CS 598 – Advanced Bayesian Modeling – Spring 2025

Assignment 5

File `pigweights.csv` contains weekly weights of 48 young pigs for each of 9 consecutive weeks.[1]. You will build and compare two different varying-coefficient hierarchical normal regression models for the weights, using JAGS and `rjags`.

(a) [2 pts] On the *same* set of axes, plot segmented lines, one for each pig, representing the weight versus the week number (1 through 9). Distinguish the lines for different pigs by using different colors or line types. (You should label the axes, but no legend is needed.)

Let y_{ij} be the weight of pig j in week i ($i = 1, \ldots, 9$ and $j = 1, \ldots, 48$). For each pig, let the weight be modeled as a simple linear regression on the *centered* week number:

$$y_{ij} \mid \beta^{(j)}, \sigma_y^2, X \quad \sim \quad \text{indep. } \mathrm{N}\big(\beta_1^{(j)} + \beta_2^{(j)}(x_i - \bar{x}),\, \sigma_y^2\big)$$

where

$$\beta^{(j)} \;=\; \begin{pmatrix} \beta_1^{(j)} \\ \beta_2^{(j)} \end{pmatrix} \qquad j = 1, \ldots, 48 \qquad\qquad x_i \;=\; i \qquad i = 1, \ldots, 9$$

Note that the coefficients are allowed to depend on the pig, but the variance is not.

(b) Let $\hat{\beta}_1^{(j)}$ and $\hat{\beta}_2^{(j)}$ be the *ordinary least squares* estimates of $\beta_1^{(j)}$ and $\beta_2^{(j)}$, estimated for pig j. You may use a function like `lm` or `lsfit` in R to compute these estimates. (For this part, the coefficient pairs are estimated completely separately for each pig.)

(i) [1 pt] Produce a scatterplot of the pairs $\big(\hat{\beta}_1^{(j)}, \hat{\beta}_2^{(j)}\big)$, $j = 1, \ldots, 48$.

(ii) [1 pt] Give the average (sample mean) of $\hat{\beta}_1^{(j)}$ and also of $\hat{\beta}_2^{(j)}$.

(iii) [1 pt] Give the sample variance of $\hat{\beta}_1^{(j)}$ and also of $\hat{\beta}_2^{(j)}$.

(iv) [1 pt] Give the sample correlation between $\hat{\beta}_1^{(j)}$ and $\hat{\beta}_2^{(j)}$.

(c) Consider the bivariate prior

$$\beta^{(j)} \mid \mu_\beta, \Sigma_\beta \quad \sim \quad \text{iid } \mathrm{N}(\mu_\beta,\, \Sigma_\beta)$$

$$\mu_\beta \;=\; \begin{pmatrix} \mu_{\beta_1} \\ \mu_{\beta_2} \end{pmatrix} \qquad\qquad \Sigma_\beta \;=\; \begin{pmatrix} \sigma_{\beta_1}^2 & \rho\,\sigma_{\beta_1}\sigma_{\beta_2} \\ \rho\,\sigma_{\beta_1}\sigma_{\beta_2} & \sigma_{\beta_2}^2 \end{pmatrix}$$

with independent hyperpriors

$$\mu_\beta \quad \sim \quad \mathrm{N}\big(0,\, 1000^2 I\big)$$
$$\Sigma_\beta^{-1} \quad \sim \quad \mathrm{Wishart}_2\big(\Sigma_0^{-1}/2\big)$$

[1] Data sourced from Wand, M. (2018). *SemiPar: Semiparametric Regression.* R package version 1.0-4.2, `https://CRAN.R-project.org/package=SemiPar`.

in the notation used in the lecture videos. For your analysis, use

$$\Sigma_0 \quad = \quad \begin{pmatrix} 15 & 0 \\ 0 & 0.5 \end{pmatrix}$$

based on preliminary analyses. Let the prior on σ_y^2 be (independently)

$$\sigma_y^2 \quad \sim \quad \text{Inv-gamma}(0.0001, 0.0001)$$

(i) [2 pts] List an appropriate JAGS model. Make sure to create nodes for Σ_β, ρ, and σ_y^2.

Now run your model using `rjags`. Make sure to use multiple chains with overdispersed starting points, check convergence, and (after convergence) monitor μ_β, Σ_β, σ_y^2, and ρ long enough to obtain effective sample sizes of at least 4000 for each parameter.

(ii) [2 pts] Display the `coda` summary of the results for the monitored parameters.

(iii) [3 pts] Give an approximate 95% central posterior interval for the correlation parameter ρ, and also produce a graph of its (estimated) posterior density. Does it seem like a good idea to allow ρ to be nonzero?

Consider a pig randomly sampled from the (perhaps hypothetical) population of pigs for which the data are representative. Under the model, this pig will have coefficient vector

$$\tilde{\beta} \quad = \quad \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix}$$

defining its individual regression line

$$\tilde{\beta}_1 \; + \; \tilde{\beta}_2 \, (x - \bar{x})$$

which represents the expected weight of this pig at week x. The model implies a *population mean* regression line of

$$\mathrm{E}\big(\tilde{\beta}_1 \; + \; \tilde{\beta}_2 \, (x - \bar{x}) \mid \mu_\beta, \Sigma_\beta\big) \quad = \quad \mu_{\beta_1} \; + \; \mu_{\beta_2} \, (x - \bar{x})$$

and the following *population variance* at x over such regression lines:

$$\mathrm{var}\big(\tilde{\beta}_1 \; + \; \tilde{\beta}_2 \, (x - \bar{x}) \mid \mu_\beta, \Sigma_\beta\big) \quad = \quad \sigma_{\beta_1}^2 \; + \; 2 \, (x - \bar{x}) \, \rho \, \sigma_{\beta_1} \sigma_{\beta_2} \; + \; (x - \bar{x})^2 \, \sigma_{\beta_2}^2$$

(iv) [1 pt] The population mean regression line represents the expected weight of an "average" pig at week x. Form an approximate 95% central posterior interval for this expected weight at week 1 ($x = 1$).

(v) [1 pt] Form an approximate 95% central posterior interval for the *population variance* of the expected weight at week 1 ($x = 1$).

(vi) [1 pt] Simple calculus shows that the population variance of the expected weight of a random pig is minimized at

$$x_{\min} \quad = \quad \bar{x} \; - \; \rho \, \sigma_{\beta_1} / \sigma_{\beta_2}$$

Approximate the posterior probability that $x_{\min} < 1$, i.e., that the minimum occurs before the week 1 time point.

(vii) [2 pts] The *prior* probability that $x_{\min} < 1$ turns out to be approximately 0.205. Approximate the Bayes factor favoring $x_{\min} < 1$ versus $x_{\min} \geq 1$. Then describe the level of data evidence for $x_{\min} < 1$.

(viii) [2 pts] Use the `rjags` function `dic.samples` to compute the effective number of parameters ("penalty") and Plummer's DIC ("Penalized deviance"). Use at least 100,000 iterations.

(d) Now consider a different model with independent "univariate" hyperpriors for the model coefficients, which do not allow for a coefficient correlation parameter:

$$\begin{aligned} \beta_1^{(j)} \mid \mu_{\beta_1}, \sigma_{\beta_1} &\sim \text{iid N}\big(\mu_{\beta_1},\, \sigma_{\beta_1}^2\big) \\ \beta_2^{(j)} \mid \mu_{\beta_2}, \sigma_{\beta_2} &\sim \text{iid N}\big(\mu_{\beta_2},\, \sigma_{\beta_2}^2\big) \end{aligned}$$

with independent hyperpriors

$$\begin{aligned} \mu_{\beta_1}, \mu_{\beta_2} &\sim \text{iid N}\big(0,\, 1000^2\big) \\ \sigma_{\beta_1}, \sigma_{\beta_2} &\sim \text{iid U}(0, 1000) \end{aligned}$$

Let the prior on σ_y^2 be the same as in the previous model.

(i) [4 pts] Draw a complete DAG for this new model.

(ii) [2 pts] List an appropriate JAGS model. Make sure that there are nodes for $\sigma_{\beta_1}^2$, $\sigma_{\beta_2}^2$, and σ_y^2.

Now run your model using `rjags`. Make sure to use multiple chains with overdispersed starting points, check convergence, and (after convergence) monitor μ_{β_1}, μ_{β_2}, $\sigma_{\beta_1}^2$, $\sigma_{\beta_2}^2$, σ_y^2 long enough to obtain effective sample sizes of at least 4000 for each parameter.

(iii) [2 pts] Display the `coda` summary of the results for the monitored parameters.

(iv) [2 pts] Recall the *expected weight* at week 1 ($x = 1$) of an "average" pig, as considered in the previous analysis. Form an approximate 95% central posterior interval for this expected weight, and compare it with the result from the previous model.

(v) [2 pts] Use the `rjags` function `dic.samples` to compute the effective number of parameters ("penalty") and Plummer's DIC ("Penalized deviance"). Use at least 100,000 iterations.

(vi) [1 pt] Compare the (Plummer's) DIC values for this model and the previous one. What do you conclude?

Total: 33 pts