MATH/STAT 455: Mathematical Statistics

Taylor Okonek

2023-12-07

Table of contents

Welcome to Mathematical Statistics!

This book contains the course notes for *MATH/STAT 455: Mathematical Statistics* at Macalester College, as taught by Prof. Taylor Okonek. These notes draw from course notes created by Prof. Kelsey Grinde, and heavily from the course textbook, *An Introduction to Mathematical Statistics and Its Applications* by Richard Larsen and Morris Marx (6th Edition). Each chapter will contain (at a minimum):

1. Learning Objectives

2. Reading Guide

3. Definitions

4. Theorems

5. Worked Examples

I will be editing and adding to these notes throughout Spring 2024, so please check consistently for updates!

If you find any typos or have other questions, please email tokonek@macalester.edu.

1 Probability: A Brief Review

MATH/STAT 455 builds directly on topics covered in *MATH/STAT 354: Probability.* You're not expected to perfectly remember everything from *Probability*, but you will need to have sufficient facility with the following topics covered in this review Chapter in order to grasp the majority of concepts covered in *MATH/STAT 455*.

1.1 Learning Objectives

By the end of this chapter, you should be able to...

- Distinguish between important probability models (e.g., Normal, Binomial)

- Derive the expectation and variance of a single random variable or a sum of random variables

- Define the moment generating function and use it to find moments or identify pdfs

1.2 Reading Guide

Associated Readings: Chapters 2-4 (pages 15-277)

1.2.1 Reading Questions

1. Which probability distributions are appropriate for *quantitative* (continuous) random variables?

2. Which probability distributions are appropriate for *categorical* random variables?

3. *Independently and Identically Distributed (iid)* random variables are an incredibly important assumption involved in many statistical methods. Why do you think it might be important/useful for random variables to have this property?

1.3 Definitions

You are expected to know the following definitions:

Random Variable

A random variable is a function that takes inputs from a sample space of all possible outcomes, and outputs real values or probabilities. As an example, consider a coin flip. The sample space of all possible outcomes consists of "heads" and "tails", and each outcome is associated with a probability (50% each, for a fair coin). For our purposes, you should know that random variables have probability density (or mass) functions, and are either discrete or continuous based on the number of possible outcomes a random variable may take. Random variables are often denoted with capital Roman letters, like X, Y, Z, etc.

Probability density function (discrete, continuous)

- Note: I don't care if you call a pmf a pdf... I will probably do this continuously throughout the semester. We don't need to be picky about this in *MATH/STAT 455*.

There are many different accepted ways to write the notation for a pdf of a random variables. Any of the following are perfectly appropriate for this class: $f(x)$, $\pi(x)$, $p(x)$, $f_X(x)$. I typically use either π or p, but might mix it up occasionally.

Key things I want you to know about probability density functions:

- $\pi(x) \geq 0$, everywhere. This should make sense (hopefully) because probabilities cannot be negative!

- $\int_{-\infty}^{\infty} \pi(x) = 1$. This should also (hopefully) makes sense. Probabilities can't be *greater* than one, and the probability of event occurring *at all (ever)* should be equal to one, if the event x is a random variable.

Cumulative distribution function (discrete, continuous)

Cumulative distribution functions we'll typically write as $F_X(x)$. or $F(x)$, for short. It is important to know that

$$F_X(x) = \Pr(X \leq x),$$

or in words, "the cumulative distribution function is the probability that a random variable lies before x." If you write $\Pr(X < x)$ instead of $\leq$, you're fine. The probability that a random variable is exactly one number (for an RV with a continuous pdf) is zero anyway, so these are the same thing. Key things I want you to know about cumulative distribution functions:

- $F(x)$ is non-decreasing. This is in part where the "cumulative" piece comes in to play. Recall that probabilities are basically integrals or sums. If we're integrating over something positive, and our upper bound for our integral *increases*, the area under the curve (cumulative probability) will increase as well.

- $0 \leq F(x) \leq 1$ (since probabilities have to be between zero and one!)

- $\Pr(a < X \leq b) = F(a) - F(b)$ (because algebra)

Joint probability density function

A joint probability density function is a probability distribution defined for more than one random variable at a time. For two random variables, X and Z, we could write their joint density function as $f_{X,Z}(x, z)$, or $f(x, z)$ for short. The joint density function encodes all sorts of fun information, including *marginal* distributions for X and Z, and conditional distributions (see next **bold** definition). We can think of the joint pdf as listing all possible pairs of outputs from the density function $f(x, z)$, for varying values of x and z. Key things I want you to know about joint pdfs:

- How to get a marginal pdf from a joint pdf:

  Suppose I want to know $f_X(x)$, and I know $f_{X,Z}(x, z)$. Then I can integrate or "average over" Z to get

$$f_X(x) = \int f_{X,Z}(x, z) dz$$

- The relationship between conditional pdfs, marginal pdfs, joint pdfs, and Bayes' theorem/rule

- How to obtain a joint pdf for *independent* random variables: just multiply their marginal pdfs together! This is how we will (typically) think about likelihoods!

- How to obtain a marginal pdf from a joint pdf when random variables are independent *without integrating* (think, "separability")

Conditional probability density function

A conditional pdf denotes the probability distribution for a (set of) random variable(s), *given that* the value for another (set of) random variable(s) is known. For two random variables, X and Z, we could write the conditional distribution of X "given" Z as $f_{X|Z}(x \mid z)$, where the "conditioning" is denoted by a vertical bar (in LaTeX, this is typeset using "\mid"). Key things I want you to know about conditional pdfs:

- The relationship between conditional pdfs, marginal pdfs, joint pdfs, and Bayes' theorem/rule

- How to obtain a conditional pdf from a joint pdf (again, think Bayes' rule)

- Relationship between conditional pdfs and independence (see next **bold** definition)

Independence

Two random variables X and Z are *independent* if and only if:

- $f_{X,Z}(x, z) = f_X(x)f_Z(z)$ (their joint pdf is "separable")

- $f_{X|Z}(x \mid z) = f_X(x)$ (the pdf for X does not depend on Z in any way)

  Note that the "opposite" is also true: $f_{Z|X}(z \mid x) = f_Z(z)$

In notation, we denote that two variables are independent as $X \perp\!\!\!\perp Z$, or $X \perp Z$. In LaTeX, the *latter* is typeset as "\perp", and the former is typeset as "\perp\!\!\!\perp". As a matter of personal preference, I (Taylor) prefer $\perp\!\!\!\perp$, but I don't like typing it out every time. Consider using the "\newcommand" functionality in LaTeX to create a shorthand for this for your documents!

Expected Value / Expectation

The expectation (or expected value) of a random variable is defined as:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Expected value is a weighted average, where the average is over all possible values a random variable can take, weighted by the probability that those values occur. Key things I want you to know about expectation:

- The relationship between expectation, variance, and moments (specifically, that $E[X]$ is the 1st moment!)

- The "law of the unconscious statistician" (see the Theorems section of this chapter)

- Expectation of linear transformations of random variables (see **Theorems** section of this chapter)

Variance

The variance of a random variable is defined as:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

In words, we can read this as "the expected value of the squared deviation from the mean" of a random variable X. Key things I want you to know about variance:

- The relationship between expectation, variance, and moments (hopefully clear, given the formula for variance)

- The relationship between variance and standard deviation: $Var(X) = sd(X)^2$

- The relationship between variance and covariance: $Var(X) = Cov(X, X)$

- $Var(X) \geq 0$. This should make sense, given that we're taking the expectation of something "squared" in order to calculate it!

- $Var(c) = 0$ for any constant, c.

- Variance of linear transformations of random variables (see **Theorems** section of this chapter)

r^{th} moment

The rth moment of a probability distribution is given by $E[X^r]$. For example, when $r = 1$, the rth moment is just the expectation of the random variable X. Key things I want you to know about moments:

- The relationship between moments, expectation, and variance

    – For example, if you know the first and second moments of a distribution, you should be able to calculate the variance of a random variable with that distribution!

- The relationship between moments and *moment generating functions* (see **Theorems** section of this chapter)

Covariance

The covariance of two random variables is a measure of their *joint* variability. We denote the covariance of two random variables X and Z as $Cov(X, Z)$, and

$$Cov(X, Z) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

Some things I want you to know about covariance:

- $Cov(X, X) = Var(X)$

- $Cov(X, Y) = Cov(Y, X)$ (order doesn't matter)

Moment Generating Function (MGF)

The moment generating function of a random variable X is defined as

$$M_X(t) = E[e^{tX}]$$

A few things to note:

- $M_X(0) = 1$, always.

- If two random variables have the same MGF, they have the same probability distribution!

- MGFs are sometimes useful for showing how different random variables are related to each other

You are also expected to know the probability distributions contained in Table 1, below. Note that you *do not* need to memorize the pdfs for these distributions, but you *should* be familiar with what types of random variables (continuous/quantitative, categorical, integer-valued, etc.) may take on different distributions. The more familiar you are with the forms of the pdfs, the easier/faster it will be to work through problem sets and quizzes.

Table 1.1: *Table 1.* Table of main probability distributions we will work with for *MATH/STAT 455.*

Distribution	PDF/PMF	Parameters
Uniform	$\pi(x) = \frac{1}{\beta - \alpha}$	$\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$
Normal	$\pi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$	$\mu \in \mathbb{R}$, $\sigma > 0$
Multivariate Normal	$\pi(\mathbf{x}) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)))$	$\mu \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$, positive semi-definite (in practice, almost always positive definite)
Gamma	$\pi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	α (shape), β (rate) > 0
Chi-square	$\pi(x) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2})$	$\nu > 0$
Exponential	$\pi(x) = \beta e^{-\beta x}$	$\beta > 0$
Student-t	$\pi(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}}(1 + \frac{x^2}{\nu})^{-(\nu+1)/2}$	$\nu > 0$
Beta	$\pi(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha, \beta > 0$
Poisson	$\pi(x) = \frac{\lambda^k e^{-\lambda}}{k!}$	$\lambda > 0$
Binomial	$\pi(x) = \binom{n}{x} p^x (1-p)^{n-x}$	$p \in [0,1], n = \{0,1,2,...\}$
Multinomial	$\pi(\mathbf{x}) = \frac{n!}{x_1!...x_k!} p_1^{x_1}...p_k^{x_k}$	$p_i > 0$, $p_1 + \cdots + p_k = 1$, $n = \{0,1,2,...\}$
Negative Binomial	$\pi(x) = \binom{k+r-1}{k}(1-p)^k p^r$	$r > 0$, $p \in [0,1]$

1.4 Theorems

- Law of Total Probability

$$P(A) = \sum_n P(A \cap B_n),$$

or

$$P(A) = \sum_n P(A \mid B_n)P(B_n)$$

- Bayes' Theorem

$$\pi(A \mid B) = \frac{\pi(B \mid A)\pi(A)}{\pi(B)}$$

- Relationship between pdf and cdf

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$$

$$\frac{\partial}{\partial y}F_Y(y) = f_Y(y)$$

- Expectation of random variables

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

  – "Law of the Unconscious Statistician"

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- Expectation and variance of linear transformations of random variables

$$E[cX + b] = cE[X] + b$$

$$Var[cX + b] = c^2 Var[X]$$

- Relationship between mean and variance

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Also, recall that $Cov[X, X] = Var[X]$.