Program in Applied Mathematics
PROBABILITY AND STATISTICS PRELIMINARY EXAMINATION
August 23, 2017

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. Show all relevant work.

1. ____
2. ____
3. ____
4. ____
5. ____

Total ____

Student Number ____________________________

1. At the beginning of the semester, an APPM student sorted alphabetically his n textbooks on a rack. As the semester went by, however, he kept placing each book back on the rack at a random location after consulting it. Let p_n be the probability that at the end of the semester no textbook ends at its original location (on the rack). Furthermore, let q_n be the conditional probability that no textbook ends at its original location given that the first textbook on the rack, say book A, does not either.

   (a) Determine p_1 and p_2.

   (b) Explain why $p_n = \frac{n-1}{n} \cdot q_n$, and $q_n = \frac{1}{n-1} \cdot p_{n-2} + p_{n-1}$.

   (c) Determine a recursion for $(p_n - p_{n-1})$ and use it to compute p_n explicitly.

   (d) Based on your findings, when n is large, what's approximately the probability that at least one book ends at its original position at the end of the semester?

2. Let X and Y be two random variables.

   (a) Show that if X and Y are independent, then $\text{Cov}(X, Y) = 0$.

   (b) Now show that the converse is false by providing a counter example. That is, provide an example of random variables X and Y that are uncorrelated but not independent.

   (c) Show that if X and Y are Bernouli random variables, $\text{Cov}(X, Y) = 0$ implies that X and Y are independent.

3. This problem consists of two parts.

   For the first part, let $(A_n)_{n \geq 0}$ and $(B_n)_{n \geq 0}$ be two sequences of random variables. Furthermore, let A denote a random variable. In what follows, $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively.

   (a) State the definitions of $A_n \xrightarrow{p} A$ and of $A_n \xrightarrow{d} A$.

   (b) Show using the above definitions one of the following lemmas.

   LEMMA 1. *If $A_n \xrightarrow{p} a$, where a is a finite constant, and the function $g : \mathbb{R} \to \mathbb{R}$ is continuous at a, then $g(A_n) \xrightarrow{p} g(a)$.*

   LEMMA 2. *If $A_n \xrightarrow{d} A$ and $B_n \xrightarrow{p} b$, where $b > 0$ is a finite constant, then $A_n/B_n \xrightarrow{d} A/b$.*

   For the second part, you are to show the following theorem. To do so you can invoke any well-known theoretical result as well as any of the previous two lemmas.

(c) THEOREM. *If $(X_n)_{n\geq 0}$ is an i.i.d. sequence of random variables such that $\mathbb{E}(X_i) = 0$ and $\mathbb{V}(X_i) = \sigma^2$, with $0 < \sigma < \infty$, then*

$$\frac{\sum\limits_{i=1}^{n} X_i}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}} \xrightarrow{d} \textit{Normal}(0,1).$$

4. An ecologist studying a particular population of grasshoppers measures the height (in cm) of n grasshoppers' initial jump under a replicable set of conditions in the field. A combination of empirical study and theoretical considerations has caused the ecologist to assume that each of these heights $H_1, ..., H_n$ are independent random samples from a Normal distribution with mean μ and standard deviation 1 cm. You should assume this too.

   (a) Show under what conditions on $\{a_1, ..., a_n\}$, where $a_i \in \mathbb{R}$, that $W = \sum_{i=1}^{n} a_i H_i$ will be an unbiased estimator of μ.

   (b) Find the unbiased estimator of this form that has minimum variance. What is the variance of this estimator?

   (c) Is this estimator the UMVUE? Show your work for full credit.

   (d) Now consider that the ecologist wants to estimate the probability that a grasshopper from the same population (under the same field conditions) will jump higher than 10 cm on its initial jump. Find the MLE of this probability. You can express your answer in terms of the c.d.f. of the standard Normal distribution, $\Phi(\cdot)$.

   (e) For her research, the ecologist wants to know if the jumping behavior of the same population of grasshoppers changes during a total solar eclipse. She measured 6 new grasshoppers' initial jumps during the latest eclipse. Recognizing that the field conditions had changed (the eclipse being the only difference in the conditions), she is no longer willing to assume that the grasshoppers' jump heights during an eclipse follow the same $\mathrm{N}(\mu, 1)$ distribution as before. She decides to use the nonparametric sign test under the null hypothesis that an equal number of heights should be above and below μ to test whether the grasshoppers' jumping heights changed during the eclipse. The ecologist's measurements showed that all 6 grasshoppers jumped lower than μ. What is your conclusion about the ecologist's research question?

5. In what follows, $(X_n)_{n\geq 0}$ is a given time-homogeneous Markov chain with state space $\mathcal{S}$ (either finite or countable), and probability transition matrix p. As usual, p^n denotes the n-th step probability transition matrix. A function $h : \mathcal{S} \to \mathbb{R}$ is called *harmonic* for the chain when

$$h(x) = \sum_{y\in\mathcal{S}} p(x,y)\cdot h(y), \text{ for each } x \in \mathcal{S}.$$

   (a) Are stationary distributions typically harmonic? Justify your answer rigorously!

   (b) Show that h *is harmonic if and only if* $h(x) = \sum\limits_{y\in\mathcal{S}} p^n(x,y)\cdot h(y)$, *for all* $x \in \mathcal{S}$ *and* $n > 0$.

   In what remains of this problem, it is assumed that $\mathcal{S} = \{0, \dots, N\}$ for a given integer $N > 0$. Furthermore, states 0 and N are *absorbing* i.e. $p(0,0) = 1$ and $p(N,N) = 1$.

   (c) Show that if $p(x, x-1) = p(x, x+1) = 1/2$, for $0 < x < N$, then the function $h(x) := x$, for $x \in \mathcal{S}$, is harmonic.

   (d) Fix $r \in [0,1]$ and suppose instead that $p(x, x-1) = (1-r)$ and $p(x, x+1) = r$, for $0 < x < N$. For each $x \in \mathcal{S}$, define $h(x)$ as the probability the chain is absorbed at state N given that $X_0 = x$. Show that the function h is harmonic for this chain.