Mathematics Prerequisites

Contents

> Math is not just a way of calculating numerical answers; it is a way of thinking, using clear definitions for concepts and rigorous logic to organize our thoughts and back up our assertions.

Cheng (2025)

These lecture notes use:

- algebra
- precalculus
- univariate calculus
- linear algebra
- vector calculus

Some key results are listed here.

1 Algebra

1.1 Elementary Algebra

Mastery of Elementary Algebra[1] (a.k.a. "College Algebra") is a prerequisite for calculus, which is a prerequisite for Epi 202 and Epi 203, which are prerequisites for this course (Epi 204). Nevertheless, each year, some Epi 204 students are still uncomfortable with algebraic manipulations of mathematical formulas. Therefore, I include this section as a quick reference.

1.1.1 Equalities

Theorem 1.1 (Equalities are transitive)**.** *If* $a = b$ *and* $b = c$*, then* $a = c$

Theorem 1.2 (Substituting equivalent expressions)**.** *If* $a = b$*, then for any function* $f(x)$*,* $f(a) = f(b)$

1.1.2 Inequalities

Theorem 1.3 (Adding to both sides of an inequality)**.** *If* $a < b$*, then* $a + c < b + c$

Theorem 1.4 (negating both sides of an inequality)**.** *If* $a < b$*, then:* $-a > -b$

Theorem 1.5 (Multiplying both sides of an inequality by a nonnegative number)**.** *If* $a < b$ *and* $c \geq 0$*, then* $ca < cb$*.*

Theorem 1.6 (Negation is multiplication by -1)**.**

$$-a = (-1) * a$$

[1] https://en.wikipedia.org/wiki/Elementary_algebra

1.1.3 Infimum and supremum

Definition 1.1 (Infimum (greatest lower bound))**.** The **infimum** of a nonempty set $A \subseteq \mathbb{R}$, written $\inf A$, is the greatest real number m satisfying $m \leq a$ for all $a \in A$:

$$\inf A \stackrel{\text{def}}{=} \max_{t \in \mathbb{R}} \{t : \forall a \in A, a \geq t\}$$

If the infimum belongs to A, it equals the minimum: $\inf A = \min A$.

Exm

Example 1.1 (Numerical examples of infimum)**.**

- $\inf\{1, 2, 3\} = 1$, since 1 is the smallest element.
- $\inf(0.5, 1] = 0.5 = \min[0.5, 1]$: for intervals open below, the infimum equals the minimum of the corresponding closed-below interval, even though $0.5 \notin (0.5, 1]$. More generally, $\inf(c, b] = \min[c, b] = c$ for any $c < b$.
- $\inf\{t \geq 0 : t > 0.5\} = 0.5$, even though 0.5 itself is not in the set.

Definition 1.2 (Supremum (least upper bound))**.** The **supremum** of a nonempty set $A \subseteq \mathbb{R}$, written $\sup A$, is the smallest real number M satisfying $M \geq a$ for all $a \in A$:

$$\sup A \stackrel{\text{def}}{=} \min_{t \in \mathbb{R}} \{t : \forall a \in A, a \leq t\}$$

If the supremum belongs to A, it equals the maximum: $\sup A = \max A$.

Exm

Example 1.2 (Numerical examples of supremum)**.**

- $\sup\{1, 2, 3\} = 3$, since 3 is the largest element.
- $\sup\{t \geq 0 : t < 0.5\} = 0.5$, even though 0.5 itself is not in the set.

1.1.4 Sums

Theorem 1.7 (adding zero changes nothing)**.**

$$a + 0 = a$$

Theorem 1.8 (Sums are symmetric)**.**

$$a + b = b + a$$

Theorem 1.9 (Sums are associative)**.**
When summing three or more terms, the order in which you sum them does not matter:

$$(a + b) + c = a + (b + c)$$

1.1.5 Products

Theorem 1.10 (Multiplying by 1 changes nothing)**.**

$$a \times 1 = a$$

Theorem 1.11 (Products are symmetric)**.**

$$a \times b = b \times a$$

Theorem 1.12 (Products are associative)**.**

$$(a \times b) \times c = a \times (b \times c)$$

1.1.6 Division

Theorem 1.13 (Division can be written as a product)**.**

$$\frac{a}{b} = a \times \frac{1}{b}$$

1.1.7 Sums and products together

Theorem 1.14 (Multiplication is distributive)**.**

$$a(b + c) = ab + ac$$

1.1.8 Quotients

Definition 1.3 (Quotients, fractions, rates)**.**
A **quotient**, **fraction**, or **rate** is a division of one quantity by another:

$$\frac{a}{b}$$

In epidemiology, rates typically have a quantity involving time or population in the denominator.
c.f. https://en.wikipedia.org/wiki/Rate_(mathematics)

Definition 1.4 (Ratios)**.** A **ratio** is a quotient in which the numerator and denominator are measured using the same unit scales.
c.f. https://en.wikipedia.org/wiki/Ratio

Definition 1.5 (Proportion)**.** In statistics, a "proportion" typically means a ratio where the numerator represents a subset of the denominator.

See https://en.wikipedia.org/wiki/Population_proportion.
See also https://en.wikipedia.org/wiki/Proportion_(mathematics) for other meanings.

Definition 1.6 (Proportional)**.** Two functions $f(x)$ and $g(x)$ are **proportional** if their ratio $\frac{f(x)}{g(x)}$ does not depend on x. (c.f. https://en.wikipedia.org/wiki/Proportionality_(mathematics))

Additional reference for elementary algebra: https://en.wikipedia.org/wiki/Population_proportion#Mathematical_definition

---

1.1.9 Exponentials and Logarithms

Theorem 1.15 (logarithm of a product is the sum of the logs of the factors)**.**

$$\log a \cdot b = \log a + \log b$$

Corollary 1.1 (logarithm of a quotient)**.**
The logarithm of a quotient is equal to the difference of the logs of the factors:

$$\log \frac{a}{b} = \log a - \log b$$

Theorem 1.16 (logarithm of an exponential function)**.**

$$\log\{a^b\} = b \cdot \log\{a\}$$

Theorem 1.17 (exponential of a sum)**.**
The exponential of a sum is equal to the product of the exponentials of the addends:

$$\exp\{a + b\} = \exp\{a\} \cdot \exp\{b\}$$

Corollary 1.2 (exponential of a difference)**.**
The exponential of a difference is equal to the quotient of the exponentials of the addends:

$$\exp\{a - b\} = \frac{\exp\{a\}}{\exp\{b\}}$$

Theorem 1.18 (exponential of a product)**.**

$$a^{bc} = (a^b)^c = (a^c)^b$$

Corollary 1.3 (natural exponential of a product)**.**

$$\exp\{ab\} = (\exp\{a\})^b = (\exp\{b\})^a$$

Exercise 1.1. For $a \geq 0,\ b, c \in \mathbb{R}$, When does $(a^b)^c = a^{(b^c)}$?

Solution

Solution 1.1. Short answer: rarely (that's all you need to know for this course).
Long answer:
If $(a^b)^c = a^{(b^c)}$, then since $(a^b)^c = a^{bc}$, we have:

$$a^{bc} = a^{(b^c)}$$

$$\log\{a^{bc}\} = \log\{a^{(b^c)}\}$$
$$bc \cdot \log\{a\} = b^c \cdot \log\{a\} \tag{1}$$

Equation 1 holds in each of the following cases:

1. $bc = b^c$ (see Exercise 1.2).
2. $a = 1$ (i.e., $\log\{a\} = 0$).
3. $a = 0$ (i.e., $\log\{a\} = -\infty$) *and* $\text{sign}\{bc\} = \text{sign}\{b^c\}$.

In particular, when $a = 0$ and $c = 0$, $bc = 0$ and $b^c = 1$ (for any $b \in \mathbb{R}$), so $\text{sign}\{bc\} \neq \text{sign}\{b^c\}$, and $(a^b)^c \neq a^{(b^c)}$:

$$\begin{aligned}(a^b)^c &= (0^b)^0 \\ &= 1\end{aligned}$$

$$\begin{aligned}a^{(b^c)} &= 0^{(b^0)} \\ &= 0^1 \\ &= 0\end{aligned}$$

Exercise 1.2. For $b, c \in \mathbb{R}$, when does $b^c = bc$?

Solution

Solution 1.2. $bc = b^c$ in each of the following cases:

1. $c = 1$.
2. $b = 0$ and $c > 0$.
3. $b = \exp\left\{\frac{\log c}{c-1}\right\}$ (for $c \geq 0$).

See the red contours in Figure 2 for a visualization.

```r
`b*c_f` <- function(b, c) b*c
`b^c_f` <- function(b, c) b^c
values_b <- seq(0, 5, by = .01)
values_c <- seq(-.5, 3, by = .01)

`b*c` <- outer(values_b, values_c, `b*c_f`)
`b^c` <- outer(values_b, values_c, `b^c_f`)
`b^c`[is.infinite(`b^c`)] = NA

opacity <- .3
z_min <- min(`b*c`, `b^c`, na.rm = TRUE)
z_max <- 5
plotly::plot_ly(
  x = ~values_b,
  y = ~values_c
) |>
  plotly::add_surface(
    z = ~ t(`b*c`),
    contours = list(
      z = list(
        show = TRUE,
        start = -1,
        end = 1,
        size = .1
      )
    ),
    name = "b*c",
    showscale = FALSE,
    opacity = opacity,
    colorscale = list(c(0, 1), c("green", "green"))
  ) |>
  plotly::add_surface(
    opacity = opacity,
    colorscale = list(c(0, 1), c("red", "red")),
    z = ~ t(`b^c`),
    contours = list(
      z = list(
        show = TRUE,
        start = z_min,
        end = z_max,
        size = .2
      )
    ),
    showscale = FALSE,
    name = "b^c"
  ) |>
  plotly::layout(
    scene = list(
      xaxis = list(
        # type = "log",
        title = "b"
      ),
      yaxis = list(
        # type = "log",
        title = "c"
      ),
      zaxis = list(
        # type = "log",
        range = c(z_min, z_max),
        title = "outcome"
      ),
      camera = list(eye = list(x = -1.25, y = -1.25, z = 0.5)),
```

```
`b^c - b*c_f` <- function(b, c) `b^c_f`(b,c) - `b*c_f`(b,c)

mat1 <- outer(values_b, values_c, `b^c - b*c_f`)
mat1[is.infinite(mat1)] = NA

opacity <- .3
plotly::plot_ly(
  x = ~values_b,
  y = ~values_c
) |>
  plotly::add_surface(
    z = ~ t(mat1),
    contours = list(
      z = list(
        show = TRUE,
        start = 0,
        end = 1,
        size = 1,
        color = "red"
      )
    ),
    name = "b^c - b*c",
    showscale = TRUE,
    opacity = opacity
  ) |>
  plotly::layout(
    scene = list(
      xaxis = list(
        # type = "log",
        title = "b"
      ),
      yaxis = list(
        # type = "log",
        title = "c"
      ),
      zaxis = list(
        title = "outcome"
      ),
      camera = list(eye = list(x = -1.25, y = -1.25, z = 0.5)),
      aspectratio = list(x = .9, y = .8, z = 0.7)
    )
  )
```

Theorem 1.19 (exp{} and log{} are mutual inverses).

$$\exp\{\log\{a\}\} = \log\{\exp\{a\}\} = a$$

2 Derivatives

Theorem 2.1 (Constant rule).

$$\frac{\partial}{\partial x}c = 0$$

Theorem 2.2 (Power rule). *If a is constant with respect to x, then:*

$$\frac{\partial}{\partial x}ay = a\frac{\partial x}{\partial y}$$

Theorem 2.3 (Power rule).

$$\frac{\partial}{\partial x}x^q = qx^{q-1}$$

Theorem 2.4 (Derivative of natural logarithm).

$$\log'\{x\} = \frac{1}{x} = x^{-1}$$

Theorem 2.5 (derivative of exponential).

$$\exp'\{x\} = \exp\{x\}$$

Theorem 2.6 (Product rule).

$$(ab)' = ab' + ba'$$

Theorem 2.7 (Quotient rule).

$$(a/b)' = a'/b - (a/b^2)b'$$

Theorem 2.8 (Chain rule).

$$\begin{aligned}\frac{\partial a}{\partial c} &= \frac{\partial a}{\partial b}\frac{\partial b}{\partial c}\\ &= \frac{\partial b}{\partial c}\frac{\partial a}{\partial b}\end{aligned}$$

or in Euler/Lagrange notation[a]*:*

$$(f(g(x)))' = g'(x)f'(g(x))$$

[a] https://en.wikipedia.org/wiki/Notation_for_differentiation#Lagrange's_notation

---

Corollary 2.1 (Chain rule for logarithms)**.**

$$\frac{\partial}{\partial x} \log f(x) = \frac{f'(x)}{f(x)}$$

Proof

Proof. Apply Theorem 2.8 and Theorem 2.4. □

---

3 Integration

Integration is the inverse operation of differentiation: it recovers a function from its derivative and accumulates quantities such as areas, totals, and probabilities. We begin with antiderivatives, then state basic integration rules, and conclude with the Fundamental Theorem of Calculus and a worked example from probability.

3.1 Antiderivatives

Definition 3.1 (Antiderivative)**.** A function F is an **antiderivative** of f on an interval I if:

$$\frac{\partial}{\partial x} F(x) = f(x), \quad \forall x \in I$$

The family of all antiderivatives of f is written as the **indefinite integral**:

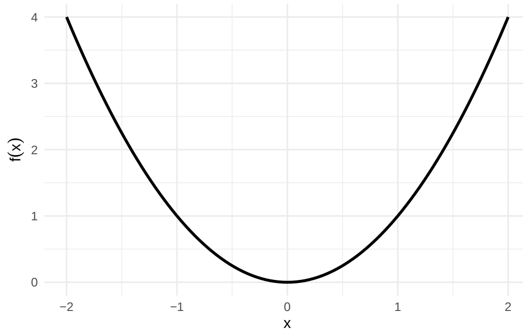
$$\int f(x)\, dx = F(x) + C$$

where C is an arbitrary constant of integration.
(Larson and Edwards 2018, sec. 4.1, pp. 248–249)

Exm

Example 3.1 (Antiderivative of a power function)**.** For $f(x) = x^2$, an antiderivative is $F(x) = \frac{x^3}{3}$, since $\frac{\partial}{\partial x}\frac{x^3}{3} = x^2 = f(x)$.
Adding any constant C gives another antiderivative; for example, with $C = 7$, $F(x) = \frac{x^3}{3} + 7$ also satisfies $F'(x) = x^2$, since adding a constant does not change the derivative. See Figure 3.

```r
ggplot() +
  geom_function(fun = \(x) x^2, xlim =
  labs(x = "x", y = expression(f(x)))
  theme_minimal()
```

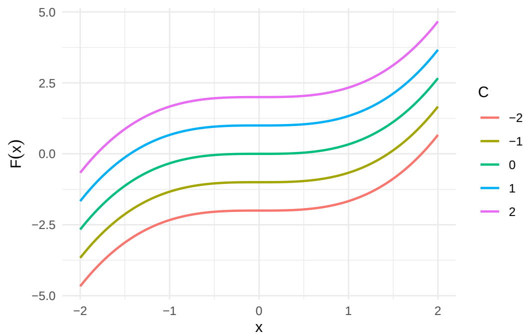


(a)

(b) The function $f(x) = x^2$.

```r
x_seq <- seq(x_lim[1], x_lim[2], length.out = 200)
df <- do.call(rbind, lapply(C_vals, \(C) {
  data.frame(
    x = x_seq,
    y = x_seq^3 / 3 + C,
    C = factor(C)
  )
}))

ggplot(df, aes(x = x, y = y, color = C)) +
  geom_line(linewidth = 0.8) +
  labs(x = "x", y = expression(F(x)), color = "C") +
  theme_minimal()
```



(c)

(d) Family of antiderivatives $F(x) = x^3/3 + C$.

Figure 3: The function $f(x) = x^2$ and five antiderivatives $F(x) = x^3/3 + C$ for $C \in \{-2, -1, 0, 1, 2\}$. Each antiderivative has the same derivative f; they differ only by a vertical shift.

Theorem 3.1 (Basic integration rules)**.** *Each antiderivative below is defined only up to an arbitrary constant C (see Definition 3.1); the table omits $+C$ from every row for brevity.*

Function $f(x)$	*Antiderivative* $F(x)$	*Condition*
c	cx	—
x^n	$\frac{x^{n+1}}{n+1}$	$n \neq -1$
$\frac{1}{x}$	$\ln\|x\|$	$x \neq 0$
e^x	e^x	*(self-antiderivative)*
e^{cx}	$\frac{1}{c}e^{cx}$	$c \neq 0$
$c \cdot f(x)$	$c \cdot F(x)$	—
$f(x) + g(x)$	$F(x) + G(x)$	—

The first two rows and the bottom two rows (linearity) are from (Larson and Edwards 2018, sec. 4.1, p. 250 "Basic Integration Rules"); $1/x$ is from (Larson and Edwards 2018, sec. 5.2, Theorem 5.5, p. 324); e^x and e^{cx} are from (Larson and Edwards 2018, sec. 5.4, Theorem 5.12, p. 346).

Exm

Example 3.2 (Antiderivative of $3x^2 - 1$)**.** By the power rule ($n = 2$) and linearity from Theorem 3.1:

$$\int (3x^2 - 1)\,dx = 3 \cdot \frac{x^3}{3} - x + C = x^3 - x + C.$$

Verify by differentiating: $\frac{\partial}{\partial x}(x^3 - x + C) = 3x^2 - 1 = f(x)$, as required.

3.2 Regularity Conditions

Definition 3.2 (Differentiable function)**.** A function f is **differentiable at** $x = c$ if the limit

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

exists and is finite. f is **differentiable on** an interval if it is differentiable at every interior point; at a closed endpoint, the appropriate one-sided derivative is used.
(Larson and Edwards 2018, sec. 2.1, p. 100)

Definition 3.3 (Continuous function)**.** A function f is **continuous at** $x = c$ if all three conditions hold:

1. $f(c)$ is defined,
2. $\lim_{x \to c} f(x)$ exists, and
3. $\lim_{x \to c} f(x) = f(c)$.

f is **continuous on** a closed interval $[a, b]$ if it is continuous at every point of $[a, b]$.
(Larson and Edwards 2018, sec. 1.4, p. 73)

Definition 3.4 (Riemann integrable)**.** A bounded function f is **Riemann integrable on** $[a, b]$ if the Riemann integral

$$\int_a^b f(x)\,dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x, \quad \Delta x = \frac{b-a}{n},$$

exists and is finite (for equal-width partitions of width $\Delta x = (b-a)/n$), where x_i^* is any point in the i-th subinterval.
(Larson and Edwards 2018, sec. 4.3, p. 272)

i General Riemann integrability

More generally, using partitions $\mathcal{P}$ of arbitrary mesh — subintervals of varying widths Δx_i — a bounded function f is Riemann integrable on $[a, b]$ if

$$\int_a^b f(x)\,dx = \lim_{\|\mathcal{P}\| \to 0} \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i$$

exists and is finite, where $\|\mathcal{P}\| = \max_i \Delta x_i$ is the mesh of the partition.

Theorem 3.2 (Equivalence of Riemann sum formulations)**.** *For continuous f on a closed interval $[a, b]$, the equal-width Riemann sum (Definition 3.4) and the arbitrary-mesh Riemann sum (in the callout above) give the same value (Rudin 1976, chap. 6). The equal-width form in Definition 3.4 is used throughout Epi 204.*

Before stating the Fundamental Theorem of Calculus, we record two prerequisite results. The FTC requires the integrand f to be continuous, and the two theorems below establish where continuity comes from (differentiability $\Rightarrow$ continuity) and what it buys us (continuity $\Rightarrow$ integrability).

Theorem 3.3 (Differentiability implies continuity)**.** *If f is differentiable at $x = c$, then f is continuous at $x = c$.*
(Larson and Edwards 2018, Theorem 2.1, p. 106)

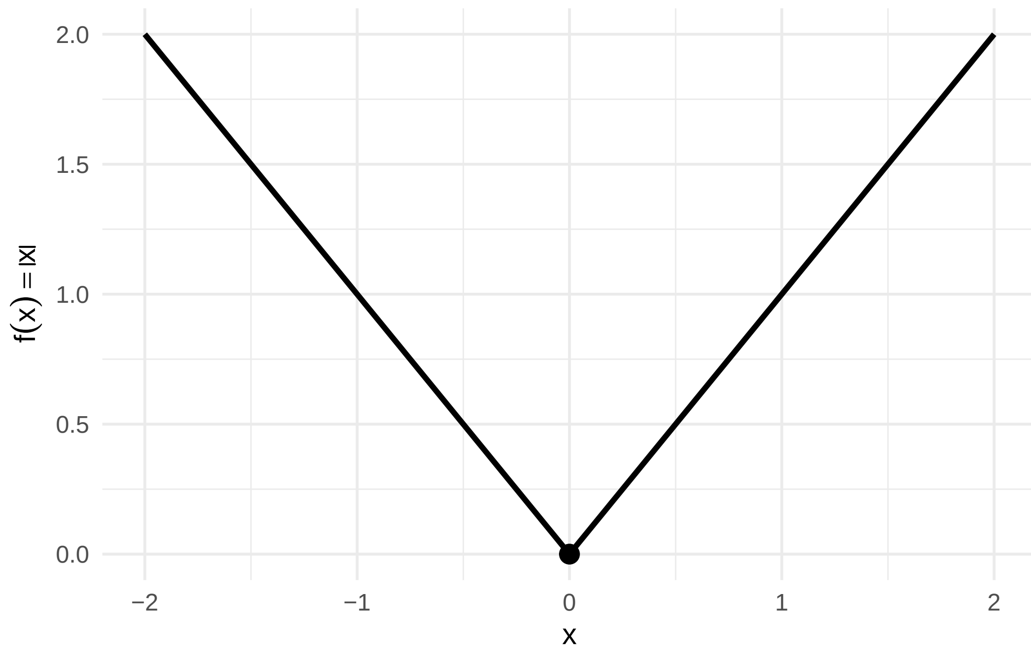
Exm

Example 3.3 (Differentiable, hence continuous: $x^3 - x$)**.** $f(x) = x^3 - x$ is differentiable everywhere (with derivative $f'(x) = 3x^2 - 1$), so by Theorem 3.3 it is continuous everywhere.

Exm

Example 3.4 (Continuous but not differentiable: $|x|$)**.** The absolute-value function $f(x) = |x|$ is continuous at $x = 0$ ($\lim_{x\to 0}|x| = 0 = |0|$), but it is not differentiable at $x = 0$: the left-derivative is -1 and the right-derivative is $+1$.
This shows that the converse of Theorem 3.3 fails: continuity does not imply differentiability. See Figure 4.

```
ggplot() +
  geom_function(fun = abs, xlim = c(-2, 2), linewidth = 1) +
  geom_point(aes(x = 0, y = 0), size = 3) +
  labs(x = "x", y = expression(f(x) == abs(x))) +
  theme_minimal()
```



(a)

Figure 4: $f(x) = |x|$ has a sharp corner at $x = 0$ (not differentiable there) but is continuous everywhere: no gaps or jumps.

Theorem 3.4 (Continuity implies integrability)**.** *If f is continuous on the closed interval $[a, b]$, then f is integrable on $[a, b]$ (i.e., the Riemann integral $\int_a^b f(x)\,dx$ exists and is finite).*
(Larson and Edwards 2018, Theorem 4.4, p. 272)

Exm

Example 3.5 (Continuous, hence integrable: polynomials)**.** Every polynomial is continuous on $\mathbb{R}$, so by Theorem 3.4 every polynomial is integrable on every closed interval $[a, b]$.

Exm

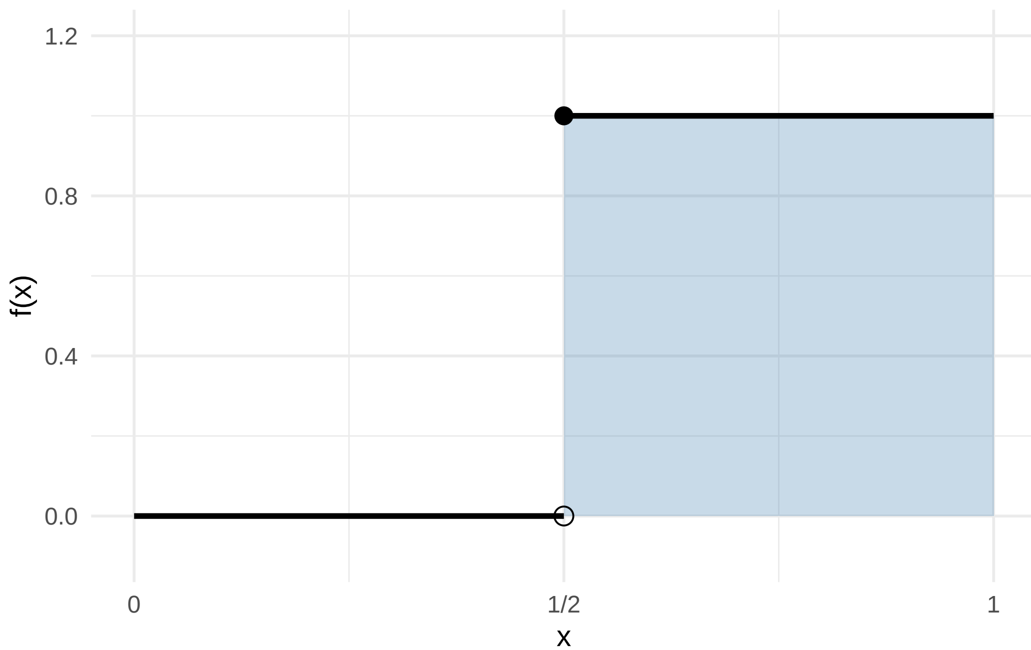
Example 3.6 (Integrable but not continuous: a step function)**.** Let $f(x) = 0$ for $x < \frac{1}{2}$ and $f(x) = 1$ for $x \geq \frac{1}{2}$. Then f is discontinuous at $x = \frac{1}{2}$, but it is integrable on $[0, 1]$:

$$\int_0^1 f(x)\,dx = \int_0^{1/2} 0\,dx + \int_{1/2}^1 1\,dx = 0 + \tfrac{1}{2} = \tfrac{1}{2}.$$

This shows that the converse of Theorem 3.4 fails: integrability does not imply continuity. See Figure 5.

```r
step_df <- data.frame(
  x = c(0, 0.5, 0.5, 1),
  y = c(0, 0, 1, 1),
  segment = c("left", "left", "right", "right")
)

ggplot() +
  geom_rect(
    aes(xmin = 0.5, xmax = 1, ymin = 0, ymax = 1),
    fill = "steelblue", alpha = 0.3
  ) +
  geom_line(
    data = step_df,
    aes(x = x, y = y, group = segment),
    linewidth = 1
  ) +
  geom_point(aes(x = 0.5, y = 0), shape = 1, size = 3) +
  geom_point(aes(x = 0.5, y = 1), shape = 16, size = 3) +
  scale_x_continuous(breaks = c(0, 0.5, 1), labels = c("0", "1/2", "1")) +
  scale_y_continuous(limits = c(-0.1, 1.2)) +
  labs(x = "x", y = "f(x)") +
  theme_minimal()
```



(a)

Figure 5: Step function: $f(x) = 0$ on $[0, \frac{1}{2})$ (open circle at the jump) and $f(x) = 1$ on $[\frac{1}{2}, 1]$ (filled circle). The shaded rectangle has area $\frac{1}{2}$, matching the integral computed above.

Together, Theorem 3.3 and Theorem 3.4 establish the chain:

$$\text{differentiable} \;\Rightarrow\; \text{continuous} \;\Rightarrow\; \text{integrable}$$

Example 3.4 and Example 3.6 show that neither implication reverses in general.

3.3 Fundamental Theorem of Calculus

Theorem 3.5 (Fundamental Theorem of Calculus)**.** *Let f be a continuous function on a closed interval $[a, b]$.*
Part 1 (Derivative of an integral). *Define $F(x) = \int_a^x f(t)\,dt$ for $x \in [a, b]$. Then F is differentiable and:*

$$\frac{\partial}{\partial x} \int_a^x f(t)\,dt = f(x) \tag{2}$$

> ***Note***
> *Continuity on all of $[a, b]$ is a sufficient condition. More generally, Part 1 holds at any individual point x where f is integrable on $[a, b]$ (see Definition 3.4) and continuous at x (see Definition 3.3), even if f has jump discontinuities elsewhere (Rudin 1976, Theorem 6.20, p. 133).*

(Larson and Edwards 2018, Theorem 4.11, p. 288)
Part 2 (Evaluation theorem). *The F here may be* any *antiderivative of f — not just the accumulation function from Part 1. If F is an antiderivative of f on $[a, b]$ (i.e., $\frac{\partial}{\partial x}F(x) = f(x)$ for all $x \in [a, b]$), then:*

$$\int_a^b f(x)\,dx = F(b) - F(a) \tag{3}$$

Equivalently, with b replaced by a variable upper limit x, integrating the derivative of F recovers the net change in F:

$$\int_a^x F'(t)\,dt = F(x) - F(a) \tag{4}$$

or equivalently in Leibniz notation:

$$\int_a^x \frac{dF}{dt}\,dt = F(x) - F(a)$$

(Banner 2007, chap. 18; Larson and Edwards 2018, Theorem 4.9, p. 282)

The two parts of the FTC together express that **differentiation and integration are inverse operations**:

- Part 1: differentiating the integral of f recovers f (Equation 2).
- Part 2: the integral of f over $[a, b]$ equals the difference of any antiderivative's values at the endpoints (Equation 3), which rearranges to "integrating the derivative of F recovers the net change in F" (Equation 4).

The standard form of the FTC assumes f is continuous on $[a, b]$; continuity is *sufficient* but not strictly necessary (see the preceding callout note for the more general statement). Since differentiability implies continuity (Theorem 3.3), the FTC applies in particular whenever f is differentiable — a common situation in Epi 204.

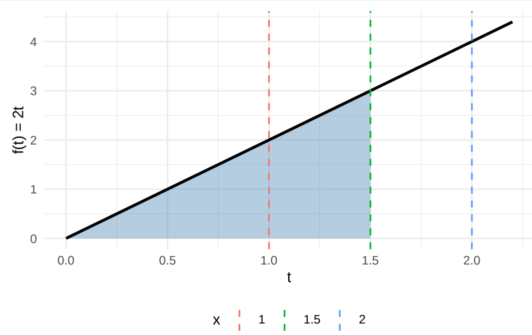
Exm

Example 3.7 (FTC Part 1 visualized: accumulation function for $f(t) = 2t$)**.** Take $f(t) = 2t$ on $[0, 2]$. The accumulation function from 0 is

$$F(x) \stackrel{\text{def}}{=} \int_0^x 2t\,dt \;=\; [t^2]_{t=0}^{t=x} \;=\; x^2 - 0^2 \;=\; x^2,$$

so $F(x) = x^2$, and indeed $F'(x) = 2x = f(x)$, as Theorem 3.5 Part 1 predicts. Figure 6 shows the integrand on the left (shaded area equals $F(x)$ at each x) and the accumulation function $F(x) = x^2$ on the right (its slope at x equals $f(x) = 2x$).

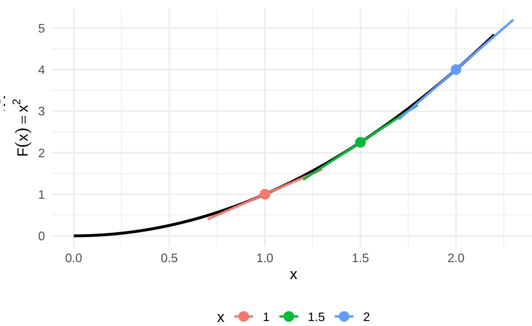
```r
ggplot() +
  geom_area(
    data = data.frame(t = seq(0, x_foc
    aes(x = t, y = 2 * t),
    fill = "steelblue", alpha = 0.4
  ) +
  geom_function(fun = \(t) 2 * t, xlim
  geom_vline(
    data = data.frame(x = x_marks),
    aes(xintercept = x, color = factor
    linetype = "dashed", linewidth = 0
  ) +
  labs(x = "t", y = "f(t) = 2t", color
  theme_minimal() +
  theme(legend.position = "bottom")
```



(a)

(b) $f(t) = 2t$; shaded area equals $F(1.5) = 2.25$; vertical lines mark $x \in \{1, 1.5, 2\}$.

```r
slope_df <- data.frame(
  x = x_marks,
  Fx = x_marks^2,
  slope = 2 * x_marks
)
ggplot() +
  geom_function(fun = \(x) x^2, xlim = c(0, 2.2), linewidth
  geom_point(
    data = slope_df,
    aes(x = x, y = Fx, color = factor(x)),
    size = 3
  ) +
  geom_segment(
    data = slope_df,
    aes(
      x = x - 0.3, xend = x + 0.3,
      y = Fx - 0.3 * slope, yend = Fx + 0.3 * slope,
      color = factor(x)
    ),
    linewidth = 0.8
  ) +
  labs(x = "x", y = expression(F(x) == x^2), color = "x") +
  theme_minimal() +
  theme(legend.position = "bottom")
```



(c)

(d) $F(x) = x^2$; tangent slope at each marked x equals $f(x) = 2x$.

Figure 6: Left: $f(t) = 2t$; the shaded area $\int_0^{1.5} 2t\,dt = F(1.5) = 2.25$; vertical lines mark $x \in \{1, 1.5, 2\}$. Right: $F(x) = x^2$; for each marked x, the tangent slope equals $f(x) = 2x$.

Exm

Example 3.8 (CDF and PDF of the exponential distribution). In what follows, f denotes the PDF and F the CDF — the same letters as the antiderivative pair in Definition 3.1, because the FTC will show F is exactly an antiderivative of f.

For the exponential distribution with rate parameter $\lambda > 0$, the probability density function (PDF) is (Kleinbaum and Klein 2012, sec. II, p. 295, "Survival and Hazard Functions for Selected Distributions"):

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0$$

FTC Part 2 gives the cumulative distribution function (CDF) from the PDF. Apply the e^{cx}

rule from Theorem 3.1 with $c = -\lambda$ to antidifferentiate the integrand:

$$\begin{aligned} F(t) &= \int_0^t \lambda \mathrm{e}^{-\lambda u}\, du \\ &= \left[\lambda \cdot \frac{1}{-\lambda} \mathrm{e}^{-\lambda u}\right]_{u=0}^{u=t} \\ &= \left[(-1)\mathrm{e}^{-\lambda u}\right]_{u=0}^{u=t} \\ &= \left[-\mathrm{e}^{-\lambda u}\right]_{u=0}^{u=t} \\ &= -\mathrm{e}^{-\lambda t} - (-\mathrm{e}^{0}) \\ &= -\mathrm{e}^{-\lambda t} - (-1) \\ &= 1 - \mathrm{e}^{-\lambda t} \end{aligned}$$

FTC Part 1 recovers the PDF from the CDF:

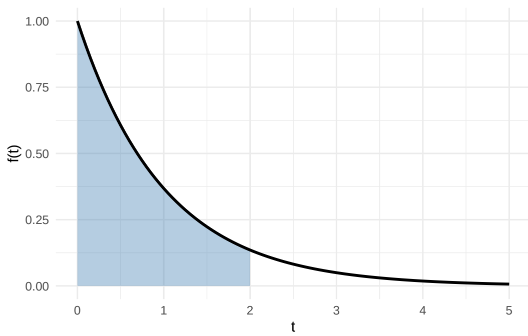
$$\begin{aligned} \frac{\partial}{\partial t} F(t) &= \frac{\partial}{\partial t}(1 - \mathrm{e}^{-\lambda t}) \\ &= 0 - (-\lambda)\mathrm{e}^{-\lambda t} \\ &= \lambda \mathrm{e}^{-\lambda t} \\ &= f(t) \end{aligned}$$

For a concrete instance: with $\lambda = 1$ (standard exponential), the probability that $T \leq 2$ is:

$$F(2) = 1 - \mathrm{e}^{-1 \cdot 2} = 1 - \mathrm{e}^{-2} \approx 1 - 0.135 = 0.865$$

See Figure 7.

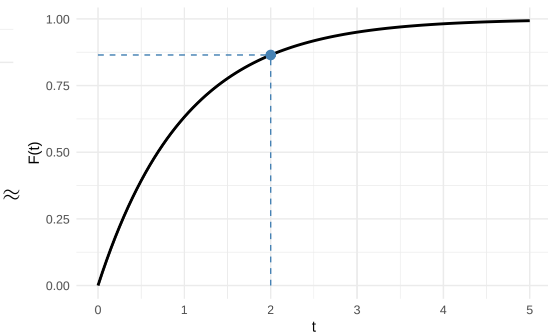
```r
ggplot() +
  geom_area(
    data = data.frame(t = seq(0, t_foc
    aes(x = t, y = lambda * exp(-lambd
    fill = "steelblue", alpha = 0.4
  ) +
  geom_function(
    fun = \(t) lambda * exp(-lambda *
    xlim = c(0, t_max), linewidth = 1
  ) +
  labs(x = "t", y = "f(t)") +
  theme_minimal()
```



(a)

(b) PDF with $\lambda = 1$; shaded area equals $F(2) \approx 0.865$.

```r
ggplot() +
  geom_function(
    fun = \(t) 1 - exp(-lambda * t),
    xlim = c(0, t_max), linewidth = 1
  ) +
  geom_point(
    aes(x = t_focus, y = F_at_focus),
    size = 3, color = "steelblue"
  ) +
  geom_segment(
    aes(x = t_focus, xend = t_focus, y = 0, yend = F_at_foc
    linetype = "dashed", color = "steelblue"
  ) +
  geom_segment(
    aes(x = 0, xend = t_focus, y = F_at_focus, yend = F_at_
    linetype = "dashed", color = "steelblue"
  ) +
  labs(x = "t", y = "F(t)") +
  theme_minimal()
```



(c)

(d) CDF with $\lambda = 1$; point marks $F(2) \approx 0.865$.

Figure 7: Exponential distribution with $\lambda = 1$. Left: the PDF $f(t) = \lambda e^{-\lambda t}$; the shaded area under the curve from 0 to 2 equals $F(2) \approx 0.865$. Right: the CDF $F(t) = 1 - e^{-\lambda t}$; the dashed lines mark the value $F(2)$ computed via FTC Part 2.

4 Double Integrals

The **Fubini–Tonelli theorem** states conditions under which the order of integration in a double integral can be exchanged. We state two versions: the Riemann version (Theorem 4.1) is what Epi 204 directly uses for double integrals of continuous functions on simple regions; the -finite measure-theoretic version (Theorem 4.2) is included to make the joint-distribution form[2] corollary in the probability chapter follow from a stated theorem rather than from an aside.

Theorem 4.1 (Fubini's theorem (Riemann version))**.** *Let f be* ***continuous*** *on a plane region $R \subseteq \mathbb{R}^2$.*

1. ***Vertically simple region.*** *If R is defined by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous on $[a, b]$, then*

$$\iint_R f(x,y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\, dy\, dx.$$

[2] probability.qmd#cor-fubini-joint

2. ***Horizontally simple region.*** *If R is defined by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are continuous on $[c, d]$, then*

$$\iint_R f(x, y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, dx\, dy.$$

When R can be described both ways, the two iterated integrals are equal — so the order of integration can be exchanged.
(Larson and Edwards 2018, Theorem 14.2, p. 982)

Theorem 4.2 (Fubini–Tonelli theorem (measure-theoretic form))**.** *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be* ***-finite*** *measure spaces, and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be measurable with respect to the product -algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$. If either*
*(a) $f \ge 0$ almost everywhere with respect to $\mu_1 \otimes \mu_2$ (**Tonelli's theorem**), or*
*(b) $\int_{\Omega_1 \times \Omega_2} |f|\, d(\mu_1 \otimes \mu_2) < \infty$ (**Fubini's theorem**),*
then both iterated integrals exist, agree with the double integral, and equal each other:

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f\, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1) \right) d\mu_2(\omega_2). \end{aligned}$$

(Billingsley 1995, Theorem 18.3; Gut 2013, Theorem 9.1, p. 65; Fubini 1907; Wikipedia contributors 2024)

This generalization is not required for Epi 204 itself, but it is what justifies the joint-distribution form[3] corollary used later in the probability chapter: probability measures are finite (hence -finite), so the -finiteness hypothesis is automatic. The integrability conditions (nonnegativity or absolute integrability) still need to be verified in each application.

Corollary 4.1 (Continuous functions on a rectangle (corollary of Theorem 4.1))**.** *If $f : [a, b] \times [c, d] \to \mathbb{R}$ is* ***continuous*** *on the closed bounded rectangle $[a, b] \times [c, d]$, then:*

$$\begin{aligned} \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx &= \int_c^d \left(\int_a^b f(x, y)\, dx \right) dy \\ &= \iint_{[a,b]\times[c,d]} f(x, y)\, dx\, dy. \end{aligned}$$

(Larson and Edwards 2018, Theorem 14.2, p. 982)

Proof

Proof. A closed bounded rectangle $[a, b] \times [c, d]$ is both vertically simple (with $g_1 \equiv c$, $g_2 \equiv d$) and horizontally simple (with $h_1 \equiv a$, $h_2 \equiv b$). Applying both parts of Theorem 4.1 to f on this rectangle gives the two iterated forms shown. □

Exm

Example 4.1 (Evaluating a double integral on a rectangle)**.** Adapted from (Larson and Edwards 2018, sec. 14.2, Example 2, pp. 982–983).
Evaluate $\iint_R \left(1 - \frac{1}{2}x^2 - \frac{1}{2}y^2\right) dA$ on the unit square $R = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le 1\}$.

[3] probability.qmd#cor-fubini-joint

The integrand is continuous on R, so Corollary 4.1 applies and either order of integration yields the same value.

Integrating y first, then x (the order Larson chooses):

$$\begin{aligned}\iint_R \left(1 - \frac{1}{2}x^2 - \frac{1}{2}y^2\right) dA &= \int_0^1 \int_0^1 \left(1 - \frac{1}{2}x^2 - \frac{1}{2}y^2\right) dy\,dx \\ &= \int_0^1 \left[\left(1 - \frac{1}{2}x^2\right) y - \frac{y^3}{6}\right]_0^1 dx \\ &= \int_0^1 \left(\frac{5}{6} - \frac{1}{2}x^2\right) dx \\ &= \left[\frac{5}{6}x - \frac{x^3}{6}\right]_0^1 \\ &= \frac{2}{3}\end{aligned}$$

Integrating x first, then y (verifying the order can be swapped):

The integrand is symmetric in x and y, so the same arithmetic with the roles swapped gives:

$$\int_0^1 \int_0^1 \left(1 - \frac{1}{2}x^2 - \frac{1}{2}y^2\right) dx\,dy = \frac{2}{3}$$

Both orders give $\frac{2}{3}$, as Corollary 4.1 guarantees.

```r
n_grid <- 41
x_seq <- seq(0, 1, length.out = n_grid)
y_seq <- seq(0, 1, length.out = n_grid)
z_mat <- outer(x_seq, y_seq, function(x, y) 1 - x^2 / 2 - y^2 / 2)

plotly::plot_ly(x = ~x_seq, y = ~y_seq, z = ~t(z_mat)) |>
  plotly::add_surface(showscale = FALSE) |>
  plotly::layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "y"),
    zaxis = list(title = "z", range = c(0, 1))
  ))
```

Figure 8: Surface $z = 1 - \frac{1}{2}x^2 - \frac{1}{2}y^2$ over the unit square $[0, 1]^2$. The double integral $\frac{2}{3}$ is the volume between this surface and the xy-plane.

Exm

Example 4.2 (Changing the order of integration for a non-rectangular region)**.** Adapted from (Larson and Edwards 2018, sec. 14.2, Example 4, pp. 984–985).
Find the volume of the solid bounded by the surface $z = \mathrm{e}^{-x^2}$ and the planes $z = 0$, $y = 0$, $y = x$, and $x = 1$.
The base of the solid in the xy-plane is the triangular region $D = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq x\}$, so the volume is

$$\iint_D \mathrm{e}^{-x^2}\, dA.$$

Order $dx\,dy$ is intractable. Re-describing D as $D = \{(x, y) : 0 \leq y \leq 1,\ y \leq x \leq 1\}$, the inner integral is

$$\int_y^1 \mathrm{e}^{-x^2}\, dx,$$

which has no elementary antiderivative.
Order $dy\,dx$ works. Applying Theorem 4.1 Part 1 (e^{-x^2} is continuous and D is the vertically simple region $0 \leq x \leq 1$, $0 \leq y \leq x$):

$$\begin{aligned}
\iint_D \mathrm{e}^{-x^2}\, dA &= \int_0^1 \int_0^x \mathrm{e}^{-x^2}\, dy\, dx \\
&= \int_0^1 \mathrm{e}^{-x^2} \left(\int_0^x dy \right) dx \\
&= \int_0^1 x\, \mathrm{e}^{-x^2}\, dx \\
&= \left[-\frac{1}{2}\, \mathrm{e}^{-x^2} \right]_0^1 \\
&= -\frac{1}{2}(\mathrm{e}^{-1} - 1) \\
&= \frac{e - 1}{2e} \\
&\approx 0.316
\end{aligned}$$

```r
n_grid <- 51
x_seq <- seq(0, 1, length.out = n_grid)
y_seq <- seq(0, 1, length.out = n_grid)

z_mat <- outer(x_seq, y_seq, function(x, y) {
  z <- exp(-x^2)
  z[y > x] <- NA
  z
})

plotly::plot_ly(x = ~x_seq, y = ~y_seq, z = ~t(z_mat)) |>
  plotly::add_surface(showscale = FALSE) |>
  plotly::layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "y"),
    zaxis = list(title = "z = exp(-x^2)"),
    camera = list(eye = list(x = 1.6, y = -1.6, z = 0.8))
  ))
```

Exm

Example 4.3 (When conditions fail: a counterexample)**.** The conditions in Theorem 4.1 are not merely technical — when they fail, iterated integrals can exist yet disagree.
Let

$$f(x,y) = \frac{x^2 - y^2}{(x^2 + y^2)^2}$$

on the unit square $R = [0,1] \times [0,1]$. Strictly, f is defined on $R \setminus \{(0,0)\}$: the denominator vanishes at the origin, so f is undefined there (we return to this point below).
Integrating y first, then x:
Using $\dfrac{\partial}{\partial y}\dfrac{y}{x^2+y^2} = \dfrac{x^2-y^2}{(x^2+y^2)^2}$:

$$\begin{aligned} \int_0^1 \int_0^1 f(x,y)\,dy\,dx &= \int_0^1 \left[\frac{y}{x^2+y^2}\right]_{y=0}^{y=1} dx \\ &= \int_0^1 \frac{1}{x^2+1}\,dx \\ &= [\arctan(x)]_0^1 \\ &= \frac{\pi}{4} \end{aligned}$$

Integrating x first, then y:
Using $\dfrac{\partial}{\partial x}\left(-\dfrac{x}{x^2+y^2}\right) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$:

$$\begin{aligned} \int_0^1 \int_0^1 f(x,y)\,dx\,dy &= \int_0^1 \left[-\frac{x}{x^2+y^2}\right]_{x=0}^{x=1} dy \\ &= \int_0^1 \left(-\frac{1}{1+y^2}\right) dy \\ &= -[\arctan(y)]_0^1 \\ &= -\frac{\pi}{4} \end{aligned}$$

Conclusion: $\dfrac{\pi}{4} \neq -\dfrac{\pi}{4}$, so the two iterated integrals are unequal. Neither Theorem 4.1 nor Corollary 4.1 applies here.
Why the hypotheses fail: Theorem 4.1 requires f to be **continuous** on R. The denominator $(x^2+y^2)^2$ vanishes at the origin $(0,0) \in R$, so f is *not even defined* there — let alone continuous — and the theorem does not apply.
For Theorem 4.2, the failure shows up as $\iint_R |f|\,dA = \infty$, which violates the absolute-integrability hypothesis (b). Switching to polar coordinates (r,θ) near the origin, the integrand satisfies $|f(x,y)| = \left|x^2 - y^2\right|/(x^2+y^2)^2 = |\cos 2\theta|/r^2$, so

$$\begin{aligned} \iint_R |f|\,dA &\geq \int_0^{\pi/2} \int_0^{\epsilon} \frac{|\cos 2\theta|}{r^2}\, r\,dr\,d\theta \\ &= \left(\int_0^{\pi/2} |\cos 2\theta|\,d\theta\right) \int_0^{\epsilon} \frac{dr}{r} \\ &= +\infty, \end{aligned}$$

since $\int_0^{\epsilon} dr/r$ diverges.
(Wikipedia contributors 2024)

```r
n_grid <- 81
eps <- 0.04
x_seq <- seq(eps, 1, length.out = n_grid)
y_seq <- seq(eps, 1, length.out = n_grid)

z_mat <- outer(x_seq, y_seq, function(x, y) (x^2 - y^2) / (x^2 + y^2)^2)
z_clip <- 50
z_mat[z_mat > z_clip] <- z_clip
z_mat[z_mat < -z_clip] <- -z_clip

plotly::plot_ly(x = ~x_seq, y = ~y_seq, z = ~t(z_mat)) |>
  plotly::add_surface(
    showscale = FALSE,
    colorscale = list(
      list(0, "#3b4cc0"),
      list(0.5, "#dddddd"),
      list(1, "#b40426")
    )
  ) |>
  plotly::layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "y"),
    zaxis = list(title = "f(x, y)", range = c(-z_clip, z_clip)),
    camera = list(eye = list(x = 1.6, y = 1.6, z = 0.6))
  ))
```

5 Linear Algebra

5.1 Vectors

Definition 5.1 (Column vector)**.** A **column vector** of length p is an ordered list of p numbers, written vertically:

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

Column vectors are the default convention in these notes and in most statistics textbooks. They are also called $p \times 1$ *matrices*.

Definition 5.2 (Transpose)**.** The **transpose** of a column vector $\tilde{x}$ is the **row vector** with the same sequence of entries, written horizontally:

$$\tilde{x}^\top \equiv \tilde{x}' \equiv [x_1,\ x_2,\ \ldots,\ x_p]$$

The transpose operation converts a column vector to a row vector, or more generally, swaps the rows and columns of a matrix (Definition 5.10).

5.1.1 Special vectors

Definition 5.3 (Zero vector)**.** The **zero vector** $\tilde{0}$ of length p has all entries equal to zero:

$$\tilde{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector is the additive identity for vector addition: $\tilde{x} + \tilde{0} = \tilde{x}$ for any vector $\tilde{x}$ of the same length.

Definition 5.4 (Ones vector)**.** The **ones vector** $\tilde{1}$ of length p has all entries equal to one:

$$\tilde{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The dot product $\tilde{1}^\top \tilde{x} = \tilde{1} \cdot \tilde{x} = \sum_{i=1}^{p} x_i$ is the sum of all entries of $\tilde{x}$.

Definition 5.5 (Indicator vector / standard basis vector)**.** The j-th **indicator vector** (or **standard basis vector**) $\tilde{e}_j$ of length p has a 1 in position j and 0s elsewhere:

$$(\tilde{e}_j)_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \tilde{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{position } j$$

They are also called **unit vectors** or **standard basis vectors**.

Theorem 5.1 (Indicator vectors select entries)**.** *For any vector $\tilde{x}$ of length p and any $j \in \{1, \dots, p\}$:*

$$\tilde{e}_j^\top \tilde{x} = x_j$$

Proof

Proof. Writing the product componentwise:

$$\begin{aligned} \tilde{e}_j^\top \tilde{x} &= \sum_{i=1}^{p} (\tilde{e}_j)_i \, x_i \\ &= \sum_{i=1}^{p} \begin{cases} 1 \cdot x_i & \text{if } i = j \\ 0 \cdot x_i & \text{if } i \neq j \end{cases} \\ &= x_j \end{aligned}$$

□

Definition 5.6 (Dot product/linear combination/inner product)**.** For any two real-valued vectors $\tilde{x} = (x_1, \dots, x_n)$ and $\tilde{y} = (y_1, \dots, y_n)$, the **dot-product**, **linear combination**, or **inner product** of $\tilde{x}$ and $\tilde{y}$ is:

$$\tilde{x} \cdot \tilde{y} = \tilde{x}^\top \tilde{y} \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i y_i$$

Note

See also the definitions in

- Dobson and Barnett (2018), §1.3 (equation 1.1, page 7)
- Kaplan (2022), here[a].
- wikipedia[b]

"Linear combination" can also refer to weighted sums of vectors, or in other words matrix-vector multiplication.

The dot-product has a different generalization for two matrices; see wikipedia[c] for more.

[a] https://www.mosaic-web.org/MOSAIC-Calculus/Textbook/Linear-combinations/28-Vectors.html#geometry-arithmetic
[b] https://en.wikipedia.org/wiki/Linear_combination
[c] https://en.wikipedia.org/wiki/Dot_product#Dyadics_and_matrices

---

Theorem 5.2 (Dot product is symmetric)**.** *The dot product is symmetric:*

$$\tilde{x} \cdot \tilde{y} = \tilde{y} \cdot \tilde{x}$$

Proof

Proof. Apply:

- Definition 5.6
- symmetry of scalar multiplication
- Definition 5.6 again

□

Exm

Example 5.1 (Dot product as matrix multiplication)**.** The dot product of two column vectors $\tilde{x}$ and $\tilde{\beta}$ can be written as a matrix product of the row vector $\tilde{x}^\top$ with the column vector $\tilde{\beta}$:

$$\begin{aligned}\tilde{x} \cdot \tilde{\beta} &= \tilde{x}^\top \tilde{\beta} \\ &= [x_1,\ x_2,\ \dots,\ x_p] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &= x_1\beta_1 + x_2\beta_2 + \cdots + x_p\beta_p\end{aligned}$$

5.1.2 Orthogonality

Definition 5.7 (Orthogonal vectors)**.** Two vectors $\tilde{x}$ and $\tilde{y}$ of the same length are **orthogonal** (written $\tilde{x} \perp \tilde{y}$) if their dot product is zero:

$$\tilde{x} \perp \tilde{y} \iff \tilde{x}^\top \tilde{y} = 0$$

Orthogonality generalizes the geometric notion of perpendicularity to arbitrary dimensions.

Definition 5.8 (Orthonormal vectors)**.** A set of vectors $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ is **orthonormal** if the vectors are mutually orthogonal and each has unit length:

$$\tilde{x}_i^\top \tilde{x}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The indicator vectors $\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_p$ (Definition 5.5) form an orthonormal set.

5.2 Matrices

Definition 5.9 (Matrix)**.** A **matrix** of dimensions $m \times n$ is a rectangular array of $m \cdot n$ numbers, arranged in m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The entry in row i and column j is denoted a_{ij} or $(\mathbf{A})_{ij}$. A column vector of length p is a special case: a $p \times 1$ matrix. A row vector of length p is a $1 \times p$ matrix.

5.2.1 Matrix transpose

Definition 5.10 (Matrix transpose)**.** The **transpose** of an $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^\top$ obtained by swapping the rows and columns of $\mathbf{A}$:

$$(\mathbf{A}^\top)_{ij} = a_{ji}$$

Theorem 5.3 (Transpose of a sum)**.**

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$

In particular, for column vectors $\tilde{x}$ and $\tilde{y}$:

$$(\tilde{x} + \tilde{y})^\top = \tilde{x}^\top + \tilde{y}^\top$$

Theorem 5.4 (Transpose of a product)**.** *For compatible matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$$

The order of the factors reverses when transposing a product.

5.2.2 Matrix addition

Definition 5.11 (Zero matrix)**.** The $m \times n$ **zero matrix** $\mathbf{0}_{m \times n}$ (or $\mathbf{0}$ when dimensions are clear from context) has all entries equal to zero:

$$\mathbf{0}_{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Definition 5.12 (Matrix addition)**.** Two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions $m \times n$ can be added element-wise:

$$(\mathbf{A}+\mathbf{B})_{ij} = a_{ij} + b_{ij}$$

Theorem 5.5 (Matrix addition is commutative)**.**

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

Theorem 5.6 (Matrix addition is associative)**.**

$$(\mathbf{A}+\mathbf{B})+\mathbf{C} = \mathbf{A}+(\mathbf{B}+\mathbf{C})$$

Theorem 5.7 (Zero matrix is the additive identity)**.**

$$\mathbf{A}+\mathbf{0} = \mathbf{A}$$

Theorem 5.8 (Additive inverse)**.** *For any matrix $\mathbf{A}$, the matrix $-\mathbf{A}$ (defined by $(-\mathbf{A})_{ij} = -a_{ij}$) satisfies:*

$$\mathbf{A}+(-\mathbf{A}) = \mathbf{0}$$

5.2.3 Scalar multiplication

Definition 5.13 (Scalar multiplication)**.** A matrix $\mathbf{A}$ can be multiplied by a scalar c:

$$(c\mathbf{A})_{ij} = c \cdot a_{ij}$$

5.2.4 Matrix multiplication

Definition 5.14 (Matrix multiplication)**.** The **product** of an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$ is the $m \times n$ matrix $\mathbf{C} = \mathbf{AB}$ with entries:

$$c_{ij} = \sum_{s=1}^{k} a_{is}\, b_{sj}$$

Matrix multiplication is only defined when the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$.

Matrix multiplication is **not** commutative in general: $\mathbf{AB} \neq \mathbf{BA}$.

Theorem 5.9 (Matrix multiplication is associative)**.**

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

Theorem 5.10 (Matrix multiplication is distributive over addition)**.**

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$$

$$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$$

5.2.5 Matrix-vector multiplication

Definition 5.15 (Matrix-vector multiplication)**.** The product of an $m \times p$ matrix $\mathbf{A}$ and a $p \times 1$ column vector $\tilde{x}$ is the $m \times 1$ column vector $\mathbf{A}\tilde{x}$ with entries:

$$(\mathbf{A}\tilde{x})_i = \sum_{j=1}^{p} a_{ij}\, x_j$$

Matrix-vector multiplication is a generalization of the dot product. Each entry of the result is a dot product of a row of $\mathbf{A}$ with the vector $\tilde{x}$.

5.3 Special Matrices

See also Definition 5.11 for the zero matrix.

Definition 5.16 (Square matrix)**.** A matrix is **square** if it has the same number of rows as columns. The number of rows (= columns) is the **order** of the matrix.

Definition 5.17 (Matrix power)**.** For a square matrix $\mathbf{A}$ of order p and a positive integer k, the k-th **power** of $\mathbf{A}$ is:

$$\mathbf{A}^k = \underbrace{\mathbf{A}\,\mathbf{A}\cdots\mathbf{A}}_{k \text{ copies}}$$

In particular, $\mathbf{A}^2 = \mathbf{AA}$.

Definition 5.18 (Identity matrix)**.** The $p \times p$ **identity matrix** $\mathbf{I}_p$ (or $\mathbf{I}$ when the size is clear from context) has ones on the main diagonal and zeros elsewhere:

$$(\mathbf{I}_p)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \mathbf{I}_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Theorem 5.11 (Identity matrix is a multiplicative identity)**.** *For any $m \times p$ matrix $\mathbf{A}$:*

$$\mathbf{A}\,\mathbf{I}_p = \mathbf{A}$$

$$\mathbf{I}_m\,\mathbf{A} = \mathbf{A}$$

Definition 5.19 (Symmetric matrix)**.** A square matrix $\mathbf{A}$ is **symmetric** if $\mathbf{A}^\top = \mathbf{A}$, i.e., $a_{ij} = a_{ji}$ for all i and j.

Covariance matrices and information matrices are symmetric.

Definition 5.20 (Diagonal matrix)**.** A square matrix $\mathbf{D}$ is a **diagonal matrix** if all off-diagonal entries are zero: $d_{ij} = 0$ whenever $i \neq j$:

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}$$

Diagonal matrices are denoted $\mathbf{D} = \operatorname{diag}(d_1, d_2, \dots, d_p)$, where $d_1, \dots, d_p$ are the diagonal entries.

Definition 5.21 (Matrix inverse)**.** For a square $p \times p$ matrix $\mathbf{A}$, the **inverse** $\mathbf{A}^{-1}$ (if it exists) is the unique matrix satisfying:

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}_p$$

A matrix that has an inverse is called **invertible** or **non-singular**.

Theorem 5.12 (Inverse of a product)**.** *For invertible matrices* $\mathbf{A}$ *and* $\mathbf{B}$*:*

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Definition 5.22 (Idempotent matrix)**.** A square matrix $\mathbf{A}$ is **idempotent** if

$$\mathbf{A}^2 = \mathbf{A}$$

Definition 5.23 (Projection matrix)**.** A square matrix $\mathbf{P}$ is a **projection matrix** (also called an **orthogonal projector**) if it is both symmetric and idempotent:

$$\mathbf{P}^\top = \mathbf{P} \qquad \text{and} \qquad \mathbf{P}^2 = \mathbf{P}$$

Theorem 5.13 (Complement of a projection matrix)**.** *If* $\mathbf{P}$ *is a projection matrix, then* $\mathbf{I} - \mathbf{P}$ *is also a projection matrix.*

ℹ Proof

Proof. We verify symmetry and idempotency.
Symmetry:

$$(\mathbf{I}-\mathbf{P})^\top = \mathbf{I}^\top - \mathbf{P}^\top = \mathbf{I}-\mathbf{P}$$

Idempotency:

$$\begin{aligned}(\mathbf{I}-\mathbf{P})^2 &= (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P}) \\ &= \mathbf{I}-\mathbf{P}-\mathbf{P}+\mathbf{P}^2 \\ &= \mathbf{I}-\mathbf{P}-\mathbf{P}+\mathbf{P} \\ &= \mathbf{I}-\mathbf{P}\end{aligned}$$

□

Theorem 5.14 (Hat matrix is a projection matrix)**.** *In a linear regression model with full-rank design matrix* $\mathbf{X}$*, the* ***hat matrix***

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

is a projection matrix.

ℹ Proof

Proof. We verify symmetry and idempotency.
Symmetry:

$$\begin{aligned}\mathbf{H}^\top &= \left(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)^\top \\ &= \left(\mathbf{X}^\top\right)^\top \cdot \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top \cdot \mathbf{X}^\top \\ &= \mathbf{X}\cdot(\mathbf{X}^\top\mathbf{X})^{-1}\cdot\mathbf{X}^\top \\ &= \mathbf{H}\end{aligned}$$

where the third line uses $\left(\mathbf{X}^\top\right)^\top = \mathbf{X}$ and the fact that $\mathbf{X}^\top\mathbf{X}$ is symmetric, so its inverse is also symmetric ($\left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top = (\mathbf{X}^\top\mathbf{X})^{-1}$).
Idempotency:

$$\begin{aligned}\mathbf{H}^2 &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\cdot\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{H}\end{aligned}$$

□

The hat matrix appears in the formula for fitted values in linear regression: $\hat{\tilde{y}} = \mathbf{X}\hat{\tilde{\beta}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{y} = \mathbf{H}\tilde{y}$. It "puts a hat" on $\tilde{y}$ — hence the name.

Theorem 5.15 (Projection matrices produce orthogonal decompositions)**.** *If* $\mathbf{P}$ *is a projection matrix and* $\tilde{v}$ *is any vector of compatible dimension, then the two components of the decomposition*

$$\tilde{v} = \underbrace{\mathbf{P}\tilde{v}}_{projected} + \underbrace{(\mathbf{I}-\mathbf{P})\tilde{v}}_{residual}$$

are orthogonal:

$$\mathbf{P}\tilde{v} \perp (\mathbf{I} - \mathbf{P})\tilde{v}$$

Proof

Proof.

$$\begin{aligned}
(\mathbf{P}\tilde{v})^\top (\mathbf{I} - \mathbf{P})\tilde{v} &= \tilde{v}^\top \mathbf{P}^\top (\mathbf{I} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top \mathbf{P} (\mathbf{I} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top (\mathbf{P} - \mathbf{P}^2)\tilde{v} \\
&= \tilde{v}^\top (\mathbf{P} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top \mathbf{0}\, \tilde{v} \\
&= 0
\end{aligned}$$

where the second line uses symmetry ($\mathbf{P}^\top = \mathbf{P}$) and the fourth line uses idempotency ($\mathbf{P}^2 = \mathbf{P}$). □

5.4 Quadratic Forms

Definition 5.24 (Quadratic form)**.** A **quadratic form** is a mathematical expression of the structure

$$\tilde{x}^\top \mathbf{S}\, \tilde{x}$$

where $\tilde{x}$ is a $p \times 1$ vector and $\mathbf{S}$ is a $p \times p$ matrix.

Quadratic forms are the matrix generalizations of the scalar expression cx^2. They occur frequently in statistics:

- The residual sum of squares in linear regression (Section 6) is a quadratic form.
- The variance of a linear combination of estimates (**?@sec-infer-LMs**) is a quadratic form: $\text{Var}\left(\tilde{x}^\top \tilde{\hat{\beta}}\right) = \tilde{x}^\top\, \text{Var}\left(\tilde{\hat{\beta}}\right) \tilde{x}$.

Theorem 5.16 (Symmetric part of a quadratic form)**.** *If* $\mathbf{S}$ *is a square matrix, then*

$$\tilde{x}^\top \mathbf{S} \tilde{x} = \tilde{x}^\top \left(\frac{1}{2}(\mathbf{S} + \mathbf{S}^\top)\right) \tilde{x}.$$

So the value of a quadratic form depends only on the symmetric part of $\mathbf{S}$*.*

5.5 Design Matrix

Definition 5.25 (Design matrix)**.** In a regression model with n observations and p predictors, the **design matrix** (or **model matrix**) $\mathbf{X}$ is the $n \times p$ matrix whose i-th row is the covariate vector $\tilde{x}_i^\top$ for observation i:

$$\mathbf{X} = \begin{bmatrix} \tilde{x}_1^\top \\ \tilde{x}_2^\top \\ \vdots \\ \tilde{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The product $\mathbf{X}\tilde{\beta}$ collects all the linear predictors $\tilde{x}_i^\top \tilde{\beta}$ into a single $n \times 1$ vector:

$$\mathbf{X}\tilde{\beta} = \begin{bmatrix} \tilde{x}_1^\top \tilde{\beta} \\ \vdots \\ \tilde{x}_n^\top \tilde{\beta} \end{bmatrix}$$

The matrix $\mathbf{X}^\top \mathbf{X}$ is a $p \times p$ symmetric matrix that appears in the OLS estimator $\hat{\tilde{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{y}$.

6 Vector Calculus

(adapted from Fieller (2016), §7.2[4])

This section covers derivatives of functions of vectors and matrices. Linear algebra prerequisites — including vectors, matrices, transpose, dot product, and quadratic forms — are covered in Section 5.

Let $\tilde{x}$ and $\tilde{\beta}$ be column vectors of length p (see Definition 5.1 and Definition 5.6).

Definition 6.1 (Vector derivative)**.** If $f(\tilde{\beta})$ is a function that takes a vector $\tilde{\beta}$ as input, such as $f(\tilde{\beta}) = x'\tilde{\beta}$, then:

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) \\ \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Definition 6.2 (Row-vector derivative)**.** If $f(\tilde{\beta})$ is a function that takes a vector $\tilde{\beta}$ as input, such as $f(\tilde{\beta}) = x'\tilde{\beta}$, then:

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) & \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) & \cdots & \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Theorem 6.1 (Row and column derivatives are transposes)**.**

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) \right)^\top$$

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) \right)^\top$$

Theorem 6.2 (Derivative of a dot product)**.**

$$\frac{\partial}{\partial \tilde{\beta}} \tilde{x} \cdot \tilde{\beta} = \frac{\partial}{\partial \tilde{\beta}} \tilde{\beta} \cdot \tilde{x} = \tilde{x}$$

[4] https://www.taylorfrancis.com/chapters/mono/10.1201/9781315370200-7/vector-matrix-calculus-nick-fieller?context=ubx&refId=c310b723-786a-4f33-ae56-720a6cccd3a1

This looks a lot like non-vector calculus, except that you have to transpose the coefficient.

ℹ Proof

Proof.

$$\begin{aligned}
\frac{\partial}{\partial \beta}(x^\top \beta) &= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p) \\ \frac{\partial}{\partial \beta_2}(x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p) \\ \vdots \\ \frac{\partial}{\partial \beta_p}(x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p) \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \\
&= \tilde{x}
\end{aligned}$$

□

Theorem 6.3 (Derivative of a quadratic form). *For a quadratic form (Definition 5.24), if S is a $p \times p$ matrix that is constant with respect to β, then:*

$$\frac{\partial}{\partial \beta}\beta' S \beta = 2S\beta$$

This is like taking the derivative of cx^2 with respect to x in non-vector calculus.

Corollary 6.1 (Derivative of a simple quadratic form).

$$\frac{\partial}{\partial \tilde{\beta}}\tilde{\beta}'\tilde{\beta} = 2\tilde{\beta}$$

This is like taking the derivative of x^2.

Theorem 6.4 (Vector chain rule).

$$\frac{\partial z}{\partial \tilde{x}} = \frac{\partial y}{\partial \tilde{x}}\frac{\partial z}{\partial y}$$

or in Euler/Lagrange notation:

$$(f(g(\tilde{x})))' = \tilde{g}'(\tilde{x}) f'(g(\tilde{x}))$$

See https://quickfem.com/finite-element-analysis/, specifically https://quickfem.com/wp-content/uploads/IFEM.AppF_.pdf

See also https://en.wikipedia.org/wiki/Gradient#Relationship_with_Fr%C3%A9chet_derivative

This chain rule is like the univariate chain rule (Theorem 2.8), but the order matters now. The version presented here is for the gradient[5] (column vector); the total derivative[6] (row vector) would

[5] https://en.wikipedia.org/wiki/Gradient
[6] https://en.wikipedia.org/wiki/Total_derivative

be the transpose of the gradient[7].

Corollary 6.2 (Vector chain rule for quadratic forms)**.**

$$\frac{\partial}{\partial \tilde{\beta}}\left(\tilde{\varepsilon}(\tilde{\beta}) \cdot \tilde{\varepsilon}(\tilde{\beta})\right) = \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{\varepsilon}(\tilde{\beta})\right)\left(2\tilde{\varepsilon}(\tilde{\beta})\right)$$

7 Additional resources

7.1 Calculus

- Kaplan (2022)
- Khuri (2003)
- Banner (2007)
- Larson and Edwards (2018)
- Miller (2016)
  - http://www.youtube.com/watch?v=xYzQL0TUtBA
  - http://www.youtube.com/watch?v=Ps2SBo_WjoE

7.2 Linear Algebra and Vector Calculus

- Fieller (2016)
- Banerjee and Roy (2014)
- Searle and Khuri (2017)

7.3 Numerical Analysis

- Hua Zhou[8]'s lecture notes for "UCLA Biostat 216 - Mathematical Methods for Biostatistics" (2023 Fall)[9]

7.4 Real Analysis

- Grinberg (2017)

References

Banerjee, Sudipto, and Anindya Roy. 2014. *Linear Algebra and Matrix Analysis for Statistics*. Vol. 181. Crc Press Boca Raton. https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388.

Banner, Adrian D. 2007. *The Calculus Lifesaver : All the Tools You Need to Excel at Calculus*. A Princeton Lifesaver Study Guide. Princeton University Press. https://press.princeton.edu/books/paperback/9780691130880/the-calculus-lifesaver.

Billingsley, Patrick. 1995. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley.

Cheng, Eugenia. 2025. "Opinion | How Math Turned Me from a D.E.I. Skeptic to a Supporter." *The New York Times*. https://www.nytimes.com/2025/09/05/opinion/math-dei.html.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

[7] https://en.wikipedia.org/wiki/Gradient#Relationship_with_total_derivative
[8] https://hua-zhou.github.io/
[9] https://ucla-biostat-216.github.io/2023fall/schedule/schedule.html

Fieller, Nick. 2016. *Basics of Matrix Algebra for Statistics with R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315370200.

Fubini, Guido. 1907. "Sugli Integrali Multipli." *Rendiconti Della Reale Accademia Dei Lincei. Classe Di Scienze Fisiche, Matematiche e Naturali* 16: 608–14.

Grinberg, Raffi. 2017. *The Real Analysis Lifesaver: All the Tools You Need to Understand Proofs*. 1st ed. Princeton Lifesaver Study Guides. Princeton University Press. https://press.princeton.edu/books/paperback/9780691172934/the-real-analysis-lifesaver.

Gut, Allan. 2013. *Probability: A Graduate Course*. 2nd ed. Springer Texts in Statistics. Springer. https://doi.org/10.1007/978-1-4614-4708-5.

Kaplan, Daniel. 2022. *MOSAIC Calculus*. Www.mosaic-web.org. www.mosaic-web.org[10].

Khuri, André I. 2003. *Advanced Calculus with Applications in Statistics*. John Wiley & Sons. https://www.wiley.com/en-us/Advanced+Calculus+with+Applications+in+Statistics%2C+2nd+Edition-p-9780471391043.

Kleinbaum, David G, and Mitchel Klein. 2012. *Survival Analysis: A Self-Learning Text*. 3rd ed. Springer. https://link.springer.com/book/10.1007/978-1-4419-6646-9.

Larson, Ron, and Bruce H. Edwards. 2018. *Calculus*. 11th ed. Cengage Learning. https://www.cengage.com/c/calculus-11e-larson/.

Miller, Steven J. 2016. *The Probability Lifesaver: Calculus Review Problems*. https://web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/index.htm#:~:text=http%3A//web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/supplementalchap_calcreview.pdf.

Rudin, Walter. 1976. *Principles of Mathematical Analysis*. 3rd ed. International Series in Pure and Applied Mathematics. McGraw-Hill.

Searle, Shayle R, and Andre I Khuri. 2017. *Matrix Algebra Useful for Statistics*. John Wiley & Sons.

Wikipedia contributors. 2024. *Fubini's Theorem — Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/wiki/Fubini%27s_theorem.

[10]https://www.mosaic-web.org