DUE: Monday, December 12, 5PM BbLearn groups (100 points)

You may discuss this assignment with whomever you wish, but please prepare and submit work in groups of ***ONE to THREE*** students, no more and no fewer. ***Each group*** will submit a copy of their group's completed assignment, via BbLearn, including the ***names and student ID numbers of all group members*** who participated on the assignment. If you discover a mistake, you may submit another version before the deadline. The last submitted (on-time) version will be graded, with all team members receiving the same score, which will be recorded in BbLearn.

GROUP MEMBERS WHO DO NOT CONTRIBUTE SUBSTANTIALLY TO AN ASSIGNMENT MAY BE REQUIRED TO WORK IN THEIR OWN GROUP OF ONE FOR THE REMAINDER OF THE SEMESTER.

While you are permitted to discuss the assignment with other groups, please prepare your own group's code/output and written answers. GROUPS WHOSE CODE AND SOLUTIONS APPEAR SUBSTANTIALLY SIMILAR MAY BE SUBJECT TO A 10% PENALTY.

Please prepare solutions in a ***neat, organized and concise fashion***! I prefer typeset presentations (e.g., cut and paste code/output into MS Word with added exposition when appropriate; knitr via EMACS and ESS; knitr or R Markdown via RStudio, the latter being the method preferred by students in recent years). At the very least, you need to ensure code and output are presented with a fixed-width font. Neatly handwritten presentations may also be appropriate for some problems. Sloppily prepared or disorganized solutions will not receive full credit.

To complete the items below, I expect you to find and use material in our lecture notes, including code/output, possibly after some modification. Some questions may be answered with code and output alone, but some exposition may be required beyond code and output for other questions. It's up to you to communicate concisely!

7.1 Introduction

In §C.6, we considered a common improper prior distribution model for the prostate data with standardized inputs; this led to a known form of posterior distribution. See also §C.4 & C.5 for more on this improper prior model.

In §C.9 and C.10, we considered a partially improper and partially proper (vaguely informative) prior distribution model for the same prostate data; this led to known full conditional posterior distributions, which we exploited using our own 3-stage Gibbs sampling algorithm programmed in R (§C.9). See also §C.7 and C.8 for a similar but fully proper independence prior and related 2-stage Gibbs sampling algorithm. The form of the entire posterior was not of known form in this case. We also analyzed this case using Hamiltonian Monte Carlo (HMC) sampling as implemented in Stan using the rstan package in R (§C.10).

We summarized these analyses in §C.11.

In this final assignment, we repeat essentially these analyses and summary using the body fat data (use file `zfat.RDS` in BbLearn), instead of the prostate data. I standardized the inputs for the body fat data so that we may elicit priors as we did for the prostate data (§C.9) and so that we may compare all three analyses in a nice summary as illustrated in §C.11.

7.1.1 Data & LS Fit

I read the data and create an initial LS fit object for use later.

```
> zfat<- readRDS(file="zfat.RDS")
> zfat.lm<- lm(brozek ~ ., data=zfat)
```

7.2 Common Improper Prior Analysis

(1) Following §C.6.2, produce a posterior summary of the regression model effects (the betas!) and a corresponding plot of the marginal posterior t distributions for these effects as in that section. Add line color (use the `col` option in the `plot` and `lines` functions) in the same fashion as the `lty` option to help distinguish effect marginal distributions. You may have to perform further edits to the plot code of §C.6.2 to get reasonable results here. For example, horizontal and vertical plot limits must be changed as well as effect names in the plot legend! Incidentally, you will want to keep some objects created here for comparison to Gibbs sampling and HMC sampling of subsequent analyses. Show your summary and plot, including the code used to produce these. Be concise. Do not include more than requested here.

7.3 Combination Improper/Proper Independence Prior Analysis: Gibbs

7.3.1 Eliciting a Prior

Following §C.9, we first elicit priors for the effects, excluding β_0 and σ^2, which are given the same improper priors of that section. Because the output/response is not on the log scale, as in the prostate example, we will consider it unlikely that the standardized covariates will change the mean body fat percentage by more than 50 percent (not 10 or log(10)) as the covariates change over their standardized range of [0,1]. This means that we believe the (absolute value of the) effects are unlikely to be more than 50. Assuming, a priori, normality and mean effects of 0 (a priori null effects), as in §C.9, this translates to a prior variance of

$$V = (50/1.96)^2$$

for the effects. (Given the results of the previous analysis, above, you might question this prior! Why? No need to answer this.)

7.3.2 Gibbs Sampling Algorithm

With above prior distribution, I repeat the 3-stage Gibbs sampling algorithm of §C.9.3, where the full conditional distribution parameters are as discussed in §C.9. (I correct an error in starting value notation, too.)

For the initially chosen values of $\sigma^2 = \sigma^{2(t=0)}$ and $\beta_0 = \beta_0^{(t=0)}$

1. sample

$$\begin{aligned}\boldsymbol{\beta}^{*(t+1)} \,|\, \beta_0^{(t)}, \sigma^{2(t)}, \mathbf{y} &\sim [\boldsymbol{\beta}^* \,|\, \beta_0^{(t)}, \sigma^{2(t)}, \mathbf{y}] \\ &= \mathrm{N}\Big((\sigma^{-2(t)}(\mathbf{X}^{*t}\mathbf{X}^*) + V^{-1}\boldsymbol{I})^{-1}(\sigma^{-2(t)}(\mathbf{X}^{*t}\mathbf{X}^*)\widehat{\boldsymbol{\beta}}^* + V^{-1}\boldsymbol{m}_0^*), \\ &\quad (\sigma^{-2(t)}(\mathbf{X}^{*t}\mathbf{X}^*) + V^{-1}\boldsymbol{I})^{-1}\Big),\end{aligned}$$

2. sample

$$\begin{aligned}\beta_0^{(t+1)} \,|\, \boldsymbol{\beta}^{*(t+1)}, \sigma^{2(t)}, \mathbf{y} &\sim [\beta_0 \,|\, \boldsymbol{\beta}^{*(t+1)}, \sigma^{2(t)}, \mathbf{y}] \\ &= \mathrm{N}\left(\bar{y}^{*(t+1)}, \frac{\sigma^{2(t)}}{n}\right)\end{aligned}$$

3. sample

$$\begin{aligned}\sigma^{2(t+1)} \,|\, \boldsymbol{\beta}^{(t+1)}, \mathbf{y} &\sim [\sigma^2 \,|\, \boldsymbol{\beta}^{(t+1)}] \\ &= \text{inv-}\chi^2\left(n, \frac{SSE + (\boldsymbol{\beta}^{(t+1)} - \widehat{\boldsymbol{\beta}})^t(\mathbf{X}^t\mathbf{X})(\boldsymbol{\beta}^{(t+1)} - \widehat{\boldsymbol{\beta}})}{n}\right)\end{aligned}$$

4. repeat 1,2 & 3 "to convergence"

All quantities are as defined in §C.9.2 of our notes.

We will run the algorithm in three chains, each chain consisting of $M = 10000$ iterations, not including starting values. I get you started with some code adapted from §C.9.4.

```
> ## Modified code from Section C.9.3 of our notes:
> y<- zfat.lm$model[,1]
> X<- model.matrix(zfat.lm)
> Xstar<- X[,-1]
> (n<- dim(X)[1])

[1] 252

> (p<- dim(X)[2])

[1] 14

> k<- p - 1
> XtX<- crossprod(X)
> XstartXstar<- XtX[-1,-1]
> ##bhat<- solve(XtX)%*%t(X)%*%y
> bhat<- solve(XtX, crossprod(X,y))
> ##prebstarhat<- solve(XstartXstar)%*%t(Xstar)
> prebstarhat<- solve(XstartXstar,t(Xstar))
> m0star<- rep(0,k) ## bstar prior mean
>
> ## Prior variance for b1-bk (i.e., for bstar not for b0)
> V<- (50/1.96)^2
> ## Prior precision for bstar
> Vinv<- V^{-1}
> B= Vinv * diag(k)
>
> ## FC df for sigma2
> nuhat<- n
> ## Intermediate computation for sigma2 FC
> SSE <- sum((y - X%*%bhat)^2)
>
> nChain<- 3
> M<- 10000
>
> sigma2 <- matrix(NA,nChain,M+1)
> beta<- array(NA,c(nChain,M+1,p))
```

7.3.3 Starting Values

Determining starting values ("the first link") in a Gibbs sampling chain (as in Markov chain Monte Carlo (MCMC)) is somewhat of an art. (We may say "initial values" or "starting values.") The instructions here for determining starting values are not beyond the realm of what is done in practice to obtain starting values. Generally speaking, we want to create "dispersed" starting values so that, hopefully, we see all of these "far-flung" starting value links converge to a common range of values, with all chains mixing among each other, which provides a necessary (not sufficient) condition for convergence to the posterior distribution. (Creating starting values that are somehow too far out in the tails of the posterior distribution may cause the algorithm to fail due to very small likelihood values or prior density/mass values creating numerical problems. Not here.)

(2) For σ^2, use the three starting values of $\widehat{\sigma}^2/10$, $\widehat{\sigma}^2$, and $10\widehat{\sigma}^2$, where $\widehat{\sigma}^2$ is the MSE from the `zfat.lm` object created in code above—the LS fit to the fat data with covariates mapped to [0,1]. Show the three starting values and the code used to produce them.

(3) You will also need three starting values for the intercept, β_0. For these three values, generate

$$\beta_0 \sim N(\widehat{\beta}_0, \sigma^{2(0)}/n),$$

one for each of the three starting values for σ^2, just mentioned, generically denoted as $\sigma^{2(0)}$. (Note that $N(\widehat{\beta}_0, \sigma^{2(0)}/n)$ is akin to the full conditional for β_0, but not exactly the same. Again, determining starting values is much an art.) Please use `set.seed(8675309 + 86011)` immediately before generating the first of these three $\beta_0^{(0)}$ starting values, with the second and third starting values generated before any other random number generation. Show the three starting values and the code used to produce them.

7.3.4 Algorithm Code

(4) Continuing to follow §C.9.4, show the (remainder of the) Gibbs sampling code that you will run to produce samples from the posterior distribution. Just report the code for the moment. Are there any changes that you made to the code?

7.3.5 Convergence Diagnostics

(5) Run the above code (you may need to get the `mvtnorm` package first) to obtain samples from the posterior distribution, and follow §C.9.5 to obtain history (trace) plots and the Brooks-Gelman-Rubin (BGR) potential scale reduction factor (psrf) (graphically or numerically) to assess convergence. Omit the density plots, please. Report your code and results and comment briefly on convergence. (You may need to modify slightly the code from §C.9.5 for this item.)

7.3.6 Posterior Summary

(6) Following §C.9.6, report the standard summary of the posterior, omitting the first 5000 iterations from each chain. Just give the code and summary results, do not use all of the code/results shown in §C.9.6. We will compare these results to those of the other analyses herein, later. No need to comment yet.

7.4 Combination Improper/Proper Independence Prior Analysis: Stan HMC

(7) Using code in §C.10.8, use HMC in Stan to sample $\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}$. Please omit (or comment out) all code for sampling $\mathbf{x}^{*t}\boldsymbol{\beta} \,|\, \mathbf{y}$ and $Y^* \,|\, \mathbf{x}^*, \mathbf{y}$; we will not do this here. Be sure to change the prior standard deviation code in the transformed data block! Show your code.

(8) Create the data list necessary to run your Stan code. Just show your code to create the data list that you will pass to Stan, shortly.

(9) We will let Stan generate starting values for us (i.e., don't follow §C.10.12). Using §C.10.13, execute your Stan model with three chains, 10000 iterations for each chain, and 5000 warmup iterations for each chain. Again, we do not obtain samples of $\mathbf{x}^{*t}\boldsymbol{\beta} \,|\, \mathbf{y}$ or $Y^* \,|\, \mathbf{x}^*, \mathbf{y}$ in this homework, so modify the code accordingly. (And, we let Stan obtain starting values for us—fingers crossed.) Add `seed = 90210` to the list of arguments in the `sampling` function. Show your code, but do not show the output yet. Don't forget to translate Stan (to C++) then compile to an executable model; see §C.10.9 and §C.10.10.

7.4.1 Convergence Diagnostics

(10) Following §C.10.14, obtain history (trace) plots and the Brooks-Gelman-Rubin (BGR) potential scale reduction factor (psrf) (graphically or numerically) to assess convergence. Omit the density plots, please. Report your code and results and comment briefly on convergence. (You may need to modify slightly the code from §C.10.14 for this item.)

7.4.2 Posterior Summary

(11) Following §C.10.15, report the standard summary of the posterior, omitting the first 5000 iterations from each chain. Just give the code and summary results, do not use all of the code/results shown in §C.10.15. We will compare these results to those of the other analyses herein, later. No need to comment yet.

7.5 Body Fat Summary

(12) Following §C.11, recreate the figure there, now for the three analyses of the body fat data done in this homework. You have to change some indicies in the code in order to create the figure correctly for the 13 effects here, not including the intercept and not including σ^2. Instead of different line types, use solid line colors black, red and green (use the `col` argument with values 1, 2 and 3). Mention two ways that you can see (if

slightly) the effect of the non-informativeness of the improper prior analysis relative to the improper/proper combination prior analyses (refer to features in the plot, of course!).