$$

$$= \exp -\frac{1}{2}\{\boldsymbol{\theta}^T A_o\boldsymbol{\theta} - 2\boldsymbol{\theta}^T b_o\} \quad (7)$$

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

Likelihood

$$p(\mathbf{X} \mid \boldsymbol{\theta}, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} \det \Sigma^{-n/2} \exp\left\{-\frac{1}{2}(x_i - \boldsymbol{\theta})^T \Sigma^{-1} (x_i - \boldsymbol{\theta})\right\} \tag{8}$$

$$\propto \exp -\frac{1}{2}\left\{\sum_i x_i^T \Sigma^{-1} x_i - 2\sum_i \boldsymbol{\theta}^T \Sigma^{-1} x_i + \sum_i \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\right\} \tag{9}$$

$$\exp -\frac{1}{2}\{-2\boldsymbol{\theta}^T \Sigma^{-1} n\bar{x} + n\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}\} \tag{10}$$

$$\exp -\frac{1}{2}\{-2\boldsymbol{\theta}^T b_1 + \boldsymbol{\theta}^T A_1 \boldsymbol{\theta}\}, \tag{11}$$

where

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

and

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

Full conditional

$$
\begin{aligned}
p(\boldsymbol{\theta} \mid \mathbf{X}, \Sigma) &\propto p(\mathbf{X} \mid \boldsymbol{\theta}, \Sigma) \times p(\boldsymbol{\theta}) && (12)\\
&\propto \exp -\frac{1}{2}\left\{-2\boldsymbol{\theta}^T b_1 + \boldsymbol{\theta}^T A_1 \boldsymbol{\theta}\right\} && (13)\\
&\times \exp -\frac{1}{2}\left\{\boldsymbol{\theta}^T A_o \boldsymbol{\theta} - 2\boldsymbol{\theta}^T b_o\right\} && (14)\\
&\propto \exp\{\boldsymbol{\theta}^T b_1 - \frac{1}{2}\boldsymbol{\theta}^T A_1 \boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T A_o \boldsymbol{\theta} + \boldsymbol{\theta}^T b_o\} && (15)\\
&\propto \exp\{\boldsymbol{\theta}^T (b_o + b_1) - \frac{1}{2}\theta^T (A_o + A_1)\theta\} && (16)
\end{aligned}
$$

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?

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

$$\begin{aligned} \mu_n = A_n^{-1} b_n &= [\Omega^{-1}\mu + n\Sigma^{-1}]^{-1}(b_o + b_1) && (17) \\ &= [\Omega^{-1}\mu + n\Sigma^{-1}]^{-1}(\Omega^{-1}\mu + \Sigma^{-1} n\bar{x}) && (18) \end{aligned}$$

$$\Sigma_n = A_n^{-1} = [\Omega^{-1}\mu + n\Sigma^{-1}]^{-1} \quad (19)$$

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.

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

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

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

$$
\begin{aligned}
& p(\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma) && (20)\\
& \propto \det(\Sigma)^{-n/2} \exp\{-\sum_i (\mathbf{X}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{X}_i - \boldsymbol{\theta})/2\} && (21)\\
& \propto \det(\Sigma)^{-n/2} \exp\{-\mathsf{tr}(S_\theta \Sigma^{-1}/2)\} && (22)
\end{aligned}
$$

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

$$
\begin{aligned}
&p(\Sigma \mid \boldsymbol{X}, \boldsymbol{\theta}) && (23)\\
&\quad = p(\Sigma) \times p(\boldsymbol{X} \mid \boldsymbol{\theta}, \Sigma) && (24)\\
&\quad \propto \det(\Sigma)^{-(\nu_o+p+1)/2} \times \exp\{-\mathsf{tr}(S_o\Sigma^{-1})/2\} && (25)\\
&\qquad \times \det(\Sigma)^{-n/2} \exp\{-\mathsf{tr}(S_\theta\Sigma^{-1})/2\} && (26)\\
&\quad \propto \det(\Sigma)^{-(\nu_o+n+p+1)/2} \exp\{-\mathsf{tr}((S_o+S_\theta)\Sigma^{-1})/2\} && (27)
\end{aligned}
$$

This implies that

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

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?

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 $\Sigma^{(s+1)}$
   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 $\Sigma^{(s+1)}$
   b) Sample $\Sigma^{(s+1)} \sim$ inverseWishart(ν_n, S_n^{-1})

Exercise

We want to model the probability of recovery for patients admitted to the hospital in severe cardiac distress.

Suppose “recovery” means the patient survived long enough to be discharged from the hospital.

You have past data from a number of patients from 6 different hospitals.
For each patient i, you have various information

- $x_i = (x_{i1}, \ldots, x_{ip})$ (e.g., gender, weight, age, blood pressure, etc.),
- and a binary outcome y_i (did the patient recover or not).

Design a Bayesian model for this problem.