Mathematics/Statistics Bootcamp

Part I: Calculus

Brian Cozzi[1] Michael Valancius[1] Graham Tierney[1]
Becky Tang[1]

[1]Department of Statistical Science
Duke University

Graduate Orientation, August 2019

Overview

Limit and Continuity

Limit

Suppose $-\infty < a, L < +\infty$ and $f(x) : X \to Y$ is a real-valued function, then

$$\lim_{x \to a} f(x) = L$$

if for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - L| < \epsilon \text{ whenever } 0 < |x - a| < \delta.$$

(The value of $f(x)$ approaches L when x approaches a.)

Left-hand limit: $\lim_{x \to a^-} f(x) = L$ if for any $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $a - \delta < x < a$.

Right-hand limit: $\lim_{x \to a^+} f(x) = L$ if for any $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $a < x < a + \delta$.

Limit: An Example

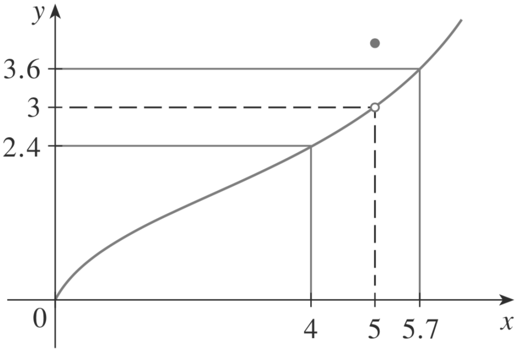


Figure: Plot of $y = f(x)$.

- What is $\lim_{x \to 5^-} f(x)$?
- What is $\lim_{x \to 5^+} f(x)$?
- What is $\lim_{x \to 5} f(x)$?

Continuity

A function f is continuous at a number a if

$$\lim_{x \to a} f(x) = f(a).$$

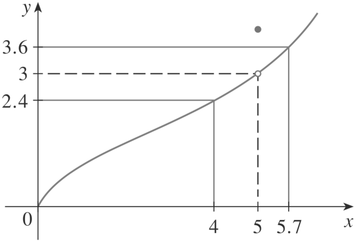
It implies 3 things:

1. $f(a)$ is defined ($a \in X$);
2. $\lim_{x \to a} f(x)$ exists;
3. $\lim_{x \to a} f(x) = f(a)$.

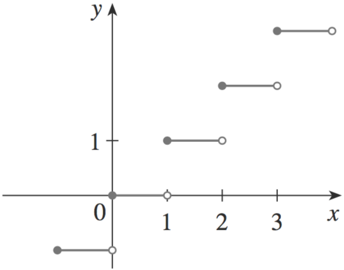
Right continuous: $\lim_{x \to a^-} f(x) = f(a)$.

Left continuous: $\lim_{x \to a^+} f(x) = f(a)$.

Continuity: Examples



This function is discontinuous at $x = 5$.



This function is discontinuous (but right continuous) at any integer x.

Derivative

Definition of Derivative

The derivative of function f at $a \in X$, denoted by $f'(a)$ is

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

if this limit exists ("differentiable").

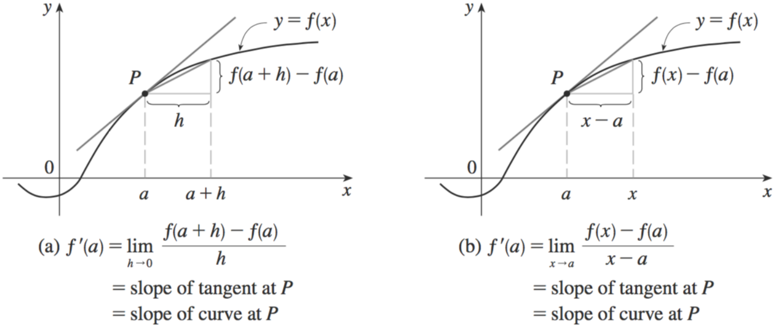


Figure: Geometric interpretations of the derivative.

Differentiation Rules

Derivatives of some common functions:

- $f(x) = const$, then $f'(x) = 0$;
- $f(x) = x^\alpha, \alpha \neq 0$, then $f'(x) = \alpha x^{\alpha-1}$;
- $(e^x)' = e^x$, $(\ln x)' = 1/x$ $(x > 0)$;
- $(\sin x)' = \cos x$, $(\cos x)' = -\sin x$, $(\tan x)' = 1/\cos^2 x$;
- $(\sin^{-1} x)' = 1/\sqrt{1-x^2}$, $(\cos^{-1} x)' = -1/\sqrt{1-x^2}$, $(\tan^{-1} x)' = 1/1+x^2$.

If both $f(x)$ and $g(x)$ are differentiable:

- $(cf(x))' = cf'(x)$, $(f(x)+g(x))' = f'(x)+g'(x)$;
- $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$;
- $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$ (assume $g(x) > 0$);
- The **chain rule**: if $F = f \circ g$, then $F'(x) = f'(g(x))g'(x)$.

Derivative: Exercises

1. Find the derivatives of the following functions
   - $f(x) = xe^x$;
   - $f(x) = 1 - \cos^2 x$;
   - $f(x) = \frac{\ln x}{x}$.

2. $f(x) = \frac{1}{\sqrt{\gamma}} \exp\left(-\frac{(x-\mu)^2}{\gamma}\right)$ where constants $\gamma > 0$ and $\mu \in \mathbb{R}$, and $x \in \mathbb{R}$. Calculate $f'(x)$ and find $x_0 \in \mathbb{R}$ such that the tangent line of $f(x)$ at x_0 is horizontal.

3. Find $\lim_{x \to 0}(1 + x)^{1/x}$.

Solution to Exercise 3

Let $f(x) = \ln x$, then

$$\begin{aligned} f'(1) &= \lim_{x \to 0} \frac{\ln(1+x) - \ln 1}{x} \\ &= \lim_{x \to 0} \frac{1}{x} \ln(1+x) \\ &= \lim_{x \to 0} \ln(1+x)^{1/x}. \end{aligned}$$

Since $f'(1) = 1$, $\lim_{x \to 0}(1+x)^{1/x} = e^1 = e$.

Minimum and Maximum

Theorem (Fermat's Theorem)

If f has a local minimum or maximum at c and $f'(c)$ exists, then $f'(c) = 0$.

Note: the converse is not true.

Theorem (The Second Derivative Test)

If f has second derivative on $(c - \epsilon_0, c + \epsilon_0)$ for a certain $\epsilon_0 > 0$, then

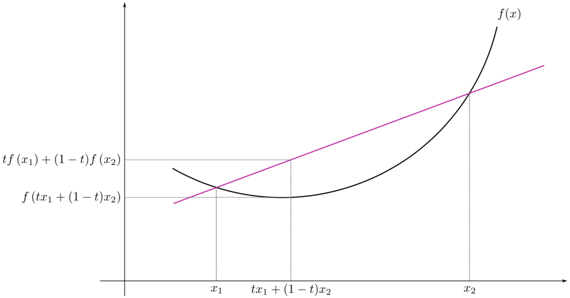
- *if $f'(c) = 0$ and $f''(c) > 0$, f has a local minimum at c;*
- *if $f'(c) = 0$ and $f''(c) < 0$, f has a local maximum at c.*

Convexity

A function defined on a convex set X, $f : X \to \mathbb{R}$ is convex if for any $x, y \in X$ and $t \in [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Visually, a convex function has a "curve up" shape:



Convexity and Derivatives

Suppose $f(x)$ is twice differentiable on interval I, then

- f is convex on I if and only if $f'(x)$ is monotonically non-decreasing on I;
- f is convex on I if and only if $f''(x) \geq 0$ for $x \in I$ (often used to test for convexity).

A nice property of convexity:
Any local minimum of a convex function is also a global minimum; a strictly convex function has at most one global minimum.
(Therefore convexity is much desired in optimization.)

Review Exercises

12. Which following functions are convex?

A $f_1(x) = |x|, x \in [-1, 1]$;

B $f_2(x) = \ln(x^2 + 1), x \in \mathbb{R}$;

C $f_3(x) = e^{-x}, x \in \mathbb{R}$.

2. Let $f(x) = \frac{1}{x}, x > 0$. For every positive integer n, find $f^{(n)}(x)$.

3. $f(x) = 4xe^{-2x}$ with $x \in (0, \infty)$ which is a Gamma(2,2) density. Find the global maximum of $f(x)$.

Taylor Expansion

Taylor Series, by *3Blue1Brown*

Integrals

Properties of Definite Integrals

Let $a \leq d \leq b \in \mathbb{R}$:

- If $c \in \mathbb{R}$ is a constant, then $\int_a^b c dx = c(b-a)$;
- $\int_a^b cf(x)dx = c\int_a^b f(x)dx$;
- $\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$;
- $\int_a^d f(x)dx + \int_d^b f(x)dx = \int_a^b f(x)dx$;
- If $f(x) \geq g(x)$ for $a \leq x \leq b$, then $\int_a^b f(x)dx \geq \int_a^b g(x)dx$;
- If $m \leq f(x) \leq M$ for $a \leq x \leq b$, then $m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$.

The Fundamental Theorem of Calculus

If f is continuous on $[a, b]$, then:

- function $g(x) = \int_a^x f(s)ds, a \leq x \leq b$ is continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) = f(x)$;
- $\int_a^b f(x)dx = F(b) - F(a)$, where F is any anti-derivative of f ($F' = f$).

Useful Rules for Integration

- **Substitution rule**: If $u = g(x)$ is continuously differentiable on $[a, b]$ and f is continuous on the range of u, then

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du.$$

- **Integration by parts**: If functions u and v are both continuously differentiable on $[a, b]$, then

$$\int_a^b u(x)v'(x)dx = [u(x)v(x)]|_a^b - \int_a^b v(x)u'(x)dx.$$

Integration: Exercises

1. Calculate $\int_1^e \frac{\ln x}{x} dx$.

2. Calculate $\int_0^\pi x \cos x dx$.

Improper Integrals

1. **Infinite intervals**: if $\int_a^t f(x)dx$ exists for every $t \geq a$ then

$$\int_a^\infty f(x)dx = \lim_{t\to\infty} \int_a^t f(x)dx$$

   provided that this limit exists (convergent); similarly, one may define $\int_{-\infty}^a f(x)dx$, and if both $\int_a^\infty f(x)dx$ and $\int_{-\infty}^a f(x)dx$ are convergent, then $\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^a f(x)dx + \int_a^\infty f(x)dx$.

2. **Discontinuous integrand**: if f is continuous on $[a, b)$ and is discontinuous at b, then

$$\int_a^b f(x)dx = \lim_{t\to b^-} \int_a^t f(x)dx$$

   if this limit exists; similarly, if f is continuous on $(a, b]$ and is discontinuous at a, $\int_a^b f(x)dx = \lim_{t\to a^+} \int_t^b f(x)dx$.

Improper Integrals: Mini-Exercise

1. For what values of $p \in \mathbb{R}$ is the integral
$$\int_1^\infty \frac{1}{x^p}\,dx$$
convergent?

2. Evaluate $\int_2^5 \frac{1}{\sqrt{x-2}}\,dx$.

Sequences and Series

Basics of Sequences

A **sequence** is a list of numbers written in a definite order. We often denote a sequence $\{a_1, a_2, a_3, \ldots\}$ by $\{a_n\}$ or $\{a_n\}_{n=1}^{\infty}$.

A sequence $\{a_n\}$ has **limit** L (written as $\lim_{n\to\infty} a_n = L$, or $a_n \to L$ as $n \to \infty$) if for every $\epsilon > 0$ there is a corresponding integer N such that $|a_n - L| < \epsilon$ whenever $n > N$.

If for every $n \in \mathbb{N}$, $a_n \leq a_{n+1}$ (increasing) or $a_n \geq a_{n+1}$ (decreasing), then the sequence $\{a_n\}$ is **monotonic**. If there exists a number $M > 0$ such that $|a_n| \leq M$ for every n then the sequence $\{a_n\}$ is **bounded**.

Monotonic Sequence Theorem: Every bounded, monotonic sequence is convergent (has a limit).

Basics of Series

A **series** can be thought of as the infinite sum of a sequence $\{a_n\}$, written as $\sum_{n=1}^{\infty} a_n$ or $\sum a_n$. More formally, it can be defined by taking the limit of partial sums $\{s_n\}$, where $s_n = \sum_{i=1}^{n} a_i$: if $\lim_{n\to\infty} s_n$ exists then the series $\sum a_n$ is convergent, otherwise it is divergent.

If $\sum_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n\to\infty} a_n = 0$.

An important example - **the geometric series**:

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + \cdots$$

is convergent if $|r| < 1$ and its sum is $\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}$. If $|r| \geq 1$, the geometric series is divergent.

Convergence of Series

Commonly used tests for convergence:

1. **The comparison test**: Suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms.
   (i) If $\sum b_n$ is convergent and $a_n \leq b_n$ for all n, then $\sum a_n$ is also convergent;
   (ii) If $\sum b_n$ is divergent and $a_n \geq b_n$ for all n, then $\sum a_n$ is also divergent.
   - Video example: **famous proof that the harmonic series diverges**, by *Khan Academy*.
2. **The integral test** (by *Khan Academy*).

Multivariate Calculus

Partial Derivatives

If u is a function of n variables, $u = f(x_1, x_2, \ldots, x_n)$, its partial derivative with respect to the ith variable x_i is

$$\frac{\partial u}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}.$$

(Strategy: treat all the other variables as constants and take the derivative with respect to the variable of interest.)

Suppose $u = f(x_1, x_2, \ldots, x_n)$ is defined on $\mathbb{R}^n$. If $\frac{\partial^2 u}{\partial x_i \partial x_j}$ and $\frac{\partial^2 u}{\partial x_j \partial x_i}$ are both continuous on $\mathbb{R}^n$, then $\frac{\partial^2 u}{\partial x_i \partial x_j} = \frac{\partial^2 u}{\partial x_j \partial x_i}$.

The Gradient Vector and Hessian Matrix

Suppose $f(x_1, x_2, \ldots, x_n)$ is a function of n variables such that all the partial derivatives exist, then the gradient vector of f is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

If all the second-order partial derivatives of f also exist, the Hessian matrix of f is

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Change of Variables

Take the two-variable case as an example:
Suppose $z = f(x, y)$ is function of x, y and $x = u(s, t)$, $y = v(s, t)$ with respect two other variables s, t, then $z = g(s, t)$ as a function of s, t, where

$$g(s, t) = f(u(s, t), v(s, t))|J|.$$

Here J is the **Jacobian** of the transformation $x = u(s, t)$, $y = v(s, t)$:

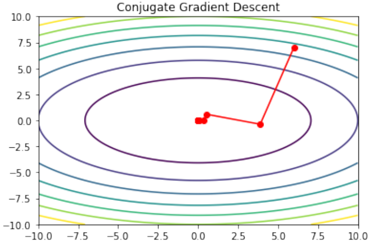
$$J = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{vmatrix} = \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} - \frac{\partial x}{\partial t}\frac{\partial y}{\partial s}.$$

Suppose that we want to integrate $f(x, y)$ over a region R. Under the transformation $x = u(s, t)$, $y = v(s, t)$ the regions becomes S and the integral becomes:

$$\iint_R f(x, y)dxdy = \iint_S f(u(s, t), v(s, t))|J|dsdt.$$

Optimization in Statistics

Often in statistics, we come across problems that involve optimization. Maximum Likelihood Estimation and Maximum A Posteriori Estimation, two commonly used methods of inference, both involve optimizing functions (the likelihood and the posterior distribution, respectively).



In more complicated settings, numerical methods are used to find minimum values in a high dimensional parameter space. An example is using stochastic gradient descent to optimize the weights of a neural network.

Lagrange Multipliers

Lagrange Multipliers are a method of optimizing a function $f(x, y, z)$ subject to a given restraint $g(x, y, z) = k$

1. Solve the following:

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$$
$$g(x, y, z) = k$$

2. Plug in (x, y, z) to $f(x, y, z)$ to find the min/max provided that the $\nabla f(x, y, z) = 0$ at that point.

Lagrange Multipliers Application

While a more in depth discussion of Maximum Likelihood Estimation will occur later, finding the MLE estimates for the cell probabilities of the multinomial distribution provides an example of Lagrange Multipliers in a statistical application.

Let $X_1 ... X_3 \sim Multinomial(p_1 ... p_3)$, with $\sum_{i=1}^{3} p_i = 1$. In MLE, we find the values of $p_1 ... p_3$ that maximize the likelihood, but here we need the constraint that they sum to 1.

$$L(p_1, p_2, p_3, \lambda) = \ell(p_1, p_2, p_3) + \alpha(1 - \sum_{i=1}^{3} p_i)$$

Taking the partial derivatives of p_1, p_2, p_3 and λ and setting them to 0 gives the maximum likelihood estimate.

MLE Multinomial Details

$$L(p_1, p_2, p_3, \lambda) = log(n!) - \sum_{i=1}^{3} log(x_i!) + \sum_{i=1}^{3} x_i log(p_i) + \lambda(1 - \sum_{i=1}^{3} p_i)$$

Taking the partial derivatives and setting them to zero gives the following system of equations to solve.

$$\frac{x_1}{\hat{p}_1} = 0, \; \frac{x_2}{\hat{p}_2} = 0$$
$$\frac{x_3}{\hat{p}_3} = 0, \; \sum_{i=1}^{3} \hat{p}_i = 1$$

Solving these equations gives $\hat{p}_i = \frac{x_i}{\sum_{j=1}^{n} x_j}$