Midterm Exam

Place all answers on the question sheet provided. The exam is open course reader, notes, handouts, and homework. Write all answers clearly and in complete sentences. All answers should be supported by analysis or an argument.

First Name: ______________________________

Last Name: ______________________________

The Stanford Honor Code

1. The Honor Code is an undertaking of the students, individually and collectively:
   (a) That they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;
   (b) that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.
2. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.
3. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work. I acknowledge and accept the Honor Code.

Signature: ______________________________

Question 1 (5 POINTS)

Suppose that we are simulating the lifetime of a system that consists of three independent components in series. Component i has a Weibull distributed lifetime with shape parameter 2 and scale parameters λ_i, with $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$. Recall that the cumulative distribution function of a Weibull rv X with scale parameter λ and shape parameter α is given by

$$F_X(x) = \begin{cases} 1 - \exp(-(\lambda x)^\alpha), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

(a) [2 PTS] Compute the distribution function of the system lifetime.

Solution: Let M be the system lifetime. Then $M = \min\{X_1, X_2, X_3\}$, so

$$\begin{aligned} P(M \geq x) &= P(X_1 \geq x, X_2 \geq x, X_3 \geq x) \\ &= P(X_1 \geq x)P(X_2 \geq x)P(X_3 \geq x) \text{ by the independence of } X_1,\ X_2,\ X_3 \\ &= \exp(-(\lambda_1 x)^2) \cdot \exp(-(\lambda_2 x)^2) \cdot \exp(-(\lambda_3 x)^2) \\ &= \exp(-14x^2) \end{aligned}$$

Hence, M also has Weibull distribution with shape parameter 2 and scale parameter $\sqrt{14}$.

(b) [3 PTS] Provide an algorithm, based on simulating a single uniform rv U, for generating the system lifetime rv.

Solution: Let $F(x)$ be the CDF of M, then $F(x) = 1 - \exp(-14x^2)$ for $x \geq 0$. Its inverse is obtained in the following way:

$$\begin{aligned} y &= 1 - \exp -14x^2 \\ \exp^{-14x^2} &= 1 - y \\ -14x^2 &= \log(1-y) \qquad \text{(natural logarithm)} \\ x &= \sqrt{-\frac{1}{14}\log(1-y)} \end{aligned}$$

Therefore,

$$F^{-1}(y) = \sqrt{-\frac{1}{14}\log(1-y)}.$$

So we may generate M by setting

$$M = \sqrt{-\frac{1}{14}\log(1-U)}$$

where U is a uniform rv on (0,1).

Question 2 (5 POINTS)

Suppose that a local company is trying to estimate the total demand for their laser printers for next year. They have designed a simulation that models next year's demand. In order to estimate next year's mean demand, they run 5 independent replications of their simulation, obtaining the demand (in millions of units): 15, 16, 11, 10, 18.

(a) [2 PTS] Construct a 99% confidence interval for the mean demand. Note for a standard normal rv,

$$P(\mathcal{N}(0,1) \le 1.65) = 0.95 \qquad P(\mathcal{N}(0,1) \le 1.96) = 0.975$$
$$P(\mathcal{N}(0,1) \le 2.33) = 0.99 \qquad P(\mathcal{N}(0,1) \le 2.58) = 0.995$$

Solution:

$$\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i = 14$$

$$s^2 = \frac{1}{4}\sum_{i=1}^{5}(X_i - \overline{X}) = 11.5$$

The 99% confidence interval is

$$(14 - \frac{2.58 \times \sqrt{11.5}}{\sqrt{5}}, 14 + \frac{2.58 \times \sqrt{11.5}}{\sqrt{5}}) \approx (10.09, 17.91)$$

(b) [3 PTS] Suppose they want their estimate of next year's mean demand to be accurate to within 1 million units. How many additional replications should they run?

Solution:

$$n \approx \frac{2.58^2 * 11.5}{1^2} = 76.55$$

so we should run about 77 additional experiments.

Question 3 (23 points)

Consider the graph shown in Figure 1. There are 7 nodes, labeled with capital letters and 8 arcs connecting some of the nodes. Consider a random walk on this graph, where we move randomly from node to node, always going to neighboring nodes with equal probabilities. For example, the probability of moving from node C to node A in one step is 1/3.

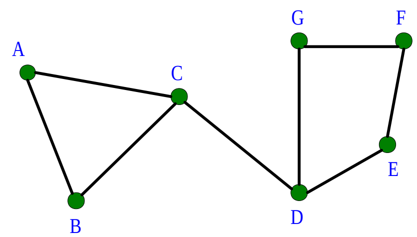

Figure 1: A random walk on graph

(a) [2 pts] What is the transition matrix for the Markov chain corresponding to the position of the random walk?

Solution:

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

(b) [3 PTS] What is the probability of going from A to D in three steps?

Solution:

$$P(A \to D \text{ in 3 steps}) = P(A \to B) \cdot P(B \to C) \cdot P(C \to D) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}.$$

(c) [2 PTS]Is the Markov chain irreducible?

Solution: Clearly, every two nodes in the graph communicate. So this Markov chain is irreducible.

(d) [4 PTS] Show that the steady-state probabilities $(\pi(x) : x \in \{A, B, \ldots, G\})$ for a random walk on such a graph are in proportion to the number of arcs that connect to a node.

Solution: Since an irreducible finite-state Markov chain is positive recurrent, we know the steady-state distribution uniquely exists. Let n_j be the number of arcs that connect to node j. It suffices to verify that

$$\sum_i n_i p_{ij} = n_j$$

for each j. Indeed, note that $p_{ij} = n_i^{-1}$ if i and j are connected and $p_{ij} = 0$ otherwise. It follows that

$$\sum_i n_i p_{ij} = \sum_{i: p_{ij} \neq 0} n_i \cdot n_i^{-1} = n_j.$$

(e) [2 PTS] What is the long-run proportion of time spent in F?

Solution: In light of part (d), we have $\pi(x) = cn_x$ where c is the normalization constant, i.e. $c = (\sum_x n_x)^{-1}$. In particular,

$$c = 2 + 2 + 3 + 3 + 2 + 2 + 2 = 16$$

and $\pi(F) = 1/8$. So the long-run proportion of time spent in F is 1/8.

For the next parts of the problem, suppose that the random walk starts in A, but stops the first time it hits either B or E. For each part, you should tell us which equations you would solve (without numerically solving them).

(f) [3 PTS] What is the expected number of visits to G before stopping, i.e., before coming to either B or E?

Solution: Let $u(x)$ be the expected number of visits to G before stopping if the random walk starts in node x. Then the first transition analysis yields the following system of linear equations.

$$\begin{aligned} u(A) &= \frac{1}{2}u(B) + \frac{1}{2}u(C) \\ u(B) &= 0 \\ u(C) &= \frac{1}{3}u(A) + \frac{1}{3}u(B) + \frac{1}{3}u(D) \\ u(D) &= \frac{1}{3}u(C) + \frac{1}{3}u(E) + \frac{1}{3}u(G) \\ u(E) &= 0 \\ u(F) &= \frac{1}{2}u(E) + \frac{1}{2}u(G) \\ u(G) &= 1 + \frac{1}{2}u(D) + \frac{1}{2}u(F) \end{aligned}$$

(g) [3 PTS] What is the expected number of steps before stopping?

Solution: Let $v(x)$ be the expected number of steps before stopping if the random walk starts in node x. Still by fist transition analysis, we have

$$\begin{aligned} v(A) &= 1 + \frac{1}{2}v(B) + \frac{1}{2}v(C) \\ v(B) &= 0 \\ v(C) &= 1 + \frac{1}{3}v(A) + \frac{1}{3}v(B) + \frac{1}{3}v(D) \\ v(D) &= 1 + \frac{1}{3}v(C) + \frac{1}{3}v(E) + \frac{1}{3}v(G) \\ v(E) &= 0 \\ v(F) &= 1 + \frac{1}{2}v(E) + \frac{1}{2}v(G) \\ v(G) &= 1 + \frac{1}{2}v(D) + \frac{1}{2}v(F) \end{aligned}$$

(h) [4 PTS] What is the probability of eventually stopping in B?

Solution: Let $q(x)$ be the probability of eventually stopping in B if the random walk starts in node x. Fist transition analysis implies

$$\begin{aligned}
q(A) &= \frac{1}{2}q(B) + \frac{1}{2}q(C) \\
q(B) &= 1 \\
q(C) &= \frac{1}{3}q(A) + \frac{1}{3}q(B) + \frac{1}{3}q(D) \\
q(D) &= \frac{1}{3}q(C) + \frac{1}{3}q(E) + \frac{1}{3}q(G) \\
q(E) &= 0 \\
q(F) &= \frac{1}{2}q(E) + \frac{1}{2}q(G) \\
q(G) &= \frac{1}{2}q(D) + \frac{1}{2}q(F)
\end{aligned}$$

Question 4 (9 POINTS)

Before Christmas of 2008, the manager of a local mill asked Bill, the analyst, to predict the annual sales for 2009 based on historical data so that he could formulate the marketing strategy. Bill went to the archive and found that the annual sales of previous years from 2005 (when the mill was built) through 2008 were 10, 15, 21, and 32 million dollars respectively. You might need the following numbers, where Y_i is the annual sales of the year 2004+i.

$$\sum_{i=1}^{4} Y_i = 78, \qquad \sum_{i=1}^{4} i = 10, \qquad \sum_{i=1}^{4} Y_i i = 231$$
$$\sum_{i=1}^{4} Y_i^2 = 1790, \qquad \sum_{i=1}^{4} i^2 = 30, \qquad \sum_{i=1}^{4} Y_i^2 i = 5969$$
$$\sum_{i=1}^{4} Y_i^3 = 46404, \qquad \sum_{i=1}^{4} i^3 = 100, \qquad \sum_{i=1}^{4} Y_i i^2 = 771$$

(a) [3 PTS] Suppose that Bill believes that a linear regression model is adequate. What is his prediction of the 2009 annual sales?

Solution: We model $(Y_i : 1 \leq i \leq n)$ as

$$Y_i = ai + b + \epsilon_i$$

where ϵ_i's are iid $\mathcal{N}(0, \sigma^2)$ rv's. Then the estimators for a and b are, respectively,

$$\widehat{a} = \frac{4\sum_{i=1}^{4} Y_i i - \left(\sum_{i=1}^{4} Y_i\right)\left(\sum_{i=1}^{4} i\right)}{4\sum_{i=1}^{4} i^2 - \left(\sum_{i=1}^{4} i\right)^2} = 7.2$$

$$\widehat{b} = \frac{1}{4}\sum_{i=1}^{4} Y_i - \widehat{a}\frac{1}{4}\sum_{i=1}^{4} i = 1.5$$

So the prediction of the 2009 annual sales is

$$5\widehat{a} + \widehat{b} = 37.5.$$

(b) [3 PTS] Suppose that the manager further asks: "Is the probability that the annual sales in 2010 are greater than 50 million dollars more than 60%?" What should Bill answer?

Solution: The standard estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{2}\sum_{i=1}^{4}(Y_i - \widehat{a}i - \widehat{b})^2 = 4.9$$

Compute

$$\widehat{\gamma} = \sqrt{\widehat{\sigma}^2\left(1 + \frac{1}{4} + \frac{(t-m)^2}{\sum_{i=1}^4 i^2 - 4m^2}\right)} \approx 4.26$$

where $t = 6$ and $m = \frac{1}{4}\sum_{i=1}^4 i = 2.5$. Then

$$Y_6 \overset{\mathcal{D}}{\approx} 6\widehat{a} + \widehat{b} + \widehat{\gamma}T_2$$

where T_2 is a Student-t rv with 2 degrees of freedom. So

$$P(Y_6 > 50) \approx P(6\widehat{a} + \widehat{b} + \widehat{\gamma}T_2 > 50) = P(T_2 > 1.2441) = 0.1697.$$

So the answer to the manager's question is "No".

(Or, simply note T_2 has symmetric distribution so the probability that T_2 is positive is less than 50%)

(c) [3 PTS] Suppose that the manager also wishes to predict the price of cotton in 2012. The price in 2008 was \$8 and the prices in 2006 and 2007 were \$9 and \$9, respectively. Bill has fit the following model to the annual prices $(Y_n : n \geq 0)$:

$$Y_{n+1} = \frac{2}{3}Y_n + \frac{1}{2}Y_{n-1} + 1 + \epsilon_{n+1}$$

where $(\epsilon_n : n \geq 1)$ is a sequence of iid $\mathcal{N}(0,1)$ rv's. What is Bill's prediction for 2012?

Solution: Let $Y_0 = 9$, $Y_1 = 9$ and $Y_2 = 8$. Then we need to predict Y_6.

$$\begin{aligned}
\widehat{Y}_3 &= \frac{2}{3}Y_2 + \frac{1}{2}Y_1 + 1 = \frac{65}{6} \\
\widehat{Y}_4 &= \frac{2}{3}\widehat{Y}_3 + \frac{1}{2}Y_2 + 1 = \frac{110}{9} \\
\widehat{Y}_5 &= \frac{2}{3}\widehat{Y}_4 + \frac{1}{2}\widehat{Y}_3 + 1 = \frac{1573}{108} \\
\widehat{Y}_6 &= \frac{2}{3}\widehat{Y}_5 + \frac{1}{2}\widehat{Y}_4 + 1 = \frac{2675}{162}
\end{aligned}$$