# Advanced Computational Statistics (STAT 7730)
# Homework 7

This homework is due by **11:30 AM on Monday, November 13**. Be sure to explain your reasoning in your solutions, label your figures and attach the code at the end. When you submit your homework, you must bring it (hard copy) to our class.

**** As in my lecture, I use the phrase "X respects G" to mean that "X is a Gibbs random field with respect to G." ****

1. (6 points) We discussed the random visitation schedule in class (the first example in the Remark 3 of Gibbs sampler), in which the step 2 selects the site according to $U\{1, 2, \cdots, d\}$,
$$s \sim U\{1, 2, \cdots, d\}.$$
Show that this algorithm is Metropolis-Hastings and specify the corresponding proposal distribution.

2. (10 points)
   (a) (Conditioning rule) Suppose that $X$ respects $G = (V, E)$ and $A$ is a subset of $V$. Show that $X_A | X_{A^c}$ respects $G' = (A, E')$, where $E' = \{\{i, j\} \in E : i, j \in A\}$ and $A^c = V - A$.
   (b) Suppose that $X$ respects $G = (V, E)$ and $A, B, C$ are subsets of $V$. Show that if any path from any vertex in $A$ to any vertex in $B$ must pass through a vertex in $C$, then $X_A$ and $X_B$ are conditionally independent given $X_C$.

3. (8 points) Consider the following joint pmf for the random variables $X_1, \dots, X_6$ over $\mathbb{Z}^6$.
$$f(x_1, \dots, x_6) \triangleq \beta\big(x_1 x_2\big)^2 \big(\sin(x_2 + x_5)\big)^4 \big(\cos(x_1 e^{x_3})\big)^2 \big(x_2 + x_4 + x_6\big)^{10} e^{-\sum_{i=1}^{6} x_i^2 - |x_1 x_6| - |x_1 x_5|}$$
where $\beta$ is a normalization constant. Prove that $X_4$ and $X_6$ are conditionally independent from $X_3$ and $X_5$ given $X_1$ and $X_2$. (Graphical proofs, with explanation, are acceptable.)

4. (12 points) Suppose that $X = (X_1, \dots, X_6)$ respects the graph $G = (V, E)$ given by $V = \{X_1, \dots, X_6\}$ and $E = \{(X_1, X_2), (X_2, X_3), (X_3, X_4), (X_4, X_5), (X_5, X_6), (X_1, X_6), (X_3, X_6)\}$.
   (a) Prove that the conditional distribution of $X_1, \dots, X_5$ given $X_6 = x_6$ is a (first-order) Markov chain. (Graphical proofs, with explanation, are acceptable.)
   (b) What graph does $X_1, \dots, X_5$ respect? (A careful drawing of the labeled graph is a fine answer. For full credit your graph should be as small as possible, but still generic for the constraint that $X$ respects $G$.)
   (c) Is the conditional distribution of $X_1, \dots, X_4$ given $X_5 = x_5$ necessarily a (first-order) Markov chain? Justify your answer for full credit.

5. (12 points) Consider the following joint pmf for the random variables $U, V, W, X, Y, Z$, each taking values in the finite set $\{1, \dots, L\}$:
$$f(u, v, w, x, y, z) = \phi_1(u, v)\phi_2(u, v, w)\phi_3(w, x, y)\phi_4(x, z)\phi_5(x, y, z)\phi_6(u, z)$$
where $\phi_1, \dots, \phi_6$ are each nonnegative functions.

(a) Do these variables respect the complete graph over 6 nodes? Why or why not?

(b) Draw the smallest possible graph these variables are guaranteed to respect.

(c) Are $U$ and $X$ necessarily independent? Could they be independent?

(d) Find a collection of variables (from the set $V, W, Y, Z$) such that $U$ and $X$ are necessarily conditionally independent given the values of these variables. Your collection should have the fewest number of variables possible. Justfy your choice or prove that no such collection necessarily exists.

6. (15 points) Consider the following hidden Markov model (HMM):

- $X = (X_1, \ldots, X_n) \in \{0, 1\}^n$ and $Y = (Y_1, \ldots, Y_n) \in \mathbb{R}^n$
- $X_1 \sim$ Bernoulli$(1/2)$
- $\mathbb{P}(X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \mathbb{P}(X_i = x_i | X_{i-1} = x_{i-1}) = \begin{cases} 1-p & \text{if } x_i = x_{i-1} \\ p & \text{if } x_i \neq x_{i-1} \end{cases}$
  [$p$ is the switching probability; when $p$ is small the Markov chain likes to stay in the same state]
- conditioned on $X$, the random variables $Y_1, \ldots, Y_n$ are independent with $Y_i | X_i \sim$ Normal$(X_i, \sigma^2)$
  [another way to say this is that $Y_i = X_i + \epsilon_i$ where the $\epsilon_i$ are iid Normal$(0, \sigma^2)$]

$X$ is the hidden Markov chain and $Y$ is the observation process. We can express the entire distribution as

$$f(x, y) = \mathbb{P}(X_1 = x_1) f(y_1 | x_1) \prod_{i=2}^{n} \mathbb{P}(X_i = x_i | X_{i-1} = x_{i-1}) f(y_i | x_i)$$

where $f(y_i | x_i)$ is the conditional distribution of $Y_i$ given $X_i = x_i$ as defined above. Note that this is indeed a HMM.

(a) Download the dataset with $y_1, \ldots, y_{1000}$ from Carmen (HW7-HMMdata). It is an observation from this HMM with $n = 1000$, $p = .01$ and $\sigma = 1$. For each $y_i$, compute the most probable $X_i$, that is,

$$x_i^* \triangleq \arg\max_{x_i \in \{0,1\}} \mathbb{P}(X_i = x_i | Y_i = y_i)$$

This does not require dynamic programming. Use Bayes Rule and note that $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 1/2$, by symmetry. Plot $x_i^*$ versus $i$ and $y_i$ versus $i$ on the same graph. Make sure the two are clearly distinguishable.

(b) Now compute the most probable sequence $X$ given the sequence $y$, that is,

$$x^{**} \triangleq \arg\max_{x \in \{0,1\}^n} \mathbb{P}(X = x | Y = y)$$

This does require dynamic programming.[1] Plot $x_i^{**}$ versus $i$ and $y_i$ versus $i$ on the same graph. Make sure the two are clearly distinuishable.[2]

(c) Briefly compare and contrast the two solutions $x^*$ and $x^{**}$. Which seems the most like an observation from a sticky Markov chain?

[1]To make your dynamic programming algorithm work, you may find it useful to maximize $\log f(x, y)$ instead of $f(x, y)$, otherwise you will run into numerical underflow/overflow problems. In the dynamic programming algorithm discussed in class, simply replace products with sums of logarithms.

[2]You may find it useful to have a way to test your code. A very easy way to do this is to solve the problem by brute force for a small dataset (say $n = 10$, for which there are $2^{10} = 1024$ different $x_1, \ldots, x_{10}$ sequences to check) and compare it to your dynamic programming solution (applied to the small dataset). Repeat for many datasets to verify that you always get the same answer.