Program in Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
January 2013

Notice. Do four of the following five problems. Place an X on the line opposite the number of the problem that you are **NOT** submitting for grading. Please do not write your name anywhere on this exam. You will be identified only by your student number, given below and on each page submitted for grading.	1. ___ 2. ___ 3. ___ 4. ___ 5. ___
Show all relevant work!	**TOTAL.** ___

STUDENT ID NUMBER: ________________

1. Consider i.i.d. random variables $X_1, X_2, X_3, \dots$ with finite mean μ and variance σ^2. **Parts (a) and (b) of this problem are unrelated.**

   (a) Explain why $\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{\gamma n} X_i = \mu\gamma\right) = 1$, for each integer $\gamma \geq 1$.

   (b) Show that if A and B are events such that $\mathbb{P}(A) = \mathbb{P}(B) = 1$ then $\mathbb{P}(A \cap B) = 1$.

   (c) Use parts (a)-(b) to justify mathematically that if $\alpha, \beta \geq 1$ are given integers then there exists a constant $c > 0$ such that

   $$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum_{i=1}^{\alpha n} X_i}{\sum_{j=1}^{\beta n} X_j} = c\right) = 1,$$

   and find c explicitly.

2. Consider a coin that comes up heads with probability p and tails with probability $(1-p)$, where $0 < p < 1$ is a known parameter. For each integer $k \geq 1$, let T_k denote the random number of tosses to observe k consecutive heads or k consecutive tails for the first time. For instance, if the first few tosses of the coin are $t, h, h, t, t, t, \dots$ then $T_1 = 1$, $T_2 = 3$ and $T_3 = 6$. Find a formula that relates $\mathbb{E}(T_k)$ to $\mathbb{E}(T_{k-1})$ and use it to determine $\mathbb{E}(T_k)$ explicitly for each $k \geq 1$.

3. Suppose $Y_1, Y_2, \dots$ are iid random variables that are uniform on $(0, \theta)$, with $\theta > 0$ unknown.

   (a) Find a minimal sufficient statistic $T \equiv T(Y_1, \dots, Y_n)$ for θ.

   (b) Suppose that θ has the Pareto distribution with fixed, known parameters $\gamma > 0$ and $\lambda > 0$. The density $\pi(\cdot)$ of θ is $\pi(\theta) = (\lambda\gamma^\lambda)/(\theta^{\lambda+1})$ for $\theta \geq \gamma$, and is 0 otherwise. Find the posterior distribution of θ given observations $y_1, \dots, y_n$. What is the posterior mean, and how does it compare to your answer from part (a)?

4. Let (X, Y) be jointly distributed

$$f_\theta(x, y) = \exp\left(-\frac{x}{\theta} - \theta y\right), \qquad x, y \geq 0$$

for $\theta > 0$.

(a) Find the Fisher information $I(\theta)$ for estimating θ based on a single observation (x, y).

(b) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an iid sample. Find a two-dimensional sufficient statistic for θ.

(c) Derive the maximum likelihood estimator (MLE), $\hat{\theta}_n$ for θ based on $(X_1, Y_1), \ldots,$ (X_n, Y_n). Give an intuitive explanation for the functional form of this MLE.

(d) Find a representation of the exact distribution of $\hat{\theta}_n^2/\theta^2$. For example, $N(0,1)/\sqrt{(1/m)\chi_m^2}$; do *not* compute the density!

5. **In what follows, Markov processes are assumed to be time-homogeneous <u>and</u> irreducible.** Consider a Markov process $X = (X_t)_{t \geq 0}$ with state space $S = \{1, 2, 3\}$ and rate-matrix:

$$Q := \begin{bmatrix} \cdot & \alpha_1 & \alpha_2 \\ \beta_1\alpha_1/\beta_2 & \cdot & \alpha_3 \\ \beta_1\alpha_2/\beta_3 & \beta_2\alpha_3/\beta_3 & \cdot \end{bmatrix}, \qquad (1)$$

where $\alpha_i \geq 0$ and $\beta_i > 0$ for $i = 1, 2, 3$ and the terms in the diagonal are appropriately chosen so that each row adds up to zero.

(a) Determine the stationary distribution π of X and explain why it is unique. Furthermore, explain why X is time-reversible if X_0 has distribution π.

(b) Conversely, show that if a stationary Markov process over the state space $\{1, 2, 3\}$ is time-reversible then its rate-matrix must be of the form stated in equation (1).

(c) Certain mutation models in Molecular Evolution use time-reversible stationary Markov processes over the state space $\{A, C, G, T\}$ to model substitutions of a DNA base at a given position in this sequence.
Based on the findings in parts (a)-(b), claim the most general form of the rate-matrix of a time-reversible stationary Markov processes over the state space $\{1, 2, 3, 4\}$. **You do not need to justify your answer!**