Stat 532 **Name:** ______________________
HW 3
Due 10/03/16

Please prepare your solutions using LaTeX or another word processing software.

1. (5 points) Describe the differences between accept-reject sampling and importance sampling.

2. (5 points) How are samples from the posterior distribution useful for inference in Bayesian statistics.

3. (5 points) Assume you are interested in obtaining a posterior predictive distribution, complete the following equation.

$$p(y^*|y_1, \ldots, y_n) = \int \qquad\qquad\qquad\qquad\qquad\qquad d\theta$$

4. (5 points) Define an improper prior and give an example of model where one might be used.

5. (20 points) Implement an accept-reject sampler to compute the area under the half-ellipse $g(x) = \sqrt{3\left(1 - \frac{x^2}{4}\right)}$. Include your R code for full credit.

6. (25 points) Verify if the prior $p(\theta|\sigma^2) \sim N(\mu_0, \tau_0^2)$ and the sampling model $p(y_1, \ldots, y_n|\theta, \sigma^2) \sim N\left(\mu, \sigma^2\right)$ that the conditional posterior is distributed $N\left(\left(\frac{\mu_0}{\tau_0^2} + \frac{\sum y_i}{\sigma^2}\right)\right)\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}, \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$. Note for full credit you will need to retain the normalizing constants and work out the integration.

7. The Bridger Ski Foundation (BSF) received a generous donation and has the ability to groom another cross country ski trail in the Bozeman area. They are considering two areas: Graf Park (south of Kagy) and the Cherry River area. Good snow is paramount to skiing so the BSF seeks your help. They have recorded snowfall measurements going back fifteen years for each area. For Graf Park the values are:
{259.5, 248.1, 235.8, 252.6, 264.1, 267.3, 231.5, 237.9, 242.0, 242.2, 228.5, 245.2, 269.6, 243.5, 234.5} and for Cherry River the values are:
{240.3, 246.3, 249.8, 236.1, 252.8, 265.7, 250.9, 225.7, 276.2, 262.3 262.2, 253.5, 254.1, 260.3, 256.6} Specifically the BSF is interested in computing three probabilistic statements.

   1. $Pr[\theta_1 > \theta_2]$ where θ_1 is the mean annual snow fall at Graf Park and θ_2 is the mean annual snow fall at Cherry River.
   2. $Pr[\theta_1 > 250]$
   3. $Pr[\theta_2 > 250]$

   (a) (5 points) How would you go about addressing the researcher's questions in a classical framework? Would you be able to compute these probabilities?

   (b) (5 points) How would you address the researcher's questions in a Bayesian framework? Can you compute these probabilities?

   (c) (25 points) Using the prior structure where $p(\sigma_1^2, \theta_1) = p(\theta_1|\sigma_1^2)p(\sigma_1^2)$ and $p(\sigma_2^2, \theta_2) = p(\theta_2|\sigma_2^2)p(\sigma_2^2)$ compute the marginal posterior distributions $p(\theta_1|y_{1,1}, \ldots, y_{1,15})$ and $p(\theta_2|y_{2,1}, \ldots, y_{2,15})$, where σ_1^2 and σ_2^2 are the variances for snowfall (in inches) at Graf Park and Cherry River respectively, and $y_{i,j}$ is the observed snowfall at location i for reading j. Then using posterior samples from each distribution compute the three values specified above.