STAT 5443 Midterm

March 17, 2017

Due: March 31, 2017

Your Name (Please print): _________________________

UAID (Please print): _________________________

Note:

1. This exam will contribute 15 points towards your final score.
2. Any attempt at academic dishonesty (e.g. using a browser during the exam) will automatically result in 0 points.
3. Append your R code to the end of your homework. In your solutions, you should just present your R output (e.g. numbers, table, figures) or snippets of R code as you deem it appropriate. Make sure to present your results (i.e., your R ouput) in a clear and readable fashion. Careless or confusing presentations will be penalized.

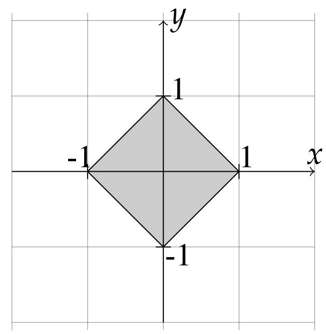
Please sign below to indicate your agreement with the following honour code.

Honour code: I promise not to cheat on this exam. I will neither give nor receive any unauthorized assistance. I promise not to share information about this exam with anyone who may be taking it at a different time. I have not been told anything about the exam by someone who has already taken it.

Signature:_________________________ Date:_________________________

Questions	Possible Points	Actual Points
1	10	
2	10	
3	20	
Total	**40**	

Problem 1 (10 pt) Derive **any one** scheme for drawing samples uniformly from the following diamond shaped area:



Explain your scheme and write a program in `R` (or a programming language of your choice) to generate 3000 samples using your algorithms and plot them.

Problem 2 (10 pt) Given $y \in \mathbb{R}^n$, consider Ridge regression with predictor matrix $X = I_{n\times n}$, i.e.,

$$\begin{aligned}\hat{\beta}^{\text{Ridge}} &= \underset{\beta\in\mathbb{R}^n}{\operatorname{argmin}} \, ||y-\beta||_2^2 + \lambda \, ||\beta||_2^2 \\ &= \underset{\beta\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_i)^2 + \lambda \sum_{i=1}^{n} \beta_i^2\end{aligned}$$

Show that the solution is:

$$\hat{\beta}_i^{\text{Ridge}} = \frac{y_i}{1+\lambda}, \; i = 1, \ldots, n.$$

Problem 3 (20 pt) A study on metabolism in 15-year-old females yielded the following data:

```
x=c(91,504,557,609,693,727,764,803,857,929,970,1043,
1089,1195,1384,1713)
```

Their energy intake, measured in megajoules, over a 24 hour period. We model the joint density as:

$$\begin{aligned}X_i &\sim \mathcal{N}(\theta, \sigma^2), \; i = 1, 2, \ldots, n \\ \theta &\sim \mathcal{N}(\theta_0, \tau^2), \quad \sigma^2 \sim \text{IG}(a, b)\end{aligned}$$

where $\text{IG}(a, b)$ is the inverted gamma distribution with density $b^a(1/x)^{a+1}e^{-b/x}/\Gamma(a)$ and with θ_0, τ^2, a, b specified. Writing $x = (x_1, \ldots, x_n)$, the posterior distribution on (θ, σ^2) is given by:

$$f(\theta, \sigma^2 \mid x) = \left[\frac{1}{(\sigma^2)^{n/2}} e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)}\right] \times \left[\frac{1}{\tau} e^{-(\theta-\theta_0)^2/(2\tau^2)}\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2}\right]$$

With a little algebra, we can get the full conditionals of θ and σ^2. (*Try to derive them on your own, it's a good practice!*)

$$\theta \mid x, \sigma^2 \sim \mathcal{N}\left(\frac{\sigma^2}{\sigma^2 + n\tau^2}\theta_0 + \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{x}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right)$$
$$\sigma^2 \mid x, \theta \sim IG\left(\frac{n}{2} + a, \frac{1}{2}\sum_i (x_i - \theta)^2 + b\right)$$

where $\bar{x}$ is the empirical average of the observations.

Using the Normal model above, implement the Gibbs sampler in `R` (or a programming language of your choice) for drawing samples from the posterior distribution of (θ, σ^2) given the data x above. Set the hyper-parameters as $a = b = 3, \tau^2 = 10$ and $\theta_0 = 5$. Plot the histograms of both $\log(\theta)$ and $\log(\sigma^2)$ and report the 90% probability intervals for $\log(\theta)$ and $\log(\sigma^2)$. Note that you can draw inverse gamma random samples using the `rigamma` function in the following package : `https://artax.karlin.mff.cuni.cz/r-help/library/pscl/html/igamma.html`.