Graduate Probability, 60850, Spring 2018, Notre Dame — Instructor: Steven Heilman

Name: ________________ ND ID: ____________ Date: __________

Signature: ________________.
(By signing here, I certify that I have taken this test while refraining from cheating.)

Mid-Term 1

This exam contains 8 pages (including this cover page) and 5 problems. Check to see if any pages are missing. Enter all requested information on the top of this page.

You may *not* use your book or any calculator on this exam. You *can* use your homeworks. You are required to show your work on each problem on the exam. The following rules apply:

- You have 60 minutes to complete the exam, starting at the beginning of class.
- **If you use a theorem or proposition from class or the notes or the book you must indicate this** and explain why the theorem may be applied. It is okay to just say, "by some theorem/proposition from class."
- **Organize your work**, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
- **Mysterious or unsupported answers will not receive full credit**. A correct answer, unsupported by calculations, explanation, or algebraic work will receive no credit; an incorrect answer supported by substantially correct calculations and explanations might still receive partial credit.
- If you need more space, use the back of the pages; clearly indicate when you have done this. Scratch paper appears at the end of the document.

Problem	Points	Score
1	30	
2	30	
3	30	
4	30	
5	30	
Total:	150	

Do not write in the table to the right. Good luck![a]

Reference sheet

Below are some definitions that may be relevant.

Let Ω be a universal set and let $\mathcal{F}$ be a σ-algebra on Ω. A **measure** μ is a function $\mu\colon \mathcal{F} \to [0,\infty]$ such that (i) $\mu(\emptyset)=0$, and (ii) if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$. A **probability measure** on Ω is a measure μ such that $\mu(\Omega)=1$.

Let $X\colon \Omega \to S$ be measurable. The **distribution** of X (or the **law** of X) is the probability measure μ_X defined for any measurable $A \subseteq S$ by

$$\mu_X(A) := \mathbf{P}(X \in A) = \mathbf{P}(\{\omega \in \Omega \colon X(\omega) \in A\}).$$

We say that a sequence of random variables $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges in probability** to a random variable $X\colon \Omega \to \mathbf{R}$ if: for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0.$$

We say that a sequence of real-valued random variables $X_1, X_2, \ldots$ **converges in distribution** to a real-valued random variable X if, for any $t \in \mathbf{R}$ such that $\mathbf{P}(X \leq t)$ is continuous at t,

$$\lim_{n\to\infty} \mathbf{P}(X_n \leq t) = \mathbf{P}(X \leq t).$$

We say that a sequence of random variables $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges in** L_2 to a random variable $X\colon \Omega \to \mathbf{R}$ if

$$\lim_{n\to\infty} \mathbf{E}\,|X_n - X|^2 = 0.$$

We say that a sequence of random variables $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges almost surely** to a random variable $X\colon \Omega \to \mathbf{R}$ if

$$\mathbf{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Let $X\colon \Omega \to [-\infty,\infty]$ be a random variable with $\mathbf{E}\,|X| < \infty$ and $\mathbf{E}X^2 < \infty$. We define the **variance** of X, denoted $\mathrm{var}(X)$, to be $\mathrm{var}(X) := \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2$.

We say a collection $\big(X_i \colon \Omega \to S_i\big)_{i\in I}$ of random variables is **independent** if the distribution of $(X_i)_{i\in I}$ is the product of the distributions of the X_i. That is, for any finite $J \subseteq I$ and for any measurable sets $A_i \subseteq S_i$, $i \in J$, we have

$$\mathbf{P}\Big(\bigcap_{i\in J}\{X_i \in A_i\}\Big) = \prod_{i\in J} \mathbf{P}(X_i \in A_i).$$

1. (a) (15 points) Let $X\colon \Omega \to \mathbf{R}$ be a random variable such that

$$\mathbf{P}(X > 0) > 0.$$

Show that there exists a positive integer n such that $\mathbf{P}(X > 1/n) > 0$.

(b) (15 points) Let $X_1, X_2, \ldots \Omega \to \mathbf{R}$ be random variables such that $\mathbf{E}X_i = 0$ and $\mathbf{E}X_i^2 = 1$ for all $i \geq 1$. Show that

$$\mathbf{P}(X_n > n \text{ for infinitely many } n \geq 1) = 0.$$

2. (30 points) Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. Let $a_1, a_2, \ldots \in \mathbf{R}$. Show that, for any $n \geq 1$,

$$\mathbf{P}\Big(\sum_{i=1}^{n} a_i X_i \geq t\Big) \leq e^{-\frac{t^2}{2\sum_{i=1}^{n} a_i^2}}, \qquad \forall\, t \geq 0.$$

(You can use the following inequality without proof: $\cosh(x) \leq e^{x^2/2}$, $\forall\, x \in \mathbf{R}$.)

3. (30 points) Let n be a positive integer. Suppose $X_1, X_2, \ldots$ are independent random variables that are uniformly distributed in the set $\{1, 2, \ldots, n\}$. We can think of $\{1, 2, \ldots, n\}$ as a set of baseball cards, and for any $i \geq 1$, X_i is a uniformly random baseball card that you have found. Your goal is to collect all of the n baseball cards.

   For any $0 \leq j \leq n$, let T_j be the first time that you have collected exactly j baseball cards. That is, T_j is the smallest integer k such that $\{X_1, \ldots, X_k\}$ consists of exactly k distinct elements. For example, $T_0 = 0$, $T_1 = 1$, T_2 is 2 when $X_2 \neq X_1$, T_2 is 3 when $X_2 = X_1$ and $X_3 \neq X_1$, and so on. (You may assume that $\mathbf{P}(T_j < \infty) = 1$ for all $0 \leq j \leq n$.)

   For any $1 \leq j \leq n$, let $Y_j := T_j - T_{j-1}$ be the time it takes to go from a collection of $j - 1$ distinct baseball cards to a collection of j distinct baseball cards.

   Show that Y_2 and Y_3 are independent, geometric random variables with parameters $\frac{n-1}{n}$ and $\frac{n-2}{n}$, respectively.

   (Recall that a geometric random variable Y with parameter $0 < p < 1$ is a positive-integer valued random variable such that $\mathbf{P}(Y = k) = (1 - p)^{k-1}p$ for any $k \geq 1$.)

4. (30 points) We continue the definitions and notation from the previous problem. In this problem, you may assume that $Y_1, \ldots, Y_n$ are independent random variables and for any $1 \leq j \leq n$, Y_j is a geometric random variable with parameter $\frac{n-j+1}{n}$.

   You may use the following fact: a geometric random variable Y with parameter $0 < p < 1$ has mean $\frac{1}{p}$ and variance $\frac{1-p}{p}$.

   Note that $T_n = Y_1 + \cdots + Y_n$.

   - Show that $\mathbf{E}T_n = n \log n + O(n)$ and $\mathrm{var}(T_n) = O(n^2)$.
   - Conclude that
   $$\frac{T_n}{n \log n}$$
   converges in probability to 1 as $n \to \infty$.

   (Hint: Can you bound $\mathbf{P}(|T_n - n \log n + O(n)| > tn)$?)

   So, if you want to complete a set of 100 distinct baseball cards, you would need to randomly sample about $100 \log 100 \approx 460$ baseball cards.

   (As usual, $O(a)$ denotes any quantity whose absolute value is bounded by a constant multiple ca of a.)

5. (30 points) Find a sequence of random variables $X_1, X_2, \ldots : \Omega \to \mathbf{R}$ such that

- $X_1, X_2, \ldots$ converges in probability to 0.
- $X_1, X_2, \ldots$ does **not** converge almost surely to 0.
- $X_1, X_2, \ldots$ does **not** converge in L_2 to 0.

Prove that your example of $X_1, X_2, \ldots$ satisfies the above three properties.

(As usual, it might be easiest to use $\Omega = [0, 1]$ with $\mathbf{P}$ Lebesgue measure on Ω.)

(Scratch paper)