Vectors, inequalities, proofs

Patrick Breheny

August 28, 2024

Introduction

- Before we get to likelihood theory, we are going to spend the first part of this course reviewing/extending/deepening our knowledge of mathematical and statistical tools
- In particular, lower-level analysis and mathematical statistics courses often focus on single-variable results
- In practice, however, statistics is almost always a multivariate pursuit
- Thus, one of the things we will focus on in this review is covering results you may have seen for single variables in terms of vectors

Asymptotic theory

- A large amount (but not all) of statistical theory is based on asymptotic, or large sample, arguments
- Exact theoretical results are often very complicated and difficult to obtain, but we can typically simplify the problem greatly by considering what happens as $n \to \infty$
- A core idea here from analysis is that of a convergent sequence: x_n converges to x if, as n gets larger, x_n gets closer and closer to x
- We'll discuss this more next week, but first, we need to take a step back and define what it means for x_n to be "close" to x

Norms: Introduction

- Throughout this course, we need to be able to measure the distance between two vectors, or equivalently, the size of a vector; such a measurement is called a *norm*
- This is straightforward for scalars: the distance from a to b is $|a - b|$
- Vectors are more complicated; as we will see, there are many ways of measuring the size of a vector
- In order to be a meaningful measure of size, however, there are certain conditions any norm must satisfy

Norm: Definition

- **Definition:** A *norm* is a function $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,
  - $\|\mathbf{x}\| \geq 0$, with $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$ (positivity)
  - $\|a\mathbf{x}\| = |a| \, \|\mathbf{x}\|$ for any $a \in \mathbb{R}$ (homogeneity)
  - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)
- The triangle inequality is also sometimes expressed as

$$\|\mathbf{x} - \mathbf{z}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\|,$$

or

$$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}),$$

where $d(\mathbf{x}, \mathbf{y})$ quantifies the distance between $\mathbf{x}$ and $\mathbf{y}$

Reverse triangle inequality

- A related inequality:
- **Theorem (reverse triangle inequality):** For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|$$

- **Corollary:** For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\begin{aligned}
\|\mathbf{x}\| - \|\mathbf{y}\| &\leq \|\mathbf{x} + \mathbf{y}\| \\
\|\mathbf{y}\| - \|\mathbf{x}\| &\leq \|\mathbf{x} + \mathbf{y}\| \\
\|\mathbf{y}\| - \|\mathbf{x}\| &\leq \|\mathbf{x} - \mathbf{y}\|
\end{aligned}$$

Examples of norms

- By far the most common norm is the Euclidean (L_2) norm:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$$

- However, there are many other norms; for example, the Manhattan (L_1) norm:

$$\|\mathbf{x}\|_1 = \sum_i |x_i|$$

- Both Euclidean and Manhattan norms are members of the L_p family of norms: for $p \geq 1$,

$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

Examples of norms (cont'd)

- Another norm worth knowing about is the L_∞ norm:

$$\|\mathbf{x}\|_\infty = \max_i |x_i|,$$

  which is the limit of the family of L_p norms as $p \to \infty$
- One last "norm" worth mentioning is the L_0 norm:

$$\|\mathbf{x}\|_0 = \sum_i 1\{x_i \neq 0\};$$

  be careful, however: this is not a proper norm! (why not?)

Inner products

- The Euclidean norm can also be thought of in terms of something called an *inner product*:

$$\mathbf{x}^\top \mathbf{y} = \sum_i x_i y_i$$

- Inner products come up all the time in statistics and mathematics, and can also be written using angle bracket notation: $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$
- Two critical things to remember:
  - The inner product $\mathbf{x}^\top \mathbf{y}$ takes two vectors and returns a scalar
  - Writing $\mathbf{x}^2$ is meaningless – never do this – because there are two ways to multiply a vector $\mathbf{x}$ with itself

Outer products

- When we change which vector is transposed, instead of getting something simpler (a scalar), we get something more complicated (a matrix)
- The operation $\mathbf{x}\mathbf{y}^\top$ results in an $n \times n$ matrix where the element in row i, column j is $x_i y_j$
- This is known as *outer product*, and can also be written $\mathbf{x} \otimes \mathbf{y}$
- We will see inner and outer products all the time in this course, so this needs to be something you understand instantly in order to read equations and formulas that will appear (we will talk more about matrix algebra in a future lecture)

Matrix norms

- There are also matrix norms, although we will not work with these as often
- In addition to the three requirements listed earlier, matrix norms must also satisfy a requirement of *submultiplicativity*:

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|;$$

  unlike the other requirements, this only applies to $n \times n$ matrices
- The simplest matrix norm is the *Frobenius* norm

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

Spectral norm

- Another common matrix norm is the *spectral norm*:

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}},$$

  where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{A}^\top\mathbf{A}$
- There are many other matrix norms

Cauchy-Schwarz

- There are several important inequalities involving norms that you should be aware of; the most important is the Cauchy-Schwarz inequality, arguably the most useful inequality in all of mathematics
- **Theorem (Cauchy-Schwarz):** For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

  $$\mathbf{x}^\top \mathbf{y} \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2,$$

  where equality holds only if $\mathbf{x} = a\mathbf{y}$ for some scalar a
- Note: the above is *the* Cauchy-Schwarz inequality, but in statistics, its probabilistic version goes by the same name:

  $$\mathbb{E}\,|XY| \le \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

  for random variables X and Y, with equality iff $X = aY$

Hölder's inequality

- The Cauchy-Schwarz inequality is actually a special case of Hölder's inequality:
- **Theorem (Hölder):** For $1/p + 1/q = 1$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\mathbf{x}^\top \mathbf{y} \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q,$$

  again with exact equality iff $\mathbf{x} = a\mathbf{y}$ for some scalar a (unless p or q is exactly 1)
- Probabilistic analogue:

$$\mathbb{E}\,|XY| \leq \sqrt[p]{\mathbb{E}\,|X|^p}\sqrt[q]{\mathbb{E}\,|Y|^q}$$

Jensen's inequality

- Another extremely important inequality is Jensen's inequality; surely you've seen it before, but perhaps not in vector form
- **Theorem (Jensen):** For $\mathbf{a}, \mathbf{x} \in \mathbb{R}^d$ with $a_i > 0$ for all i, if g is a convex function, then

$$g\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right) \leq \frac{\sum_i a_i g(x_i)}{\sum_i a_i}$$

- Probabilistic analog:

$$g(\mathbb{E}X) \leq \mathbb{E}g(X)$$

- The inequalities are reversed if g is concave

Relationships between norms

- Getting back to the different norms, there are many important relationships between norms that are often useful to know
- **Theorem:** For all $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{d}\|\mathbf{x}\|_2$$

- Obvious, but useful:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq d\|\mathbf{x}\|_\infty$$
$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{d}\|\mathbf{x}\|_\infty$$

Equivalence of norms

- The relationships on the previous slide suggest the following statement, which is in fact always true: for any two norms a and b, there exist constants c_1 and c_2 such that

$$\|\mathbf{x}\|_a \leq c_1 \|\mathbf{x}\|_b \leq c_2 \|\mathbf{x}\|_a$$

- This result is known as the *equivalence of norms* and means that we can often generalize results for any one norm to all norms
- For example, we will often encounter results that look like:

$$A = B + \|\mathbf{r}\|$$

and show that $\|\mathbf{r}\| \to 0$, so $A \to B$

Equivalence of norms (cont'd)

- By the equivalence of norms, if, say, $\|\mathbf{r}\|_1 \to 0$, then $\|\mathbf{r}\|_2 \to 0$ and so on for all norms (except not the L_0 "norm"!)
- In this course, we will almost always be working with the Euclidean norm, so much so that I will often write $\|\mathbf{x}\|$ to mean the Euclidean norm and not even bother with the subscript
- That said, it is important to note that with these relationships, we can always derive corollaries that extend results to other norms

Equivalence of matrix norms

- Like vector norms, matrix norms are also equivalent
- For example,

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{r}\|\mathbf{A}\|_2,$$

where r is the rank of $\mathbf{A}$

Inclusion

- Many of the inequalities we will work with in this class involve random variables, so it is worth covering a simple but important point that doesn't seem to have a name (that I know of) but I will call the principle of inclusion
- **Theorem (Inclusion):** Suppose $a < b$ and Y is a random variable. Then
$$\mathbb{P}\{Y < a\} \leq \mathbb{P}\{Y < b\}$$
- **Corollary:** Suppose $a < b$ and Y is a random variable. Then
$$\mathbb{P}\{b < Y\} \leq \mathbb{P}\{a < Y\}$$

Intuition

- In words, if the "big side" gets bigger or the "small side" gets smaller, the probability goes up
- For example, combining this with the triangle inequality:

$$\mathbb{P}\{X < |a + b|\} \leq \mathbb{P}\{X < |a| + |b|\}$$

  (the "big side" got bigger, so the probability went up)
- This is not hard to understand, but also a common source of mistakes, because it's easy to wind up with the inequality going the wrong way:

$$\mathbb{P}\{|a + b| < X\} \leq \mathbb{P}\{|a| + |b| < X\}$$

Introduction

- We will spend a lot of time in this course (both in-class and on homework/exams) proving things
- It is useful to discuss best practices for structuring proofs up front, especially if you have never taken a course on proofs in the past
- We will start with the easiest (most straightforward) type of proof, the proof by calculation

Introduction

- Let's start with an extremely simple proof: Suppose that $2x + 3 = x$; prove that $x = -3$
- One way to approach this problem would be as follows:

$$\begin{aligned} 2x + 3 &= x \\ 2x + 3 - x &= x - x \\ x + 3 &= 0 \\ x &= -3 \end{aligned}$$

- This approach may be familiar, and is adequate for a simple problem like this, but isn't ideal – the approach is actually quite limiting and does not extend well to more complicated types of proofs

Linear/chain proofs

- Instead, I would encourage you to get in the habit of writing proofs that form a linear chain of steps from beginning to end:

$$\begin{aligned} x &= x + (x - x) + (3 - 3) \\ &= (2x + 3) - x - 3 \\ &= x - x - 3 \\ &= -3 \end{aligned}$$

- This may be less familiar, but
  - The steps here are much more representative of what we will be doing in future lectures
  - It extends to more complex proof structures
  - It is easier to check the logic and ensure that the result holds

Proofs involving inequalities

- For example, suppose we know that $x + 3 \geq 2$ and $y + 2x < 3$; prove that $y < 5$
- Forming a linear chain:

$$\begin{aligned} y &= y + 2x - 2x \\ &< 3 - 2x \\ &= 9 - 2(x+3) \\ &\leq 9 - 2 \cdot 2 \\ &= 5 \end{aligned}$$

- Is it possible to prove this in other ways? Of course, but here we have a clear chain that allows us to immediately see how the inequalities relate to one another

More complex structures

- In this course, proofs will obviously be more complicated than this:
    - Multiple steps (as opposed to a single chain of equations/inequalities)
    - More abstract results (there exists a number such that… or this is true for all such… as opposed to a result about a specific instance)
- However, equation/inequality chains are often central components of these more complex proofs

Sequential order is absolutely critical

Furthermore, all proofs, no matter how complicated, must follow a *sequential* chain of logic in which each statement is unquestionably true based on what has come before

- **Never** refer to a quantity or a variable that has not yet been defined
- **Never** write something down that may or may not be true

Contradiction

- Arguably, the one exception to this rule is proof by contradiction: if we suppose that something is true and show that it leads to an impossibility, therefore the original premise must be false
- For example, show that there is no integer k such that $k^2 = 2$
- It is critical, however, that the *scope* is clearly defined here: within the contradiction block, the supposition is *unquestionably true*, and outside the contradiction block, the supposition is *unquestionably false* – it is never "maybe true?"

Convergence

- To get a sense of what these more complex structures look like, let's consider the definition of convergence, which we will discuss further next week
- **Definition:** A sequence of scalar values x_n is said to converge to x, which we denote $x_n \to x$, if for every $\epsilon > 0$, there is a number N such that $n > N$ implies that $|x_n - x| < \epsilon$
- Pay very close attention to the wording here, because we are *not* saying that there is a single N that always works

Convergence (cont'd)

- Instead, we are saying that if you:
  (1) pick an ϵ, then
  (2) you can always find an N that works, where N is allowed to depend on ϵ (and typically, must)
- For example, show that $1/n$ converges to 0
- These ideas of *for every* ($\forall$) and *there exists* ($\exists$) are fundamental to mathematical analysis and statistical theory, so make sure you know exactly what they mean and how they are different from each other