NOTES ON MATH FOR DATA SCIENCE

BEN HORVATH

LAST REVISION: JANUARY 7, 2019

The following collection was compiled while studying for the degree of masters of data science at City University of New York over the years 2018–present.

1 ALGEBRA

ARITHMETIC OPERATIONS

$$a\left(\frac{b}{c}\right) = \frac{ab}{c} \qquad \frac{\frac{a}{b}}{c} = \frac{a}{bc}$$

$$\frac{a}{\frac{b}{c}} = \frac{ac}{b} \qquad \frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}$$

$$\frac{a-b}{c-d} = \frac{b-a}{d-c} \qquad \frac{ab+ac}{a} = b+c,\ \forall a \neq 0$$

$$\frac{a+bc}{c} = \frac{a}{c} = \frac{b}{c} \qquad \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

EXPONENTS

$$x^a \cdot x^b = x^{a+b} \qquad \frac{x^a}{x^b} = x^{a-b}$$

$$(xy)^a = x^a y^a \qquad \frac{x}{y}^a = \frac{x^a}{y^a}$$

$$x^0 = 1^{\dagger} \qquad x^{-n} = \frac{1}{x}^{\dagger}$$

$$x^{\frac{1}{a}} = \sqrt[a]{x}^{\dagger} \qquad (xy)^a = x^a y^a$$

$$x^{-a} = \frac{1}{x^a} \qquad \left(\frac{x}{y}\right)^{-a} = \left(\frac{y}{x}\right)^a = \frac{y^a}{x^a}$$

$^{\dagger}\forall x \neq 0$

RADICALS

$$\sqrt[a]{x} = x^{\frac{1}{a}} \qquad \sqrt[a]{xy} = \sqrt[a]{x}\,\sqrt[a]{y}$$

LOGARITHMS

Defined as $y = \log_b x \equiv x = b^y$, with domain of $x > 0$

$$\log_b b = 1 \qquad \log_b 1 = 0$$

$$\log_b b^x = x \qquad b^{\log_b x} = x$$

$$\log_b (x^r) = r \log_b x$$

$$\log_b \left(\frac{x}{y}\right) = \log_b x - \log_b y$$

$$\log_b (xy) = \log_b x + \log_b y$$

FACTORING

$$x^2 - a^2 = (x+a)(x-a)$$

$$x^2 \pm 2ax + a^2 = (x \pm a)^2$$

$$x^2 + (a+b)x + ab = (x+a)(x+b)$$

FUNCTIONS

- A rule for a relationship between an input and an output quantity where each input *uniquely determines* an output value
- Must be of the form $y = f(x)$

TYPE	FORM
Constant	$f(x) = c$
Identity	$f(x) = x$
Absolute	$f(x) = \lvert x \rvert$
Quadratic	$f(x) = x^2$
Cubic	$f(x) = x^3$
Reciprocal	$f(x) = \frac{1}{x^2}$
Square root	$f(x) = \sqrt{x}$

TRANSFORMATIONS OF FUNCTIONS

Transformation	Form
Vertical shift	$f'(x) = f(x) + k$
Horizontal shift	$f'(x) = f(x - k)$
Horizontal reflection	$f'(x) = f(-x)$
Vertical reflection	$f'(x) = -f(x)$
Vertical stretch	$f'(x) = kf(x),\ k > 1$
Vertical compression	$f'(x) = kf(x),\ k0 < k < 1$

DISTANCE FORMULA

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

is the distance between points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$

LINEAR FUNCTIONS

- Linear functions have a constant rate of change, m

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

- Estimate horizontal line by solving $f(x) = 0$
- Find the point where two non-parallel lines meet: set $f(x) = g(x)$ and solve for x
- Estimate slope from two points by solving for m:

$$y_1 - y_2 = m(x_1 - x_2)$$

QUADRATIC FUNCTIONS

Standard form	$f(x) = ax^2 + bx + c$
Vertex or transformation form	$f(x) = a(x - h)^2 + k$
Quadratic eq.	$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

- $b^2 - 4ac > 0 \Rightarrow$ Two real unequal solutions
- $b^2 - 4ac = 0 \Rightarrow$ Repeated real solution
- $b^2 - 4ac < 0 \Rightarrow$ Two complex solutions

POLYNOMIAL FUNCTIONS

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^{n-k}$$

a_0	Constant term
a_n	Polynomial coefficient
$a_n x^m$	Term (m is called *degree*)

RATIONAL FUNCTIONS

Can be written as a quotient of two polynomials $P(x)$ and $Q(x)$:

$$f(x) = \frac{P(x)}{Q(x)} = \frac{\sum_{k=0}^{n} a_k x^{n-k}}{\sum_{k=0}^{n} b_k x^{n-k}}$$

- Horizontal Intercept is the inputs where the output is 0
- Vertical Intercept is where input is 0 (if defined)

ASYMPTOTES

The "line" a function approaches but never touches

vertical A vertical line $x = a$ where the graph tends towards positive or negative infinity as the inputs approach a:

$$x \to a,\ f(x) \to \pm\infty$$

horizontal A horizontal line $y = b$ where the graph approaches the line as the input gets larger:

$$x \to \pm\infty,\ f(x) \to b$$

ASYMPTOTES OF RATIONAL FUNCTIONS

vertical Where denominator = 0 but numerator $\neq 0$

horizontal Determined by respective degrees of numerator and denominator:

- Degree of denominator $>$ degree of numerator $\Rightarrow$ Horizontal asymptote at $y = 0$
- Degree of denominator $<$ degree of numerator $\Rightarrow$ No horizontal asymptote
- Degree of denominator = degree of numerator $\Rightarrow$ Horizontal asymptote is ratio of leading coefficients

EXPONENTIAL FUNCTIONS

Rate of change is as a percent, i.e., not an constant (absolute) rate. Takes the form:

$$f(x) = a(1+r)^x$$

or,

$$f(x) = ab^x,\ b = 1 + r$$

Can always be rewritten in terms of logarithms:

$$b^a = c \equiv \log_b c = a$$

continuous growth Use $e = 2.718282\ldots$ for continuous growth, often natural phenomena

$$f(x) = ae^{rx}$$

where

- $a \equiv$ initial quantity
- $r \equiv$ continuous growth rate

ALMOST ALWAYS USE CONTINUOUS e FORM!

Solve by:

1. Isolate exponential expression, where possible
2. Take log on both sides
3. Use *exponent property* of logs to pull variables of out exponent
4. Use algebra to solve for variable

GRAPHING EXPONENTIAL FUNCTIONS

- $a \equiv$ vertical intercept
  - $a > 0 \Rightarrow$ concave up
  - $a < 0 \Rightarrow$ concave down
- $b \equiv$ rate of growth
  - $b > 1 \Rightarrow$ growing
  - $0 < b < 1 \Rightarrow$ decaying
- Horizontal asymptote is where $y = 0$

LOGARITHMIC FUNCTIONS

The inverse of exponential functions; use to solve exponential functions. Commonly used to express quantities that vary widely in size. Form:

$$\log_b x$$

which can be rewritten in terms of exponents:

$$b^a = c \equiv \log_b c = a$$

inverse property of logs

$$\begin{aligned} \log_b b^x &= x \\ b^{\log_b x} &= x \end{aligned}$$

exponent property

$$\log_b A^r = r \log_b A$$

2 PROBABILITY

ASSUMPTIONS

$$0 \leqslant P(E) \leqslant 1$$
$$P(S) = 1$$
$$P(S) = \sum_{i=1}^{\infty} P(E_i)$$

Sample space S contains each event E_i, e.g., $E = \{$all outcomes in S starting with a 3$\}$

UNIONS AND INTERSECTIONS

$$E \cup F$$

is the union of the two sets E and F, i.e., the event where either E or F occurs. The intersection of two events, the outcomes in both E and F is

$$E \cap F.$$

Commutative	$E \cup F = F \cup E$
	$E \cap F = F \cap E$
Associative	$(E \cup F) \cup G = E \cup (F \cup G)$
	$(E \cap F) \cap G = E \cap (F \cap G)$
Distributive	$(E \cup F) \cap G = (E \cap G) \cap (F \cap G)$
	$(E \cap F) \cup G = (E \cup G)(F \cup G)$

INDEPENDENT EVENTS

Two events are independent if knowing the outcome of one provides no useful information about the outcome of the other.

MUTUALLY EXCLUSIVE EVENTS

$$P(A \cap B) = 0$$

are disjoint events, when A and B are mutually exclusive and there is no intersection—it is not possible for both to happen.

UNION AND ADDITION RULE

$$P(A \cup B) \equiv P(A \vee B) \equiv \{x : x \in A \vee x \in B\}$$

Independent	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Mutually exclusive	$P(A \cup B) = P(A) + P(B)$

INTERSECTION AND MULTIPLICATION RULES

$$P(A \cap B) \equiv P(A \wedge B) \equiv \{x : x \in A \wedge x \in B\}$$

Independent	$P(A \cap B) = P(A) \cdot P(B)$
Mutually exclusive	$P(A \cap B) = 0$
Dependent	$P(A \cap B) = P(A) \cdot P(B \vert A)$

COMPLEMENT AND SUBTRACTION RULE

$$\begin{aligned} P(A') \equiv P(A^c) \equiv P(\neg A) &\equiv \{x : x \notin A\} \equiv 1 - P(A) \\ P(A') &= 1 - P(A) \end{aligned}$$

Some implications:

$$P(A \cup A^c) = 1$$
$$P(A) = 1 - P(A^c)$$
$$P(A|B) = 1 - P(A^c|B)$$

CONDITIONAL PROBABILITY RULE

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

is the probability of the outcome of event A given condition B.

BAYES THEOREM

$$P(A|B) = \frac{P(A)\ P(B|A)}{P(A)\ P(B|A) + P(A)\ P(B|A)}$$

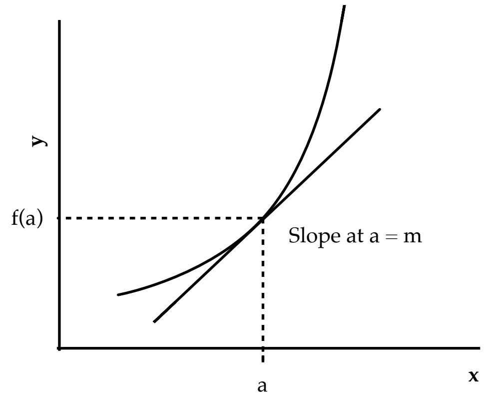
---

3 CALCULUS

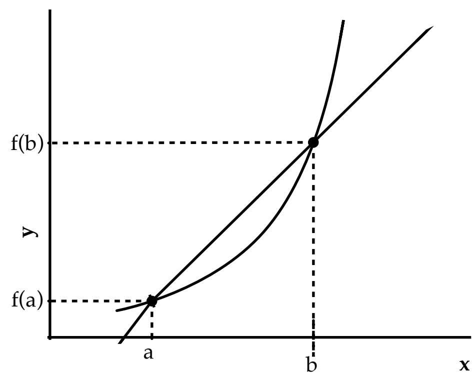
TANGENT AND SECANT LINES

- Tangent line is the line on $f(x)$ at point $x = a$ that just touches the graph of the function

- The slope of a non-linear function at one point



- Secant line is the line between A and B on a curve



Note: As A gets closer to B the secant slope approaches the tangent line

---

AVERAGE RATE OF CHANGE

$$\text{ARC} = f(a, a+\Delta) = \frac{\Delta y}{\Delta x} = \frac{f(a+\Delta x) - f(a)}{\Delta x}$$

over interval $[a, a+\Delta x]$

DERIVATIVE

The instantaneous rate of change of a function $f(x)$ at point $x = a$

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

"the derivative of y w.r.t. x" $\equiv \frac{dy}{dx} \equiv \lim_{\delta \to 0} \frac{\delta y}{\delta x}$

SIMPLE DERIVATIVES

y	y'
k	0
x	1
x^n	nx^{n-1}
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\sqrt[n]{x}$	$\frac{1}{n\sqrt[n]{x^{n-1}}}$
$\frac{1}{x}$	$\frac{-1}{x^2}$
e^x	e^x
a^x	$a^x \ln(a)$
x^x	$xx^{x-1} + x^x \ln(x)$
$\ln(x)$	$\frac{1}{x}$
$\log_a(x)$	$\frac{1}{x}\log_a(e)$

COMPOSITE DERIVATIVES

y	y'
u^n	$nu^{n-1}u'$
$\sqrt{u}$	$\frac{u'}{2\sqrt{u}}$
$\sqrt[n]{u}$	$\frac{u'}{n\sqrt[n]{u^{n-1}}}$
$\frac{1}{u}$	$\frac{-u'}{u^2}$
e^u	$e^u u'$
a^u	$a^u \ln(a)u'$
u^v	$vu^{v-1}u' + u^v \ln(u)v'$
$\ln(u)$	$\frac{u'}{u}$
$\log_a(u)$	$\frac{u'}{u}\log_a(e)$

DERIVATIVE OPERATIONS

Sum	$(f(x)+g(x))' = f'(x)+g'(x)$
Difference	$(f-g)'(x) = f'(x) - g'(x)$
Product	$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$
Quotient	$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}$
Chain rule	$(f(g))'(x) = f'(g(x))g'(x)$
Inverse	$(f^{-1})'(x) = \frac{1}{f'(x)}$

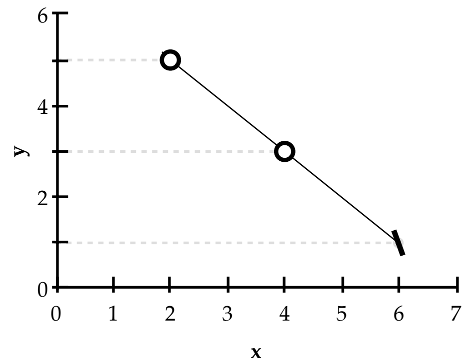
LIMITS

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

or, "the limit of $f(x)$, as x approaches c, is L." $\lim_{x\to c} f(x)$ is a *single number* that describes the behavior of the function $f(x)$ *near* but not *at* the point $x = c$.

Introduced to make calculating rate of change at 0 feasible, by making the Δ so infinitesimal the difference is between it and 0 is negligible—"allows" division by 0

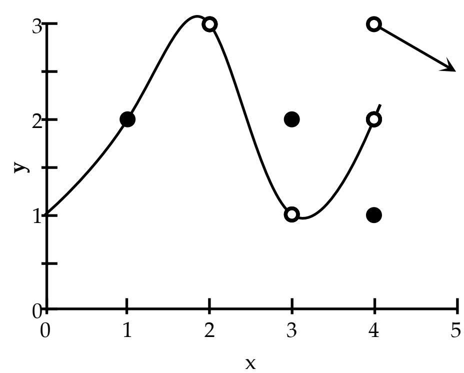
Example



where hollow points are undefined and solid points are defined

1. $\lim_{x\to 6} f(x) = 1$
2. $\lim_{x\to 4} f(x) = 3$
   - *Note*: even though $x = 4$ is undefined, we're only concerned with the area *around* 4, so we can still find the limit
3. $\lim_{x\to 2} f(x) = 5$

Example: Determining Limits of Non-Linear Functions



where hollow points are undefined and solid points are defined

1. $\lim_{x\to 1} f(x) = 2$—values where x is close to but not equal to 1 are near 2
2. $\lim_{x\to 2} f(x) = 3$—even though $f(2)$ is undefined, only values *near* $f(2)$ are important
3. $\lim_{x\to 3} f(x) = 1$—even though $f(3)$ is actually 2
4. $\lim_{x\to 4} f(x)$ = *does not exist*: Can't determine a single number because $f(4)$ from the right is about 2, and from the left about 3

Example: Determining Limits Using Algebra

Factor equation to simplest form and plug in c (assuming function at c is defined):

$$\lim_{x\to 5} \frac{x^2-6x+8}{x-4} = \frac{25-30+8}{1} = 3$$

LIMITS OF BROKEN FUNCTIONS

Some functions are continuous but in an unusual way—they appear "broken" when graphed—and so there is a *left* and *right* limit

left "The limit coming from the left"; values of $f(x)$ as $f(x)$ nears x and left of c, $x < c$

$$\lim_{x\to c^-} f(x) = L$$

right "This limit coming from the right"; values of $f(x)$ as $f(x)$ nears x and right of c, $x > c$

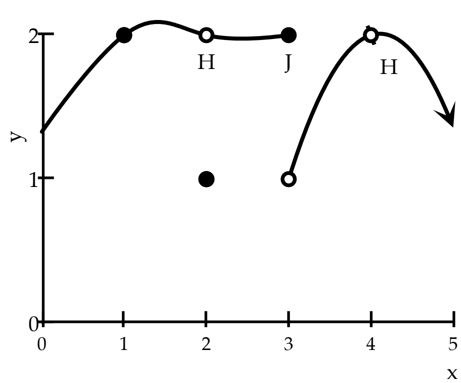
$$\lim_{x\to c^+} f(x) = L$$

Note: If left and rights limits are not the same, limit *doesn't exist*.

CONTINUITY

A function f is continuous at $x = a$ iff $\lim_{x\to a} f(x) = f(a)$, i.e., if the limit of x at a is equals $f(a)$, i.e., no breaks or jumps

Example



where H indicates a hole—where the graph is defined but could be made continuous by changing the point—and J a jump—where the left and right limits are not the same.

- Continuous at 1 since $\lim_{x\to 1} f(x) = f(1) = 2$
- Not continuous at 2, 3, or 4
  - $\lim_{x\to 2} f(x) = 2 \neq f(2) = 1$
  - $\lim_{x\to 3} f(x) = \text{doesn't exist} \neq f(3) = 2$
  - $\lim_{x\to 4} f(x) = 2 \neq f(4) = \text{undefined}$

CALCULATING DERIVATIVES

Example: Using Formal Definition

Find derivative of $f(x) = 2x^2 - 16x + 35$.

1. Assemble using the formal definition

$$\frac{\left[2(x-h)^2 - 16(x+h) + 35\right] - \left[2x^2 - 16x - 35\right]}{h}$$

2. Factor—*cannot plug $h = 0$ because no division by zero!*

$$= \frac{2x^2 + 4xh + 2h^2 - 16x - 16h + 35 - 2x^2 + 16x - 35}{h}$$
$$= \lim_{h\to 0} \frac{4xh + 2h^2 - 16h}{h}$$

3. Factor out h in numerator to cancel h in denominator

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{h(4x+2h-16)}{h} \\ &= \lim_{h\to 0} 4x + 2h - 16 \\ &= 4x - 16 \end{aligned}$$

IMPLICIT DIFFERENTIATION

The process to find $y' = f'(x)$ when $f(x)$ is difficult or impossible to use with explicit differentiation, by assuming y is a function of x:

Example

Implicitly differentiate $x^2 + y^2 = 25$

1. Differentiate each side, treating y as a function

$$\frac{d}{dx}\left(x^2 + y^2\right) = \frac{d}{dx}25$$
$$\Rightarrow \quad \frac{d}{dx}x^2 + \frac{d}{dx}y^2 = 0$$
$$\Rightarrow \quad 2x + 2y\frac{dy}{dx} = 0$$
$$\Rightarrow \quad 2x + 2yy' = 0$$

2. Algebraically solve for y'

$$
\begin{aligned}
2yy' &= -2x \\
\Rightarrow \quad y' &= \frac{-2x}{2y} \\
\Rightarrow \quad y' &= -\frac{x}{y}
\end{aligned}
$$

DEFINITE INTEGRAL

The definite integral of a positive function $f(x)$ over an interval $[a, b]$ is the area between f, the x-axis, $x = a$, and $x = b$.

$$\int_a^b f(x)\,dx$$

where a and b are the "limits of integration" and $f(x)$ is the integrand

$$
\begin{aligned}
\int_a^a f(x)\,dx &= 0 \\
\int_a^b f(x)\,dx &= -\int_b^a f(x)\,dx \\
\int_a^b f(x)\,dx &= \int_a^c f(x)\,dx + \int_c^b f(x)\,dx
\end{aligned}
$$

where $a < c < b$

INTEGRATION OPERATIONS

Sum	$\int u + v\,dx = \int u\,dx + \int v\,dx$
Difference	$\int u - v\,dx = \int u\,dx - \int v\,dx$
Product	$\int af(x)\,dx = a \int f(x)\,dx$
Parts	$\int u\,dv = uv - \int v\,du$
Substitution	$\int f(u)u'\,dx = \int f(u)\,du$

GROWTH, CONCAVITY, AND EXTREMA

growth

- $\forall x \in I\ f'(x) \geqslant 0 \Rightarrow f$ is increasing in I.
- $\forall x \in I\ f'(x) \leqslant 0 \Rightarrow f$ is decreasing in I.

concavity

- $\forall x \in I\ f''(x) \geqslant 0 \Rightarrow f$ is concave up in I.
- $\forall x \in I\ f''(x) \leqslant 0 \Rightarrow f$ is concave down in I.

extrema If $f'(a) = 0$ (critical point)

- $f''(a) < 0 \Rightarrow f$ has a local maximum at $x = a$.

Figure 1: Concave up (left) and concave down (right)

- $f''(a) > 0 \Rightarrow f$ has a local minimum at $x = a$.

INCREASING AND DECREASING FUNCTIONS

- $f(x)$ is *increasing* iff $\forall\ x_1, x_2$ in interval I is such that $x_1 < x_2$ and $f(x_1) < f(x_2)$
- $f(x)$ is *decreasing* iff $\forall\ x_1, x_2$ in interval I is such that $x_1 > x_2$ and $f(x_1) > f(x_2)$
- Determine all intervals where $f(x)$ is in/decreasing:
  1. Find all critical points (via first deriative)
  2. For each critical point, select a number a in that range and see if $f'(a)$ is positive or negative

---

INFLECTION POINTS

- Where the second derivative changes signs
- The point(s) on a graph where the concavity of a function changes from up to down
- functions can be increasing (positive derivative) or decreasing (negative derivative) regardless of concavity,

---

MAXIMA AND MINIMA

A *critical point* is a point where either $f'(a) = 0$ or $f'(a)$ is undefined, and is a *candidate* of being a local or global extreme

Local

maximum at a if $f(a) \geqslant f(x)\ \forall x$ near a

minimum at a if $f(a) \leqslant f(x)\ \forall x$ near a

extreme at a if $f(a)$ is a local maximum or minimum

Global

maximum at a if $f(a) \geqslant f(x)\ \forall x$ in domain of f

minimum at a if $f(a) \leqslant f(x)\ \forall x$ in domain of f

extreme at a if $f(a)$ is a global maximum or minimum

Example

Find the critical point of $f(x) = x^3 - 6x^2 + 9x + 2$

1. Find $f'(x)$

$$f'(x) = 3x^2 - 12x + 9$$
$$\begin{aligned} &= 3(x^2 - 4x + 3) \\ &= 3(x-1)(x-3) \end{aligned}$$

2. Find where $f'(x) = 0$, which is 1 and 3
3. Put $x = 1$ and $x = 3$ into $f(x)$ to find the critical points

$$\begin{aligned} (1, f(1)) &= (1, 6) \\ (3, f(2)) &= (3, 2) \end{aligned}$$

ANTIDERIVATIVES

- *An* antiderivative of a function $f(x)$ is any function $F(x)$ such that $F'(x) = f(x)$
- *The* antiderivative is an entire family of functions, written $F(x) + c$
- Also known as the *indefinite integral* (with no limit markers):

$$\int f(x)\ dx$$

Example

An antiderivative of $\int 2x\ dx$ is $x^2 - 5.2$; *the* antiderative is $x^2 + C$

DEFINITE V. INDEFINITE INTEGRALS

Indefinite integrals do not have limits to integration where definite integrals do

---

INTEGRATION BY SUBSTITUTION

A method to algebraically manipulate an integrand so it is amenable to antiderivative rules; especially useful when there is a product in the integral.
Substitute u for $g(x)$ where necessary, making $\frac{du}{dx} = g'(x)$, so $du = g'(x)\ dx$. Since

$$\frac{du}{dx} = g'(x) \equiv du = g'(x)\ dx$$

we can substitute so that

$$\int f'(g(x))g'(x)\ dx \equiv \int f'(u)\ du$$

Now integrate $f'(u)\ du$. (Note that $g(x) \equiv u$ and $g'(x)\ dx \equiv du$.)

1. Set one part of the integrand to u, one "level" into the integral
2. Compute $du = \frac{du}{dx}\ dx$ (the derivative of u)
3. Convert x's to u's in original integral, even including in dx
4. Integrate new u integral
5. Substitute u's back to x's in integral

Example

Integrate $\int (x+1)^3\ dx$

1. Rearrange so that $u = x + 1$ and $du = 1\ dx$

$$= \int (x+1)^3 \cdot 1\ dx$$

2. Substitute in u and du

$$\int u^3\ du$$

3. Integrate

$$= \frac{u^4}{4} + C$$

4. Add u back in

$$= \frac{(x+1)^4}{4} + C$$

INTEGRATION BY PARTS

Integrate a complex function by rewriting it as a product of two simpler functions u and du, using two possible forms:

$$\begin{aligned}\int u\ dv &= uv - \int v\ du \\ \int_a^b u\ dv &= uv|_a^b - \int_a^b v\ du\end{aligned}$$

3.1 *Example: First Form*

Integrate $\int xe^x\ dx$:

1. Break into two parts: $u = x$ and $dv = e^x\ dx$
2. Calculate the derivative of u, du, and v, the integral of dv

$$\begin{aligned}du &= \left(\frac{d}{dx}x\right)\ dx = 1\ dx \\ v &= \int dv = \int e^x\ dx = e^x\end{aligned}$$

3. Using the first formula, noting the prior forms from 1 and 2

$$\begin{aligned}\int u\ dv &= \int xe^x\ dx \\ &= xe^x - \int e^x\ dv \\ &= xe^x - e^x + C\end{aligned}$$

3.2 *Example: Second Form*

Integrate $\int_1^4 6x^2 \ln x\ dx$:

1. Break into two parts: $u = \ln x$ and $dv = 6x^2$
2. Calculate the derivative of u, du, and the integral of dv, v:

$$\begin{aligned}du &= \frac{d}{dx}\ln x = \frac{1}{x}\ dx \\ v &= \int 6x^2\ dx = 6\int x^2\ dx = 6 \cdot \frac{x^3}{3} = 2x^3\end{aligned}$$

3. Use the second formula

$$\begin{aligned}\int_1^4 6x^2 \ln x\ dx &= 2x^3 \ln x|_1^4 - \int_1^4 2x^3 \frac{1}{x}\ dx \\ &= 2x^3 \ln x|_1^4 - 3x^2|_1^4\end{aligned}$$

4. Find the integral the usual way:

$$\begin{aligned}&\left[(2\cdot 4^3 \ln(4)) - (2\cdot 1^3 \ln(1))\right] - \left[(3\cdot 4^2) - (3\cdot 1^2)\right] \\ &= 128 \cdot \ln(4) - 45 \\ &\approx 132.446\end{aligned}$$

ANTIDERIVATIVE RULES

$\int a\, dx$	$ax + C$
$\int x^n\, dx$	$\frac{u'}{2\sqrt{u}}$
$\int e^x\, dx$	$e^x + C$
$\int a^x\, dx$	$\frac{a^x}{\ln(a)} + C$
$\int \frac{1}{x}\, dx$	$\ln\lvert x\rvert + C$

---

4 STATISTICS

EXPECTED VALUE

$$E(X) = \mu = \sum_{i=1}^{k} x_i \; P(X = x_i)$$

for a discrete random variable with k possible values.

GENERAL VARIANCE FORMULA

$$Var(X) = \sigma^2 = \sum_{j=1}^{k} (x_j - \mu)^2 \; P(X = x_j),$$

or, the sum of the squared deviations $(x_j - \mu)^2$ weighted by the corresponding probabilities $P(X = x_1), \dots, P(X = x_k)$.

GENERAL STANDARD DEVIATION

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)}$$

LINEAR COMBINATIONS OF VARIABLES

$$Z = aX + bY$$

is a linear combination of the independent, random variables X and Y (often a and b are 1 or -1).

$$\begin{aligned} E(Z) &= a \times E(X) + b \times E(Y) \\ Var(Z) &= a^2 \times Var(X) + b^2 \times Var(Y) \end{aligned}$$

PROBABILITY DENSITY FUNCTION (PDF)

$$P(a \leqslant X \leqslant b) = \int_a^b f(x) \; dx$$

is a PDF of X, for any two numbers a and b where $a \leqslant b$. I.e., the probability that X takes on a value in the interval $[a, b]$ is the area above this interval and below the graph of the density curve.

- $P(X = c) = 0$ for any constant (bins are infinitesimally small)
- $\sum P(x_i) = 1$

NORMAL DISTRIBUTION V. STANDARD NORMAL

There is any entire family of distributions that can be called normal, but the prototypical distribution with mean of 0 and standard deviation of 1 is called the standard normal. Formally defined by its PDF as:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Properties

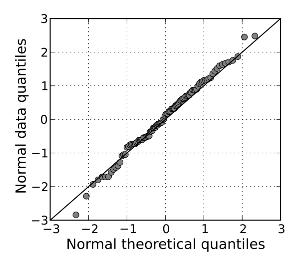
1. Symmetric around mean
2. Mean = mode = median
3. Denser at center than in tails

Consequently,

- 68 percent of distribution is within one standard deviation of the mean
- 95 percent of distribution is within approximately two standard deviations of the mean

---

EVALUATING NORMALITY



- A normal probability plot using quartiles can be used to evaluate how closely a given distribution adheres to normality, where the straight line is a perfect normal curve
- As N increases, the deviation from normality will decrease

Z SCORES

$$Z = \frac{x - \mu}{\sigma}$$

converts any value from a normal distribution to its corresponding value on the standard normal distribution

- Describes the number of standard deviations a point is from the mean μ
- Z scores to the left of μ are negative, and positive to the right of μ

Z SCORES: PROBABILITIES ON NORMAL DISTRIBUTION

Ex. What is the probability $X > A$, *given* $X \sim N(\mu = 1500, \sigma = 300)$?

$$Z = \frac{x - \mu}{\sigma} = \frac{1630 - 1500}{300} = 0.43$$

This is 0.6664 on Z table, so 66.64 percent of X is to the left of A so:

$$1 - 0.6664 = 0.3336$$

The probability $X > A$ is 33.36 percent.

Ex. Given $A = 1400$ *and* $X \sim N(\mu = 1500, \sigma = 300)$, *what is the percentile corresponding to A?*

$$Z = \frac{x - \mu}{\sigma} = \frac{1400 - 1500}{300} = -0.33$$

The corresponding value on the Z table is 0.3707, so A is the 37th percentile.

Ex. Given $p = .40$ *and* $X \sim N(\mu = 70, \sigma = 3.3)$, *what is the value corresponding to percentile* p?

Lookup p on Z table, getting a $Z = -0.25$. Work backwards:

$$-0.25 = Z = \frac{x - \mu}{\sigma} = \frac{x - 70}{3.3}$$

and solve for $x = 69.18$.

Ex. What is the probability X *is between* A *and* B, *given* $X \sim N(\mu, \sigma)$?

Using Z-scores method, find the area to the left of A and to the right of B, then $A - B = 1-$ area left of $A-$ are to right of B.

BERNOULLI DISTRIBUTION

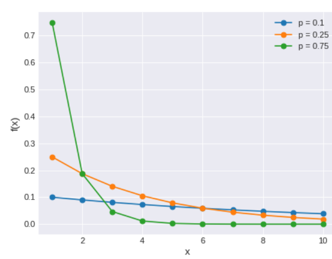
$$P(X = x) = \begin{cases} p \text{ for } x = 1 \\ 1 - p \text{ for } x = 0 \end{cases}$$

describes the distribution of individual trials with two possible outcomes, success or failure, described by proportion of successes $0 \leqslant p \leqslant 1$:

$$\begin{aligned} \hat{p} &= \frac{|\,\text{successes}\,|}{|\,\text{failures}\,|} \\ \mu &= p \\ \sigma^2 &= p(1-p) \end{aligned}$$

- The probability of success after n trials is $(1-p)^{n-1} \times p$

BERNOULLI: GEOMETRIC DISTRIBUTION

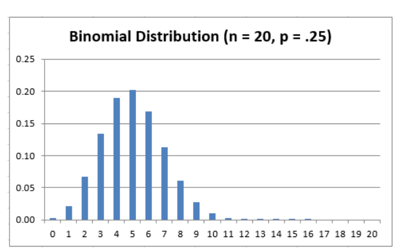


describes the wait time until a success for *independent* Bernoulli random variables; or, the probability of observing the k-th success by the n-th trial

$$\begin{aligned} \mu &= \frac{1}{p} \\ \sigma^2 &= \frac{1-p}{p^2} \end{aligned}$$

- Higher p means fewer trials until success
- Can never be approximated by a normal distribution

BINOMIAL DISTRIBUTION



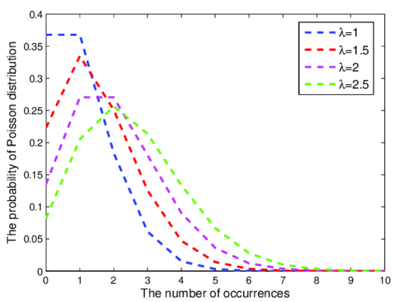
describes the probability of having exactly k successes in n independent Bernoulli trials (with probability of success p):

$$
\begin{aligned}
P(x = k \mid n, \mu, \sigma) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}
\end{aligned}
$$

Parameters, can be used to approximate to normal when n is sufficiently large and np and $n(1-p)$ are both greater than or equal to 10:

$$
\begin{aligned}
\mu &= np \\
\sigma^2 &= np(1-p)
\end{aligned}
$$

POISSON DISTRIBUTION



$$P(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Describes the number of events in a larger population over a unit of time with rate λ:

$$
\begin{aligned}
\mu &= \lambda \\
\sigma^2 &= \lambda
\end{aligned}
$$

INFERENTIAL STATISTICS

The body of thought governing the inferences of populations from samples, and how these sample statistics can vary.

STANDARD ERROR

$$SE = \frac{\sigma}{\sqrt{n}}$$

One-Sided	Two-Sided
$H_0 : x = A$	$H_0 : x = A$
$H_A : x > / < A$	$H_A : x \neq A$

is the standard deviation of distributions of sample statistics, when population σ is known. If it is unknown, and if $n > 30$, substitute sample standard deviation s

- SE decreases as n increases
- SE decreases as σ (or s) decreases

CONFIDENCE INTERVALS

$$\bar{x} \pm z \times SE$$

- $\bar{x}$ is the sample statistic, such as sample mean
- $z \times SE$ is the *margin of error*
- z is the desired confidence level, e.g., $z = 1.96$ for a 95 percent confidence interval

Interpretation. "We are Z percent confident the true population *statistic* is between A and B"; or, "Z percent of samples will have a *sample statistic* between A and B."

CENTRAL LIMIT THEOREM

Given a population with a finite mean μ and a finite non-zero variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean of μ and a variance of $\frac{\sigma^2}{N}$, as N, the sample size, increases—regardless of the shape of the parent population.

HYPOTHESIS TESTING

The process of comparing two point estimates, to determine if any difference between them is "real" or the result of natural variance in samples.

- *Type I* errors, or false positives, occur when the null hypothesis is true, but rejected
- *Type II* errors, or false negatives, occur when the alternative hypothesis is true, and the null hypothesis is not rejected

Quantifying Risk

- The risk of Type I errors is quantified by α, i.e., the probability the point estimate is more than z^* standard deviations away from the true population parameter
- The p-value is $1-\alpha$, the probability of observing data at least as favorable to the alternative hypothesis as the present data set, if the null hypothesis is true

One- and Two-Sided Hypotheses

SAMPLE STATISTICS: MEAN AND VARIANCE

- $\mu_M = \mu$ is the mean of the sampling distribution of means
- $\sigma_M^2 = \frac{\sigma^2}{N}$ is the variance of the sampling distribution of the mean
- $\sigma_M = \frac{\sigma}{\sqrt{N}}$ is the standard error of the sampling distribution of the mean

As N increases, variance of sample mean decreases

SAMPLE STATISTICS: DIFFERENCE IN MEAN

Two samples from a population the size n_1 and n_2, calculate the means M_1 and M_2, and the difference is $M_1 - M_2$

$$\begin{aligned}
\mu_{M_1-M_2} &= M_1 - M_2 \\
\sigma^2_{M_1-M_2} &= \sigma^2_{M_1} + \sigma^2_{M_2} \\
\sigma_{M_1-M_2} &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
\end{aligned}$$

When variance and sample size are the same, standard error becomes:

$$\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}i} = \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = \sqrt{\frac{2\sigma^2}{n}}$$

If $n_1 \neq n_2$ then variance becomes:

$$\sigma^2_{M_1-M2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

What is the probability that the mean of sample 1 will exceed that of sample 2 by N *or more?*

1. Find mean: $\mu_{M_1-M_2} = M_1 - M_2$
2. Find standard error: $\sigma_{M_1-M_2}$
3. Find area underneath distribution of sample 1 to the right of the mean of sample 2 plus N

SAMPLE STATISTICS: r AND ρ

- Not normally distributed—right-skewed—because correlation cannot exceed 1
- As ρ increases, the more right-skewed the distribution

SAMPLE STATISTICS: PROPORTION π

Sampling proportion is closely related to the binomial distribution—the total number of successes—where p is the distribution of the mean number of successes

$$\begin{aligned}
\mu_p &= \pi \\
\sigma_p &= \frac{\sqrt{N\pi(1-\pi)}}{N} = \sqrt{\frac{\pi(1-\pi)}{N}}
\end{aligned}$$

Find probability p *is greater than* A

Given N and population proportion π:

1. Find mean of $p = \pi$
2. Calculate standard error as above
3. Conduct as normal distribution given N is sufficently large and π is not too close to 0 or 1

ESTIMATION

The process of estimating population parameters from sample statistics. Usually results in a point estimate as well as interval estimates called confidence intervals.

DEGREES OF FREEDOM

4.1 *Chapter 3 questions*

What percent of the standard normal distribution is found in Z ¿/¡ A? A: Use pnorm() or a Z table. If A is negative and we want to know Z ¿ A, we know the percent is greater than 15. If —Z— ¡ 0.5 – the bounds are -.5 and 0.5, so Z(.5) - Z(-.5).
Normal distributions described as $m\ N(\mu, \sigma)$.
Problem 3.4. Given two distributions, and two points in each, which does each compare relative to their groups? 1. Convert both points into Z scores, and compare them. You can also convert the Z-scores into probabilities to say the percent they beat/were beaten by.
Geometric examples:
https://www.redbrick.dcu.ie/~minisham/CA/Notes/CA266/10_Geometric_Distribution.pdf
Problem No. 3.18. 68-95-99.7% rule.
Problem No. 3.22. Given a probability distribution of a sequence, what is the probability the Xth item meets some condition (number of coin tosses til heads, number of products produced until defect, etc.)? Use geometric distribution.
Problem No. 3.22.d. What is the probability the sequence produces no defects after n? Uses factorials
Problem No. 3.22.c. How many in the sequence before the first defect? This is mean or expected value of geometric distribution.
Problem No. 3.38. Binomial model to calculate 2/3 kids will be boys.
Problem No. 3.42. Probability on the Nth trial will have the Mth success? Solved similar as above with binomial model? Referring to a probability distribution, not a discrete event (like gambler's fallacy).

4.2 *Chapter 4 Questions*

Problem No. 4.4e. Measure uncertainty around a point estimate by computing a 95% confidence interval