

STA 601 - Homework 8

Kedar Prabhudesai

September 27, 2013

Missing Data:

To simulate rejecting $p\%$ of data, I created a matrix of random values from $U[0, 1]$ and ran a threshold of $p\%$ to get a random indicator matrix O_{ij} . I also made sure that for a given sample we have at least one value present. Using this approach I got a rejection percentage close to $p\%$.

This can be summarized as follows:

- $O_{ij} \in [0, 1]$, $i = 1, 2, \dots, 100$, $j = 1, 2$.
- If $O_{ij} \leq p/100 = 1$, else 0.
- 1: Reject Data, 0: Keep Data.
- Make sure that, $\forall i$ at least one j is present.

Gibbs Sampling:

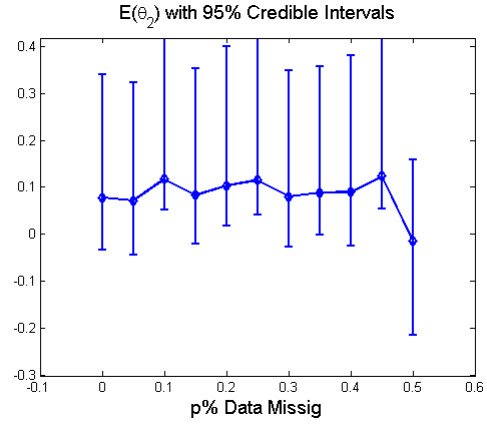
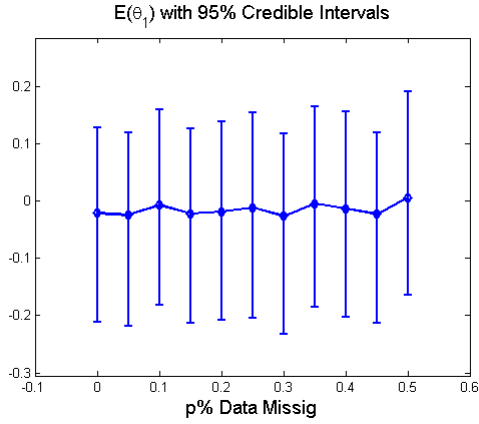
- Priors for θ and Σ :
 $\mu \sim \mathcal{N}_2(\mu_0, \Lambda_0) \cdot \mu_0 = \begin{pmatrix} 0.2 \\ 0.2 \end{pmatrix}, \Lambda_0 = \begin{pmatrix} 1.25 & 0.6 \\ 0.6 & 1.25 \end{pmatrix}$
 $\Sigma \sim \text{Inverse-Wishart}(\nu_0, S_0^{-1}) \cdot \nu_0 = 4, S_0 = \begin{pmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{pmatrix}.$
- Missing Data: We know that our data is drawn from a Bivariate Normal. We can partition each sample as $\{y_{obs}, y_{miss}\}$
- Gibbs Sampler: Given the starting values of $\{\Sigma^{(0)}, y_{miss}^{(0)} = 2\}$, we generate $\{\theta^{(s+1)}, \Sigma^{(s+1)}, y_{miss}^{(s+1)}\}$, as follows:
 - $\theta^{(s+1)} \sim p(\theta | y_{obs}, y_{miss}^{(s)}, \Sigma^{(s)})$
 - $\Sigma^{(s+1)} \sim p(\Sigma | y_{obs}, y_{miss}^{(s)}, \theta^{(s+1)})$
 - $y_{miss}^{(s+1)} \sim p(y_{miss} | y_{obs}, \theta^{(s+1)}, \Sigma^{(s+1)})$

- Full Conditionals:

- Update \bar{y}
- $\theta^{(s+1)} \mid y_{obs}, y_{miss}^{(s)}, \Sigma^{(s)} \sim \mathcal{N}_2 \left[\left(\frac{n\Sigma^{(s)-1}\bar{y} + \Lambda_0^{-1}\mu_0}{n\Sigma^{(s)-1} + \Lambda_0^{-1}} \right), (n\Sigma^{(s)-1} + \Lambda_0^{-1})^{-1} \right]$
- $\Sigma^{(s+1)} \mid y_{obs}, y_{miss}^{(s)}, \theta^{(s+1)} \sim IW \left[\nu_0 + n, (S_0 + S_\theta)^{-1} \right], S_\theta = \sum_{i=1}^n (y_i - \theta^{(s+1)})(y_i - \theta^{(s+1)})'$
- $y_{miss}^{(s+1)} \mid y_{obs}, \theta^{(s+1)}, \Sigma^{(s+1)} \sim \mathcal{N} \left(\theta_{b|a}^{(s+1)}, \Sigma_{b|a}^{(s+1)} \right)$
 $\theta_{b|a}^{(s+1)} = \theta_b^{(s+1)} + \Sigma_{b,a}^{(s+1)} \left(\Sigma_{a,a}^{(s+1)} \right)^{-1} (y_a - \theta^{(s+1)}_a)$
 $\Sigma_{b|a}^{(s+1)} = \Sigma_{b,b}^{(s+1)} - \Sigma_{b,a}^{(s+1)} \left(\Sigma_{a,a}^{(s+1)} \right)^{-1} \Sigma_{a,b}^{(s+1)},$
 where, b is the index of 'Data Missing' and a is index of 'Data Present'.

Results:

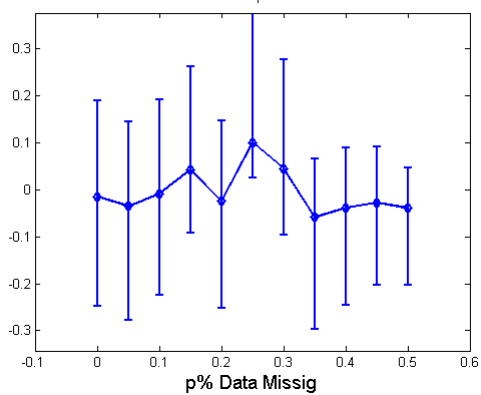
Gibbs Samples = 10000, Burn-In = 1000. Given here are estimates of $E(\theta_1)$ and $E(\theta_2)$ with their 95% Credible Intervals. We see that as the % of rejected data increases, our estimate of mean is less accurate and has a larger credible interval. Note that 0% point represents no data was rejected.



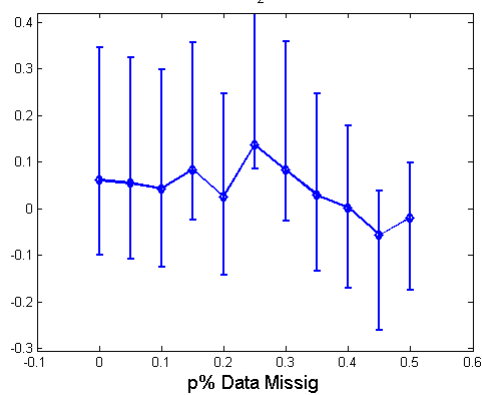
Complete Case Analysis:

For this we reject all samples which contain missing data. By doing this our estimate wanders even more about the true estimate and the intervals are much bigger than when we impute the missing data.

Complete Case Analysis - $E(\theta_1)$ with 95% Credible Intervals



Complete Case Analysis - $E(\theta_2)$ with 95% Credible Intervals



Appendix:

```
1 %% STA 601 — Homework 8
2 % Author: Kedar Prabhudesai
3 % Created on: 9/27/2013
4
5 close all;
6 clear all;
7
8 %% Setup Data and Distributions
9 % Create Bivariate Distribution
10 nSamples = 100;
11 rho = 0.8;
12 mu = [0 0];
13 SIGMA = [1 rho; rho 1];
14
15 % Get Data
16 rng('shuffle');
17 yTruth = mvnrnd(mu, SIGMA, nSamples);
18
19 pctMissing = 0:0.05:0.5;
20
21 y1Bayes = [];
22 y1BayesConf = [];
23 y2Bayes = [];
24 y2BayesConf = [];
25
26 %% Missing Data
27 for iPct = 1:numel(pctMissing)
28     % Simulate Missing Data
29     yTruthMissing = yTruth;
30     yTruthMissing = yTruthMissing';
31
32     % Get Random indices to reject data
33     OInds = rand(size(yTruthMissing));
34     % Indicator which tells us what is missing: 1—Miss, 0—Present
35     Oij = (OInds ≤ pctMissing(iPct));
36     disp(['Processing ', num2str(pctMissing(iPct)), ' Actual Percent ', num2str(sum(Oij(:))/...
37         numel(OInds))]);
38
39     % Reject Data
40     yTruthMissing(Oij) = NaN;
41     % Make sure we do not have both y1 and y2 missing
42     y1AndY2Missing = isnan(yTruthMissing(1,:)) & isnan(yTruthMissing(2,:));
43
44     for iSample = 1:nSamples
45         idxRnd = randperm(2,1);
46         yTruthMissing(idXRnd,iSample) = yTruth(iSample,idxRnd);
47         Oij(idXRnd,iSample) = 0;
48     end
49     idxMissing = find(sum(Oij,1)==1);
50
51     %% Gibbs Sampler
52     nGibbs = 10000;
53     nBurnIn = 1000;
54
55     mu0 = [0.2 0.2]';
56     L0 = [1.25 0.6; 0.6 1.25];
57
58     nu0 = 4;
59     S0 = [1.2 0.4; 0.4 1.3];
```

```

60 thetaSamples = zeros(2,nGibbs);
61 sigmaSamples = zeros(2,2,nGibbs);
62
63 % Initialize first value of Sigma and missing data
64 sigmaSamples(:, :, 1) = S0;
65 for iMiss = 1:numel(idxMissing)
66     % b: Missing Index
67     b = find(Oij(:, idxMissing(iMiss)));
68     yTruthMissing(b, idxMissing(iMiss)) = 2;
69 end
70
71 for iGibbs = 2:nGibbs
72     % Update ybar
73     ybar = mean(yTruthMissing, 2);
74
75     % Update theta
76     Ln = inv(inv(L0) + nSamples.*inv(sigmaSamples(:, :, iGibbs-1)));
77     mun = Ln*(inv(L0)*mu0 + nSamples.*inv(sigmaSamples(:, :, iGibbs-1))*ybar);
78     thetaHere = mvnrnd(mun, Ln);
79     thetaSamples(:, iGibbs) = thetaHere;
80     % Update Sigma
81     Sn = S0 + (bsxfun(@minus, yTruthMissing, thetaSamples(:, iGibbs)) * (bsxfun(@minus, ...
82         yTruthMissing, thetaSamples(:, iGibbs))));
83     Z = mvnrnd([0 0], inv(Sn), nu0+nSamples);
84     sigmaHere = inv(Z'*Z);
85     sigmaSamples(:, :, iGibbs) = sigmaHere;
86
87     % Update Missing Data
88     for iMiss = 1:numel(idxMissing)
89         % a: Present Index
90         % b: Missing Index
91         b = find(Oij(:, idxMissing(iMiss)));
92         if b == 1
93             a = 2;
94         else
95             a = 1;
96         end
97         ThetaMissGivenPres = thetaHere(b) + (sigmaHere(b, a) * (yTruthMissing(a, idxMissing...
98             (iMiss)) - thetaHere(a))) / sigmaHere(a, a);
99         SigmaMissGivenPres = sqrt(sigmaHere(b, b) - (sigmaHere(b, a) * sigmaHere(a, b)) / ...
100             sigmaHere(a, a));
101         yTruthMissing(b, idxMissing(iMiss)) = ThetaMissGivenPres + SigmaMissGivenPres*...
102             randn;
103     end
104 end
105
106 % Burn-In
107 thetaSamples(:, 1:nBurnIn) = [];
108 sigmaSamples(:, :, 1:nBurnIn) = [];
109
110 % Estimate mu and Sigma
111 muBayes = mean(thetaSamples, 2)';
112 sigmaBayes = mean(sigmaSamples, 3);
113
114 % Organize to plot
115 y1Bayes = [y1Bayes; muBayes(1)];
116 y1BayesConf = [y1BayesConf; quantile(thetaSamples(1, :), [0.025 0.975])];
117 y2Bayes = [y2Bayes; muBayes(2)];
118 y2BayesConf = [y2BayesConf; quantile(thetaSamples(2, :), [0.025 0.975])];
119 end
120 figure;
121 errorbar(pctMissing, y1Bayes, y1BayesConf(:, 1), y1BayesConf(:, 2), 'Marker', 'diamond', 'Linewidth...
122     ', 2);
123 title('E(\theta_1) with 95% Credible Intervals', 'FontSize', 14);

```

```

119 xlabel('p% Data Missig','FontSize',14);
120 ylim([min(y1BayesConf(:,1))-0.1 max(y1BayesConf(:,2))+0.1]);
121 xlim([-0.1 0.6]);
122
123 figure;
124 errorbar(pctMissing,y2Bayes,y2BayesConf(:,1),y2BayesConf(:,2),'Marker','diamond','Linewidth...
    ',2);
125 title('E(\theta_2) with 95% Credible Intervals','FontSize',14);
126 xlabel('p% Data Missig','FontSize',14);
127 ylim([min(y2BayesConf(:,1))-0.1 max(y2BayesConf(:,2))+0.1]);
128 xlim([-0.1 0.6]);
129
130 %% Complete Case Analysis
131 y1Bayes = [];
132 y1BayesConf = [];
133 y2Bayes = [];
134 y2BayesConf = [];
135
136 for iPct = 1:numel(pctMissing)
137     % Simulate Missing Data
138     yTruthMissing = yTruth;
139     yTruthMissing = yTruthMissing';
140
141     % Get Random indices to reject data
142     OInds = rand(size(yTruthMissing));
143     % Indicator which tells us what is missing: 1=Miss, 0=Present
144     Oij = (OInds <= pctMissing(iPct));
145     disp(['Processing ',num2str(pctMissing(iPct)),' Actual Percent ',num2str(sum(Oij(:))/...
        numel(OInds))]);
146
147     % Reject Data
148     yTruthMissing(Oij) = NaN;
149     % Make sure we do not have both y1 and y2 missing
150     y1AndY2Missing = isnan(yTruthMissing(1,:)) & isnan(yTruthMissing(2,:));
151
152     for iSample = 1:nSamples
153         idxRnd = randperm(2,1);
154         yTruthMissing(idXRnd,iSample) = yTruth(iSample,idxRnd);
155         Oij(idXRnd,iSample) = 0;
156     end
157     idxMissing = find(sum(Oij,1)==1);
158
159     %% Get rid of samples where data is missing
160     for iMiss = 1:numel(idxMissing)
161         yTruthMissing(:,idxMissing(iMiss)) = [0;0];
162     end
163
164     %% Gibbs Sampler
165     nGibbs = 10000;
166     nBurnIn = 1000;
167
168     mu0 = [0.2 0.2]';
169     L0 = [1.25 0.6;0.6 1.25];
170
171     nu0 = 4;
172     S0 = [1.2 0.4;0.4 1.3];
173
174     thetaSamples = zeros(2,nGibbs);
175     sigmaSamples = zeros(2,2,nGibbs);
176
177     % Initialize first value of Sigma
178     sigmaSamples(:, :, 1) = S0;
179     ybar = mean(yTruthMissing,2);
180

```

```

181     for iGibbs = 2:nGibbs
182         % Update theta
183         Ln = inv(inv(L0) + nSamples.*inv(sigmaSamples(:, :, iGibbs-1)));
184         mun = Ln*(inv(L0)*mu0 + nSamples.*inv(sigmaSamples(:, :, iGibbs-1))*ybar);
185         thetaHere = mvnrnd(mun, Ln);
186         thetaSamples(:, iGibbs) = thetaHere;
187
188         % Update Sigma
189         Sn = S0 + (bsxfun(@minus, yTruthMissing, thetaSamples(:, iGibbs))) * (bsxfun(@minus, ...
190             yTruthMissing, thetaSamples(:, iGibbs)))';
191         Z = mvnrnd([0 0], inv(Sn), nu0+nSamples);
192         sigmaHere = inv(Z'*Z);
193         sigmaSamples(:, :, iGibbs) = sigmaHere;
194     end
195
196     % Burn-In
197     thetaSamples(:, 1:nBurnIn) = [];
198     sigmaSamples(:, :, 1:nBurnIn) = [];
199
200     % Estimate mu and Sigma
201     muBayes = mean(thetaSamples, 2)';
202     sigmaBayes = mean(sigmaSamples, 3);
203
204     % Organize to plot
205     y1Bayes = [y1Bayes; muBayes(1)];
206     y1BayesConf = [y1BayesConf; quantile(thetaSamples(1, :), [0.025 0.975])];
207     y2Bayes = [y2Bayes; muBayes(2)];
208     y2BayesConf = [y2BayesConf; quantile(thetaSamples(2, :), [0.025 0.975])];
209 end
210 figure;
211 errorbar(pctMissing, y1Bayes, y1BayesConf(:, 1), y1BayesConf(:, 2), 'Marker', 'diamond', 'Linewidth...
212     ', 2);
213 title('Complete Case Analysis - E(\theta_1) with 95% Credible Intervals', 'FontSize', 14);
214 xlabel('p% Data Missig', 'FontSize', 14);
215 ylim([min(y1BayesConf(:, 1))-0.1 max(y1BayesConf(:, 2))+0.1]);
216 xlim([-0.1 0.6]);
217
218 figure;
219 errorbar(pctMissing, y2Bayes, y2BayesConf(:, 1), y2BayesConf(:, 2), 'Marker', 'diamond', 'Linewidth...
220     ', 2);
221 title('Complete Case Analysis - E(\theta_2) with 95% Credible Intervals', 'FontSize', 14);
222 xlabel('p% Data Missig', 'FontSize', 14);
223 ylim([min(y2BayesConf(:, 1))-0.1 max(y2BayesConf(:, 2))+0.1]);
224 xlim([-0.1 0.6]);

```