Math 30530, Introduction to Probability

Spring 2019

Notes for final

Organizational details

The final exam will be

Tuesday, May 7, 8am-10am, 109 Pasquerilla Center.

In preparation for the final, I will have office hours as follows:

- Thursday — 3.30pm-5pm, Hayes-Healy 132 (my usual office)
- Friday — noon-1.30pm, Pasquerilla 102 (**not** my usual office!)
- Monday — 4.30pm-6pm, Hayes-Healy 132 (my usual office).

Topics

The final will be a cumulative exam (so, covering *all* topics from the semester). Here is a list of topics that we have covered:

- Chapter 1:
  - Experiments, sample spaces, events and probability functions
  - Elementary consequences of the laws of probability
  - Principle of inclusion-exclusion
  - Spaces where all outcomes are equally likely
  - Multiplication principle and addition principle
  - Sampling with and without replacement
  - Permutations and combinations
  - Factorials and binomial coefficients
  - Counting derangements
  - Distributing indistinguishable objects among distinguishable boxes ("stars and bars")
  - Conditional probability and the multiplication principle for probability
  - Conditioning gives a new probability function

- The law of total probability
- Bayes' formula, tree diagrams

- Chapter 2:
  - Discrete random variables & probability mass functions
  - Continuous random variables & density functions
  - Independence of random variables
  - Cumulative distribution functions
  - Relationship between density and distribution functions in continuous case
  - Expectation and variance
  - Linearity and monotonicity of expectation
  - Markov's and Chebychev's inequalities
  - Law of unconscious statistician
  - Probability generating function of discrete random variables
  - Extracting statistics of a random variable from its generating function
  - The seven basic families of discrete random variables (uniform, Bernoulli, binomial, geometric, hypergeometric, negative binomial, Poisson) — for each of these you know when it is appropriate to use it as a model, what its parameters are and what they mean, what is the mass function, what is the mean, and, in some cases (all but hypergeometric and negative binomial), what is the variance
  - The four basic families of continuous random variables (uniform, exponential, normal, gamma) — same as for the basic discrete random variables listed above (except, for "mass function" read "density function", and you shouldn't be concerned too much about the Gamma)
  - The normal random variable, standard and general, connection between the two, z-scores, reading a normal table
  - Finding the density function of a function of a continuous random variable
  - Using a uniform random variable to simulate any continuous random variable

- Chapter 3:
  - Joint mass function and marginal mass functions
  - Law of unconscious statistician for joint masses
  - Linearity of expectation
  - Covariance and correlation coefficient
  - Variance of sum
  - Independence of discrete random variables, and variance of sum in this case
  - Independence and mass function

- Expectation of a conditioned random variable
- Expectation of a random variable as a weighted sum of expectations of conditioned random variables

- Chapter 4:
  - Joint density function and marginal density functions
  - Law of unconscious statistician for joint densities
  - Linearity of expectation
  - Covariance and correlation coefficient
  - Variance of sum
  - Independence of continuous random variables, and variance of sum in this case
  - Independence and density functions, distribution functions
  - Computing probabilities related to combinations of continuous random variables, via double integrals
  - Computing the density function of a function of a pair of random variables, by first computing the distribution function

- Chapter 5:
  - Moment generating function
  - Extracting statistics of a random variable from moment generating function
  - The moment generating function of a sum of independent random variables

- Chapter 6:
  - The weak law of large numbers, including quantitative version
  - Using law of large numbers to estimate integrals
  - The central limit theorem, informally and formally

Practice questions

For general preparation for the exam, it will be helpful to look over your class notes, homework, quizzes, and midterm exams. For a more practical preparation, here are the questions that appeared on the Fall 2011 final of Math 30530.

1. There are twelve parking spaces in a row outside the main building - reserved (in no particular order) for Provost Burrish and his eleven associate provosts.
   (a) One day, six of the twelve drive to work, and when each one arrives they choose a random empty parking spot to park in. When the sixth person arrives, what is the probability that he finds that the spots at both ends are free?

(b) On another day, eight of the twelve drive, and again they each choose a random empty parking spot to park in. What is the probability that after all eight have parked, there are four consecutive free parking spaces?

2. I buy three bags of holiday themed maltballs from the South Bend Chocolate Company. By the time I get home with them, all that's left in bag I is 3 red maltballs and 2 green ones; all that's left in bag II is 2 red and 1 green, and all that's left in bag III is 1 red and 3 green. I give the three bags to my wife.

   (a) If she selects one maltball from each bag, what is the probability that she selects three red balls?

   (b) If she instead selects one bag at random, and selects a ball at random from it, what is the probability that she selects a red ball?

   (c) It turns out that she did select a red ball in part b) above. Given that information, what is the probability that she made her selection from bag 1?

3. There are three hiking trials that lead up to the summit of Hoosier hill, Indiana's highest point. On a typical Saturday afternoon in summer, hikers arrive at the summit from the (easy) Colfax trail at a rate of one every 12 minutes, from the (moderate) LaSalle trail on average once an hour, and from the (strenuous) Marquette trail on average once every 90 minutes. Hikers in Indiana are rugged individualists, and trek independently of each other. One Saturday at noon I take myself up to the summit to watch the hikers arrive.

   (a) Using an appropriate random variable to model the situation, calculate the probability that over the course of an hour I see no more than 8 hikers arrive from the Colfax trail. (You may leave your answer in the form of a sum.)

   (b) What is the probability that over the course of two hours, I see no more than 8 hikers reaching the summit in total (from all three trails)? (You may leave your answer in the form of a sum.)

   (c) What is the probability that in exactly three of the five hours that I am at the summit (noon to 1pm, 1 to 2pm, etc.), I see at least one hiker arriving at the summit from the Marquette trail? (You need not fully simplify your answer, but for full credit it should not contain a summation.)

4. Let X be an exponential random variable with parameter λ. Let $Y = \sqrt{X}$.

   (a) Compute the distribution function of Y.

   (b) Compute the density function of Y.

5. A new life form, ET2, has been discovered in a remote part of the galaxy. The lifetime of a randomly chosen ET2 is an exponentially distributed random variable X with parameter λ.

   (a) Compute the cumulative distribution function F of X.

   (b) The distribution function satisfies the equation $F(1000) = 1/2$. (We say that 1000 is the *median* of the distribution.) Find λ.

(c) How does $E(X)$ compare to the median, 1000?

(d) Let $a > 0$. Find $b > a$ such that

$$P(X > b | X > a) = \frac{1}{2}.$$

6. A joint probability density function of random variables X and Y is given by the formula

$$f(x,y) = \begin{cases} \frac{2}{3}(2x+y) & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal density $f_Y(y)$.

(b) Write down (but don't evaluate!) an integral whose value equals $P(3Y > X + 1)$.

(c) Write down (but don't evaluate!) an integral whose value equals $E((X - Y)^2 \ln Y)$.

7. In Flatland, a three-sided die (marked with the numbers 1, 2, 3, one on each side) is a very popular toy. Two such dice are rolled. Let X be the number rolled on the first dice, and Y the maximum of the two numbers rolled.

(a) Find the joint probability mass function of X and Y.

$X \backslash Y$	**1**	**2**	**3**
1			
2			
3			

(b) Find $E(Y)$.

(c) Find the mass function of the random variable XY.

(d) Find $\mathrm{Cov}(X, Y)$.

8. Martin's supermarket sells grapefruits in plastic bags (four to a bag). The weight of a single grapefruit is a normal random variable with mean 0.5 lb and standard deviation 0.15 lb. The weight that can be sustained by a bag is also a normal random variable, with mean 2.2 lb and standard deviation 0.4 lb.

(a) What is the probability that a single grapefruit weighs more than .55 lbs?

(b) What is the probability that four grapefruits weight in total more than 2.2 lbs?

(c) What is the probability that a bag of four grapefruits will break when you try to lift it?

9. Let N be a positive integer, and let X be a discrete random variable with probability mass function given by

$$p(k) = \begin{cases} \frac{1}{N} & \text{if } k \in \{1, \ldots, N\} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the moment generating function of X. (You may leave it in summation form).

(b) Use the moment generating function to find $E(X)$. (You may leave your answer in summation form, but there's a bonus point for writing it in closed form).

(c) Let $X_1, \ldots, X_n$ be independent random variables, each having the same mass function p (as given above). What is the moment generating function of $X_1 + \ldots + X_n$?

10. When I make a phone call to one of my friends in Ireland, the length of the call (measured in minutes) is a random variable X with density function

$$f(x) = \begin{cases} \frac{3}{x^4} & \text{if } x \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute $E(X)$, $E(X^2)$ and $\mathrm{Var}(X)$.

(b) I make 25 calls in a row. Assuming that call lengths are independent, use the central limit theorem to estimate the probability that I am on the phone for more than 42 minutes.