# Multivariate Normal Model

## (1) semi conjugate prior for the mean

Sampling model: $\vec{y}_i \sim MVN(\vec{\theta}, \Sigma)$ — $\vec{y}_i$ is a ~~(1xp)~~ $(p\times 1)$ vector

Prior $\vec{\theta} \sim MVN(\vec{\mu}_0, \Lambda_0)$ — $\Sigma$ is a $[p\times p]$ covariance matrix.

derive $p(\vec{\theta} \mid \vec{y}_1 .. \vec{y}_n, \Sigma)$ — $\vec{\theta}$ is a ~~(1xp)~~ $(p\times 1)$ vector)

We have

- $p(\theta) = (2\pi)^{-p/2} |\Lambda_0|^{-1/2} \exp\left(-\frac{1}{2}(\vec{\theta}-\vec{\mu}_0)^T \Lambda_0^{-1}(\vec{\theta}-\vec{\mu}_0)\right)$

$\propto \exp\left(-\frac{1}{2}\theta^T \Lambda_0^{-1}\theta + \theta^T \Lambda_0^{-1}\mu_0\right) = \exp\left(-\frac{1}{2}\theta^T A_0 \theta + \theta^T b_0\right)$ $\quad A_0 = \Lambda_0^{-1}$, $b_0 = \Lambda_0^{-1}\mu_0$.

- $p(\vec{y}_1 .. \vec{y}_n \mid \theta, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\vec{y}_i - \vec{\theta})^T \Sigma^{-1}(\vec{y}_i - \vec{\theta})\right)$

$$\propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}S_0)\right)$$

$$\propto \exp\left(-\frac{1}{2}\theta^T A_1 \theta + \theta^T b_1\right) \qquad A_1 = n\Sigma^{-1}, \quad b_1 = n\Sigma^{-1}\vec{\bar{y}}$$

$\left(S_0 = \sum_{i=1}^{n}(\vec{y}_i - \vec{\theta})(\vec{y}_i - \vec{\theta})^T\right)$

$\bar{y} = \left(\frac{1}{n}\sum_{i=1}^{n} y_{i1}, \dots, \frac{1}{n}\sum_{i=1}^{n} y_{in}\right)$

So we have

$$p(\vec{\theta} \mid \vec{y}_1 .. \vec{y}_n, \Sigma) \propto \exp\left(-\frac{1}{2}\left[(\vec{\theta}-\vec{\mu}_0)^T \Lambda_0^{-1}(\vec{\theta}-\vec{\mu}_0) + \sum_{i=1}^{n}(\vec{y}_i - \vec{\theta})^T \Sigma^{-1}(\vec{y}_i - \vec{\theta})\right]\right)$$

$$= \exp\left(-\frac{1}{2}\theta^T A_0 \theta + \theta^T b_0 - \frac{1}{2}\theta^T A_1 \theta + \theta^T b_1\right)$$

$$= \exp\left(-\frac{1}{2}\theta^T A_n \theta + \theta^T b_n\right). \qquad A_n = A_0 + A_1, \quad b_n = b_0 + b_1.$$

density of a MVN distribution. $(\mu_n, \Lambda_n)$

$\mu_n = E(\vec{\theta} \mid \vec{y}_1 .. \vec{y}_n, \Sigma) = A_n^{-1} b_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y})$.

$\Lambda_n = Cov(\vec{\theta} \mid \vec{y}_1 .. \vec{y}_n, \Sigma) = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}$

We can show $\left.\begin{array}{l} E(\vec{y} \mid \vec{y}_1 .. \vec{y}_n) = \mu_n \\ Var(\vec{y} \mid \vec{y}_1 .. \vec{y}_n) = \Sigma + \Lambda_n \end{array}\right\}$ Similar to univariate case.

# Non informative prior density for $\theta$

$p(\vec{\theta}) \propto k$. (obtained in the limit as $|\Lambda_0^{-1}| \to 0$

(prior mean is irrelevant with infinite prior variance)

So $p(\vec{\theta} \mid \Sigma, \vec{y}_1 \dots \vec{y}_n) \propto p(\vec{y}_1 \dots \vec{y}_n \mid \Sigma, \vec{\theta})$.

$\vec{\theta} \mid \Sigma, \vec{y}_1 \dots \vec{y}_n \propto MVN(\bar{\vec{y}}, \frac{\Sigma}{n})$. (proper distrib. only if $n \geq p$)

($n < p \Rightarrow$ high dimensional statistics).

---

Simulation from a multivariate normal distribution.

$\rightarrow$ Cholesky decomposition

$\Lambda_0 = AA^T$ A: Cholesky factor.

Let $z_1 \dots z_p$ be 'p' independent standard normal random variables.

then $\vec{\theta} = \mu_0 + Az$ is a random draw from a multivariate normal distribution with covariance matrix $\Lambda_0$, mean vector $\mu_0$ (dimension 'p').

# Multivariate normal model with unknown mean and variance.

Inverse Wishart (conjugate) prior on $\Sigma$. (Multivariate analogue of Gamma distribution).

$$\Sigma \sim \text{Inv Wishart}(\nu_0, S_0^{-1})$$

$$E(\Sigma) = \frac{1}{\nu_0 - p - 1} S_0. \rightarrow \text{ can set. } S_0 = (\nu_0 - p - 1)\Sigma_0).$$

If $\nu_0 = p+2$ (or some other small value) then prior given on $\Sigma$ is loosely centred on $\Sigma_0$ (corresponds to weakly informative prior)

Derive.

$$p(\Sigma \mid \vec{y}_1 \ldots \vec{y}_n, \vec{\theta}) \propto |\Sigma|^{-(\nu_0+p+1)/2} \exp\left(-\text{tr}(S_0\Sigma^{-1})/2\right) \times \left(|\Sigma|^{-n/2} \exp\left[\text{tr}(S_\theta \Sigma^{-1})/2\right]\right)$$

$$\left(\left(S_\theta = \sum_{i=1}^{n} (\vec{y}_i - \vec{\theta})(\vec{y}_i - \vec{\theta})^T\right)\right.$$ residual sums of squares (SS) matrix for vectors $\vec{y}_1 \ldots \vec{y}_n$. )

$$= |\Sigma|^{-(\nu_0+n+p+1)/2} \exp\left(-\text{tr}(S_0 + S_\theta)\Sigma^{-1}/2\right).$$

$$\therefore \Sigma \mid \vec{y}_1 \ldots \vec{y}_n, \vec{\theta} \sim \text{Inv Wishart}\left(\nu_0 + n, (S_0 + S_\theta)^{-1}\right).$$

$\nu_0 + n$: (prior sample size + data sample size

$S_0 + S_\theta$: prior SS plus data SS

$$E(\Sigma \mid \vec{y}_1 \ldots \vec{y}_n, \vec{\theta}) = \frac{1}{\nu_0 + n - p - 1}(S_0 + S_\theta)$$

$$= \left(\frac{\nu_0 - p - 1}{\nu_0 + n - p - 1}\right) \times \frac{1}{\nu_0 - n - p - 1} S_0 + \frac{n}{\nu_0 + n - p - 1} \times \frac{1}{n} S_\theta.$$

prior expectation

unbiased estimator conditional on $\theta$.

## Gibbs sampler of the mean and covariance

$$\vec{\theta} \mid \vec{y}_1 \ldots \vec{y}_n, \Sigma \sim MVN(\mu_n, \Lambda_n)$$

$$\Sigma \mid \vec{y}_1 \ldots \vec{y}_n, \vec{\theta} \sim \text{InvWishart}(\nu_n, S_n^{-1})$$

## Missing Data and Imputation

- Multiple Imputation $Y_{mis}$ are unknown quantities (like $\theta$) to be estimated as well
- add an additional step to the Gibbs sampler for Multivariate normal data

## Health data example:

- where do you think the R code may fail?
  - entire row is missing so no observed information for unit 'i' to update with.

- What if MVN assumption is not appropriate.

For example, there is an indicator variable

eg $\begin{cases} = 1 \text{ if women is on chronic medication} \\ = 0 \text{ otherwise.} \end{cases}$

with missing values

How would your Bayesian be modified for this binary variable