# PHP2530 Problem Set 4

Due April. 21st online

1. The Importance Sampling Algorithm
   a. Write a function which implements the importance sampling algorithm:

```
# Arguments:
#
# logTargetDensityFunc is a function of one argument:
# logTargetDensityFunc(xVal)
#
# logProposalDensityFunc is a function of one arguments:
# logProposalDensityFunc(xVal)
#
# proposalNewFunc is a function of zero arguments:
# proposalNewFunc( )
#
# Return Value:
#
# This function should return nSamples many samples, the
# corresponding log weights, as well as the estimated ESS.
#
# When the rejectionControlConstant is non-null and a positive
# quantity then perform rejection control and return only the
# accepted samples with their modi_ed log weights, the
# acceptance rate of the samples and the estimated ESS.
#
ImpSampler(nSamples, logTargetDensityFunc,
logProposalDensityFunc,
proposalNewFunc, rejectionControlConstant = NULL)
```

   b. Consider the following mixture of two Normals set up in one dimension:

      f(x) =1/3 Normal(-2; $1^2$; x) + 2/3Normal(2; $1^2$; x)
      g(x) = Normal(0; $3^2$; x)

   c. Plot the two density functions f(x) and g(x) on the same plot. Take a look at it and observe that g(x) is a reasonable importance density function.
   d. Let X ~ f(.). Theoretically compute
      i. $\mu_1^{'} = E(X)$,
      ii. $\mu_2^{'} = E(X^2)$,
      iii. $\theta = E(\exp(X))$
   e. Use the function you wrote ImpSampler( ), get the samples and the weights, compute the weighted average of the relevant h(.) functions and

thus estimate $\mu_1^{'}, \mu_2^{'}$ and $\theta$. Use m = 5000 samples without any rejection control and report the estimated ESS from your runs.

f. Now vary the rejection control constant $c \in \{1, \ldots, 10\}$. For each such c, as in the previous part, use the function you wrote ImpSampler( ) and estimate $\mu_1^{'}, \mu_2^{'}$ and $\theta$ from m = 5000 samples with rejection control. For each c, report estimated ESS ($e_C$, say), acceptance rate of the samples ($a_C$, say) from your runs. Let the error in the estimates be $\left|\hat{\mu}_1^{'} - \mu_1^{'}\right|, \left|\hat{\mu}_2^{'} - \mu_2^{'}\right|$, and $\left|\hat{\theta} - \theta\right|$. Now plot $(c, \left|\hat{\mu}_1^{'} - \mu_1^{'}\right|), (c, \left|\hat{\mu}_2^{'} - \mu_2^{'}\right|)$, $(c, \left|\hat{\theta} - \theta\right|)$, (c, $e_C$) and (c, $a_C$). Now share your observations on the error in the estimators, acceptance rate, ESS, recommendations on the choice of c.

2. Chapter 10 Question 5.
3. Chapter 10 Question 8
4. Chapter 11 Question 2
5. Chapter 11 Question 3
6. Chapter 11 Question 4
7. Chapter 13 Question 5
8. Generate / simulate n = 120 observations from a three component mixture Gaussian model with

   $$(p_1, p_2, p_3) = (0.1, 0.3, 0.6)$$
   $$(\mu_1, \sigma_1^2) = (0, 1)$$
   $$(\mu_2, \sigma_2^2) = (-2, 2)$$
   $$(\mu_3, \sigma_3^2) = (3, 16)$$

   Use these n data points to estimate $p_i, \sigma_i^2, \mu_i$ using:

   a. EM: implement EM to find their MLE
   b. MCMC (Gibbs): use proper diffused prior $p_i, \sigma_i^2, \mu_i$, implement Gibbs algorithm to find the posterior distribution of $p_i, \sigma_i^2, \mu_i$.
   c. Summarize the results and compare between the two approaches, any problems observed during the sampling or maximum likelihood estimator.
9. Chapter 15 question 3.