Final exam study guide – Probability Theory (235A), Fall 2013

The final exam will be held on Thursday, Dec. 5 from 1:35 to 3:00 in 1344 Storer Hall. Please come on time! It will be a closed-book exam. The special distributions handout sheet (similar to the table on page 54 of the lecture notes) will be included with the exam.

The final exam will cover all material covered in the lectures, discussion sections and homework, with an emphasis on applications, practical computations and understanding (and being able to explain) definitions of concepts and the statements of major results. You do not need to memorize proofs.

This study guide has a short list of sample questions that are designed to be indicative of the style of questions that may appear on the final exam. (These questions are not subject to the time restriction of the exam so they may take a bit longer to solve than actual exam questions.) It is provided to aid you in studying for the final but is not meant as a substitute for going over the full material. **Do not assume that if a topic is not mentioned in these questions then it will not appear on the exam.**

1. Compute the c.d.f. or density function of $X + Y$ when X, Y are independent random variables such that:
   (a) $X \sim \mathrm{Exp}(a),\ Y \sim \mathrm{Exp}(a)$.
   (b) $X \sim \mathrm{Exp}(a),\ Y \sim \mathrm{Exp}(b)$ with $a \neq b$.
   (c) $X \sim \mathrm{Exp}(a),\ -Y \sim \mathrm{Exp}(b)$.
   (d) $X \sim \mathrm{Beta}(2, 1)$, $Y \sim \mathrm{Beta}(1, 2)$.
2. Let $X, Y \sim \mathrm{Geom}(p)$ be independent random variables. Compute the distribution of:
   (a) $X + Y$
   (b) $\min(X, Y)$
   (c) $\max(X, Y)$

   Note: when you are asked to compute the distribution of a random variable Z taking only integer values, the answer should be a formula for the probabilities $\mathbf{P}(Z = k)$, or if the distribution is one of the standard special distributions then an identification of this distribution and the relevant parameters.

3. A sequence of experiments is performed for $n = 1, 2, \ldots$ as follows: in the nth experiment, n unbiased coins are tossed. All coin tosses are independent. Let E_n be the event that at least one of the coins tossed in the nth step came up "heads".

   (a) Compute $\mathbf{P}(E_n)$.

   (b) Let B be the event that infinitely many of the events E_n occurred. Express B in terms of set operations on the events E_n. Find $\mathbf{P}(B)$.

   (c) Let C be the event that for infinitely many values of n, both E_n and E_{n+1} occurred. Express C in terms of set operations on the events E_n. Find $\mathbf{P}(C)$.

   (d) Let D be the event that for infinitely many values of n, either E_n^c or E_{n+1}^c occurred. Express D in terms of set operations on the events E_n. Find $\mathbf{P}(D)$.

   Note: in a problem of this type, be prepared to explain why $\mathbf{P}(B), \mathbf{P}(C), \mathbf{P}(D)$ have the value that you claim they do, by quoting a relevant result and providing any additional necessary arguments.

4. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables satisfying $\mathbf{P}(X_k = 1) = 2/3$, $\mathbf{P}(X_k = -2) = 1/3$. Denote $S_n = \sum_{k=1}^n X_k$.

   (a) Find $\lim_{n\to\infty} \mathbf{P}(S_n > n)$. Explain your reasoning.

   (b) Find $\lim_{n\to\infty} \mathbf{P}(S_n > 0)$. Explain your reasoning.

   (c) Find $\lim_{n\to\infty} \sqrt{n}\mathbf{P}(S_{3n} = 0)$. Explain your reasoning. (**Hint.** Look for a binomial r.v. hiding here.)

   (d) Find $\lim_{n\to\infty} n^{-2}\mathbf{E}(S_n^4)$. Explain your reasoning.

5. Let $a, b \geq 1$ be some integers. In the **Polyá urn** experiment, the urn initially has a red balls and b green balls. We successively sample a uniformly random ball from the urn, observe its color, put it back and add a new ball of the same color. Denote by E_n the event that on the nth step a red ball was sampled. For example, $\mathbf{P}(R_1) = a/(a+b)$. Denote $I_n = 1_{E_n}$, the indicator random variable of E_n. Denote $S_n = \sum_{k=1}^n I_k$, the number of red balls added to the urn in the first nth steps.

(a) Prove that for any sequence $x_1, \ldots, x_n \in \{0, 1\}$ we have the formula

$$\mathbf{P}(I_1 = x_1, I_2 = x_2, \ldots, I_n = x_n) = \frac{a(a+1)\ldots(a+k-1) \cdot b(b+1)\ldots(b+n-k-1)}{(a+b)(a+b+1)\ldots(a+b+n-1)},$$

where $k = \sum_{j=1}^{n} x_j$ (the number of x_j's for $1 \leq j \leq n$ that are equal to 1).

Hint. A key thing to figure out is why the left-hand side in this equation is independent of the ordering of $x_1, \ldots, x_n$.

(b) Deduce from part (a) a formula for $\mathbf{P}(S_n + a = k)$, the distribution of the number of red balls immediately after the nth step.

(c) Note that $\frac{S_n+a}{n+a+b}$ represents the *fraction* of red balls in the urn immediately after the nth step. Using part (b), show the convergence in distribution

$$\frac{S_n + a}{n + a + b} \implies \text{Beta}(a, b) \quad \text{as } n \to \infty.$$

(This is a bit more difficult than a realistic final exam question; if you are having difficulties, try showing this first in the case $a = b = 1$ and then in the case $a = 2$, $b = 1$.)

6. You are given a random number generator that can produce a sequence $X_1, X_2, \ldots$ of i.i.d. samples from the $U(0, 1)$ distribution. Suggest a simple algorithm that uses these samples to simulate a Poi(λ) random variable. Explain why the algorithm works, and what facts about special distributions you're relying on.

7. n college students (labelled with the numbers $1, \ldots, n$) are attending a party. Each pair i, j of students are friends with probability p, independently of each other pair.

(a) Find the expected value and variance of the number F_n of friend pairs i, j.

(b) Find the expected value and variance of the number T_n of "friendship triangles" i, j, k (triples of students each two of whom are friends).

Hint. Represent T_n as a sum of indicators $T_n = \sum_{1 \leq i < j < k \leq n} 1_{E(i,j,k)}$, where $E(i, j, k)$ is the event that i, j, k are in a friendship triangle. Then use the formulas for the sum and covariance of a sum of random variables. To compute the covariances of $1_{E(i,j,k)}$ and $1_{E(i',j',k')}$, divide into cases according to the number of students involved (3, 4, 5 or 6).

(c) Find a constant $c = c(p)$ such that we have the convergence in probability

$$\frac{T_n}{n^3} \xrightarrow[n\to\infty]{P} c,$$

and explain why this convergence holds.

Hint. _________'s inequality...

8. Let X be a positive r.v. with associated c.d.f. $F = F_X$ and density function $f = f_X$. We think of X as modeling the **time to failure** of a piece of equipment or machinery (e.g., the human body!). The **survival function** of X, denoted $R_X(t)$, is defined as the probability that the equipment survived up to time t, that is, $R_X(t) = \mathbf{P}(X > t)$. The **hazard rate function** of X, denoted $h_X(t)$, represents the instantaneous probability per unit time that the equipment will fail between time t and $t + \Delta t$ where Δt is infinitesimally small, conditioned on the knowledge that it survived up to time t. That is,

$$h_X(t) = \lim_{\Delta t \downarrow 0} \frac{\mathbf{P}(t < X < t + \Delta t \,|\, X > t)}{\Delta t} \qquad (t \geq 0).$$

(a) Express $h_X(t)$ in terms of the survival function R_X, and in terms of the c.d.f. F_X and/or density f_X.

(b) If $h_X(t)$ is the constant function $h_X(t) \equiv \rho$ (where $\rho > 0$), what is the distribution of X? Explain.

(c) Assume that the machine in question has two subsystems, and that in order for the machine to fail, both subsystems have to fail. (For example, a twin-engine plane can continue to fly with just one engine operational.) Assume that the two subsystems are independent of each other. If the hazard rate functions of the two subsystems are $h_1(t) = h_2(t) = 1$, find the hazard rate function of the full system. Explain your reasoning.

(d) Answer part (c) above if the machine fails as soon as *one* of its two subsystems fails. (This is much easier than (c); try to guess the answer based on intuition alone before computing it.)

9. Compute the volume (a.k.a. Lebesgue measure) of the subset $\{(x, y, z) \in [0, 1]^3 : xyz \leq t\}$ of the unit cube in $\mathbb{R}^3$, as a function of t.

Hint: of course this can be solved by integration, but if you rely on some known facts about special distributions you can immediately write this volume as a single integral instead of a triple integral. Furthermore, this probabilistic approach is easily generalized to n dimensions.

10. (a) Compute the characteristic function (a.k.a. Fourier-Stieltjes transform) $\varphi_X(t) = \mathbf{E}(e^{itX})$ when:

    i. $X \sim \mathrm{Bin}(n, p)$

    ii. $X \sim \mathrm{Poi}(\lambda)$

    iii. $X \sim \mathrm{Gamma}(\alpha, \lambda)$

    (b) Explain how the computations of part (a) can be used to prove the convolution identities:

    i. $\mathrm{Bin}(n, p) \boxplus \mathrm{Bin}(m, p) \stackrel{d}{=} \mathrm{Bin}(n + m, p)$

    ii. $\mathrm{Poi}(\lambda) \boxplus \mathrm{Poi}(\mu) \stackrel{d}{=} \mathrm{Poi}(\lambda + \mu)$

    iii. $\mathrm{Gamma}(\alpha, \lambda) \boxplus \mathrm{Gamma}(\beta, \lambda) \stackrel{d}{=} \mathrm{Gamma}(\alpha + \beta, \lambda)$

    (Here, "$\boxplus$" denotes convolution, i.e., sum of independent samples; "$\stackrel{d}{=}$" denotes equality of distributions.)