Exercises

1. Fitting a Multiple Regression Model Using the MCMC Procedure

In exercise physiology, an objective measure of aerobic fitness is how fast the body can absorb and use oxygen (oxygen consumption). Subjects participated in a predetermined exercise run of 1.5 miles. Measurements of oxygen consumption as well as several other continuous measurements such as age, pulse, and weight were recorded. The researchers are interested in determining whether any of these other variables can help predict oxygen consumption. This data are found in Rawlings (1998) but certain values of **Run_Pulse** were changed for illustration.

The **sasuser.fitness** data set contains the following variables:

Runtime	time to run 1.5 miles (in minutes)
Age	age of the member (in years)
Weight	weight of the member (in kilograms)
Oxygen_Consumption	a measure of the ability to use oxygen in the blood stream
Run_Pulse	pulse rate at the end of the run
Rest_Pulse	resting pulse rate

a. Fit a multiple regression model in PROC MCMC for the fitness data set. Specify the initial values for the parameter estimates as 0 and the variance as 1. Specify a normal prior distribution with a mean of 0 and a variance of 10000 for the parameter estimates and an inverse gamma distribution with a shape of 2.001 and a scale of 1.001 for the variance. Use an assignment statement to define **mu** and a MODEL statement to indicate that the response variable, **oxygen_consumption**, is normally distributed with parameters **mu** and **sigma2**. Display a fitted penalized B-spline curve for each trace plot, the Markov chain sampling history, and all the diagnostic statistics including the DIC criterion.

1) Do any of the diagnostic statistics show a problem with the convergence of the Markov chain?

2) Do the diagnostic plots show good Markov chain mixing?

3) What steps can be taken to improve the Markov chain mixing?

b. Refit the Bayesian model for the fitness data but increase the number of iterations to 200,000, the number of iterations to use in each proposal tuning phase to 5,000, the number of burn-in iterations to 5,000, the thinning rate to 5, and specify the quasi-Newton optimization in constructing the initial covariance matrix for the Metropolis-Hastings algorithm.

1) Do any of the diagnostic statistics show a problem with the Markov chain convergence?

2) Do the diagnostic plots show good Markov chain mixing?

c. Fit the same model as before, but create a data set that contains random samples from the posterior predictive distribution from a new data set **sasuser.new_fitness** and create a data set of the posterior summaries of the predicted means, medians, and standard deviations of the response variable. Do not create any diagnostic plots or output.

1) What is the predicted mean of the first scored observation?

2. Fitting a Mixed Model Using the MCMC Procedure

An investigator wants to study the effect of three surfactants on the specific volume of bread loaves. Four flours from different sources are used as blocking factors. The data are stored in the SAS data set **sasuser.bakery**, which includes the following variables:

flour the source of flour (1, 2, 3, or 4)

surf the type of surfactant (1, 2, or 3)

volume the volume of the bread loaf

a. Fit a general linear mixed model in PROC MCMC for the bakery data set. In PROC MCMC, use an ARRAY statement to define the fixed effects for **surf**, use a RANDOM statement to define the random effect **gamma** for **flour** with a normal prior distribution with a mean of 0 and a variance of s2g, and use the BEGINNODATA and ENDNODATA statements to estimate the contrast between the fixed effect parameters of surfactant 1 versus 2, surfactant 1 versus 3, and surfactant 2 versus 3. Also use noninformative priors and the monitor option to monitor the parameters of interest, use the NAMESUFFIX=POSITION option in the RANDOM statement to construct the random effect parameter names using position number. Request all the diagnostic statistics and the DIC fit statistic, display the Markov chain sampling history, and create diagnostic plots with a smooth spline in the trace plot.

1) Do any of the diagnostic statistics show a problem with the convergence of the Markov chain?

2) Do the diagnostic plots show good Markov chain mixing?

3) What steps can be taken to improve the Markov chain mixing?

b. Refit the Bayesian model for the bakery data but increase the number of iterations to 750,000, the number of iterations to use in each proposal tuning phase to 5,000, the number of burn-in iterations to 5,000, the thinning rate to 10, and specify the quasi-Newton optimization in constructing the initial covariance matrix for the Metropolis-Hastings algorithm. Furthermore, create an output data set of the posterior samples of all model parameters.

1) Do any of the diagnostic statistics show a problem with the Markov chain convergence?

2) Do the diagnostic plots show good Markov chain mixing?

c. Using the output data set with the generated posterior samples from the Bayesian model of the bakery data set, generate the posterior probability that the contrast between surfactant 1 and 2, surfactant 1 and 3, and surfactant 2 and 3 are greater than 0.

1) What is the posterior probability that the contrast between surfactant 1 and 2 is greater than 0?

d. Use the output data set for the posterior samples and create a box plot of the population volume means by surfactant.

1) Which means appear to be different on the box plot?

End of Exercises