Study Guide – Math 240, Midterm II

(This study guide has been adapted with permission from a study guide created by Joseph Blitzstein)

1 General Information

The second midterm will be on Wednesday, October 28. No books, notes, computers, or cell phones are allowed, except that you may bring one page of standard-sized paper (8.5" $\times$ 11") with anything you want written (or typed) on both sides. Material from the first midterm is fair game on this midterm since *probability is cumulative*. That being said, think about what results have cropped up on the homework since the first exam and you have identified prime candidates for "cumulative material".

To study, I recommend carefully going through class notes, homework problems, and this handout actively (intermixing reading, thinking, solving problems, and asking questions). After reviewing those materials I recommend solving lots and lots practice problems. (The Ⓢ problems are prime candidates!)

2 Topics

First Principles

- Combinatorics: multiplication rule, tree diagrams, binomial coefficients, permutations and combinations, sampling with/without replacement when order does/doesn't matter, inclusion-exclusion
- Basic Probability: sample spaces, events, axioms of probability, equally likely outcomes, inclusion-exclusion, unions, intersections, and complements
- Conditional Probability: definition and meaning, Bayes' Rule, Law of Total Probability, thinking conditionally, independence vs. conditional independence

Univariate Distributions

- Random Variables: definition and interpretations, stories, discrete vs. continuous, distributions, CDFs, PMFs, functions of a RV, indicator RVs, memorylessness, Poisson approximation, select sums of independent RVs
- Expected Value: linearity, indicator RVs, variance, standard deviation, LOTUS
- Important Discrete Distributions: Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric, Poisson, Uniform
- Jointly Distributed Random Variables: joint, conditional, and marginal distributions, independence

Overall Strategies

- Complements
- Conditioning
- Symmetry
- Linearity

- Indicator RVs
- Checking whether answers make sense (e.g., looking at simple and extreme cases and avoiding category errors).

3 Important Distributions

3.1 Table of Distributions

The table below *will be provided* on the exam (included as the last page). This is meant to help avoid having to memorize formulas for the distributions (or having to take up a lot of space on your page of notes). Here $0 < p < 1$. The parameters for Gamma and Beta are positive real numbers; n, r, m, and N are positive integers.

Name	Param.	PMF or PDF	Mean	Variance
Bernoulli	p	$P(X=1)=p$ $P(X=0)=1-p$	p	$p(1-p)$
Binomial	n, p	$\binom{n}{x}p^x(1-p)^{n-x}$ $x \in \{0,1,\dots,n\}$	np	$np(1-p)$
Geometric	p	$p(1-p)^{x-1}$ $x \in \{1,2,\dots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	r, p	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$ $x \in \{r, r+1, \dots\}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric	m, n, N	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$ $x \in \{0,1,\dots m\}$	$\frac{nm}{N}$	$\frac{nm(N-m)}{N^2}\left(1-\frac{n-1}{N-1}\right)$
Poisson	λ	$\frac{e^{-\lambda}\lambda^x}{x!}$ $x \in \{0,1,2,\dots\}$	λ	λ
Uniform	$a < b$	$\frac{1}{b-a}, \ x \in \{a, ..., b\}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

3.2 Connections Between Distributions

The table above summarizes the PMFs of the important distributions, and their means and variances, but it does not say where each distribution comes from (stories), or how the distributions interrelate. Some of these connections between distributions are listed below.

Also note that some of the important distributions are special cases of others. Bernoulli is a special case of Binomial; and Geometric is a special case of Negative Binomial.

1. Binomial: If $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p), then $X_1 + \cdots + X_n \sim$ Binomial(n, p).
2. Negative Binomial: If $G_1, \ldots, G_n$ are i.i.d. Geometric(p), then $G_1 + \cdots + G_n \sim$ NegBinom(r, p)
3. Symmetry: If $X \sim$ Binomial($n, 1/2$), then $n - X \sim$ Binomial($n, 1/2$).

4 Sums of Independent Random Variables

Let $X_1, \ldots, X_n$ be independent random variables. The table below shows the distribution of their sum, $X_1 + \cdots + X_n$, for various important cases depending on the distribution of X_i. The central limit theorem says that a sum of a large number of i.i.d. RVs will be approximately Normal, while these are exact distributions.

X_i	$\sum_{i=1}^{n} X_i$
Bernoulli(p)	Binomial(n, p)
Binomial(n_i, p)	Binomial($\sum_{i=1}^n n_i, p$)
Geometric(p)	NegBin(n, p)
NegBin(r_i, p)	NegBin($\sum_{i=1}^n r_i, p$)
Poisson(λ_i)	Poisson($\sum_{i=1}^n \lambda_i$)

5 Review of Some Useful Results

5.1 De Morgan's Laws

$$(A_1 \cup A_2 \cup \cdots \cup A_n)^c = A_1^c \cap A_2^c \cap \cdots \cap A_n^c$$
$$(A_1 \cap A_2 \cap \cdots \cap A_n)^c = A_1^c \cup A_2^c \cup \cdots \cup A_n^c$$

5.2 Complements

$$P(A^c) = 1 - P(A)$$

5.3 Unions

$$P(A \cup B) = P(A) + P(B) - P(A \cup B)$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P\left(A_i\right), \text{ is the } A_i \text{are disjoint;}$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \left((-1)^{k+1} \sum_{i_1 < i_2 < \cdots < i_n} P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) \right) \text{ (Inclusion-Exclusion)}$$

5.4 Intersections

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$
$$P\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) = P\left(A_1\right) P\left(A_2 \middle| A_1\right) P\left(A_3 \middle| A_1, A_2\right) \cdots P\left(A_n \middle| A_1, \ldots, A_{n-1}\right)$$

5.5 Law of Total Probability

If $E_1, E_2, \ldots, E_n$ are a partition of the sample space S(i.e., they are disjoint and their union is all of S) and $P(E_i) \neq 0$ for all i, then

$$P(A) = \sum_{i=1}^{n} P\left(A \middle| E_i\right) P\left(E_i\right)$$

5.6 Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Often the denominator $P(B)$ is then expanded by the Law of Total Probability. For continuous RVs X and Y, Bayes' Rule becomes

$$f_{Y|X}(y|x) = \frac{f_{Y|X}(x|y) f_Y(y)}{f_X(x)}.$$

5.7 Expected Value and Variance

Expected value is *linear*: for any random variables X and Y and constant c,

$$E(X+Y) = E(X) + E(Y)$$
$$E(cX) = cE(X)$$

Variance can be computed in two ways:

$$Var(X) = E\left[(X - EX)^2\right] = E\left(X^2\right) - (EX)^2$$

Constants come out from variance as the constant squared:

$$Var(cX) = c^2 Var(X).$$

5.8 Law of the Unconscious Statistician (LOTUS)

Let X be a discrete random variable and h be a real-valued function. Then $Y = h(X)$ is a random variable. To compute EY using the definition of expected value, we would need to first find the PMF of Y and use $EY = \sum_y y \cdot P(Y = y)$. The Law of the Unconscious Statistician says we can use the PMF of X directly:

$$E\left[h(X)\right] = \sum_x h(x) \cdot P(X = x).$$

5.9 Indicator Random Variables

Let A and B be events. Indicator RVs bridge between probability and expectation: $P(A) = E(I_A)$, where I_A is the indicator RV for A. It is often useful to think of a "counting" RV as a sum of indicator RVs. Indicator RVs have many pleasant properties. For example, $(I_A)^k = I_A$ for any positive number k, so it's easy to handle moments of indicator RVs. Also note that

$$\begin{aligned} I_{A\cap B} &= I_A I_B \\ I_{A\cup B} &= I_A + I_B - I_A I_B \end{aligned}$$

5.10 Symmetry

There are many beautiful and useful forms of symmetry in probability and statistics. For example:

1. If X and Y are i.i.d., then $P(X < Y) = P(Y < X)$. More generally, if $X_1, \ldots X_2$ are i.i.d., then $P\left(X_1 < X_2 < \cdots < X_n\right) = P\left(X_n < X_{n-1} < \cdots < X_1\right)$, and likewise all $n!$ orderings are equally likely (in the continuous case it follows that $P\left(X_1 < X_2 < \cdots < X_n\right) = \frac{1}{n!}$, while in the discrete case we also have to consider ties).
2. If we shuffle a deck of cards and deal the first two cards, then the probability is $1/52$ that the second card is the Ace of Spades, since by symmetry it's equally likely to be any card; it's not necessary to do a law of total probability calculation conditioning on the first card.
3. Consider the Hypergeometric, thought of as the distribution of the number of white balls, where we draw n balls from a jar with w white balls and b black balls (without replacement). By symmetry and linearity, we can immediately get that the expected value is $n\frac{w}{w+b}$, even though the trials are not independent, as the jth ball is equally likely to be any of the balls, and linearity still holds with dependent RVs.

6 Common Mistakes in Probability

6.1 Category Errors

A category error is a mistake that not only happens to be wrong, but also it is wrong in every possible universe. If someone answers the question "How many students are enrolled at Lawrence?" with "50" that is wrong, but there is no logical reason the enrollment couldn't be 50. But answering the question with "-42" or "π" or "pink elephants" would be a category error. To help avoid being categorically wrong, always think about what type an answer should have. Should it be an integer? A nonnegative integer? A number between 0 and 1? A random variable? A distribution?

- Probabilities must be between 0 and 1

- Variances must be nonnegative
- The range of possible values must make sense
- Units should make sense
- A number can't equal a random variable (unless the RV is actually a constant). Quantities such as $E(X)$, $P(X > 1)$, $F_X(1)$, are numbers. We often use the notation "$X = x$", but this is shorthand for an event (it is the set of all possible outcomes of the experiment where X takes the value x).
- Don't replace a RV by its mean, or confuse$E(g(X))$ with $g(EX)$.
- An event is not a random variable.
- Dummy variables in an integral can't make their way out of the integral.
- A random variable is not the same thing as its distribution.

6.2 Notational Paralysis

Another common mistake is a reluctance to introduce notation. This can be both a symptom and a cause of not seeing the structure of a problem. Be sure to define your notation clearly, carefully distinguishing between constants, random variables, and events.

- Give objects names if you want to work with them.

  Example: Suppose that we want to show that

  $$E\left(\cos^4\left(X^2+1\right)\right) \geq \left(E\left(\cos^2\left(X^2+1\right)\right)\right)^2.$$

  The essential pattern is that there is a RV. on the right and its square on the left; so let $Y = \cos^2\left(X^2+1\right)$, which turns the desired inequality into the statement $E\left(Y^2\right) \geq (EY)^2$, which we know is true because variance is nonnegative.
- Introduce clear notation for events and RVs of interest.

6.3 Common Sense and Checking Answers

Whenever possible (i.e., when not under severe time pressure), look for simple ways to check your answers, or at least to check that they are plausible. This can be done in various ways, such as using the following methods.

1. *Miracle checks.* Does your answer seem intuitively plausible? Is there a category error? Did asymmetry appear out of nowhere when there should be symmetry?
2. *Checking simple and extreme cases.* What is the answer to a simpler version of the problem? What happens if $n = 1$ or $n = 2$, or as $n \to \infty$, if the problem involves showing something for all n?
3. Looking for alternative approaches and connections with other problems. Is there another natural way to think about the problem? Does the problem relate to other problems we've seen?
   - Probability is full of counterintuitive results, but not impossible results!
   - Check simple and extreme cases whenever possible.
   - Check that PMFs are nonnegative and sum to 1, and PDFs are nonnegative and integrate to 1 (or that it is at least plausible), when it is not too messy.

6.4 Random Variables vs. Distributions

A random variable is not the same thing as its distribution! Some people call this confusion sympathetic magic, and the consequences of this confusion are often disastrous. Every random variable has a distribution (which can always be expressed using a CDF, which can be expressed by a PMF in the discrete case, and which can be expressed by a PDF in the continuous case).

Every distribution can be used as a blueprint for generating RVs (for example, one way to do this is using the Probability Integral Transformation). But that doesn't mean that doing something to a RV corresponds to doing it to the distribution of the RV. Confusing a distribution with a RV with that distribution is like confusing a map of a city with the city itself, or a blueprint of a house with the house itself. The word is not the thing, the map is not the territory.

- A function of a RV is a RV
- Avoid sympathetic magic
- A CDF $F(x) = P(X \leq x)$ is a way to specify the distribution of X, and is a function defined for all real values of x. Here X is the RV, and x is any number; we could just as well have written $F(t) = P(X \leq t)$.

6.5 Conditioning

It is easy to make mistakes with conditional probability so it is important to think carefully about what to condition on and how to carry that out.

- Condition on *all* the evidence!
- Don't destroy information.
- Independence shouldn't be assumed without justification, and it is important to be careful not to implicitly assume independence without justification.
- Independence is completely different from disjointness!
- Independence is a symmetric property: if A is independent of B, then B is independent of A. *There's no such thing as unrequited independence.*
- The marginal distributions can be extracted from the joint distribution, but knowing the marginal distributions does not determine the joint distribution if the RVs are independent.
- Don't confuse $P(A|B)$ with $P(B|A)$.
- Don't confuse $P(A|B)$ with $P(A, B)$.