Analysis review, Part 2

Patrick Breheny

August 31

Introduction

Our analysis review continues today with three more topics:

- O, o notation: An extremely useful tool
- Taylor series expansions: Probably the single most useful mathematical tool in all of statistics
- Uniform convergence: An often poorly understood topic that not everyone is familiar with

O-notation: Introduction

- When investigating the asymptotic behavior of functions, it is often convenient to replace unwieldy expressions with compact notation
- For example, suppose we have a term like

$$\frac{\exp\{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2\}}{2\sqrt{n}\theta\int_0^\infty g(s)ds};$$

  if we're investigating what this looks like asymptotically (with respect to n), maybe we can just replace this with $c/\sqrt{n}$, where c is a constant
- If the term ends up going away as $n \to \infty$, why bother writing down all those extra terms over and over again?

O-notation

- A very useful companion of o-notation is O-notation, which denotes whether or not quantities are bounded as $n \to \infty$
- **Definition:** A sequence of numbers X_n is said to be $O(1)$ if there exist M and n_0 such that

  $$|X_n| < M$$

  for all $n > n_0$. Likewise, X_n is said to be $O(r_n)$ if there exist M and n_0 such that for all $n > n_0$,

  $$\left|\frac{X_n}{r_n}\right| < M.$$

- Note that $X_n = O(1)$ does not necessarily mean that X_n is bounded, just that it is eventually bounded

o-notation

- Its companion is o-notation
- **Definition:** A sequence of numbers X_n is said to be $o(1)$ if it converges to zero. Likewise, X_n is said to be $o(r_n)$ if

$$\frac{X_n}{r_n} \to 0$$

as $n \to \infty$.

- For example, the expression on slide 3 is $o(1)$, and $O(n^{-1/2})$

Algebra of O, o notation

O, o-notation are useful in combination because simple rules govern how they interact with each other

Theorem: For $a \leq b$:

$$
\begin{aligned}
O(1) + O(1) &= O(1) & O\{O(1)\} &= O(1) \\
o(1) + o(1) &= o(1) & o\{O(1)\} &= o(1) \\
o(1) + O(1) &= O(1) & o(r_n) &= r_n o(1) \\
O(1)O(1) &= O(1) & O(r_n) &= r_n O(1) \\
O(1)o(1) &= o(1) & O(n^a) + O(n^b) &= O(n^b) \\
\{1 + o(1)\}^{-1} &= O(1) & o(n^a) + o(n^b) &= o(n^b)
\end{aligned}
$$

Remarks

- O, o "equations" are meant to be read left-to-right; for example, $O(\sqrt{n}) = O(n)$ is a valid statement, but $O(n) = O(\sqrt{n})$ is not
- **Exercise:** Determine the order of

$$n^{-2}\left\{(-1)^n \sqrt[n]{2} + (1 + \tfrac{1}{n})^n\right\}.$$

- As we will see in a week or two, there are stochastic equivalents of these concepts, involving convergence in probability and being bounded in probability
- As such, we won't do a great deal with O, o-notation right now, but will use the stochastic equivalents extensively

Taylor series: Introduction

- It is difficult to overstate the importance of Taylor series expansions to statistical theory, and for that reason we are now going to cover them fairly extensively
- In particular, Taylor's theorem comes in a number of versions, and it is worth knowing at least two of them, since they both come up in statistics quite often
- Furthermore, students often have not seen the multivariate versions of these expansions

Taylor's theorem

- **Theorem (Taylor):** Suppose n is a positive integer and $f : \mathbb{R} \to \mathbb{R}$ is n times differentiable at a point x_0. Then

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n(x, x_0),$$

where the remainder R_n satisfies

$$R_n(x, x_0) = o(|x - x_0|^n) \text{ as } x \to x_0$$

- You could also say that R_n is $O(|x - x_0|^{n+1})$
- This form of the remainder is sometimes called the *Peano* form

Taylor's theorem: Lagrange form

- **Theorem (Taylor):** If $f^{(n+1)}$ exists on the open interval and $f^{(n)}$ is continuous on the closed interval between x and x_0, then there exists $x^* \in (x, x_0)$:

$$R_n(x, x_0) = \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - x_0)^{(n+1)}.$$

- This is also known as the *mean-value form*, as the mean value theorem is the central idea in proving the result
- Note that we have a simpler expression, but at the cost of stronger assumptions: $f^{(n+1)}$ must exist along the entire interval from x to x_0, not just at the point x_0

Multivariable forms of Taylor's theorem

- We now turn our attentions to the multivariate case
- For the sake of clarity, I'll present the first- and second-order expansions for each of the previous forms, rather than abstract formulae involving $f^{(n)}$
- Lastly, I'll provide a form that goes out to third order, although higher orders are less convenient as they can't be represented compactly using vectors and matrices

Taylor's theorem

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable at a point $\mathbf{x}_0$. Then

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|)$$

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable at a point $\mathbf{x}_0$. Then

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + \\ \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2) \end{aligned}$$

Taylor's theorem: Lagrange form

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\mathbf{x}^*$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}_0)$$

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\mathbf{x}^*$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}_0)$$

- "$\mathbf{x}^*$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$" means that there exists $w \in [0, 1]$ such that $\mathbf{x}^* = w\mathbf{x} + (1 - w)(\mathbf{x}_0)$

Taylor's theorem: Third order

Theorem (Taylor): Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is three times differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\mathbf{x}^*$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{x}_0) &+ \sum_{j=1}^d \frac{\partial f(\mathbf{x}_0)}{\partial x_j}(x_j - x_{0j}) \\ &+ \frac{1}{2}\sum_{j=1}^d \sum_{k=1}^d \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_j \partial x_k}(x_j - x_{0j})(x_k - x_{0k}) \\ &+ \frac{1}{6}\sum_{j=1}^d \sum_{k=1}^d \sum_{\ell=1}^d \frac{\partial^3 f(\mathbf{x}^*)}{\partial x_j \partial x_k \partial x_\ell}(x_j - x_{0j})(x_k - x_{0k})(x_\ell - x_{0\ell}), \end{aligned}$$

where $\partial f(\mathbf{x}_0)/\partial x_j$ is shorthand for $\partial f(\mathbf{x})/\partial x_j$ evaluated at $\mathbf{x}_0$

Statistical convergence: Motivation

- Convergence is a very important concept in theoretical statistics; for example, we often know that

$$f_n(\theta) \to f(\theta);$$

here, I am using $\to$ in an intentionally vague sense – we will talk more about probabilistic convergence in a few weeks

- For example, we might know that

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \theta)^2 - \sigma^2 \to 0$$

- From this, we often want to know: suppose $\hat{\theta} \to \theta$, does

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta})^2 - \sigma^2 \to 0?$$

Framing the issue

- In other words, does $f_n(\hat{\theta}) \to f(\theta)$ as $\hat{\theta} \to \theta$?
- We'll return to the probabilistic question later in the course; for now, let's discuss the problem in deterministic terms
- Suppose we have a sequence of functions $f_1, f_2, \ldots$ such that for all values of x, we have $f_n(x) \to f(x)$
- Our central question is whether the following holds or not:

$$\lim_{x \to x_0} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to x_0} f_n(x)$$

Counterexample

- Unfortunately, the answer is no – in general, this is not true
- For example:

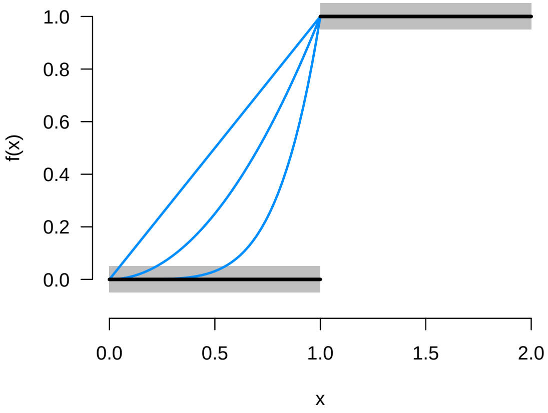
$$f_n(x) = \begin{cases} x^n & x \in [0,1] \\ 1 & x \in (1,\infty) \end{cases}$$

- We have

$$\lim_{x \to 1^-} \lim_{n \to \infty} f_n(x) = 0$$
$$\lim_{n \to \infty} \lim_{x \to 1^-} f_n(x) = 1$$

Illustration

The underlying issue is that f_n doesn't really converge to f in the sense of always lying within $\pm\epsilon$ of it:



Uniform convergence

- The relationship between f_n and f is one of *pointwise convergence*; we need something stronger
- **Definition:** A sequence of function $f_1, f_2, \ldots$, *converges uniformly* on a set E to a function f if for every $\epsilon > 0$ there exists N such that $n > N$ implies

$$|f_n(x) - f(x)| < \epsilon$$

  for all $x \in E$
- **Corollary:** $f_n \to f$ uniformly on E if and only if

$$\sup_{x \in E} |f_n(x) - f(x)| \to 0.$$

Why this is useful

- This is useful because with uniform convergence, we can reach the kind of conclusion we originally sought
- **Theorem:** Suppose $f_n \to f$ uniformly, with f_n continuous for all n. Then $f_n(\mathbf{x}) \to f(\mathbf{x}_0)$ as $\mathbf{x} \to \mathbf{x}_0$.
- Note that this argument does not work without uniform convergence

Extensions

- The theorem on the previous page can actually be made somewhat stronger:
- **Theorem:** Suppose $f_n \to f$ uniformly on E and that $\lim_{x \to x_0} f_n(x)$ exists for all n. Then for any limit point x_0 of E,

$$\lim_{x \to x_0} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to x_0} f_n(x).$$

- **Corollary:** If $\{f_n\}$ is a sequence of continuous functions on E and if $f_n \to f$ uniformly on E, then f is continuous on E.

Related concepts

- There are number of related concepts similar to uniform convergence
- **Definition:** A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *uniformly continuous* if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{y}\| < \delta$, we have $|f(\mathbf{x}) - f(\mathbf{y})| < \epsilon$.
- For example, $f(x) = x^2$ is uniformly continuous over $[0, 1]$ but not over $[0, \infty)$
- **Definition:** A sequence $X_1, X_2, \ldots$ of random variables is said to be *uniformly bounded* if there exists M such that $|X_n| < M$ for all X_n.