Statistical Methods for Analysis with Missing Data

Lecture 11 setup: examples of Gibbs sampler, data augmentation for proper multiple imputation, MICE

Mauricio Sadinle

Department of Biostatistics

UNIVERSITY of WASHINGTON

Previous Lectures

- Gibbs sampling
- Data augmentation to handle missing data in Bayesian inference
- Multiple imputation as a Monte Carlo approximation of proper Bayesian procedure
- Uncongeniality generally leads to invalidity of inferences based on Rubin's combining rules
- MICE: practical implementation of multiple imputation that builds on Gibbs sampling ideas

Today's Lecture

R implementations of

- Gibbs sampler
- Data augmentation for proper multiple imputation
- MICE

Outline

Example of Gibbs Sampler

Data Augmentation for Proper Multiple Imputation

MICE

Summary

Bhattacharyya's Distribution

Consider real-valued random variables X and Y having a joint distribution with density[1]

$$p_{X,Y}(x,y) = \exp\left\{ \begin{bmatrix} 1, x, x^2 \end{bmatrix} \begin{bmatrix} m_{00}, m_{01}, m_{02} \\ m_{10}, m_{11}, m_{12} \\ m_{20}, m_{21}, m_{22} \end{bmatrix} \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix} \right\},$$

where either

(a) $m_{22} = m_{21} = m_{12} = 0;\ m_{20}, m_{02} < 0;\ m_{11}^2 < 4m_{20}m_{02};$

(b) $m_{22} < 0,\ 4m_{22}m_{02} > m_{12}^2,\ 4m_{22}m_{20} > m_{21}^2.$

m_{00} is determined by the other m_{ij}'s so that $p_{X,Y}$ integrates to 1.

[1] Distribution credited to Anil Kumar Bhattacharyya, who was a professor at the Indian Statistical Institute. See, e.g., https://projecteuclid.org/download/pdf_1/euclid.ss/1009213728

Bhattacharyya's Distribution

From $p_{X,Y}(x,y)$ it is easy to see that

$$p_{X|y}(x|y) \propto \frac{1}{\sigma_X(y)} \exp\left\{-\frac{[x-\mu_X(y)]^2}{2\sigma_X^2(y)}\right\},$$

where

$$\mu_X(y) = -\frac{m_{10}+m_{11}y+m_{12}y^2}{2(m_{20}+m_{21}y+m_{22}y^2)},$$

and

$$\sigma_X^2(y) = -\frac{1}{2(m_{20}+m_{21}y+m_{22}y^2)}$$

Bhattacharyya's Distribution

And analogously, it is easy to see that

$$p_{Y|x}(y|x) \propto \frac{1}{\sigma_Y(x)} \exp\left\{-\frac{[y-\mu_Y(x)]^2}{2\sigma_Y^2(x)}\right\},$$

where

$$\mu_Y(x) = -\frac{m_{01} + m_{11}x + m_{21}x^2}{2(m_{02} + m_{12}x + m_{22}x^2)},$$

and

$$\sigma_Y^2(x) = -\frac{1}{2(m_{02} + m_{12}x + m_{22}x^2)}$$

Bhattacharyya's Distribution

- In fact, Bhattacharyya's distribution characterizes *all* bivariate distributions with normal conditionals[2]

- Gibbs sampler to draw from $p_{X,Y}$ is easy to implement:
  - Choose starting point $(x^{(0)}, y^{(0)})$
  - At iteration t draw

$$X^{(t)} \sim \text{Normal}[\ \mu_X(y^{(t-1)}),\ \sigma_X^2(y^{(t-1)})\ ]$$
$$Y^{(t)} \sim \text{Normal}[\ \mu_Y(x^{(t)}),\ \sigma_Y^2(x^{(t)})\ ]$$

[2]Arnold, Castillo and Sarabia (Statistical Science, 2001):
`https://projecteuclid.org/download/pdf_1/euclid.ss/1009213728`

R Time!

Open file `Lecture11code.R`, part 1

Outline

Example of Gibbs Sampler

Data Augmentation for Proper Multiple Imputation

MICE

Summary

Example: Multivariate Normal

- Distribution of the data

$$\mathbf{Z} = \{Z_i\}_{i=1}^{n} \mid \mu, \Lambda \overset{i.i.d.}{\sim} \text{Normal}(\mu, \Lambda^{-1})$$

where $Z_i \in \mathbb{R}^K$, μ is the vector of means, Λ^{-1} is the covariance matrix, and Λ is the inverse covariance matrix (the *precision matrix*)

- Conjugate prior is constructed in two steps

$$\mu \mid \Lambda \sim \text{Normal}(\mu_0, (\kappa_0 \Lambda)^{-1})$$
$$\Lambda \sim \text{Wishart}(\upsilon_0, W_0)$$

Joint distribution of (μ, Λ) is called *Normal-Wishart*. The parameterization is such that $E(\Lambda) = \upsilon_0 W_0$

Example: Multivariate Normal

- Distribution of the data

$$\mathbf{Z} = \{Z_i\}_{i=1}^{n} \mid \mu, \Lambda \overset{i.i.d.}{\sim} \text{Normal}(\mu, \Lambda^{-1})$$

where $Z_i \in \mathbb{R}^K$, μ is the vector of means, Λ^{-1} is the covariance matrix, and Λ is the inverse covariance matrix (the *precision matrix*)

- Conjugate prior is constructed in two steps

$$\mu \mid \Lambda \sim \text{Normal}(\mu_0, (\kappa_0 \Lambda)^{-1})$$
$$\Lambda \sim \text{Wishart}(\upsilon_0, W_0)$$

Joint distribution of (μ, Λ) is called *Normal-Wishart*. The parameterization is such that $E(\Lambda) = \upsilon_0 W_0$

Example: Multivariate Normal

Posterior is also Normal-Wishart

$$\mu \mid \Lambda, \mathbf{z} \sim \text{Normal}(\mu', (\kappa'\Lambda)^{-1})$$
$$\Lambda \mid \mathbf{z} \sim \text{Wishart}(\upsilon', W')$$

where

$$\begin{aligned}
\mu' &= (\kappa_0\mu_0 + n\bar{z})/\kappa' \\
\kappa' &= \kappa_0 + n \\
\upsilon' &= \upsilon_0 + n \\
W' &= \{W_0^{-1} + n[\hat{\Sigma} + \frac{\kappa_0}{\kappa'}(\bar{z} - \mu_0)(\bar{z} - \mu_0)^T]\}^{-1} \\
\bar{z} &= \sum_{i=1}^{n} z_i/n \\
\hat{\Sigma} &= \sum_{i=1}^{n} (z_i - \bar{z})(z_i - \bar{z})^T/n
\end{aligned}$$

Example: Multivariate Normal

HW3: write down and implement a data augmentation algorithm under ignorability and multivariate normality

R Time!

Open file `Lecture11code.R`, part 2

Outline

Multivariate Imputation by Chained Equations

Multivariate Imputation by Chained Equations (MICE)[3] is an ad-hoc multiple imputation procedure that builds on Gibbs sampling ideas

- If each $Y_1, \ldots, Y_K$ is subject to missingness, we can posit K different regression models

$$
\begin{gathered}
p_1(y_1 \mid y_{-1}, \theta_1) \\
p_2(y_2 \mid y_{-2}, \theta_2) \\
\vdots \\
p_K(y_K \mid y_{-K}, \theta_K)
\end{gathered}
$$

- θ_k: parameters of the kth conditional distribution
- $y_{-k} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_K)$
- Key idea: use these models to sequentially impute, one variable at a time. Repeat this over a number of iterations

[3] https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$
- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i: r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i: r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

R Time!

Open file `Lecture11code.R`, part 3

Outline

Summary

Main take-aways from today's lecture:

- Example of Gibbs sampler
- Data Augmentation and Proper Multiple Imputation using 'norm' package
- Multivariate Imputation by Chained Equations with the 'mice' package

Next lecture:

- Inverse Probability Weighting