math-240-notes

Katie St. Clair

2024-01-01

Table of contents

Math 240 Probability

This is a reading guide resource for students in my Math 240 Probability class taught at Carleton College (Northfield, MN). Our textbook is Wagaman (2021). You can find this book online in our library.

1 First principles of probability

First principles cover sections in chapter 1 of Wagaman (2021). Make sure to review our week 1 schedule.

Day 1

- What to read: Read sections 1.1, 1.2, 1.3, 1.4, and 1.8 to see the generalization of property #3.
- Learning objectives: These sections will help you
  - define basic probability and set theory terminology
  - define fundamental properties of probability

Section 1.1

Key terminology to know:

- [ ] Sample space Ω
- [ ] Outcome ω: this can also be called an element
- [ ] Event A, B, ...

Comments:

The sample space for a random experiment is the list of all possible outcomes. To define a sample space, start by thinking about how to define a couple individual outcomes and then generalize this to all outcomes. If there are a large or infinite number of outcomes, then you can use a "…" to show that a pattern will continue (like Examples 1.3 and 1.4).

There may not be a unique way to define a sample space and outcomes for a random experiment. For example, one may chose to define the sample space for Example 1.3 as just the number of votes for Yolanda

$$\Omega = \{0, 1, 2, \dots, 999, 1000\}$$

. But all the outcomes in a sample space must define unique possibilities for a random experiment (i.e. they are mutually exclusive, section 1.4).

"Complex" random experiments are often composed of simpler experiments. Example 1.2 is an example of this as we define one outcome for two dice rolls as the joint outcome of two individual die rolls. Similar for the sample space for three coin flips. In scenarios where the individual simpler experiments have equally likely outcomes, the outcomes in the more complex sample space (e.g. two dice rolls, three coin flips) are also equally likely outcomes.

Section 1.2-1.3

Key ideas to know:

- ☐ relative frequency interpretation of probability
- ☐ probability function and its essential properties

Comments:

Make sure you can put the probability function properties into words: (1) means probabilities can't be negative, (2) means something in the sample space has to happen with probability 1 and (3) means the probability of an event is just the sum of the probabilities of outcomes that make up that event.

Section 1.4 + 1.8

Key set theory ideas to know

- ☐ complement A^c (not A)
- ☐ union $A \cup B$ (at least one A or B or both)
- ☐ intersection $A \cap B = AB$ (both A and B)
- ☐ $(AB)^c = A^c \cup B^c$ (at most one)
- ☐ $(A \cup B)^c = A^c B^c$ (neither)
- ☐ AB^c (A but not B)
- ☐ subset $\subseteq$
- ☐ mutually exclusive/disjoint events
- ☐ Venn diagram
- ☐ empty set $\emptyset$

Comments:

The first six ideas create a new event out of one or more events (think addition, subtraction, etc). Subset, mutually exclusive and Venn diagrams tells us relational information about events (think less than, etc). The empty set is a set that has no outcomes (think zero).

Key probability properties to know

- ☐ Addition rule for mutually exclusive events (only add probability of events when they are mutually exclusive)
- ☐ General addition rule (events do not need to be mutually exclusive)
- ☐ Complement rule

Comments:

The addition rule properties are used to find the probability of a *union* of two or more events. When events are mutually exclusive, we just add the individual event probabilities. Make sure that you assess whether events are mutually exclusive before simply adding their probabilities.

When events are not mutually exclusive, you start by adding event probabilities but then you need to subtract out the probability of the overlap (intersection) between events. The inclusion/exclusion rule is an extension of property (3) (eq. 1.3).

The complement rule along with the addition rule is useful for computing the probability of neither A nor B:

$$P(A^c B^c) = 1 - P(A \cap B)$$

Day 2

- What to read: Read sections 1.5 and 1.6
- Learning objectives: These sections will help you
  - use the multiplication principle to count the number of outcomes in a sample space or event
  - compute event probabilities using counting when outcomes are equally likely
  - define and count permutations
- After reading, take reading quiz 1

Section 1.5

Key idea to know:

- ☐ equally like outcomes

Comments:

Our definition of the probability function Section 1.3 means that the probability of any event A is equal to the number of outcomes in A divided by the total number of outcomes when all outcomes are equally likely.

Section 1.6

Key ideas to know:

- ☐ multiplication principle
- ☐ permutation
- ☐ counting the number of permutations of n unique/distinct objects
- ☐ counting the number of permutations of size k from n unique/distinct objects
- ☐ sampling with vs without replacement

Comments:

The outcomes counted by the multiplication principle describe a specific ordering, or arrangement, of a random experiment. For example 1.13, the outcomes in the sample space are found by fixing the exam (eg exams 1-4) and randomly assigning one of five grades to each exam. There are $5 \times 5 \times 5 \times 5 = 5^4$ such outcomes in the sample space. The complement of the event of interest is getting no A's and we must describe outcomes in this set the same way that we described them in the sample space (eg assigning one of four non-A grades to each exam). There are 4^4 such ways to assign a non-A grade to each exam.

Permutations are always ordered arrangements of unique/distinct objects or individual outcomes.

Example 1.14 shows a common technique in counting problems. One outcome describes a specific (unique) arrangement of 15 books (eg positions 1-5 are top shelf, 6-10 middle, and 11-15 bottom). There are 15! such arrangements in the sample space. The event of interest, all math books on the bottom, uses both permutation counting and the multiplication principle. Permutations counts the number of ways to arrange math books (5!) and novels (10!). Each unique arrangement of math books can be combined with each unique arrangement of novels. Hence the multiplication principle is used to count the number of ways to get 10 novels on the top and middle shelves and 5 math books on the bottom: $10! \times 5!$.

One twist to problem 1.14: suppose the event of interest was simply that the 5 math books were on the same shelf (ie they could be on the top, middle or bottom shelves). Each of these three options, $A_{top}, A_{middle}, A_{bottom}$ for math book placement has $10! \times 5!$ ways to arrange the books and each of these shelf positions A_i is mutually exclusive. Hence the probability that the 5 math books are on the same shelf is found using the addition rule for mutually exclusive events:

$$P(A_{top} \cup A_{middle} \cup A_{bottom}) = P(A_{top}) + P(A_{middle}) + P(A_{bottom}) = 3 \times \frac{10! \times 5!}{15!}$$

Day 3

- What to read: Read sections 1.7
- Learning objectives: These sections will help you
  - define and count combinations using the binomial coefficient
  - use the binomial coefficient to count permutations of two types of objects (a binary sequence)
  - distinguish between a permutation and a combination
- After reading, take reading quiz 2

Section 1.7

Key ideas to know:

- ☐ combination
- ☐ correspondence between a binary sequence/list and an unordered subset/sample of size k from n unique objects
- ☐ binomial coefficient

Comments:

We will be considering counting problems where outcomes are either *ordered* and *unordered.* Make sure that your strategy for solving a problem is consistent when counting sample space and event outcomes (eg don't use an ordered strategy for one and an unordered strategy for the other).

While you need to describe outcomes in an event and sample space of interest in the same manner, you don't need to use the same counting method to count outcomes in each. e.g. Example 1.26 uses the multiplication principle to count all possible ways to flip a coin 20 times while the binomial coefficient is used to count how many of these outcomes contain example 10 H and 10 T.

The binomial coefficient, which counts both the number of unordered subsets of unique objects and the number of binary sequences, can be generlized to a multinomial coefficient. Example 1.15(ii) involves picking 4 subsets of 13 cards from a deck of 52: $\frac{52!}{13!13!13!13!}$. Example 1.27(ii) also uses a multinomial coefficient in the numerator: $\frac{20!}{4!5!3!8!}$ counts the number of ways to arrange 4 As, 5 Gs, 3 Ts, and 8 Cs.

Don't worry about the binomial theorem (and its proof) for now.

2 Thinking conditionally

Chapter 2 of Wagaman (2021) covers conditional probabilities and the concept of independence.

Make sure to review Chapter 1 before this chapter and review our week 2 schedule.

Day 4

- What to read: Read sections 2.1, 2.2, 2.3
- Learning objectives: These sections will help you
  - conceptualize and compute a conditional probability
  - use the general multiplication rule to compute the probabilities of intersections
  - use tree diagrams to organize conditional information and compute probabilities
- After reading, take reading quiz 3

Section 2.1 and 2.2

Key terminology to know:

- ☐ conditional probability function and its essential properties

Comments:

Don't get bogged down in the notation choice used to define a conditional probability, $P(A \mid B)$. The big idea is that to compute a conditional probability we compute the ratio of the intersection probability (chance that both events occur) relative to the probability of the event we are conditioning on (ie the event that has occurred).

Don't worry about the simulation code in example 2.3.

The conceptual idea in 2.2 is most important, that a conditional probability function $P(\cdot \mid B)$ should have the same properties as an unconditional probability $P(\cdot)$. The only difference is that the sample space for the conditional probability is restricted to only outcomes that agree with the conditional information.

Section 2.3

Key terminology to know:

- ☐ general multiplication rule for two or more events
- ☐ tree diagram

Comments:

Tree diagrams are a useful visual tool for organizing information that is conditional or sequential in nature. Venn diagrams are not useful for organizing such information.

Day 5

- What to read: Read sections 2.4 and 2.5
- Learning objectives: These sections will help you

  - use the law of total probability to compute an unconditional probability from conditional information
  - use Bayes rule to compute a "flipped/inverted" conditional probability

- After reading, take reading quiz 4

Section 2.4

Key terminology to know:

☐ sample space partition
☐ law of total probability (LoTP)

Comments:

Recognize that the LoTP uses both the multiplication rule (to compute intersections) and the addition rule for mutually exclusive events.

Section 2.5

Key terminology to know:

☐ Bayes formula/rule
☐ tree diagram

Comments:

Bayes formula is useful when we want to flip the direction of a conditional probability, eg. we want $P(B \mid A)$ but we are given $P(A \mid B)$ along with unconditional info about B. But the formula didn't appear from nothing, it is based on the original definition of a conditional probability from 2.1: the numerator is the multiplication rule and the denominator is the LoTP.

Day 6

- What to read: Read sections 2.6
- Learning objectives: These sections will help you
  - determine whether events are independent or dependent
  - compute intersection probabilities *if* events are independent
- After reading, take reading quiz 5

Section 2.6

Key terminology to know:

☐ independent and dependent events
☐ mutual independence
☐ multiplication rule for independent events

Comments:

Independence is proved (or disproved) by thinking conditionally: if the occurrence of B doesn't affect the probability of A, then these events are independent.

If we know events are independent, then we can multiply unconditional probabilities to compute intersections. A common error is that I see is when the multiplication rule for independent events is used without first checking the essential assumption of independence.

Don't worry about A before B results (ie they are interesting but not a general rule you need to know).

3 Introduction to discrete random variables

Chapter 3 of Wagaman (2021) introduces us to discrete random variables and some common probability models used to describe these random variables.

Make sure to review our week 3 schedule.

Day 7

- What to read: Read sections 3.1, 3.2, 3.3, 3.4, plus the pmf definition on page 126
- Learning objectives: These sections will help you
  - understand what a "random variable" (RV) is
  - define a probability mass function and support set for a discrete RV
  - understand whether discrete RV are independent
  - get familiar with some common discrete RV types (uniform, Bernoulli, binomial, Poisson)
- After reading, take reading quiz 6

Section 3.1 and page 126

Key terminology to know:

- ☐ random variable
- ☐ discrete random variable
- ☐ probability mass function (pmf) of a RV
- ☐ the (support) set S of a RV
- ☐ uniform RV (know shorthand notation, pmf and support set)

Comments:

Pay special attention to the notation used for random variables (RV). Uppercase letters (typically at the end of the English alphabet) are used to denote the RV (which is random and doesn't have a fixed value) while lower case are used to denote a fixed numeric value (which is fixed even though there may not be a value specified).

The "distribution" of a random variable describes the probability structure of the RV, so think pmf if you are asked to describe a distribution. (ie what values of the RV are most likely, which ones are less likely, etc) You can also describe a distribution of a RV by name if it has a common pmf (eg "The distribution of X is discrete uniform.")

The subsection "Random variables as function" explains how a RV $X(\cdot)$ takes in one or more outcomes ω from the sample space of an experiment and outputs a numeric value. This is our formal definition of a RV but we typically won't be using the $X(\omega)$ notation for a RV. Instead we will just refer to RV as X, Y, etc. But keep this underlying connection with the sample space in mind as the term progresses (especially for discrete RV).

The uniform random variable box on page 96 is an example of a probability mass function (pmf) and support set S. Even though pmf aren't formally defined until ch. 4, I find it useful to use that language at the start of discrete RV discussions.

Section 3.2

Key terminology to know:

- ☐ independent discrete random variables

Comments:

We can use either Equations 3.1 or 3.2 to define two independent discrete RV. The general definition of independent random variables on page 98 extends beyond discrete RV to continuous RV (which we will cover soon). Skim this idea but our main focus right now is discrete RV.

Section 3.3 and 3.4

Key terminology to know:

- ☐ Bernoulli RV (know shorthand notation, what it counts, pmf and support set)
- ☐ Independent and identically distributed (i.i.d.)
- ☐ Binomial RV (know shorthand notation, what it counts, pmf and support set)

Comments:

Bernoulli and binomial RV are extremely common types of RV so make sure to carefully review these sections.

Your book describes a binomial RV as the sum of an [i.i.d.] Bernoulli "sequence" of length n. Another common way to describe this is the use the phrasing "trials" instead of "sequence": a binomial RV is the sum of n i.i.d. Bernoulli trials.

The R code on pages 102 and 104 will be talked about on day 8.

Skim Example 3.14 but we will focus on solving problems involving two RV later on this term.

Day 8

- What to read: Read sections 3.5
- Learning objectives: These sections will help you
  - get familiar with the Poisson distribution
  - review or get introduced to infinite series
- After reading, take reading quiz 7

Section 3.5

Key terminology to know:

☐ Poisson RV (know shorthand notation, what it counts, pmf and support set)

Comments:

A Poisson RV is another RV that counts "things", make sure you can distinguish settings where we should be modeling a count RV as a Poisson RV vs. a binomial RV (vs. something else).

Series are a Calc BC or Math 210 topic that not all of you have seen. If you haven't had prior exposure to these ideas please stop by drop-in hours (or make an appointment) if you have questions! Appendix C at the end of your book reviews some useful math/calc facts, including some common series. You can use these facts without deriving them from scratch.

Skim 3.5.1 and 3.5.2. Time permitting, I'll cover the proof connecting the binomial and Poisson distributions from 3.5.2 class (it's fun!) but it isn't essential course material.

4 More with discrete random variables

Chapters 4 and 5 of Wagaman (2021) cover more properties and types of discrete random variables.

Make sure to review Chapter 3 before these chapters and review our week 4 schedule.

Day 10

- What to read: Read sections 4.1, 4.2, 4.4, 4.5 but review the comments below on what to skim/skip.
- Learning objectives: These sections will help you
  - conceptualize and compute the expected value of a discrete random variable and a function of a discrete random variable
  - conceptualize and compute the expected value of a linear function of **two or more** discrete random variable
  - conceptualize and compute the expected value of a product of **two independent** discrete random variable
- After reading, take reading quiz 8

Section 4.1 and 4.2

Key terminology to know:

- ☐ Expectation, or expected value, of a random variable, denoted $E(X)$ where X is the random variable
- ☐ A function of a random variable, often denoted $g(X)$, and its expectation
- ☐ A *linear* function of a random variable, denoted $aX + b$, and its expectation

Comments:

- We will skip 4.3 for now and we'll formally cover joint pmf in a couple weeks.
- Examples 4.3, 4.4, 4.8 use Series facts shown in Appendix C. Be comfortable using these facts to compute expectations.
- The expectation of a linear function of a random variable is an *extremely* useful fact so make sure you spend some time absorbing it. For example, suppose X counts the number of red lights in 5 days (ie it is Binomial). We might care more about $X/5$, which is the *proportion* of read lights in a week. If we know $E(X)$, the average *number* of red lights in a week, then we can easily compute $E(X/5)$ using this rule.
- The R code shows simulation examples that can be used to approximate an expected value. We won't be doing simulations this term but feel free to run the code if you want to learn how to do this for selected examples.

Section 4.4 and 4.5

Key terminology to know:

- ☐ Expectation of the product of two independent random variables

☐ Expectation of a linear combination of two or more random variables

Comments:

- For section 4.4, focus on the expectation rules for indep. RV in Equations 4.4 and 4.5. We will formally define and work with joint pmf later this term.
- Skim Section 4.4.1. We will prove that sums of independent Poisson RVs is Poisson using a different method.
- For section 4.5, focus on the linearity result at the start of the section but don't worry about the proof (for now)
- The expectation of a linear function of *two or more* random variable is an *extremely* useful fact so make sure you spend some time absorbing it.

Day 11

- What to read: Read sections 4.6, 5.2
- Learning objectives: These sections will help you
  - conceptualize and compute the variance and standard deviation of a discrete random variable
  - compute the variance and standard deviation of a linear function of a discrete random variable
  - compute the variance and standard deviation of a linear function of two independent discrete random variables
  - understand how moments of a random variable can be computed via the moment generating function
  - understand how to use moment generating functions to determine the distribution of a function of one or more random variables
- After reading, take reading quiz 9

Section 4.6

Key terminology to know:

☐ variance
☐ standard deviation

Comments:

- Equation 4.10 will be the typical way we compute variance/standard deviation since it is less mathematically intensive to compute compared to the expected value used to define it.
- Example 4.24 use Series facts shown in Appendix C.

Section 5.2

Key terminology to know:

☐ the kth moments of a random variable
☐ moment generating function (mgf)

Comments:

- We jump ahead a bit to cover moment generating functions (mgf) which can be used to find expectations of the form $E(X^k)$ (the kth moment).
- An *extremely* useful fact about mgf's is that they uniquely identify a random variable, eg if two random variables have the same mgf then they have the same distribution. This combined with the properties described on pages 195-6 make mgf's a very useful tool for determining the distribution of linear combinations of one or more random variables.

Day 12

- What to read: Read sections 5.1, 5.3, 5.4
- Learning objectives: These sections will help you
  - get familiar with a few more common discrete RV types (geometric, negative binomial, and hypergeometric)
- After reading, take reading quiz 10

Sections 5.1, 5.3, 5.4

Key terminology to know:

- ☐ geometric RV (know shorthand notation, pmf and support set)
- ☐ memorylessness of the geometric distribution
- ☐ negative binomial RV (know shorthand notation, pmf and support set)
- ☐ hypergeometric RV (know shorthand notation, pmf and support set)

Comments:

- skim 5.1.2
- Take care to understand what the geometric and negative binomial random variables count. While the Binomial counts the number of "successes" in a fixed number of trials, the geometric/negative binomail fix the number of successes while the number of trials that it takes to get those successes is random.
- Some probability books or online sources define geometric and negative binomial random variables as the number *failures* until some fixed number of successes. Eg. if X is the number of trials until the 1st success, then $Y = X - 1$ is the number of failures until the 1st success. We will stick with the number of *trials* definition in this course.

- A hypergeometric random variable is similar in spirit as the binomial, but the hypergeometric counts the number of successes in a fixed sample size taken *without replacement* from a group of unique individuals. A binomial random variable is similar but it assumes with replacement, or independent sampling draw-to-draw.

5 Introduction to continuous random variables

Chapter 6 of Wagaman (2021) introduces us to continuous random variables.

Make sure to review our week 5 schedule.

Day 13

- What to read: Read intro to ch. 6, sections 6.1, 6.2, 6.4, 7.1.3
- Learning objectives: These sections will help you
  - understand what a continuous random variable is
  - use a probability mass function (pdf) to compute probabilities for events involving a continuous RV
  - use a cumulative distribution function (cdf) to compute probabilities for events involving a continuous RV
  - understand how to get a pdf from a cdf and vice versa
  - compute quantiles for a continuous distribution
  - get familiar with a (continuous) uniform RV
- After reading, take reading quiz 11

Section 6.1

Key terminology to know:

☐ continuous random variable
☐ probability density function (pdf) of a continuous RV

Comments:

- Pay particular attention to the comment box on page 230. I am very particular when asking for pdf (or cdf) functions, especially on exams. If you are asked to define a pdf/cdf *over the entire real ine*, then you must get both the support S and "0 otherwise" parts correct.
- Yes, you will be integrating functions in this class! While I may ask you do to more involved calculations, eg integration by parts, on homework, I don't ask these longer calculation-focused questions on exams. Often (but not always) on exams, I'll ask you to appropriately "set-up" an integral but not actually complete the calculation.
- You can check integration calculations using technology, but you can't let technology do all the work. See the AI/tech homework policy on the syllabus.

Section 6.2 and 7.1.3

Key terminology to know:

☐ cumulative distribution function (cdf) of a continuous RV
☐ quantile of a continuous RV

Comments:

- Take some time to make sure you see how pdf's and cdf's are related for continuous random variables.

- Take a look at the cdf example for a discrete RV, but I find cdf most useful for continuous RV. It is essential that you can use/derive cdf for continuous RV, but not as important for discrete (in my class).
- In 7.1.3, I just want you to review the definition of a quantile and how it relates to a cdf (6.2). You can skim the examples in 7.1.3, but we will formally introduce the normal distribution on another day.
- I've seen quantile described either as a percentage or a proportion/probability. E.g. the 50th quantile or the 0.5 quantile. Your book defines it as the former, "50th", but be comfortable with either description.

Section 6.4

Key terminology to know:

☐ uniform RV (know shorthand notation, basic properties)

Comments:

- You can just read up to Expectation and Variance on page 240. Review these after day 14's reading.
- Be comfortable computing probabilities for uniform RV using both a calculus approach and geometric approach. There are pros/cons to each, depending on the problem.

Day 14

- What to read: Read sections 6.3, 6.5, 7.1, 7.2, 7.4
- Learning objectives: These sections will help you
  - compute the expected value, variance and sd of a continuous RV

- review important properties of expected values and variance/sd
- get familiar with some common continous RV types (exponential, normal, gamma, beta)

- After reading, take reading quiz 12

Section 6.3

Key terminology to know:

- ☐ expected value of a continuous RV
- ☐ variance/sd of a continuous RV

Comments:

- We'll cover 6.6-6.8 later this term

Section 6.5 and 7.2

Key terminology to know:

- ☐ exponential RV (know shorthand notation, basic properties)
- ☐ gamma RV (know shorthand notation, basic properties)
- ☐ gamma function Γ

Comments:

- Exponential and gamma are common models for "waiting times", eg time until something happens.
- Make sure you recognize the exponential as a special case of the gamma
- We will cover the connection between these distributions and the Poisson distribution on another day

Section 7.1

Key terminology to know:

- ☐ normal RV (know shorthand notation, basic properties)

Comments:

- Skip 7.1.2 (pages 278-282)

Section 7.4

Key terminology to know:

- ☐ beta RV (know shorthand notation, basic properties)

Comments:

- The beta model is very useful for random variables that have a support between 0 and 1.

Day 15

- What to read: Read sections 8.1
- Learning objectives: This section will help you
  - use cdf's to derive the distribution (cdf, then pdf) of a function of a continuous RV
- No reading quiz on section 8.1

Section 8.1

Key terminology to know:

☐ cdf and pdf (make sure these ideas are solid)

Comments:

- Skim 8.1.1 and skip 8.1.2
- This section outlines how to derive the pdf of a function of a RV using the cdf method.
- This method always “works”, though often the mgf method is simpler *if* the function is linear and the pdf distribution is a common one (eg normal, exponential, etc).

6 Introduction to continuous random variables

Chapter 6 of Wagaman (2021) introduces us to continuous random variables.

Make sure to review our week 5 schedule.

Day 13

- What to read: Read section 7.3 through page 299.
- Learning objectives: These sections will help you
  - understand what a Poisson process is
  - see how "interarrival" or waiting times between Poisson events can be modeled by exponential or gamma distributed random variables
- After reading, take reading quiz 13

Section 7.3

Key terminology to know:

☐ Poisson process with $N_t \sim Pois(\lambda t)$
☐ Interarrival times $E_k \sim Exp(\lambda)$

☐ Time until the *nth* event $S_n \sim Gamma(n, \lambda)$

Comments:

- You can skim/skip the proof on page 296 (up to example 7.15). This proof of the Poisson process model relies on joint pdf which we haven't covered yet.
- There is a **typo** on page 298: The *Stationary increments* property should state that "For all $0 < s, t$, ..." instead of $0 < s < t$

References

Wagaman, & Dobrow, A. S. 2021. *Probability: With Applications and r*. 2nd ed. Wiley.