Homework Assignment 3

Statistical Methods for Analysis with Missing Data, Winter 2019

Instructor: Mauricio Sadinle, Department of Biostatistics, U. of Washington – Seattle

Submit your solutions via Canvas. Due by 12:00pm (noon) on March 6, 2019.

From this assignment you can get a maximum of 20 points. The assignment contains a list of problems, each worth a different number of points. You may choose any combination of problems that you like. I recommend that you solve a combination of problems that is worth more than 20 points as a way of gaining insurance against errors in some of your problem solutions. If you are submitting solutions to theoretical problems, feel free to hand-write them and submit a scanned copy.

For problems 1–8 we consider the changepoint detection problem presented by Carlin, Gelfand and Smith (1992).[1] The data are counts generated over discrete time as

$$X_s \sim \text{Poisson}(\mu), \text{ if } s = 1, \ldots, \tau \tag{1}$$

$$X_s \sim \text{Poisson}(\lambda), \text{ if } s = \tau + 1, \ldots, T \tag{2}$$

where τ is unknown. The vector of parameters is $\theta = (\mu, \lambda, \tau)$. Consider the independent priors $\mu \sim \text{Gamma}(a_1, b_1)$, $\lambda \sim \text{Gamma}(a_2, b_2)$, $\tau \sim \text{Uniform}(\{1, \ldots, T\})$.

1. (1 point) Derive the posterior $p(\mu, \lambda, \tau \mid x_1, \ldots, x_T)$ up to a proportionality constant.

2. (1 point) Derive the posterior conditional distributions $p(\mu \mid \lambda, \tau, x_1, \ldots, x_T)$, $p(\lambda \mid \mu, \tau, x_1, \ldots, x_T)$, $p(\tau \mid \mu, \lambda, x_1, \ldots, x_T)$.

3. (3 points) Implement a Gibbs sampler in R that iteratively samples μ, λ, τ from the conditional posteriors, and illustrate its use with data generated from the model given by (1) and (2). Submit a report with your results and code.

4. (1 point) Show that / explain why $\mu \perp\!\!\!\perp \lambda \mid \tau, x_1, \ldots, x_T$.

[1] `www.jstor.org/stable/2347570`

5. (1 point) Compute the proportionality constant involved in the posterior $p(\mu, \lambda, \tau \mid x_1, \ldots, x_T)$. Fun fact: we know that $\int x^{\alpha-1} e^{-\beta x} dx = \Gamma(\alpha)/\beta^{\alpha}$, for $\alpha, \beta > 0$, which can be seen from the fact that the density function for a gamma random variable integrates to 1.

6. (1 point) Compute the marginal posterior $p(\tau \mid x_1, \ldots, x_T)$ in closed form.

7. (1 point) Use $p(\tau \mid x_1, \ldots, x_T)$ and the fact that $\mu \perp\!\!\!\perp \lambda \mid \tau, x_1, \ldots, x_T$ to propose an approach to sample from the posterior $p(\mu, \lambda, \tau \mid x_1, \ldots, x_T)$ that does not rely on Gibbs sampling.

8. (1 point) Implement the posterior sampling approach proposed in the previous problem, illustrate its use with data generated from the model given by (1) and (2), and compare with the results obtained from Gibbs sampling. Submit a report with your results and code.

For problems 9–10, let $Z_1, Z_2 \in \{1, 2\}$, $R \in \{0, 1\}^2$. Say the full-data probability density is given by $p(r, z) \equiv p(r_1, r_2, z_1, z_2) \equiv \pi_{r_1 r_2 z_1 z_2}$. In HW2 you derived the observed-data probability density $p(r, z_{(r)})$ for all elements $(r, z_{(r)})$ in the sample space of $(R, Z_{(R)})$. Now, let the random variable $W_j \in \{1, 2, *\}$ be defined as $W_j = Z_j$ if $R_j = 1$, $W_j = *$ if $R_j = 0$, $j = 1, 2$. The probability density of (W_1, W_2) is given by $p(w_1, w_2) \equiv \kappa_{w_1 w_2}$.

9. (2 points) Explain the relationship between $p(w_1, w_2)$ and $p(r, z_{(r)})$.

10. (3 points) Propose a Bayesian approach to estimate the probabilities $\kappa = (\kappa_{11}, \ldots, \kappa_{**})$ based on a random sample $\{(W_{i1}, W_{i2})\}_{i=1}^{n} \overset{iid}{\sim} \text{Categorical}(\kappa)$. What is this procedure estimating in terms of the original Z and R?

For problems 11–12, let $Z_i = (Z_{i1}, Z_{i2})$, $Z_{i1}, Z_{i2} \in \{1, 2\}$, Z_i's are i.i.d. Denote

$$p(Z_{i1} = z_{i1}, Z_{i2} = z_{i2} \mid \theta) = \pi_{z_{i1} z_{i2}},$$

and the likelihood of the study variables as $L(\theta) = \prod_i \pi_{z_{i1} z_{i2}}$. Let $R_i = (R_{i1}, R_{i2})$, $R_{i1}, R_{i2} \in \{0, 1\}$, R_i's are i.i.d. Assume ignorability and a Dirichlet prior for $\theta = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$.

11. (2 points) Note that part (b) in the algorithm presented in slide 31 of Lecture 9 only uses $n_{kl}^{(t)}$ from part (a) of the same algorithm. Find a way of simplifying part (a) so that we don't need to sample each $z_i^{(t)}$ individually but still obtain each $n_{kl}^{(t)}$.

12. (5 points) Implement a Data Augmentation / Gibbs sampler to obtain posterior samples of θ. Illustrate its use with data generated under MCAR. Submit a report with your results and code.

For problems 13–14 we have the setup of a two-class mixture model. Think about the following generative process for the data:

- Each individual i is randomly assigned to one of two classes. Let $C_i \sim \text{Bernoulli}(\pi)$ represent the class assigned to individual i.
- Given the value of C_i, the individual gets assigned a univariate measurement $Z_i \mid C_i \sim \text{Poisson}(\gamma_{C_i})$. Here γ_j represents the parameter for class j, where $j = 0, 1$.
- Individuals are generated independently from each other.
- Assume that none of the C_i's are observed.

We will consider Bayesian estimation of the parameters of this model $(\pi, \gamma_0, \gamma_1)$, using the priors $\gamma_0 \sim \text{Gamma}(a_0, b_0)$, $\gamma_1 \sim \text{Gamma}(a_1, b_1)$, $\pi \sim \text{Beta}(a_\pi, b_\pi)$.

13. (5 points) Derive a Data Augmentation / Gibbs sampler to estimate the parameters in this model.

14. (5 points) Code this Data Augmentation / Gibbs sampler in R, and test it with some data generated using the generative process above. Submit a report with your results and code.

For problems 15–16, the distribution of the data is

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

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*). A realized value of Z_i is denoted z_i, and a realized value of the data $\mathbf{Z}$ is $\mathbf{z}$. The conjugate prior for this model is constructed in two steps

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

The joint distribution of (μ, Λ) is called *Normal-Wishart*.

Important: in this problem set we think of multivariate normals as being parameterized in terms of covariances, and the parameterization of the Wishart is such that $E(\Lambda) = \upsilon_0 W_0$. This differs from what was initially presented in class, but the lecture notes are now updated to match the parameterization used here.

Under a normal-Wishart prior, the posterior given data $\mathbf{z}$ 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 $\mu' = (\kappa_0 \mu_0 + n\bar{z})/\kappa'$, $\kappa' = \kappa_0 + n$, $\upsilon' = \upsilon_0 + n$,

$$W' = \{W_0^{-1} + n[\hat{\Sigma} + \frac{\kappa_0}{\kappa'}(\bar{z} - \mu_0)(\bar{z} - \mu_0)^T]\}^{-1},$$

$\bar{z} = \sum_{i=1}^n z_i/n$, and $\hat{\Sigma} = \sum_{i=1}^n (z_i - \bar{z})(z_i - \bar{z})^T/n$.

15. (2 points) Take $K = 2$ and assume that the Z_i's are subject to ignorable nonresponse. Derive a Data Augmentation / Gibbs sampler to obtain posterior samples from (μ, Λ).

16. (5 points) Implement the Data Augmentation / Gibbs sampler derived above. Illustrate its use with bivariate normal data with MCAR missing data. Submit a report with your results and code.

The following are computational problems that build on R session 3.

17. (10 points) Design and run a simulation study with the goal of exploring the performance of proper multiple imputation under the assumption of multivariate normality

when the data are *clearly* non-normal, and in terms of estimation of regression coefficients. Submit your R code and a pdf report with your results. If you plan to work on this problem, consult with me for guidance.

18. (10 points) Design and run a simulation study with the goal of exploring the performance of MICE in terms of estimation of regression coefficients. Your data generation should be realistic. Submit your R code and a pdf report with your results. If you plan to work on this problem, consult with me for guidance.