Homework 2: Making Inferences from the Posterior Distribution

Solutions to this assignment are to be submitted in myCourses via Assignment (formerly known as Dropbox). The submission deadline is **Wednesday October 16, 2019 at 11:59pm**. You should submit a zipped file containing a pdf (this could be a scanned handwritten document or a latex/Word generated pdf) with your written answers and the Jupyter notebook with any code needed for the assignment. Use comments to explain your code. All code and plots should be in the notebook while descriptions/explanations/derivations should be in the PDF.

Question: Inferring the posterior for a Gaussian likelihood example [PDF] The goal of this assignment is to review Module 2 in class where we discussed different techniques for deriving the parameters of the posterior distribution. We reviewed the direct estimation of parameters using conjugate priors, point estimations (MAP and MLE) of the posterior, and lastly, simulating and testing parameters in order to eventually generate samples directly (Metropolis-Hastings Sampling a form of Markov Chain Monte Carlo - MCMC). Statistics of the samples can then be calculated.

If $x_1, x_2, \ldots, x_n$ are independent observations of a random variable x in a dataset of size n, then the likelihood for the model (or the joint probability of all $x_i's$) is:

$$f(X|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

Because the Gaussian distribution is used quite a bit in behavior modeling, we will dive into working with the Gaussian likelihood function in the following exercises:

(a) (20 points) Derive the form of the posterior distribution if the likelihood is a Gaussian with known variance σ^2, but unknown mean μ, where the conjugate prior is also of the Gaussian form. This is a contrived example since we generally do not know σ^2, but it keeps the mathematics simpler, while still making the concepts clear. Use $f(\mu|X) \propto f(X|\mu)f(\mu)$

What can you say about the relationship between the parameters of the posterior, prior and likelihood functions?

(b) (30 points) For the example described above, derive the expressions for the maximum likelihood θ_{MLE} and *maximum a posteriori* θ_{MAP}

(c) (35 points) Suppose we have a class of 30 students who are about to take an exam, their heart rates (HR) are measured and the mean HR is $\bar{x} = 75$, with a standard deviation $\sigma = 10$ (in line with the derivations above, variance is known). Heart rate can give a measure of how stressed the students are going in to an exam. But having taken similar measurements before, over different semester exams, the past HR means have given us an overall mean μ of 70. The past means have varied from semester to semester giving us a standard deviation of the means of $\tau = 5$, i.e. τ reflects how much our past means have varied but does not really reflect the variability of the individual heart rates.

You goal is ultimately to update the knowledge of μ in $f(\mu|x)$. Using the expressions obtained above, find the value of θ_{MAP}. Be careful when substituting the different values of means and variances/std_dev in your formula.

Which function has more influence on the posterior in this problem? The prior or the likelihood? Why do you conclude this?

(d) (15 points) Using the Metropolis-Hastings algorithm, write your own sampler to simulate points from the posterior. The steps to accomplish this are:

1. Establish the starting value for parameter $\theta^{j=0}$; set the counter j=1
2. Draw a "candidate" parameter (or proposal) θ^c from a proposal distribution (usually another Gaussian)
3. Compute the ratio $\rho = \min(1, \frac{f(\theta^c) f(X|\theta^c)}{f(\theta^{j-1}) f(X|\theta^{j-1})})$
4. Compare ρ with a random draw u from $U(0,1)$. If $\rho > u$, then accept the proposal by setting $\theta^j = \theta^c$; otherwise set $\theta^j = \theta^{j-1}$; Record the number of accepted proposals. Efficiency of the algorithm will be computed as $\frac{\#\text{num\_accepts}}{\#\text{num\_iterations}}$
5. Set $j = j + 1$ and return to step 2 until enough draws are obtained

   You are provided with a Jupyter notebook `Sampling.ipynb` that was written for a binomial likelihood and beta prior. Modify this sample code to (i) write your MCMC sampler for the problem described in part (c) and (ii) plot the true posterior, the distribution of your simulated samples as well as the distribution of prior samples, all on the same figure. Note that your likelihood is a Gaussian with known variance and there is a lot more code here than is required for your homework.