Stat 515: Stochastic Processes I — Spring 2012
Take-Home Final Exam — April 21–May 2, 2012

This exam is worth 10 points. You have 11 days. Electronic submission is required. *Show all of your work for full credit; answers submitted without supporting work will receive little or no credit.* The rules of the exam are as follows:

A. You may not communicate about this exam with anyone other than the instructor, not even the grader. You may not receive help of any kind on this exam from anyone else except the instructor. You may not give help of any kind on this exam to anyone else.

B. You must submit your writeup and your code to ANGEL, in the appropriate dropboxes, before 5:00pm on Wednesday, May 2. Do not include your code in your writeup; the code must be easy to read (with comments as appropriate) and it should be clear how I can run it. You may use R, Matlab, or Python. I strongly encourage the use of a sensible text editor in preparing your code.

Problem 1. [2 points] Suppose that a player has five dollars and wishes to play craps until she either runs out of money or doubles her money (i.e., until the first time she has either zero dollars or some amount greater than or equal to ten dollars), at which point she stops playing. Betting in craps works as follows: If the player has x dollars and bets y dollars (where $y \leq x$), then with probability 244/495, she wins and her new total is $x + y$; with probability 251/495, she loses and her new total is $x - y$.

(a) Suppose that the player always bets one dollar. What is the probability that she will eventually double her money?

(b) Suppose that for each new game, the player bets 3, 2, or 1 dollar with probabilities 1/6, 2/6, and 3/6, respectively. (If she only has two dollars left, she bets 2 or 1 dollar with probabilities 1/3 and 2/3, respectively. If she only has one dollar left, she bets one dollar.) What is the expected number of games she will be able to play before stopping (at either zero dollars or ten or more dollars)?

Don't forget to explain all of your work in your writeup.

Problem 2. [8 points] Suppose we have parameters distributed as follows:

$$\begin{aligned} \theta_0, \theta_1 &\stackrel{\text{iid}}{\sim} N(0,1), \\ \lambda &\sim \text{beta}(2,2), \quad \text{independently of } \theta_0 \text{ and } \theta_1. \end{aligned}$$

Furthermore, suppose that, conditional on the parameters,

$$Z_1, \ldots, Z_{10} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\lambda).$$

(In other words, $P(Z_i = 1) = 1 - P(Z_i = 0) = \lambda$.) Finally, assume that $X_1, \ldots, X_{10}$ are conditionally independent—conditional on the parameters and the Z_i—with mass function

$$p(x_i \mid Z_i, \theta_0, \theta_1, \lambda) = \binom{20}{x_i} \left(\frac{e^{x_i \theta_0}}{(1 + e^{\theta_0})^{20}} \right)^{1 - Z_i} \left(\frac{e^{x_i \theta_1}}{(1 + e^{\theta_1})^{20}} \right)^{Z_i} \quad \text{for } i = 1, \ldots, 10.$$

Intuitively, this means that X_i is conditionally distributed as binomial$(20, p_i)$, where

$$p_i = \frac{\exp\{\theta_{Z_i}\}}{1 + \exp\{\theta_{Z_i}\}}.$$

(a) [**4 points**] Here are the data:

i	1	2	3	4	5	6	7	8	9	10
X_i	18	9	12	9	14	5	18	12	8	9
Z_i	1	0	0	0	1	0	1	0	0	0

Using these data:

(i) Demonstrate that λ, θ_0, and θ_1 are independent of one another in the posterior distribution.

(ii) Implement three separate importance samplers to estimate the posterior means of θ_0, θ_1, and λ, respectively. You may implement your samplers using any q distributions that you think are appropriate, but please explain what your choice is in each case.

(iii) Based on your samplers, give 95% confidence intervals for each of the three true posterior means. Make sure that you have sampled enough to ensure that your confidence intervals are no wider than 0.01.

Don't forget to explain all of your work in your writeup. Also, submit code that I could run to test your importance samplers.

(b) [**4 points**] Now, suppose that not all of the data have been observed. We only know the following:

i	1	2	3	4	5	6	7	8	9	10
X_i	18	9	12	9	14	5	18	12	8	9
Z_i	1	0	0	0	1	??	??	??	??	??

Using these data, in which $Z_6, \ldots, Z_{10}$ may now be considered to be parameters:

(i) Derive the full conditional densities (up to multiplicative constants) for λ, θ_0, and θ_1. Also derive the full conditional mass function for Z_i, where i can be any value from 6 to 10.

(ii) Implement a variable-at-a-time Metropolis-Hastings algorithm to sample from the posterior distribution of $(\theta_0, \theta_1, \lambda)$. Describe the proposal distributions you use for this purpose and how you decided how long to run the chain. For the updates of $Z_6, \ldots, Z_{10}$, use Gibbs sampling together with the fact that for any Bernoulli variable Y with mass function proportional to $\alpha^y \beta^{1-y}$,

$$P(Y = 1) = 1 - P(Y = 0) = \frac{\alpha}{\alpha + \beta}.$$

(iii) Give 95% credible intervals for the two binomial proportions $\exp\{\theta_0\}/(1 + \exp\{\theta_0\})$ and $\exp\{\theta_1\}/(1 + \exp\{\theta_1\})$. Base these intervals on the 0.025 and 0.975 quantiles of the θ_0 and θ_1 parameters, respectively.

(iv) Based on your MCMC run, give estimates of the posterior means of $Z_6, \ldots, Z_{10}$ along with corresponding confidence intervals.

Don't forget to explain all of your work in your writeup. Also, submit code that I could run to test your Metropolis-Hastings algorithm. I should be able to use your code to easily obtain an MCMC-based sample from the approximate posterior distribution.