Missing Data and Imputation

Rebecca C. Steorts

Bayesian Methods and Modern Statistics: STA 360/601

Lecture

- Missing data for real applications
- Common approaches
- Bayesian approach
- Lab: Example

Health related measurements on women from Pima Indian heritage village

- glu: glucose concentration
- bp: diastolic blood pressure
- skin: skin fold thickness
- bmi: body mass index

	glu	bp	skin	bmi
1	86	68	28	30.2
2	195	70	33	NA
3	77	82	NA	35.8
4	NA	76	43	47.9
5	107	60	NA	NA
6	97	76	27	NA

How would we do parameter estimate with missing values?

Remove the missing values (throwing away data).

Impute these with the population mean (statistically incorrect).

Notation and Missing at Random

- $\boldsymbol{Y}$ are observed covariates.
- $\boldsymbol{O}_i = (O_1, \ldots, O_p)^T$
- $O_i = 1$ implies Y_{ij} is not missing, 0 otherwise.
- We assume data is missing data random, meaning that $\boldsymbol{O}_i$ are $\boldsymbol{Y}_i$ are statistically independent
- We assume also that $\boldsymbol{O}_i$ does not depend on $\boldsymbol{\theta}$ or Σ.
- For when it's not missing at random, see Gelman and Rubin, Chapter 21.

Missing at Random

$$
\begin{aligned}
& p(\{y_{ij} : o_{ij} = 1\} \mid \boldsymbol{\theta}, \Sigma) && (1)\\
& = p(\boldsymbol{o}_i) \times p(\{y_{ij} : o_{ij} = 1\} \mid \boldsymbol{\theta}, \Sigma) && (2)\\
& = p(\boldsymbol{o}_i) \times \int \left\{ p(y_{i,1}, \ldots, y_{i,p} \mid \boldsymbol{\theta}, \Sigma) \prod_{y_{ij} : o_{ij} = 0} dy_{ij} \right\} && (3)
\end{aligned}
$$

Main point: integrate out all the missing y's.

Simple Example

Let $\boldsymbol{y}_i = (y_{i1}, \texttt{NA}, y_{i3}, \texttt{NA})^T$.

Then $\boldsymbol{o}_i = (1, 0, 1, 0)^T$.

$$
\begin{aligned}
& p(\{y_{ij} : o_{ij} = 1\} \mid \boldsymbol{\theta}, \Sigma) && (4) \\
& = p(\boldsymbol{o}_i, y_{i1}, y_{i3} \mid \boldsymbol{\theta}, \Sigma) && (5) \\
& = p(\boldsymbol{o}_i) \times \int \{p(\boldsymbol{y_i} \mid \boldsymbol{\theta}, \Sigma)\ dy_2\ d_4\} && (6)
\end{aligned}
$$

Missing data with Gibbs Sampling

- Let $\boldsymbol{Y}$ be the observed data.
- Let $\boldsymbol{Y}_{obs}$ be the data we do observe (not missing).
- Let $\boldsymbol{Y}_{miss}$ be the data we do not observe (missing).

For any observed data, we want to estimate

$$p(\boldsymbol{\theta}, \Sigma, \boldsymbol{Y}_{miss} \mid \boldsymbol{Y}_{obs}).$$

We do this via Gibbs sampling.

Suppose starting values $\Sigma^{(o)}, \boldsymbol{Y}_{miss}^{(o)}$.

We generate

$$\boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)}, \boldsymbol{Y}_{miss}^{(s+1)}$$

and

$$\boldsymbol{\theta}^{(s)}, \Sigma^{(s)}, \boldsymbol{Y}_{miss}^{(s)}$$

by

1. Sampling $\boldsymbol{\theta}^{(s)}$ from $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(s)}, \Sigma^{(s)})$
2. Sampling $\Sigma^{(s+1)}$ from $p(\Sigma \mid \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{miss}^{(s)}, \boldsymbol{\theta}^{(s+1)})$
3. Sampling $\boldsymbol{Y}_{miss}^{(s+1)}$ from $p(\boldsymbol{Y}_{miss} \mid \boldsymbol{Y}_{obs}, \boldsymbol{\theta}^{(s+1)}, \Sigma^{(s)})$

Using steps 1 and 2, we obtain a full matrix $\boldsymbol{Y}$.

Then $\boldsymbol{\theta} \mid \boldsymbol{Y}, \Sigma \sim MVN(\mu_n, \Sigma_n)$
(Slide 10, MVN and Wishart Lecture, Hoff eqn 7.6)

Also, $\Sigma \mid \boldsymbol{Y}, \theta \sim$ inverseWishart$(\nu_o + n, S_n)$
(Slide 15, MVN and Wishart Lecture, Hoff eqn 7.9).

Consider

$$\begin{aligned}
p(\boldsymbol{Y}_{miss} \mid \boldsymbol{Y}_{obs}, \boldsymbol{\theta}, \Sigma) &\propto p(\boldsymbol{Y}_{miss}, \boldsymbol{Y}_{obs} \mid \boldsymbol{\theta}, \Sigma) && (7)\\
&\propto \prod_i p(\boldsymbol{y}_{i,miss}, \boldsymbol{y}_{i,obs} \mid \boldsymbol{\theta}, \Sigma) && (8)\\
&\propto \prod_i p(\boldsymbol{y}_{i,miss} \mid \boldsymbol{y}_{i,obs}, \boldsymbol{\theta}, \Sigma) && (9)
\end{aligned}$$

How can we compute this? There's clever little result (eqn 7.11, Hoff) that makes it possible. (It's just another MVN result).

Illustration from Hoff's example, p. 120

The prior mean is set at $(120, 64, 26, 26)^T$

We use Hoff's code given in the book and run 1,000 iterations of the GS.

Posterior mean is $(123.4671.0329.3532.18)^T$

We convert the covariance matrix into a correlation matrix (see Hoff for formula).

Then we look at the marginal posterior of 95 percent quantile based confidence infervals.

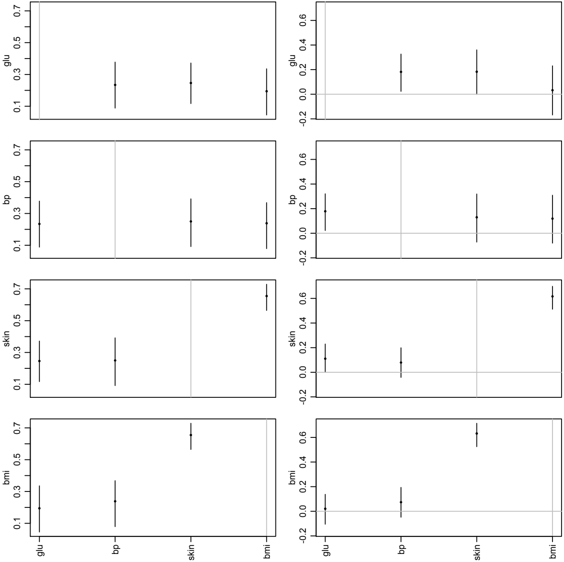

Figure 1: 95 percent posterior confidence intervals for correlations and regression coefficients.

Then do the prediction example from Hoff