Stats 545: Midterm exam

This is a 75-minute exam for 32 points. Write your name and PUID on each sheet, and also include the number of answer sheets. Attempt all questions.

1 Miscellaneous [8 pts]

1. What is a conjugate prior? Give an example of a conjugate prior. [2pts]
2. What is the difference between gradient descent and *stochastic* gradient descent. List a few pros and cons. [1pts]
3. Explain Newton's method to find the root of a function, and derive its update rule. What are its pros and cons? [3pts]
4. Briefly explain the Wolfe conditions and why they are necessary. [2pts]

2 Monte Carlo estimation [8 pts]

1. Explain the pros and cons of importance sampling versus rejection sampling. [1pts]

Below, you only have access to Gaussian and uniform random number generators. Provide `R` code or pseudocode when asked.

2. For a random X, $p(X) \propto \log(1 + X)$ if $X \in [0, 2]$, and 0 else. Give a rejection sampling algorithm to sample X. [2pts]
3. You want to calculate the mean of a standard Gaussian conditioned on it being larger than 4. Provide code to calculate this using simple Monte Carlo sampling. [2pts]
4. What is the problem with this estimator? [1pts]
5. Provide a better estimator using importance sampling. [2pts]

3 MCMC [7 pts]

1. Let $\mathcal{K}(x_{old}, x_{new})$ be a transition kernel producing a new sample x_{new} from x_{old}. What does it mean when:
   (a) $\mathcal{K}$ has a prob. distribution π as its stationary distribution? Why is stationarity not sufficient for MCMC? [2pts]
   (b) $\mathcal{K}$ satisfies detailed balance with respect to a probability distribution π? [1pts]
2. Show that detailed balance implies stationarity. [1pts]
3. What is effective sample size (ESS)? Why is a large ESS necessary but not sufficient for good MCMC mixing? [2pts]
4. For a fixed computational budget, what are the pros and cons of one vs multiple MCMC chains? [1pts]

4 Metropolis-Hastings and Gibbs [9 pts]

1. Consider a Metropolis-Hastings (MH) algorithm where you propose x_{new} from a $N(x_{old}, 1)$ distribution (where x_{old} is the current sample) and accept with probability $\min(1, \frac{2+x_{old}^2}{2+x_{new}^2})$. What is the stationary distribution? [2pts]
2. What is the stationary distribution if you proposed x_{new} from $N(0, 1)$ instead? [1pts]
3. What are the pros and cons of low and high MH acceptance probabilities? [1pts]

You want to sample from $p(x, y|\sigma) \propto \exp(-\frac{(x-y)^2}{y\sigma^2})$, where x is real-valued and $y \in \{1, 2, 3, 4, 5\}$ and σ is known.

4. What is the conditional distribution $p(y|x)$? And $p(x|y)$? What families do these belong to? [3pts]
5. Describe the overall Gibbs sampling algorithm briefly, and how you would use it to calculate the mean of x/y. [1pts]
6. Explain why this will perform poorly for large values of σ. [1pts]