Logic

Sentence - statement that can be determined to be either true or false; $3 > 2$
Sentential Function - contains a variable which means it may or may not be true (e.g., $x > 2$); can be transformed into sentence by substituting value for variable
Quantifiers
- Universal - "for all", $\forall$
- Existential - "there exists", $\exists$

Negation - "not", ~
Conjunction - "and", Λ, both parts have to be true; e.g., $3 > 2 \Lambda 3$ is an integer is true
Disjunction, "or", V, one part has to be true; e.g., $3 > 2$ V 3 is negative is true
- Inclusive - no restrictions; both parts can be true: $(p \vee q) \vee (p \Lambda q)$
- Exclusive - if one part is true, both parts can't be true: $(p \vee q) \Lambda \sim(p \Lambda q)$

Implications - "if... then"... $p \Rightarrow q$
- Antecedent - p
- Consequent - q

p	q	$p \Rightarrow q$
T	T	T
T	F	F
F	T	T*
F	F	T*

* Can't prove anything if antecedent is false

- Sufficient - if $(p \Rightarrow q)$ is true, we know if p is true, then q is true so p is sufficient for q
- Necessary - if $(p \Rightarrow q)$ is true, if we know q is true, we can't say anything about p; q is necessary for p, but not sufficient
- Necessary and Sufficient - $p \Leftrightarrow q$; "if and only if", "iff"; means $(p \Rightarrow q) \Lambda (q \Rightarrow p)$

Truth Table - look at every possible combination

p	q	$p \Lambda q$	$p \vee q$	$p \Leftrightarrow q$	~p
T	T	T	T	T	F
T	F	F	T	F	F
F	T	F	T	F	T
F	F	F	F	F	T

Laws - sentences that are always true; used to derive theorems
- Identity - $\forall p, p \Rightarrow p$
- Syllogism - transitivity; $(\forall\ p,q)\ ((p \Rightarrow q) \Lambda (q \Rightarrow r)) \Rightarrow (p \Rightarrow r)$
- Noncontradiction * - $\sim(p \Lambda \sim p)$
- Excluded Middle * - $p \vee \sim p$; no such thing as half true or half false
- Tautology - $p \Lambda p \Rightarrow p$
  $p \vee p \Rightarrow p$
- Commutative - $p \Lambda q \Leftrightarrow q \Lambda p$
  $p \vee q \Leftrightarrow q \vee p$
- Associative - $p \Lambda (q \Lambda r) \Leftrightarrow (p \Lambda q) \Lambda r$; $p \vee (q \vee r) \Leftrightarrow (p \vee q) \vee r$
- Distributive - $[p \Lambda (q \vee r)] \Leftrightarrow [(p \Lambda q) \vee (p \Lambda r)]$
  $[p \vee (q \Lambda r)] \Leftrightarrow [(p \vee q) \Lambda (p \vee r)]$
- DeMorgan's * - $\sim(p \Lambda q) \Leftrightarrow (\sim p \vee \sim q)$
  $\sim(p \vee q) \Leftrightarrow (\sim p \Lambda \sim q)$
- * used to do proof by contradiction
  - Example: $\forall\ p,q\ \ p \Lambda q \Rightarrow p$; use truth table to show this statement is a law (always true)

p	q	$p \wedge q$	$p \wedge q \Rightarrow p$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	T

$p \Rightarrow q$

- Converse - $q \Rightarrow p$
- Inverse - $\sim p \Rightarrow \sim q$
- Contrapositive - $\sim q \Rightarrow \sim p$
- $\sim(p \Rightarrow q) \Leftrightarrow p \wedge (p \Rightarrow \sim q)$
- Double Negation - $\sim\sim p \Rightarrow p$
- Original statement and contrapositive are equivalent (as are converse and inverse)

Rules of Inference

- Substitution - $\forall x, \exists y \; x + y = 5$; plug in any constant for x and statement is still true
- Detachment - $p \Rightarrow q$, if you know p is true, then q has to be true

Proof - start with true statement and use various laws to show your desired conclusion is true

- Deductive - look only at logical rules and structure of argument (truth depends on truth of initial assumption)
- Contradiction - assume ~p and try to get contradiction which implies assumption is wrong so p must be true
- Construction - use rules of inference to get p without making assumptions; also called direct proof
- Inductive - look at past and draw conclusions; not used in formal proofs

Assumptions must be:

- Independent - assumptions shouldn't interfere with each other; should be able to prove an assumption from subset of assumptions
- Consistent - if assumptions prove p, they shouldn't be able to prove ~p
- Complete - should be able to prove true or false; if you can't, something is missing
- Exists - can prove it's true, but you may need to prove it exists

To prove $p \Rightarrow q$

- a) Assume p, show q (conditional proof)
- b) Assume ~q, show ~p (contrapositive proof)

To prove $p \Leftrightarrow q$

- a) Show $p \Rightarrow q$ and $q \Rightarrow p$
- b) Show $p \Rightarrow q$ and $\sim p \Rightarrow \sim q$
- c) Show $p \Rightarrow a \Rightarrow q \Rightarrow b \Rightarrow p$
- d) Show $p \Leftrightarrow a \Leftrightarrow b \Leftrightarrow q$

To prove $\forall x$, p

- Pick an arbitrary x and show p

To prove $\exists x$, p

- a) Proof by contradiction (assume $\forall x$, ~p and find some x where ~p is false)
- b) Construct p from some x

Proof by Induction - (not accepted by everyone)
- Show $\forall n, p_n$
- Start by showing p_1
- Show $\forall k \, (P_k \Rightarrow P_{k+1})$

Example: sum of first n odd numbers if n^2
- If $n = 1$, $1 = 1^2$
- Assume $1 + 3 + 5 + ... + (2k-1) = k^2$
- $1 + 3 + 5 + ... + (2k - 1) + 2k + 1 =? (k + 1)^2$
- $k^2 + 2k + 1 = (k + 1)^2$... this is true

Example: the largest number n* is 1
- Assume n* > 1, then n*2 > n* so n* can not be greater than 1
- Do same thing for n* < 1
- Valid proof, but antecedent assumes largest number exists... with false antecedent, you can prove anything

Real Numbers

Natural Numbers - all positive whole numbers; 1, 2, 3, ...
Integers - zero and all positive and negative whole numbers; 0, ±1, ±2, ...
Rational Numbers - can be written a/b, where a & b are integers
Irrational Numbers - can't be written as rational numbers (e.g., sqrt(2), π)
Real Numbers - include both rational and irrational numbers

Intervals

(a,b) - open interval; any number between a and b, excluding a and b ($a < x < b$)
$[a,b]$ - closed interval; any number between a and b, including end points ($a \le x \le b$)
$(a,b]$ - half open; any number between a and b, only includes one end point ($a < x \le b$)
$[a,b)$ - $a \le x < b$

Absolute Value

$$|a| = \begin{cases} a & \text{if } a \ge 0 \\ -a & \text{if } a < 0 \end{cases}$$

$|x| \le a \Rightarrow -a \le x \le a$

Powers

$a^n = a \cdot a \cdot ... \cdot a$ (n factors)
$a^{-n} = 1/a^n$
$a^r \cdot a^s = a^{r+s}$
$(a^r)^s = a^{rs}$
$a^{1/2} = \sqrt{a}$ (valid if $a \ge 0$)
$\sqrt{ab} = \sqrt{a}\sqrt{b}$
$\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$
$a^{1/n} = \sqrt[n]{a}$
$a^{p/q} = (a^{1/q})^p = (a^p)^{1/q} = \sqrt[q]{a^p}$ (p an integer, q a natural number)
$(abc)^s = a^s \cdot b^s \cdot c^s$

Algebra Rules

$a + b = b + a$
$(a + b) + c = a + (b + c)$
$a + 0 = a$
$a + (-a) = 0$
$ab = ba$
$(ab)c = a(bc)$
$1 \cdot a = a$
$aa^{-1} = 1$ (for $a \neq 0$)

$(-a)b = a(-b) = -ab$
$(-a)(-b) = ab$
$a(b + c) = ab + ac$
$(a + b)c = ac + bc$
$(a + b)^2 = a^2 + 2ab + b^2$
$(a - b)^2 = a^2 - 2ab + b^2$
$(a + b)(a - b) = a^2 - b^2$

Inequalities

$a > b$ and $b > c \Rightarrow a > c$

$a > b$ and $c > 0 \Rightarrow ac > bc$

$a > b$ and $c < 0 \Rightarrow ac < bc$

$a > b$ and $c > d \Rightarrow a + c > b + d$

$a > b \Rightarrow -a < -b$ (direction of inequality reversed if both sides multiplied by negative number)

Relations

Relations (R = "relates to")

x R y (x is at least as good as y) and y R x (y is at least as good as x)

Preferences (P = "preferred to")

x P y means $y \not\mathrm{P}\, x$

$\{(x,y) : x \text{ R } y\}$... order is important; designates preferences

Sets

Set - grouping or collection of objects, $S = \{a, b, c\}$ or S = {typical member : defining properties}, e.g., $B = \{(x,y) : px + qy \leq m, x \geq 0, y \geq 0\}$

Element - an object in a set; also called members; $x \in S$

Subsets - $A \subseteq B$ if every member of A is also a member of B

Proper Subset - $A \subset B$ means $A \subseteq B$ and $B \not\subset A$)

Equal Sets - each element of set A is an element of set B and each elopement of B is an element of A; i.e., $A = B \Leftrightarrow A \subset B$ and $B \subset A$

Null Set ($\emptyset$) - empty set; set with no elements; $\emptyset \subset$ of all sets

Universal set (U) - set that contains everything; can define your own universal set in the context of what you're doing (e.g., can only have 0 and 1 in U)

Complement (S^C) - $x \in S^C$ iff $x \notin S$; sometime written as $S^C{}_U$, complement of set S with respect to universal set U

Union - elements that belong to at least one of the sets A and B; like AND in logic

$A \cup B = \{x : x \in A$ or $x \in B\}$

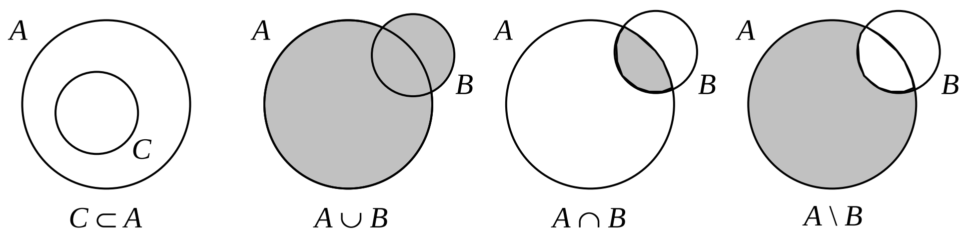
Intersection - elements that belong to both A and B; like OR in logic

$A \cap B = \{x : x \in A$ and $x \in B\}$

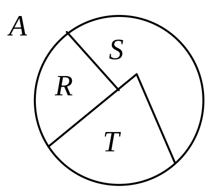
Disjoint Sets - $A \cap B = \emptyset$

Minus - elements that belong to A, but not to B

$A \setminus B = \{x : x \in A$ and $x \notin B\} = \{x : x \in (A \cap B^C)\}$



Partition - sets form a partition if they are pairwise disjoint; the union of the partitions makes up the original set and the intersection of any two partitions is $\emptyset$; used for probabilities in econometrics



R, S, T partition A if:

$R \cup S \cup T = A$

$R \cap S = R \cap T = S \cap T = \emptyset$

Product - gives ordered pairs $(x,y) \in S \times T$ such that $x \in S$ and $y \in T$; e.g., $S = \{1,2\}$ and $T = \{3,4\}$, then $S \times T = \{(1,3), (1,4), (2,3), (2,4)\}$

Rules

$S \cup S = S$

$S \cap S = S$

$S \cup SC = U$

$S \cap SC = \varnothing$ (law of contradiction)

$(S \cap T) \cap Z = S \cap (T \cap Z) = (S \cap Z) \cap T$ (commutative law)

$S \cap (A \cup B) = (S \cap A) \cup (S \cap B)$ (distributive law)

If $S \subset T$ and $T \subset V$ then $S \subset V$ (transitivity)

Any two sets must relate to each other in one and only one of the following ways:

1) Identical ($S = T$)
2) Disjoint ($S \cap T = \varnothing$)
3) $S \subset T$ (proper subset)
4) $T \subset S$ (proper subset)
5) $S \cap T \neq \varnothing$, but $S \not\subset T$ and $T \not\subset S$ (intersect, but neither is a subset of the other)

Functions

Function - rules that assigns a unique real number, y, to each number, x

$y = f(x)$ ("y is a function of x")

$f: x \to y$ ("function f maps x to y")

x = independent variable, exogenous variable, or argument of the function

y = dependent variable, endogenous variable

Domain - all values of x for which the function gives a meaningful value

Range - the set of values that a function assumes

$y = x^2$ is a function

y = sqrt(x^4) is not a function

Graphs - shows all ordered pairs (x,y) that satisfy the function

Implicit Functions - determined by some property

$x^2 + y^2 = 27$

Explicit Functions - y is explicitly defined as a function of the independent variable

y = sqrt($27 - x^2$)

Constant Functions - y doesn't change based on x; horizontal line on a graph

Linear Functions - $y = ax + b$

Slope - $a = (y_2 - y_1)/(x_2 - x_1)$

y-Intercept - b

Linear Inequalities - $B = \{(x,y) : px + qy \leq m\}$

Increasing - $\forall\ x_1, x_2, x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$

Weakly Increasing - use ≥ instead of >; also called non-decreasing

Decreasing - $\forall\ x_1, x_2, x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$

Weakly Decreasing - use ≤ instead of <; also called non-increasing

Correspondence - domain points are mapped to a subset of the range

Solving Equations - find all values of the variables for which the equation is satisfied

Two Techniques - do the following to both sides of the equality

(a) add (or subtract) the same number

(b) multiply (or divide) by the same number (≠ 0)

Parameters - constants

Polynomials

Quadratic Formula - for $b^2 - 4ac \geq 0$ and $a \neq 0$, $ax^2 + bx + c = 0 \iff x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

n^{th} Order - $a_n^{xn} + a_{n-1}x^{n-1} + ... + a_1x + a_0$; can have up to n distinct roots

Summation

$$\sum_{i=p}^{q} a_i = a_p + a_{p+1} + \cdots + a_q$$

$$\sum_{i=1}^{n} (a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i \qquad \text{(additivity property)}$$

$$\sum_{i=1}^{n} c a_i = c \sum_{i=1}^{n} a_i \qquad \text{(homogeneity property)}$$

Useful Formulas

$$\sum_{i=1}^{n} i = 1 + 2 + \cdots + n = \frac{1}{2} n(n+1)$$

$$\sum_{i=1}^{n} i^2 = \frac{1}{6} n(n+1)(2n+1)$$

$$\sum_{i=1}^{n} i^3 = \left[\frac{1}{2} n(n+1)\right]^2$$

$(a^n - b^n) = (a - b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + ... + b^{n-1})$

$$(a+b)^m = a^m + \binom{m}{1} a^{m-1} b + \cdots + \binom{m}{m-1} a b^{m-1} + \binom{m}{m} b^m \qquad \text{(Newton's binomial formula)}$$

$$\binom{m}{k} = \frac{m!}{(n-k)!k!}, \quad \binom{m}{0} = 1, \quad \binom{m}{1} = m, \quad \binom{m}{m} = 1$$

Limits and Continuity

Sequence - function from positive integers to real numbers; e.g., $x_n = n+1$, $n = 1, 2, ...$
Limit - sequence has a limit if it converges to a real number L

$\lim_{n\to\infty} x_n = L$

Example - $x_n = 1/n$; $\lim_{n\to\infty} x_n = 0$

Rigorous Definition

$f(x)$ has limit (or tends to) A as x tends to a, and write $\lim_{x\to a} f(x) = A$, if for each number $\varepsilon > 0$ there exists a number $\delta > 0$ such that $|f(x) - A| < \varepsilon$ for every x with $0 < |x - a| < \delta$

In English - $\lim_{x\to a} f(x) = A$ means that we can make $f(x)$ as close to A as we want for all x sufficiently close to (but not equal to) a

Using It - if asked to use the definition to prove a limit exists, you first assume any $\varepsilon > 0$ and solve $|f(x) - A| < \varepsilon$ for x. Then use $0 < |x - a| < \delta$ to get a value for δ in terms of ε.

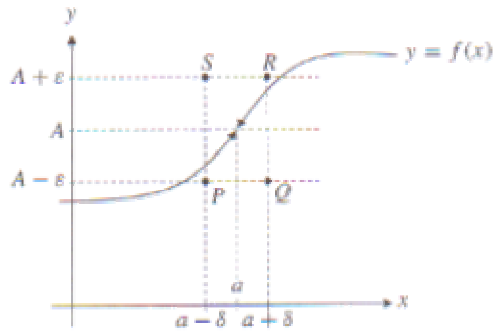


Figure 6 For every ϵ, there is a δ, so $\lim_{x\to a} f(x) = A$

Theorems

1) If a sequence $\{x_n\}$ is non-decreasing ($\forall\ n\ x_{n+1} \geq x_n$) and bounded from above ($\exists\ L$ s.t. $L \geq x_n\ \forall\ n$), then the sequence $\{x_n\}$ must converge

2) If a sequence $\{x_n\}$ is non-increasing ($\forall\ n\ x_{n+1} \leq x_n$) and bounded from below ($\exists\ L$ s.t. $L \leq x_n\ \forall\ n$), then the sequence $\{x_n\}$ must converge

3) If a sequence is not monotonic, but has bounds, then the sequence may not converge, but it has convergent subsequences

Example: x_n = 1 when n is even and -1 when n is odd; can't find a δ to satisfy definition for ε < 1, but the subsequences are bounded at 1 and -1

3a) f all the subsequences have the same limit, then the sequence has a limit

Example: $x_n = 1/n$ when n is even and $-1/n$ when n is odd; converges to 0

4) If a sequence is not bounded, it will diverge (but can't say a sequence that is bounded necessarily converges... see #3)

Rules for Limits

If $\lim_{x\to a} f(x) = A$ and $\lim_{x\to a} g(x) = B$, then

a) $\lim_{x\to a} A = A$

b) $\lim_{x\to a} (f(x) \pm g(x)) = A \pm B$

c) $\lim_{x\to a} (f(x) \cdot g(x)) = A \cdot B$

d) $\lim_{x\to a} (f(x) / g(x)) = A / B$ (if $B \neq 0$)

e) $\lim_{x\to a} (f(x))^{p/q} = A^{p/q}$ (if $A^{p/q}$ is defined)

f) If functions f and g are equal for all x close to a (but not necessarily at $x = a$), then $\lim_{x\to a} f(x) = \lim_{x\to a} g(x)$ whenever either limit exists

Special Cases

Don't Exist - vertically asymptotic functions ($\pm\infty$)

One-Sided - value depends on which side you approach the limit from

Infinite Limits - horizontally asymptotic functions

Vector Notation

$x = (x_1, x_2, \ldots , x_k)$

Sequence of vectors - converge when $|L - x_n|$ gets smaller

Euclidean Distance - $d(x,y) = \text{sqrt}((x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots)$

Taxi Distance - $d(x,y) = |x_1 - y_1| + |x_2 - y_2| + \ldots$

?? Distance - $d(x,y) = \max(|x_1 - y_1|, |x_2 - y_2|, \ldots)$

Neighborhood

Neighborhood of x is a region around that point with certain distance, $\aleph(x,\varepsilon) = \{y: d(x,y) < \varepsilon\}$...

i.e., a circle centered on x with radius ε

Limit Point - number x is a limit point of a set S if every ε neighborhood of x contains a point of S other than x

Finite sets never have limit points

Interior Point - x is interior to S if $\exists\ \varepsilon > 0$, such that if $y \in \aleph(x,\varepsilon)$ then $y \in S$ (i.e., $\aleph(x,\varepsilon) \subset S$)

Every interior point is a limit point, but not the other way around

Open Set - S is an open set if every element is an interior point

Closed Set - a set is closed if it contains all of its limit points (points on border are limit points of open sets even though the points aren't in the set)

Special Cases - only two sets can be both open and closed at the same time ($\varnothing$ and U)

If S is open, then S^C is closed

Limits of Functions (using neighborhoods)

$f(x)$ has limit L at a if for each number $\varepsilon > 0$ there exists a number $\delta > 0$ such that if $x \in \aleph(a,\delta)$ then $f(x) \in \aleph(L,\varepsilon)$

Continuity

Continuous - graph of the function has no breaks; formal definition:

f is continuous at $x = a$ if $\lim_{x\to a} f(x) = f(a)$

Conditions:

1) function f must be defined at $x = a$
2) the limit of $f(x)$ as x tends to a must exist
3) this limit must be exactly equal to $f(a)$

If only condition 1 isn't satisfied, it is a "removable" discontinuity

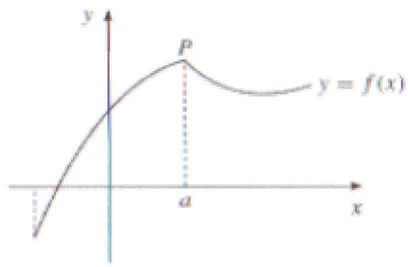


Figure 1 A continuous function

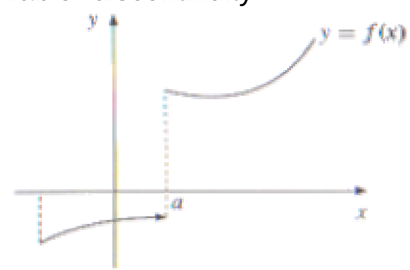


Figure 2 A discontinuous function

Some continuous functions

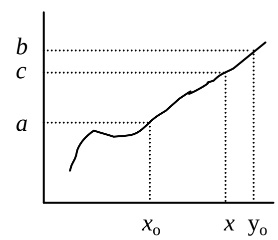
$f(x) = c$ (a constant)

$f(x) = x$

Polynomials (they're a sum of continuous functions)

$R(x) = P(x)/Q(x)$ (where $P(x)$ and $Q(x)$ are polynomials and $Q(x) \neq 0$)

Intermediate Value Theorem



Let f be a continuous function for all x. Let $f(x_o) = a$ and $f(y_o) = b$ where $a < b$, then for any c between a and b, $\exists$ x between x_o and y_o such that $f(x) = c$

Proof 1 (outline) -

a) Create two sequences by if $f[(x_o + y_o)/2] < c$ then $x1 = (x_o + y_o)/2$, else $y1 = ($

b) Show the sequences converge at c

Proof 2 (outline) -

a) Define $A = \{x\colon f(x) \geq c$ and $x_o \leq x \leq y_o\}$ and $B = \{x\colon f(x) \leq c$ and $x_o \leq x \leq y_o\}$

b) Show A and B are closed sets

c) Show $A \cap B \neq \varnothing$

d) $\exists\, x \in A \cap B$... $f(x) \leq c$ and $f(x) \geq c$ so $f(x) = c$

Properties of Continuous Functions

If f and g are continuous at a, then

a) $f + g$ and $f - g$ are continuous at a

b) $f \cdot g$ and f / g (if $g(a) \neq 0$) are continuous at a

c) $[f(x)]^{p/q}$ is continuous at a if $[f(x)]^{p/q}$ is defined

d) $f(g(x))$ is continuous at a if both $f(x)$ and $g(x)$ are continuous at a (composites)

Limits of Continuous Functions

Just plug in value rather than taking the limit

Continuity and Differentiability

If f is differentiable at $x = a$, then f is continuous at $x = a$

Differentiation

Derivative - use to describe rate of change (to study of how quickly quantities change over time)

Slope - $f'(a)$ = slope of the tangent to the curve $y = f(x)$ at the point $(a,f(a))$

Secant - straight line connecting two points on the graph of a function

Tangent - limiting straight line toward which the secant tends as you hold one point constant and move the other one closer; say first point is $(a,f(a))$ and second point is $(a + h, f(a + h))$; find slope of tangent by taking limit as $h \to 0$ (eqn below)

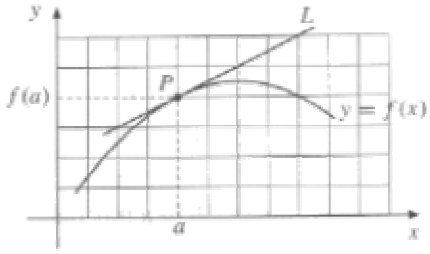


Figure 1 $f'(a) = 1/2$

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

Newton (differential) quotient of f

Equation for tangent at $(a,f(a))$ is

$$y - f(a) = f'(a)(x - a)$$

Derivative only exists at a point if there is a unique tangent (i.e., no kinks)

Hard Way (using definition of tangent to compute a derivative)

1) Add h ($h \neq 0$) to a and compute $f(a + h)$
2) Compute the corresponding change in the function value: $f(a + h) - f(a)$
3) For $h \neq 0$, form the Newton quotient (eqn above)
4) Simplify the fraction as much as possible; should cancel h from the denominator
5) Take limit of fraction as $h \to 0$

Example:

$f(x) = x^2$

$$\frac{f(a+h) - f(a)}{h} = \frac{(a+h)^2 - a^2}{h} = \frac{a^2 + 2ah + h^2 - a^2}{h} = \frac{h^2 + 2ah}{h} = \frac{h(h+2a)}{h} = h + 2a$$

$$f'(a) = \lim_{h\to 0} h + 2a = 2a$$

Notation

$$f'(x) = \frac{dy}{dx} = dy/dx = \frac{df(x)}{dx} = df(x)/dx = \frac{d}{dx} f(x)$$

Increasing & Decreasing (revisited)

$f'(x) \geq 0\ \forall\ x \in$ interval $I \Leftrightarrow f$ is increasing in I (strictly increasing if > 0)

$f'(x) \leq 0\ \forall\ x \in$ interval $I \Leftrightarrow f$ is decreasing in I (strictly decreasing if < 0)

$f'(x) = 0\ \forall\ x \in$ interval $I \Leftrightarrow f$ is constant in I

Other Interpretations

Rate of Change - change in y per unit change in x

Instantaneous rate of change of f at a is $f'(a)$; (e.g., anything labeled "marginal" such as "marginal cost")

Relative (or proportional) rate of change of f at a is $f'(a) / f(a)$; usually quoted as a percentage per unit time

Continuity

Continuity is necessary but not sufficient for derivatives to exist

If a function is differentiable then it is continuous (diff $\Rightarrow$ continuity, but not continuity $\Rightarrow$ diff)

Rules for Differentiation

$f(x) = A$ (constant) $\Rightarrow f'(x) = 0$

$y = A + f(x) \Rightarrow y' = f'(x)$ (additive constants disappear)

$y = Af(x) \Rightarrow y' = Af'(x)$ (multiplicative constants are preserved)

$f(x) = x^a \Rightarrow f'(x) = ax^{a-1}$ (power rule)

$F(x) = f(x) \pm g(x) \Rightarrow F'(x) = f'(x) \pm g'(x)$

$F(x) = f(x) \cdot g(x) \Rightarrow F'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$

$$F(x) = \frac{f(x)}{g(x)} \Rightarrow F'(x) = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{\left(g(x)\right)^2}$$

$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$ (chain rule)

$y = u^a \Rightarrow y' = au^{a-1}u'$ (generalized chain rule)

$F(x) = f(g(x)) \Rightarrow F'(x) = f'(g(x)) \cdot g'(x)$ (still the chain rule)

Examples:

$f(x) = (2x + 1)^2$, let $a = 2x + 1$ so $f(x) = a^2$

$f'(x) = f'(a^2) \cdot g'(2x + 1) = 2a \cdot (2) = 4a = 4(2x + 1)$

$z = a(x^3 + 2x)^3 + 4(x^3 + 2x)$, let $y = x^3 + 2x$ so $z = ay^3 + 4y$

$dz/dx = dz/dy \cdot dy/dx = (3ay^2 + 4) \cdot (3x^2 + 2) = 3a(x^3 + 3x)^2 + 4)(3x^2 + 2)$

$f(x) = ((2x + 1)^5 + 2(2x + 1)^4 + 3)(2x^2 + 1)$, let $y = 2x + 1$

$f'(x) = [5(2x + 1)^4(2) + 8(2x + 1)^3(2)](2x^2 + 1) + 4x((2x + 1)^5 + 2(2x + 1)^4 + 3)$

Implicit Differentiation

$F(x,y) = c$.... implicit function g allows you to write y in terms of x (i.e., $y = g(x)$ so $F(x,g(x)) = c$)

If two variables x and y are related by an equation, to find y':

a) Differentiate each side of the equation with respect to x, considering y as a function of x. (Usually, you'll need the chain rule)

b) Solve the resulting equation for y'

Example: $y^3 + 3x^2y = 13$

a) $3y^2y' + (6xy + 3x^2y') = 0$

b) $y' = -2xy/(x^2 + y^2)$

Example: $m(x, y) = x\sqrt{y} = 2$

Let f(x) = y

$m(x, f(x)) = x\sqrt{f(x)} = 2$

$$\text{diff both sides wrt } x: 1 \cdot \sqrt{fx)} + x \cdot \frac{1}{2\sqrt{f(x)}} f'(x) = 0$$

$$\text{mult both sides by } 2\sqrt{f(x)}:\ 2f(x) + x \cdot f'(x) = 0$$

$$f'(x) = -2f(x)/x = -2y/x$$

Example: $x^2 - 12x + y^2 + 14y + 85 = 0$

diff both sides: $2x - 12 + 2y\cdot(dy/dx) + 14\cdot(dy/dx) = 0$

$(2y + 14)(dy/dx) = 12 - 2x$

$dy/dx = (6 - x)/(y + 7)$

Check for validity:

Rewrite it as $x^2 - 12 + 36 + y^2 + 14y + 49 = 0 = (x - 6)^2 + (y + 7)^2 = 0$

Only solution is $x = 6$ and $y = -7$

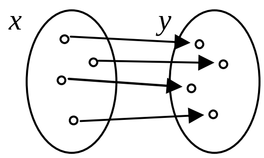
Substitute that solution into $dy/dx = (6 - 6)/(-7 + 7) = 0/0$... so dy/dx is not valid

(If there is only one point that is a solution, you have an infinite number of tangents so there isn't a valid derivative.)

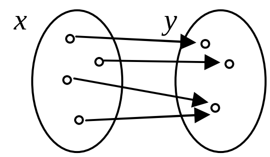
If it was $x^2 - 12x + y^2 + 14y + 84 = 0$, you can get $(x - 6)^2 + (y + 7)^2 = 1$, a circle with center (6,-7) and radius 1. You can differentiate that so now $dy/dx = (6 - x)/(y + 7)$ is valid

if it was $x^2 - 12x + y^2 + 14y +86 = 0$, you get $(x - 6)^2 + (y + 7)^2 = -1$... a function with no solution so dy/dx again is meaningless.

Inverse Functions
$y = f(x)$, maps x to y
Inverse function $f^{-1}(y) = x$, maps y to x, but only exists if $f(x)$ is one-to one



1-to-1



Not 1-to-1

NOTE: $f^{-1}(x) \neq \dfrac{1}{f(x)}$, write that as $\left(f(x)\right)^{-1}$

Finding inverse - swap variables in function (i.e., use y for x and x for y) and solve for x

Theorem: If f is <u>continuous</u> and <u>strictly increasing</u> (or strictly decreasing) in an interval I, then f has an inverse function g, which is continuous and strictly increasing (strictly decreasing) in the interval $f(I)$. If x_0 is an interior point of I and $f'(x_0) \neq 0$, then g is differentiable at $y_0 = f(x_0)$ and $g'(y_0) = 1/f'(x_0)$

Given $y = f(x)$ and $f^{-1}(y) = x$, we have identities $y \equiv f(f^{-1}(y))$ and $x \equiv f^{-1}(f(x))$ so you can take a derivative and solve for derivative of f^{-1}

$x \equiv f^{-1}(f(x))$

Differentiate both sides: $1 = (f^{-1})'(f'(x)$

Solve for $(f^{-1})'$: $(f^{-1})' = 1/f'(x)$

Example

$f(x) = 2x + 5$... $f'(x) = 2$ so $(f\text{-}1)' = 1/2$

Check it: $f\text{-}1(x) = (x - 5)/2$... $(f\text{-}1)' = 1/2$

Approximations (when x is close to a)

$f(x) \approx f(a) + f'(a)\cdot(x - a)$ (Linear)

$f(x) \approx f(a) + f'(a)\cdot(x - a) + 1/2\, f''(a)\cdot(x - a)^2$ (Quadratic)

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$ (Polynomial or Taylor)

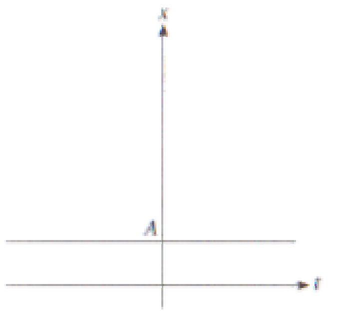
Exponents and Logarithms

Exponential Functions

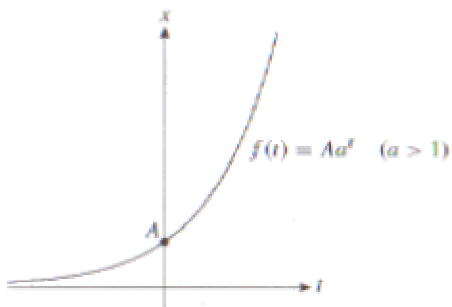
Quantity that increases (or decreases) by a fixed factor per unit of time is said to increase (or decrease) exponentially; if fixed factor is a, then

$$x = Aa^t$$

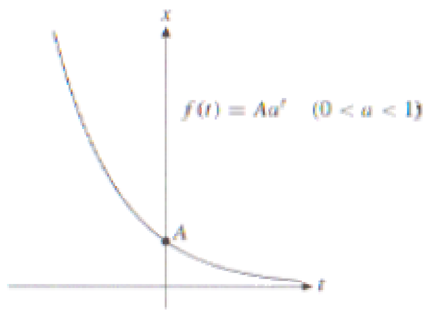
Variable (factor or base) is a
Power is t
Cases of a:
- $a = 1$ means x is constant
- $a > 1$, x is increasing
- $a < 1$, x is decreasing



Graph of $f(t) = Aa^t$ $(a = 0)$



Graph of $f(t) = Aa^t$ $(a > 1)$



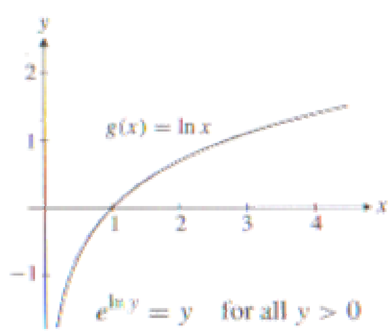
Graph of $f(t) = Aa^t$ $0 < a < 1)$

Logarithms

Logarithms is inverse of exponential function

$$y = \log_b x \quad (x > 0)$$

If you leave out the b, assumption is $\log_{10}$



Rules

$b^x b^y = b^{(x+y)}$

$b^x / b^y = b^{x-y}$

$(b^x)^y = b^{xy}$

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

$\log_b x/y = \log_b x - \log_b y$

$\log_b x^k = \mathrm{k}\cdot\log_b x$

$\log_b b^x = x$

$$\log_b x = \frac{\log_c x}{\log_c b} = \frac{\ln c}{\ln b}\cdot\frac{\ln x}{\ln c} = \frac{\ln x}{\ln b}$$

$\log_b c = 1/\log_c b$

**** NOTE: $\log_b (x + y) \neq \log_b x + \log_b y$ ****

Why use it:

Questions of form $a^x = b$ (e.g., "at present rate of inflation, how long will it take the price level to triple", growth, compound interest)

Solving $b^y = a^x$

Take log of both sides: $\log_b b^y = \log_b a^x \Rightarrow y = x\log_b a$

Cobb-Douglas functions: $y = x^a \cdot z^b$

Take log of both sides: $\ln y = \ln(x^a \cdot z^b) = a \cdot \ln x + b \cdot \ln z$ (easier to work with)

Elasticities - if $y = f(x)$, $\varepsilon_{y,x}$: elasticity of y with repect to x

$$\varepsilon_{y,x} = \frac{x}{y} \cdot \frac{dy}{dx} = \frac{x}{f(x)} \cdot f'(x)$$

$$\frac{dy/y}{dx/x} = \frac{\%\text{ change in } y}{\%\text{ change in x}}$$

$$\frac{d\ln y}{d\ln x} = \frac{d\ln y}{dy} \cdot \frac{dy}{dx} \cdot \frac{dx}{d\ln x} = \frac{x}{y} \cdot \frac{dy}{dx}$$

Natural Logarithm

$e = \lim_{n\to\infty}(1 + 1/n)^n = \sum_{j=0}^{\infty} \frac{1}{j!} \approx 2.7182818...$ (irrational number)

$\ln = \log_e$

Special Cases

$\ln 1 = 0$

$\ln e = 1$

$e^{\ln x} = x \quad (x > 0)$

$\ln e^x = x$

$$\log_b x = \frac{\ln x}{\ln b}$$

Derivatives

$$\frac{d}{dx}(\ln x) = \frac{1}{x} \qquad \frac{d}{dx}(\ln f(x)) = \frac{f'(x)}{f(x)}$$

$$\frac{d}{dx}(e^x) = e^x \qquad \frac{d}{dx}\left(e^{f(x)}\right) = e^{f(x)} \cdot f'(x)$$

Logarithmic Differentiation

Sometimes easier to use $\ln$ or e to take derivatives

Examples:

$y = \log_b x$... use conversion $\log_b x = \dfrac{\ln x}{\ln b}$

$$\frac{dy}{dx} = \frac{d\left(\frac{\ln x}{\ln b}\right)}{dx} = \frac{1}{\ln b} \cdot \frac{d(\ln x)}{dx} = \frac{1}{\ln b} \cdot \frac{1}{x}$$

$y = b^x$.... take $\ln$ of both sides: $\ln y = x \ln b$

$$\frac{d(\ln y)}{dy} = \frac{dx}{dy} \cdot \ln b = \frac{1}{y}$$

Use inverse function rule

$$\frac{dy}{dx} = y \cdot \ln b = b^x \ln b$$

Examples using the chain rule:

$$y = e^{e^x}$$

Let $u = e^x$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = e^u \cdot e^x = e^{\left(e^x + x\right)}$$

$y = \ln(\ln(\ln x)))$

Let $z = \ln x$, $w = \ln z$, $y = \ln w$

$$\frac{dy}{dx} = \frac{dy}{dw} \cdot \frac{dw}{dz} \cdot \frac{dz}{dx} = \frac{1}{w} \cdot \frac{1}{z} \cdot \frac{1}{x} = \frac{1}{\ln z} \cdot \frac{1}{\ln x} \cdot \frac{1}{x} = \frac{1}{\ln(\ln x)} \cdot \frac{1}{\ln x} \cdot \frac{1}{x}$$

$$y = \left(ax^2 + bx\right) \cdot e^x \cdot \ln\left(x^5 + 3x\right)$$

Take $\ln$ of both sides because $\ln(ABC) = \ln A + \ln B + \ln C$ (must have $a, b, x > 0$)

$$z = \ln y = \ln\left(ax^2 + bx\right) + x \ln e + \ln\left[\ln\left(x^5 + 3x\right)\right]$$

$$\frac{dz}{dx} = \frac{2ax + b}{ax^2 + bx} + 1 + \frac{\frac{5x^4 + 3}{x^5 + 3x}}{\ln\left(x^5 + 3x\right)}$$

Proofs

Prove: $\frac{d}{dx}\left(\ln x\right) = \frac{1}{x}$

$$\frac{d(\ln x)}{dx} = \lim_{h \to 0} \frac{\ln(x+h) - \ln x}{h} = \lim_{h \to 0} \frac{\ln\left(\frac{x+h}{x}\right)}{h} = \lim_{h \to 0} \frac{\ln(1 + h/x)}{h}$$

Multiply by $1 = x/x$

$$\lim_{h \to 0} \frac{1}{x} \cdot \frac{x}{h} \ln\left(1 + \frac{1}{x/h}\right)$$

Now substitute $m = x/h$; note limit $h \to 0$ is equivalent to $m \to \infty$

(Could use L'Hopital's Rule instead)

$$\lim_{m \to \infty} \frac{1}{x} \cdot m \ln\left(1 + \frac{1}{m}\right) = \frac{1}{x} \lim_{m \to \infty} \ln\left(1 + \frac{1}{m}\right)^m$$

Because $\ln$ is a continuous function, rewrite as

$$\frac{1}{x}\ln\left(\lim_{m\to\infty}\left(1+\frac{1}{m}\right)^{m}\right)=\frac{1}{x}\ln e=\frac{1}{x}$$

Prove: $\frac{d}{dx}\left(\ln f(x)\right)=\frac{f'(x)}{f(x)}$

Use chain rule (same as proof for $e^{f(x)}$)

Prove: $\frac{d}{dx}\left(e^{x}\right)=e^{x}$

$$\frac{d\left(e^{x}\right)}{dx}=\lim_{h\to 0}\frac{e^{x+h}-e^{x}}{h}=\lim_{h\to 0}\frac{e^{x}\left(e^{h}-1\right)}{h}$$

Use the inverse function rule: $y = e^x$, take $\ln$ of both sides: $\ln y = \ln e^x = x\ln e = x$

Take derivative of both sides with respect to y

$$\frac{d\ln y}{dy}=\frac{dx}{dy}$$

Use derivative of $\ln$

$$\frac{1}{y}=\frac{dx}{dy}$$

Use inverse function rule:

$$\frac{dy}{dx}=y \quad \text{so} \quad \frac{\mathrm{d}\left(\mathrm{e}^{x}\right)}{dx}=e^{x}$$

Prove: $\frac{d}{dx}\left(e^{f(x)}\right)=e^{f(x)}\cdot f'(x)$

$$y=e^{f(x)}$$

Version 1:

Take $\ln$ of both sides: $\ln y = f(x)$

Take derivative of both sides: $d(\ln y) = d(f(x))$

$$\frac{1}{y}dy = f'(x)\cdot dx \Rightarrow \frac{dy}{dx}=y\cdot f'(x)=e^{f(x)}\cdot f'(x)$$

Version 2: (chain rule)

Let $u = f(x)$, so $y = e^u$

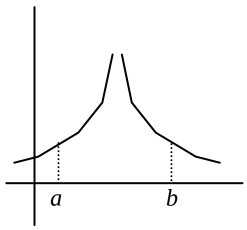
$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx}=e^{u}\cdot f'(x)=e^{f(x)}\cdot f'(x)$$

L'Hopital's Rule (and EVT & MVT)

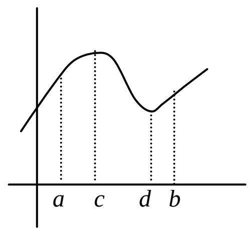
Extreme Value Theorem

Aim is to find points in domain at which function reaches its max and min values

- Max - $c \in D$ (domain) is a max iff $f(x) \leq f(x)\ \forall\ x \in D$
- EVT - if f is a <u>continuous</u> function in a <u>closed</u> and bounded interval $[a,b]$, then f attains a maximum and a minimum value on $[a,b]$... proof is really hard



Not continuous (EVT doesn't apply); can find min, but no max



max is $f(c)$
min is $f(d)$

Theorem 7.4 (Necessary First-Order Condition)

Let f be defined in an interval I and let x_0 be an interior point of I. If x_0 is a max or min point and $f'(x_0)$ exists, then $f'(x_0) = 0$

Proof:

Assume x_0 is a max point interior to I and $f'(x_0)$ exists.
If absolute value of h is sufficiently small, $(x_0 + h) \in I$ because x_0 is an interior point
Since x_0 is a max, $f(x_0 + h) \leq f(x_0)\ \forall\ h$

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0,\ \forall\ h > 0$$

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0,\ \forall\ h < 0$$

Therefore $f'(x_0) = 0$
Proof for min is similar

Rolle's Theorem

Let $f(x)$ be continuous on the interval $[a,b]$ and $f(a) = f(b) = 0$. If f is differentiable on (a,b), then $\exists\ c \in (a,b)$ such that $f'(c) = 0$

Proof:

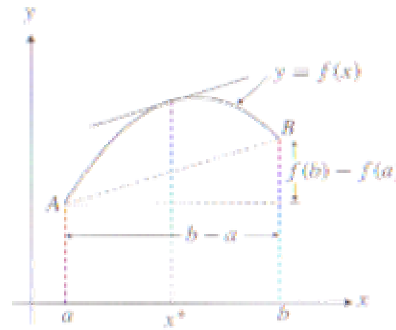
Case 1: $f(x) = 0\ \forall\ x \in [a,b]$. f is constant so $f'(x) = 0$

Case 2: $\exists\ x_0 \in (a,b)$ such that $f(x_0) > 0$ and $f'(x_0) \neq 0$ (can't assume "then" portion of theorem)

From EVT, f should take a maximum value $x^* \in [a,b]$
$f(x^*) \geq f(x_0)$, and $x^* \neq a$ and $x^* \neq b$ (i.e., x^* not on the boundary)
Since x^* is an interior point and is a max, then $f'(x^*) = 0$ (Thm 7.4)

Case 3: $\exists\ x_0 \in (a,b)$ such that $f(x_0) < 0$ and $f'(x_0) \neq 0$ (similar to case 2)

Mean Value Theorem



Let f be a continuous function on $[a,b]$ and have a finite derivative at every $x \in (a,b)$ (i.e., function is differentiable), then $\exists\, c \in (a,b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

(i.e., $f'(c)$ = slope of line connecting a and b)

Proof: consider the function

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Observe $g(a) = g(b) = 0$

g is differentiable and continuous because it is a function of f (which is differentiable and continuous)

$\exists\, c \in (a,b)$ such that $g'(c) = 0$ (by Rolle's Theorem)

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

Therefore at c

$$g'(c) = 0 = f'(x) - \frac{f(b) - f(a)}{b - a}, \text{ so } f'(c) = \frac{f(b) - f(a)}{b - a}$$

L'Hopital's Rule

Application of MVT used to examine limit when both numerator and denominator tend to zero

If (1) $f(x)$ and $g(x)$ are continuously differentiable (derivatives are also continuous);

(2) $f(x_0) = g(x_0) = 0$ *** also works for $f(x_0) = g(x_0) = \infty$ or combinations thereof ***

(3) $g'(x) \neq 0$ in some neighborhood of x_0

(4) $\lim_{x \to x_0} \frac{f'(x)}{g'(x)}$ exists

Then

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

Proof:

Let $y = g(x)$ and $z = f(x)$; these functions define a curve $z = h(y)$

$z = f(x) = h(g(x))\ \forall\, x$

Take derivative of both sides and use chain rule on right side

$$\frac{df}{dx} = \frac{dh}{dg} \cdot \frac{dg}{dx} \Rightarrow \frac{dh}{dg} = \frac{f'(x)}{g'(x)}$$

Consider any x_n near x_0 so $h(y_n) = z_n$ (i.e., point on the curve)

By MVT $\exists\, p \in (0,y_n)$ such that $h(y_n) = h(0) + h'(p)(y_n - 0)$ (i.e., slope from 0 to y_n equals $y'(p)$)

Therefore, $h'(p) = h(y_n)/y_n = f(x_n)/g(x_n)$ since $y_n = g(x_n)$ and $h(y_n) = z_n = f(x_n)$

$$h'(p) = \frac{f'(x(p))}{g'(x(p))} = \frac{f(x_n)}{g(x_n)}$$ (right side from line above; middle when x_n close to x_0)

Note as $x_n \to x_0$ then $x(p) \to x_0$ since $x(p) \in (x_n,x_0)$

$$\lim_{x_n \to x_0} \frac{f(x_n)}{g(x_n)} = \lim_{p \to 0} \frac{f'(x(p))}{g'(x(p))} = \lim_{x_n \to x_0} \frac{f'(x_n)}{g'(x_n)}$$

Using L'Hopital's Rule
Usually need it when you're taking the limit of something to a power

$\lim_{m\to\infty}(1+f(m))^m = ?$ where $\lim_{m\to\infty} f(m) = 0$
Trick used in these types of problems is:
$\lim_{m\to\infty}(1+f(m))^m = \lim_{m\to\infty} e^{\ln[1+f(m)]^m} = e^{\lim_{m\to\infty}\ln[1+f(m)]^m}$
Look at the exponent first:
$\lim_{m\to\infty}(m\cdot\ln[1+f(m)]) = \lim_{m\to\infty}\dfrac{\ln[1+f(m)]}{1/m}$ (little trick to get L'Hopital's Rule to work)
Use L'Hopital's Rule, first check assumptions:
1) f & g continuously differentiable... in this case $f = \ln[1+f(m)]$, $g = 1/m$
2) $f(x_0) = g(x_0) = 0$ (where x_0 is the limit; ∞ in this case)
   Since $\lim_{m\to\infty} f(m) = 0$ (given), then $f(x_0 = \infty) = \ln(1+0) = \ln(1) = 0$
   $g(x_0 = \infty) = 0$
3) $g'(x) \neq 0$... need to check after applying rule
4) $\lim_{x\to x_0}\dfrac{f'(x)}{g'(x)}$ exists... need to check after applying rule

So now $\lim_{m\to\infty}\dfrac{\ln[1+f(m)]}{1/m}$ = limit of ratio of derivatives (L'Hopital's Rule)

$$= \lim_{m\to\infty}\frac{\dfrac{f'(m)}{1+f(m)}}{\dfrac{-1}{m^2}} = \lim_{m\to\infty} -m^2\frac{f'(m)}{1+f(m)}$$

Solve this limit and plug the exponent back into eqn to get answer

<u>Example</u> using $f(m) = 1/\sqrt{m}$

Note $f'(m) = \dfrac{-1}{2}m^{-3/2}$

$$\lim_{m\to\infty}\ln\left[1+\frac{1}{\sqrt{m}}\right]^m = \lim_{m\to\infty} -m^2\frac{\dfrac{-1}{2}m^{-3/2}}{1+m^{-1/2}} = \lim_{m\to\infty}\frac{1}{2}\frac{\overbrace{m^{1/2}}^{\infty}}{1+\underbrace{\dfrac{1}{\sqrt{m}}}_{0}} = \infty$$

so now $e^{\infty} = \infty$

<u>Example</u> using $f(m) = 1/m$

$$\lim_{m\to\infty}\left(1+\frac{1}{m}\right)^m = \lim_{m\to\infty} e^{\ln(1+1/m)^m} = e^{\lim_{m\to\infty}\ln(1+1/m)^m}$$

Look at exponent

$$\lim_{m\to\infty} \ln(1+1/m)^m = \lim_{m\to\infty} m\ln(1+1/m)^m = \lim_{m\to\infty} \frac{\ln(1+1/m)}{1/m}$$

Use L'Hopital's Rule

$$\lim_{m\to\infty} \frac{\ln(1+1/m)}{1/m} = \lim_{m\to\infty} \frac{\frac{-1}{m^2}\Big/(1+1/m)}{-1/m^2} = \lim_{m\to\infty} \frac{1}{1+1/m} = 1$$

Check validity of assumptions 3 & 4 (good)

So now $e1 = e$

0

<u>Example</u> using $f(m) = 1/m^2$

$$\lim_{m\to\infty}\left(1+\frac{1}{m^2}\right)^m = \lim_{m\to\infty} e^{\ln\left(1+\frac{1}{m^2}\right)^m} = e^{\lim_{m\to\infty}\ln\left(1+\frac{1}{m^2}\right)^m}$$

Look at exponent:

$$\lim_{m\to\infty} \ln\left(1+\frac{1}{m^2}\right)^m = \lim_{m\to\infty} m\ln\left(1+\frac{1}{m^2}\right) = \lim_{m\to\infty} \frac{\ln\left(1+\frac{1}{m^2}\right)}{1/m}$$

0

Use L'Hopital's Rule

$$\lim_{m\to\infty} \frac{\frac{-2}{m^3}\Big/\left(1+\frac{1}{m^2}\right)}{\frac{-1}{m^2}} = \lim_{m\to\infty} \frac{2/m}{1+1/m^2} = \frac{0}{1+0} = 0$$

0

Plug exponent back in

$e^0 = 1$

<u>Example</u> using Constant Elasticity of Substitution (CES)

Q = $\left[\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right]^{-1/\rho}$

Find $\lim_{\rho\to 0} Q$

$$\lim_{\rho\to 0} e^{\ln Q} = \lim_{\rho\to 0} e^{\ln\left[\alpha K^{-\rho}+(1-\alpha)L^{-\rho}\right]^{-1/\rho}} = e^{\lim_{\rho\to 0}\ln\left[\alpha K^{-\rho}+(1-\alpha)L^{-\rho}\right]^{-1/\rho}}$$

Look at exponent

$$\lim_{\rho\to 0} \frac{-1}{\rho}\ln\left[\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right]$$

Use L'Hopital's Rule

$$-\lim_{\rho\to 0} \frac{\ln\left[\alpha K^{-\rho} + (1-\alpha)L^{-\rho}\right]}{\rho} = \frac{\frac{-\alpha K^{-\rho}\ln K - (1-\alpha)L^{-\rho}\ln L}{\alpha K^{-\rho} + (1-\alpha)L^{-\rho}}}{1}$$

all = 1

$$\lim_{\rho\to 0} \frac{\alpha K^{-\rho}\ln K + (1-\alpha)L^{-\rho}\ln L}{\alpha K^{-\rho} + (1-\alpha)L^{-\rho}} = \frac{\alpha\ln K + (1-\alpha)\ln L}{\alpha + (1-\partial)} = \alpha\ln K + (1-\alpha)\ln L$$

Plug exponent back in

$$e^{\alpha \ln K + (1-\alpha)\ln L} = e^{\ln K^{\alpha} + \ln L^{1-\alpha}} = e^{\ln K^{\alpha} L^{1-\alpha}} = K^{\alpha} L^{1-\alpha}$$

NOTE: if $y = K^{-\rho}$

Use $\ln y = -\rho \ln K$

Differentiate both sides

$$\frac{1}{y}\frac{dy}{d\rho} = -\ln K$$

$$\frac{dy}{d\rho} = -K^{-\rho} \ln K$$

<u>Example</u>

$\lim_{x\to 0} \frac{x^2}{1+x} = \frac{0}{1+0} = 0$ (Don't need to use L'Hopital's rule)

<u>Example</u>

Consider $\quad y = a^x \quad a > 1$

$\qquad\qquad z = x^t \quad t > 0$

Both y and z go to ∞ as x → ∞, but which one is faster?

$$\lim_{x\to\infty} \frac{x^t}{a^x} = e^{\lim_{x\to\infty} \ln\left(\frac{x^t}{a^x}\right)} = e^{\lim_{x\to\infty}(t \ln x - x \ln a)}$$

Look at exponent

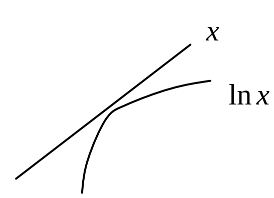
$\ln\left(\lim_{x\to\infty} \frac{x^t}{a^x}\right)$ (more elegant, but violates assumption 2 of L'Hopital's Rule

$$\lim_{x\to\infty} \ln(t \ln x - x \ln a) = \lim_{x\to\infty} \frac{x}{x}(t \ln x - x \ln a) = \lim_{x\to\infty} x\left[\frac{t \ln x}{x} - \ln a\right] =$$

$$\lim_{x\to\infty} x \cdot \lim_{x\to\infty}\left[\frac{t \ln x}{x} - \ln a\right] = \infty \cdot \lim_{x\to\infty}\left[\frac{t \ln x}{x} - \ln a\right] = \infty \cdot \left(\lim_{x\to\infty} \frac{t \ln x}{x} - \lim_{x\to\infty} \ln a\right) =$$

$$\infty\left(t \lim_{x\to\infty} \frac{\ln x}{x} - \ln a\right)$$

Solve limit



$\lim_{x\to\infty} \frac{\ln x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0$, using L'Hopital's Rule... makes sense from graph:

$$\infty\left(t \lim_{x\to\infty} \frac{\ln x}{x} - \ln a\right) = \infty(t \cdot 0 - \ln a) = \infty(-\ln a) = -\infty$$

Plug exponent back in

$e^{-\infty} = 0$ so a^x is faster

Trick:

$\lim_{x \to x_0} f(x) - g(x) = \lim_{x \to x_0} f(x) \cdot \lim_{x \to x_0} g(x)$ (if f(x) is continuous)

Undefined terms:

- $\infty \cdot 0$, $\infty \cdot \infty$
- ∞/∞, $0/\infty$, $\infty/0$, $0/0$
- $\infty \pm \infty$

Aside:

$x^2 = x \cdot x = (x + x + ... + x)$... there are x number of x's

Take derivatives

$2x = (1 + 1 + ... + 1)$... there are x number of 1's so

$2x = x$

$2 = 1$

Now can show all positive integers are = 1

$3 = 2 + 1 = 1 + 1 = 2 = 1$

Can show $0 = 1$

$1 = 2 - 1 = 2 - 2 = 0$

Can show fractions = 1

$1/n = 1/1 = 1$

Problem was derivative of product is not equal to derivative of sum... that's why we have the multiplication rule for taking derivatives

$a = b$

$ab = b^2$

$ab - a^2 = b^2 - a^2$

$a(b - a) = (b + a)(b - a)$

$a = b + a$

$a = 2a$

$1 = 2$

Problem... divided by $(b - a)$, but can't do that because $a = b$ so $(b - a) = 0$ (division by zero)

Higher Order Derivatives

First derivative is slope of function

Second derivative is slope of first derivative $\quad f''(x) = \lim_{h \to 0} \dfrac{f'(x+h) - f'(x)}{h}$

n-th derivative is slope of (n-1)th derivative $\quad f^{(n)}(x) = \lim_{h \to 0} \dfrac{f^{(n-1)}(x+h) - f^{(n-1)}(x)}{h}$

<u>Example</u>

$$f(x) = \frac{x+7}{x-3}$$

$$f'(x) = \frac{1 \cdot (x-3) - (x+7) \cdot 1}{(x-3)^2} = \frac{-10}{(x-3)^2}$$

$$f''(x) = 20(x-3)^{-3}$$

Graphing

$y' \geq 0 \Rightarrow$ y is increasing

$y'' \geq 0 \Rightarrow$ y is convex  (think of the "v" in convex)

$y'' \leq 0 \Rightarrow$ y is concave  (think of a cave)

Inflection Point

$y'' = 0$ and y'' changes sign; changes from convex to concave or vice versa

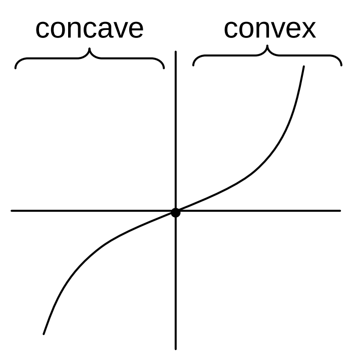
<u>Example</u>

$y = x^3$

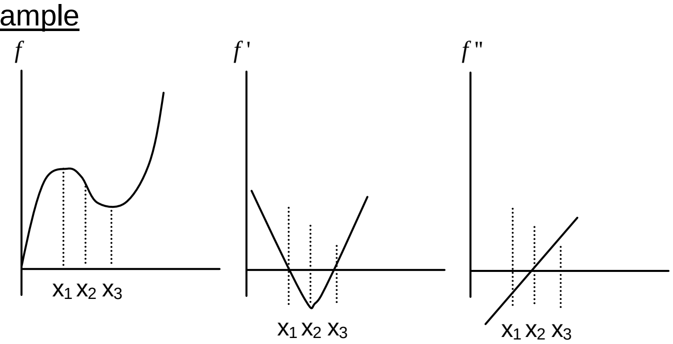
$y' = 3x^2 > 0 \; \forall$ x; y' = 0 $\Leftrightarrow$ x = 0

$y'' = 6x > 0$ for x > 0; y'' < 0 for x < 0; y'' = 0 $\Leftrightarrow$ x = 0

$y''' = 6$ (constant)



<u>Example</u>



Min and Max

If y' = 0 we have a local min if y'' > 0 (i.e., convex)

If y' = 0 we have a local max if y'' < 0 (i.e., concave)

Taylor Series Approximations

$y = f(z)$ and $z = g(x)$

Use chain rule

$$\frac{dy}{dx} = \frac{df}{dz} \cdot \frac{dz}{dx} = f'(z) \cdot g'(x)$$

Use multiplication rule

$$\frac{d^2 y}{dx^2} = \frac{d(f'(z))}{dx} \cdot g'(x) + f'(z) \cdot \frac{d(g'(x))}{dx} = f''(z)g'(x) + f'(z) + g''(x)$$

Taylor Series

Approximation of functions (used in maximization or minimization)

If $f(x)$ is infinitely differentiable for any x. Take a point a

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

will be exact (=) is you add "$+R^n(x,a)$"

where $R^n(x,a)$ is the remainder and $\lim_{n\to\infty} R^n(x,a) = 0$

Lagrange Form of Remainders

$\exists\, p \in (x,a)$ such that $R^n(x,a) = \dfrac{f^{(n+1)}(p)(x-a)^{n+1}}{(n+1)!}$

Proof (by induction)

$n = 0$, need to show $R^0(x,a) = \dfrac{f'(p)(x-a)}{1}$

Mean Value Theorem says $\exists\, p \in (x,a)$ such that $f'(p) = \dfrac{f(x) - f(a)}{x-a}$, which can be

rewritten as $f(x) = f(a) + f'(p)(x-a)$

(This is Taylor series with $n = 0$ and the R^0 we wanted to show)

$n = 1$, $f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}s(x)(x-a)^2$, need to show $s(x) = f''(p)$

Keep x fixed and define a function g for $t \in (a,x)$ by

$$g(t) = f(x) - \left[f(t) + f'(t)(x-t) + \frac{1}{2}s(x)(x-t)^2 \right]$$

$$g(x) = f(x) - \left[f(x) + f'(x)(x-x) + \frac{1}{2}s(x)(x-x)^2 \right] = f(x) - f(x) = 0$$

$$g(a) = f(x) - \left[f(a) + f'(a)(x-a) + \frac{1}{2}s(x)(x-a)^2 \right] = f(x) - f(x) = 0$$

By Rolle's Theorem, $\exists\, p \in (a,x)$ such that $g'(p) = 0$ so take derivative of g with respect to t, then plug in p and set it equal to 0

$g'(t) = 0 - \left[f'(t) + f''(t)(x-t) + f'(t)(-1) + s(x)(x-t)(-1) \right] = -(f''(t) - s(x))(x-t)$

$g'(p) = -(f''(p) - s(x))(x-p) = 0$

Can't have x = p (since $p \in (x,a)$) so that means $f''(p) = s(x)$

Taylor series is just a polynomial $p(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + \cdots + a_n(x-x_0)^n$

1) $p(x_0) = f(x_0)$ i.e., $a_0 = f(x_0)$

2) $p'(x_0) = f'(x_0) = a1 + 2a_2(x-x_0) + \cdots + na_n(x-x_0)^{n-1} = a_1$ i.e., $a_1 = f'(x_0)$

3) $p''(x_0) = f''(x_0) = 2a_2 + 6a_3(x-x_0) + \cdots + n(n-1)a_n(x-x_0)^{n-2} = 2a_2$ i.e., $a_2 = f''(x_0)/2$

$\vdots$

n) $p^{(n)}(x_0) = n^1 a_n$ i.e., $a_n = \dfrac{f^{(n)}(x_0)}{n!}$

<u>Example</u>

$f(x) = \dfrac{1}{1-x}$

$x_0 = 0$

Find Taylor approximation for best polynomial

$f(x_0) = f(0) = 1$

$f'(x_0) = -(1-x)^2(-1) = (1-x)^2 = 1$

$f''(x_0) = -2(1-x)^{-3}(-1) = 2(1-x)^{-3} \Rightarrow f''(0) = 2$

$f^{(3)}(x_0) = -6(1-x)^{-4}(-1) = 6(1-x)^{-4} \Rightarrow f^{(3)}(0) = 6$

$\vdots$

$f^{(n)}(x) = f^{(n)}(0) = n!$

So use Taylor series: $f(x) \approx f(0) + \dfrac{f'(0)x}{1!} + \dfrac{f''(0)x^2}{2!} + \dfrac{f^{(3)}x^3}{3!} + \cdots + \dfrac{f^{(n)}x^n}{n!}$

Plug in values for derivatives: $f(x) \approx 1 + \dfrac{1x}{1!} + \dfrac{2x^2}{2!} + \dfrac{6x^3}{3!} + \cdots + \dfrac{n!x^n}{n!} = \sum_{i=0}^{n} x^i$

NOTE: this is definition of a geometric expansion: $\sum_{i=0}^{\infty} x^i = \dfrac{1}{1-x}$

<u>Example</u>

$f(x) = e^x$

$x_0 = 0$

Find Taylor approximation for best polynomial

$f(x_0) = f(0) = 1$

$f'(x_0) = e^x \Rightarrow f'(0) = 1$

etc... (all derivatives = 1)

Using Taylor series: $f(x) \approx 1 + \frac{1x}{1!} + \frac{1x^2}{2!} + \frac{1x^3}{3!} + \cdots + \frac{1x^n}{n!} = \sum_{i=0}^{n} \frac{x^i}{i!}$

NOTE: this is definition of e^x: $\sum_{i=0}^{\infty} \frac{x^i}{i!} = e^x$

Unconstrained Optimization

Min/Max

Let $f(x)$ have domain D.

Then $c \in D$ is a <u>maximum</u> point for f in $D \Leftrightarrow f(x) \le f(c)\ \forall\ x \in D$

Then $d \in D$ is a <u>minimum</u> point for f in $D \Leftrightarrow f(x) \ge f(d)\ \forall\ x \in D$

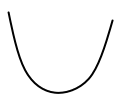
For strict max/min, use $<$ or $>$.

Stationary Point

x_0 is a stationary point for $f(x)$ if $f'(x_0) = 0$

<u>Example</u>

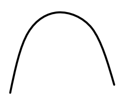
$f(x) = x^2$



$f'(x) = 2x = 0 \Rightarrow x_0 = 0$

x_0 is min

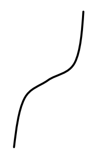
$f(x) = -x^2$



$x_0 = 0$

x_0 is min

$f(x) = x^3$



$f'(x) = 3x^2 = 0 \Rightarrow x_0 = 0$

x_0 is min

First derivative test for max/min

If $f'(x) \ge 0$ for $x \le c$ and $f'(x) \le 0$ for $x \ge c \Rightarrow x = c$ is a max.

" " $\le$ " $\ge$ " " min.

Proof:

Take a Taylor series around x_0

$$f(x) = f(x_0) + f'(x_0)\cdot(x - x_0) + \cdots + \frac{f^{(n)}(x_0)(x-x_0)^n}{n!} + R_{n+1}(x)$$

$$R_{n+1}(x) = \frac{f^{(n+1)}(\bar{x})(x-x_0)^{n+1}}{(n+1)!} \text{ for } \bar{x} \in (x_0, x) \text{ or } \bar{x} \in (x, x_0)$$

For $n = 0$, $f(x) = f(x_0) + f'(\bar{x})(x - x_0) = f(x_0) + R_1(x)$

Case 1: $x < x_0 \Rightarrow \bar{x} \in (x, x_0)$

Now if $f'(y) \ge 0\ \ \forall y \in (x, x_0) \Rightarrow f'(\bar{x}) \ge 0$

$f(x) = f(x_0) + f'(\bar{x})(x - x_0)$, and $(x - x_0) < 0$ so $f(x) \le f(x_0)\ \ \forall x \le x_0$

Case 2: $x > x_0 \Rightarrow \bar{x} \in (x_0, x)$

Now if $f'(y) \le 0\ \ \forall y \in (x_0, x) \Rightarrow f'(x) \le 0$

$f(x) = f(x_0) + f'(\bar{x})(x - x_0)$, and $(x - x_0) > 0$ so $f(x) \le f(x_0)\ \ \forall x \ge x_0$

Therefore $f(x_0)$ is max point

Proof for min is similar

Extreme Values

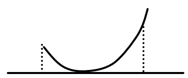
If $f'(x)$ exists for every point in I

1) Find all the stationary points: $f' = 0$ interior to I
2) Evaluate the function f at the boundary and stationary points.
3) Identify the largest and smallest values of f

NOTE: If $f'(x)$ does not exist for some point(s), then we have to evaluate the function at these points also

Example

$y = x^2$, $I = [-1,2]$



Stationary points... $f' = 0$ @ $x = 0$
Evaluate $f(-1) = 1$, $f(2) = 4$, $f(0) = 0$
Largest value @ $x = 2$
Smallest value @ $x = 0$

Second Order Conditions

If x_0 is an interior point and $f'(x_0) = 0$, then $f''(x_0) > 0$ (i.e., convex) $\Rightarrow f'(x_0 + h) > 0$ and $f'(x_0 - h) < 0$ (i.e., x_0 is min point)

" " $f''(x_0) < 0$ (i.e., concave)

Example

$f(x) = x^4$
$f'(x) = 4x^3$
$4x^3 = 0 \Rightarrow x = 0$
$f'(x) < 0$ if $x < 0$
$f'(x) > 0$ if $x > 0$... therefore, by first derivative test, this is a min point
$f''(x) = 12x^2$
$f''(0) \geq 0 \; \forall \, x$ (and only $= 0$ at $x = 0$)... by second order conditions, this is a min point

Exception - if $f''(x_0) = 0$ when $f'(x_0) = 0$, we have to find the first positive derivative. If it is an odd number derivative, then x_0 is an inflection point

Example

$y = x^3$
$y' = 3x^2$
$3x^2 = 0 \Rightarrow x = 0$
$y'' = 6x$
$6x = 0 \Rightarrow x = 0$
$y''' = 6 > 0$... 3rd derivative is odd number so 0 is an inflection point

Inflection Point

c is an inflection point for a twice differentiable function f if there exists an interval (a,b) such that either

a: $f''(x) \geq 0 \quad a < x < c$
$f''(x) \leq 0 \quad c < x < b$

b: $f''(x) \leq 0 \quad a < x < c$
$f''(x) \geq 0 \quad c < x < b$

(i.e., if f'' changes sign at c, then c is an inflection point); at inflection point, function changes from concave to convex (or vice versa)

Convexity & Concavity

Def 1: Assume f is continuous on interval I and twice differentiable on I^0 interior of I,

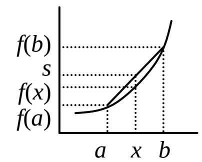
f is concave in $I \Leftrightarrow f''(x) \le 0\ \forall\ x \in I^0$

f is convex on $I \Leftrightarrow f''(x) \ge 0\ \forall\ x \in I^0$

Def 2: f is concave (convex) if any line segment joining any 2 points is never above (below) the graph of the function

Concave Convex

Let $x \in (a,b) \Rightarrow x = (1 - \lambda)a + \lambda b$ for $\lambda \in (0,1) \Rightarrow \lambda = \dfrac{x-a}{b-a} = \dfrac{\text{distance form } x \text{ to } a}{\text{distance from } b \text{ to } a}$ (defining a line segment between a and b)



Equation for line passing through $(a,f(a))$ and $(b,f(b))$ is

$$y - f(a) = \frac{f(b) - f(a)}{b-a}(x-a)$$

For point (x,s) on the line,

$$s - f(a) = \frac{f(b) - f(a)}{b-a}(x-a)$$

$$s - f(a) = \frac{f(b) - f(a)}{b-a}((1-\lambda)a + \lambda b - a)$$

$$s - f(a) = \frac{f(b) - f(a)}{b-a}\lambda(b-a) = \lambda\left[f(b) - f(a)\right]$$

$s = (1-\lambda)f(a) + \lambda f(b)$ this is the point on the line connecting $(a,f(a))$ and $(b,f(b))$ at x

$f(x) = f((1-\lambda)a + \lambda b)$ this is the point on the original function at x

Def 3: $f(x)$ is convex if $\forall\ a,b \in I,\ \forall\ \lambda \in (0,1)$ $f([1-\lambda]a + \lambda b) \le (1-\lambda)f(a) + \lambda f(b)$

" concave " $\ge$ "

(strictly convex (concave) if $<$ ($>$) rather than $\le$ or $\ge$)

Remark... f is concave if $-f$ is convex

Proof

Since $-f$ is convex $-f((1-\lambda)a + \lambda b) \le (1-\lambda)(-f(a)) + \lambda(-f(b))$

Multiply both sides by -1 (changes direction of inequality)

$f((1-\lambda)a + \lambda b) \ge (1-\lambda)f(a) + \lambda f(b)$ which is definition of f is concave

Prove test of concavity: $f''(x) < 0\ \forall\ x \in (a,b) \Rightarrow f$ is strictly concave by using Taylor series

WLOG (without loss of generality) assume that $b > a \Rightarrow (a,b)$

WLOG let $f''(x) < 0\ \forall\ x \in (a,b)$

$$f(b) = f(a) + f'(a)(b-a) + \underbrace{\frac{1}{2}\underbrace{f''(p)}_{-}\underbrace{(b-a)^2}_{+}}_{-} \text{ for } p \in (a,b)$$

By assumption (2nd line) $f''(p) < 0$

$f(b) = f(a) + f'(a)(b-a) +$ (negative term)

$$\frac{f(b)-f(a)}{b-a} = f'(a) + \text{(negative term)}$$

So $f'(a) > \frac{f(b)-f(a)}{b-a}$, which means slope at point a is greater than slope of line connecting a and b so f is concave based on Def 2

Jensen's Inequality

f is concave on $I \Leftrightarrow \forall\, x_1, x_2, ..., x_n \in I$ and $\forall\, \lambda_1, \lambda_2, ...\, \lambda_n \geq 0$ with $\sum_{i=1}^{n} \lambda_i = 1$, and

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) \geq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n)$$

(concave if ≤)

<u>Example</u>

Let λ_i = probability that $x = x_i$ for $I = 1, 2, ..., n$

$$E(x) = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n = \sum_{i=1}^{n} \lambda_i x_i$$

$$E(f(x)) = \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n) = \sum_{i=1}^{n} \lambda_i f(x_i)$$

If f is concave $\Rightarrow$ $f(E(x)) \geq E(f(x))$

If f is convex $\Rightarrow$ $f(E(x)) \leq E(f(x))$

<u>Example</u>

Show that $f(x) = |x|$ is convex in $(-\infty,\infty)$

Let $a,b \in (-\infty,\infty)$ and for $\lambda \in (0,1) \Rightarrow f((1-\lambda)a + \lambda b) = |(1-\lambda)a + \lambda b|$

Use triangular inequality: $|a+b| \leq |a| + |b|$: $|(1-\lambda)a + \lambda b| \leq |(1-\lambda)a| + |\lambda b|$

Since $(1 - \lambda)$ and λ are positive, we can say:

$|(1-\lambda)a| + |\lambda b| = (1-\lambda)|a| + \lambda|b| = (1-\lambda) f(a) + \lambda f(b)$, so f is convex

Probability

Definitions

Outcomes - mutually exclusive results of a random process

Example: heads or tails (can't have both)

Example: your computer might never crash, it might crash once, twice, etc. Only one of these outcomes will occur; i.e., the outcomes are mutually exclusive

Probability - of an outcome is the proportion of the time that outcome occurs in the long run

Example: if the probability of your computer not crashing while you are writing a paper is 80%, then over the course of writing "many" papers, you will complete 80% of the time without a crash

Sample Space - set of all possible outcomes

Event - subset of the sample space

Random Variable (<u>r.v.</u>) - numerical summary of a random outcome

Example: the number of crashes while writing a paper is numerical and random, so it is a random variable

Probability Distribution - of a <u>discrete</u> r.v. is the list of all possible values of the variable and the probability that each value will occur. The probabilities sum up to 1.

Example: Let M be the number of times your computer crashes while you are writing a paper. The probability distribution is:

M	0	1	2	3	4
Prob Dist	0.80	0.10	0.06	0.03	0.01
Cum Dist	0.80	0.90	0.96	0.99	1.00

Example: P(1 or 2 crashes) = $P(M = 1 \text{ or } M = 2) = P(M = 1) + P(M = 2) = 0.06 + 0.10 = 0.16$

Cumulative Distribution - probability that the r.v. is less than or equal to a particular value

Example: $P(M \leq 1) = 0.90$

Bernoulli r.v. - a binary r.v.; can only take on one of two values

Example: Let r.v. $G = \begin{cases} 1 & \text{if it is going to rain} \\ 0 & \text{if it isn't} \end{cases}$

The outcomes of G and their probabilities are $G = \begin{cases} 1 & \text{with probability p} \\ 0 & \text{with probability 1 - p} \end{cases}$

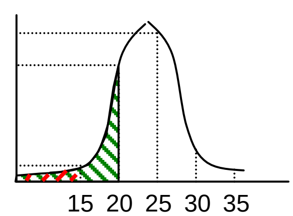
Probability Distribution - of a <u>continuous</u> r.v. takes on a continuum of possible values and the probability distribution is summarized by a pdf (probability density function)

Example: Let the driving time to school be a r.v. D and let its pdf be given by:

$P(D \leq 15)$ = red area

$P(D \leq 20)$ = green area

$P(15 \leq D \leq 20) = P(D \leq 20) - P(D \leq 15)$



Expected Value (mean) - of r.v. y is denoted by $E(y)$; long run average value of the r.v. over many repeated trials or occurrences

If y takes on k possible values $y_1, \ldots, y_k \Rightarrow E(y) = \sum_{i=1}^{k} p_i y_i$

Example: using data for M, $E(M) = 0.80(0) + (0.10)(1) + 0.06(2) + 0.03(3) + 0.01(4) = 0.35$

Example: Bernoulli r.v. $E(G) = 1p + 0(1 - p) = p$

Variance - of r.v. y is denoted by Var(y)

$E\left[(y - E(y))^2\right] = \sum_{i=1}^{k} p_i (y_i - E(y_i))^2$ (rhs if y is discrete)

Example: using data for M,

$Var(M) = 0.80(0\text{-}0.35)^2 + 0.10(1\text{-}0.35)^2 + 0.06(2\text{-}0.35)^2 + 0.03(3\text{-}0.35)^2 + 0.01(4\text{-}0.35)^2 = 0.6475$

Example: Var of Bernoulli = $(1 - p)(0 - p)^2 + p(1 - p)^2 = p(1 - p)$

Standard Deviation - denoted by s_y or $\sigma_y = \sqrt{Var(y)}$

Example: using data for M, s_y = sqrt(0.6475) = 0.8

Moments - expected values of various powers of a function

First Moment - $E(Y)$

Nth Moment - $E(Y^n)$

Covariance - measure of the extent to which two r.v.'s move together

$$Cov(x,y) = \sigma_{x,y} = E\left[(x-\mu_x)(y-\mu_y)\right] = \sum_{i=i}^{k}\sum_{j=1}^{l}(x_j-\mu_x)(y_i-\mu_y)P(x=x_j \,\&\, y=y_i)$$

Joint probability distribution - combines probabilities for two events that occur at same time

Example:

	x=0 (rain)	x=1 (no rain)	Total
y=0 (long drive)	0.15	0.07	0.22
y=1 (short drive)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Correlation Coefficient -

$$Corr(x,y) = \rho = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Properties of Expected Value

If b is a constant, $E(b) = b$

If a, b are constants, $E(ax+b) = aE(x)+b$

If x & y are independent r.v.s, $E(xy) = E(x)E(y)$

Mean & variance of a linear function of a r.v.

Let x & y be two revs that are related by $y = a + bx$, then

$E(y) = E(a+bx) = a + bE(x)$

$Var(y) = E(y-E(y))^2 = E\left[(a+bx)-(a+bE(x))\right]^2 = E\left[bx - bE(x)\right]^2$ =

1 way: $E\left[b^2x^2 - 2bxE(x) + b^2(E(x))^2\right]$ =

2nd way: $E\left[b(x-E(x))\right]^2 = E\left[b^2(x-E(x))^2\right]$ (can now use rules above

$b^2E(x-E(x))^2 = b^2Var(x)$

Marginal Probability Density Function - original distributions of r.v.s in joint distribution function

$f(x,y)$

$f(x) = \sum_y f(x,y)$ and $f(y) = \sum_x f(x,y)$ ... seen in total column and row in table above

Conditional Probability

Probability that x takes a specific value given that y has a specific value, $P(x = x_0 \mid y = y_0)$

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

Example:

$$f(x = -2 \mid y = 3) = \frac{f(x = -2, y = 3)}{f(y = 3)} = \frac{0.27}{0.51} = 0.53$$

Note, $f(x = -2) = 0.27$

Independence

x & y are statistically independent if $f(x, y) = f(x)f(y)$

Example: $f(x = -2, y = 3) = 0.27 \neq f(x = -2)f(y = 3) = 0.27(0.51)$ so x & y are not ind

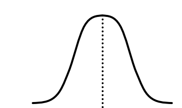
Example: a bag contains 3 balls numbered 1, 2, 3. Two balls are drawn at random with replacement. Let x denote the # of the first ball and y denote the # of the 2nd ball.

$f(x = x_0, y = y_0) = 1/9 = f(x = x_0)f(y = y_0)$ so x & y are independent

Normal Distribution

With mean μ and variance σ^2; denoted by $N(\mu,\sigma^2)$

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



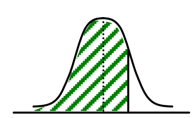
Properties

1) Symmetric around the mean
2) Approx 95% of area under the curve lies between $\mu \pm 2\sigma$
3) $x_1 \sim N(\mu_1, \sigma_1^2)$ and $x_2 \sim N(\mu_2, \sigma_2^2)$, then $y = ax_1 + bx_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
4) Standard normal dist (z) has $\mu = 0$ and $\sigma^2 = 1$; used to calculate the are under the curve by converting regular normal distribution to std norm by:

$$z = \frac{y - \mu}{\sigma}$$

Example: $y \sim N(1,4)$, $P(y \leq 2) = ?$, convert to standard normal

$$P\left(\frac{y-1}{2} \leq \frac{2-1}{2}\right) = P(z \leq 1/2) = 0.69$$



Example: $P(z > 1.12) = 0.5 - 0.3686 = 0.1314$

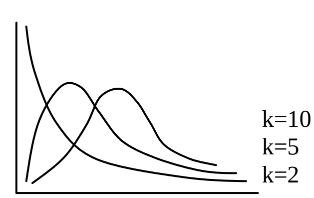
Rules using z

1) $P(z \geq c) = 1 - P(z < c) = 1 - \Phi(c)$ ($\Phi(c)$ is the cumulative density function)
2) $P(c \leq z \leq d) = P(z \leq d) - P(z \leq c)$

Chi-Squared Distribution

Sum of k squared independent standard normal r.v.s

$$\sum_{i=1}^{k} z_i^2 \sim \kappa_k^2$$ (k = degrees of freedom, df)



Properties

1) Skewed to the right (not symmetric); degree of skewness depends on k; large k approaches normal distribution
2) mean is k, variance is $2k$
3) $z_1 \sim \kappa^2_{k_1}$ and $z_2 \sim \kappa^2_{k_2}$ then $z_1 + z_2 \sim \kappa^2_{k_1+k_2}$

Example

$P(\kappa^2_{20} > 40) \approx 0.005$ (using table)

$P(\kappa^2_{20} < 8.26) = 1 - P(\kappa^2_{20} > 8.76) \approx 1 - 0.99 = 0.01$

F Distribution

If z_1 and z_2 are independently distributed κ^2 r.v.s with k_1 and k_2 df then

$$\frac{z_1 / k_1}{z_2 / k_2} \sim F_{k_1,k_2}$$

Properties

1) Skewed right; approaches normal as $k_1, k_2 \to \infty$
2) $\mu = \dfrac{k_2}{k_2 - 2}$ $(k_2 > 2)$, $\sigma = \dfrac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)(k_2 - 4)}$ $(k_2 > 4)$
3) If k_2 is fairly large then $k_1 F = \kappa^2_{k_1}$

Example

$P(F_{10,8} > 3.4) \approx 0.05$ (using table)

Student's t Distribution

If $z_1 \sim N(0,1)$ and $z_2 \sim \kappa^2_k$ are independent, then

$$\frac{z_1}{\sqrt{z_2 / k}} \sim t_k$$

Properties

1) Symmetric like normal, but flatter (thicker tails)
2) $\mu = 0$, $\sigma = \dfrac{k}{k-2}$

Example

$P(t > 3) \approx 0.005$

$P(|t| > 3) = 2P(t > 3) \approx 0.01$ (because it's symmetric)

Central Limit Theorem

Let $y_1, y_2, ..., y_n$ be <u>iid</u> (independently and identically distributed) with $E(y_i) = \mu_y < \infty$ and

$Var(y_i) = \sigma_y^2 < \infty$, then $\dfrac{\bar{y} - \mu_y}{\sigma_{\bar{y}}} \xrightarrow[n\to\infty]{} N(0,1)$

Systems of Linear Equations and Matrices

Linear Algebra

Study of systems of linear equations

Note: $x_1 x_2 = 5$ is not a linear equation, but you can make it linear by taking $\ln$ of both sides:

$\ln x_1 + \ln x_2 = \ln 5$ (which is linear)

System - m independent equations; n variables

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

- Coefficients - a_{ij}
- Right-hand Sides - b_j
- Solution - ordered set or list of numbers $(s_1, s_2, \ldots, s_n)$ that satisfies all the equations simultaneously
- Consistent - a linear system that has at least one solution
  - $m < n$ - many solutions
  - $m = n$ - at most 1 solution
  - $m > n$ - no solution
- Inconsistent - linear system with no solution

Matrix

Rectangular array of numbers considered as an entity; m rows & n columns; denoted by bold capital letters

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

- Order - dimensions of a matrix: m x n
- Elements - entries; each a_{ij}
- Row Vector - matrix with only one row
- Column Vector - matrix with only one column
- Square Matrix of order n - matrix where $m = n$
  - Main Diagonal - elements $a_{11}, a_{22}, \ldots, a_{nn}$

Vector Operations

Vectors typically denoted by lower case bold letters

Components - coordinates; elements of a vector

Equal - all corresponding components of each vector are equal

$$\mathbf{a} = \mathbf{b} \Leftrightarrow a_i = b_i \ \forall \ i = 1, 2, \ldots, n$$

Sum - add corresponding components of $1 \text{ x } n$ or $n \text{ x } 1$ vectors

$$\mathbf{a} + \mathbf{b} = (a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

Scalar Multiplication - multiply each component by t

$$t\mathbf{a} = t(a_1, a_2,..., a_n) = (ta_1, ta_2,..., ta_n)$$

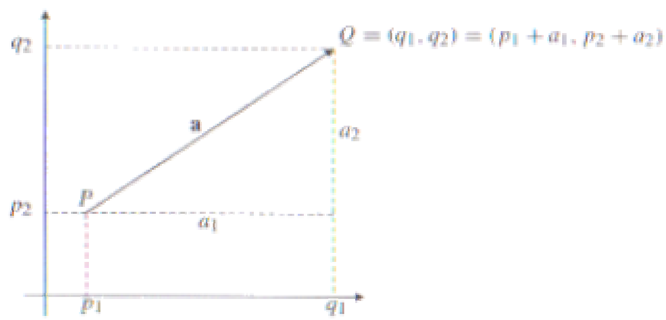
Difference - $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$
Linear Combination - $r\mathbf{a} + s\mathbf{b}$
Dot Product - also called inner product or scalar product; adds the product of corresponding components of two vectors:

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + ... + a_nb_n = \sum_{i=1}^{n} a_i b_i$$

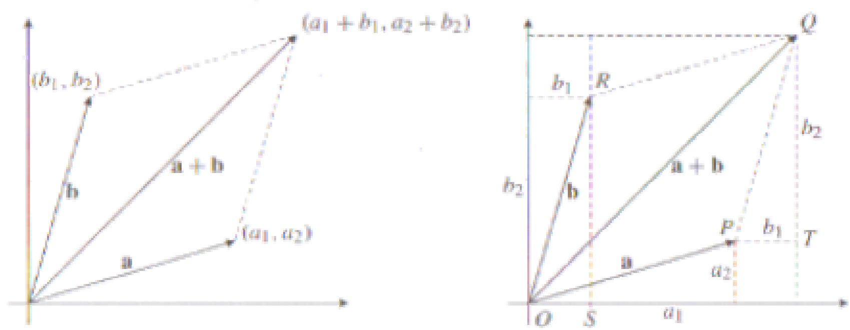
$$\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\cdot\|\mathbf{b}\|\cos\theta \text{ (see orthagonality in next section)}$$

<u>Rules for Dot Product</u>
$\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$
$\mathbf{a}\cdot(\mathbf{b} + \mathbf{c}) = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}$
$(\alpha\mathbf{a})\cdot\mathbf{b} = \alpha(\mathbf{a}\cdot\mathbf{b}) = \mathbf{a}\cdot(\alpha\mathbf{b})$
$\mathbf{a}\cdot\mathbf{a} = \sum_{i=1}^{n} a_i^2 \geq 0$

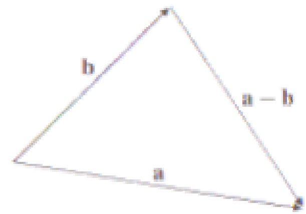
Geometry of Vectors
Vector $\mathbf{a}$ describes the movement from point $P = (p_1, p_2)$ to $Q = (q_1, q_2)$



Sum $\mathbf{a} + \mathbf{b}$ is the diagonal in the parallelogram determined by the two sides $\mathbf{a}$ and $\mathbf{b}$



Difference a - b is the diagonal in the parallelogram determined by the two sides a and -b



b + (a − b) = a

Length - also called norm

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} = \sqrt{a \cdot a}$$

Euclidean Distance

$$\| a - b \| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

Cauchy-Schwarz Inequality

$$|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \cdot \|\mathbf{b}\|$$

Proof:

$|\mathbf{a} \cdot \mathbf{b}| = |a_1 b_1 + a_2 b_2 + \cdots + a_n b_n| \leq (?)\ \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \cdot \sqrt{b_1^2 + b_2^2 + \cdots + b_n^2}$

Square both sides:

$(a_1 b_1 + a_2 b_2 + \cdots + a_n b_n)^2 \leq (?)\ (a_1^2 + a_2^2 + \cdots + a_n^2)\ (b_1^2 + b_2^2 + \cdots + b_n^2)$

Define $g(x) = (a_1 x + b_1)^2 + (a_2 x + b_2)^2 + \cdots + (a_n x + b_n)^2 \geq 0$

Expand: $g(x) = (a_1^2 x^2 + 2a_1 b_1 x + b_1^2) + \cdots + (a_n^2 x + 2a_n b_n x + b^2) \geq 0$

Combine terms: $(a_1^2 + a_2^2 + \cdots + a_n^2)x^2 + (2a_1 b_1 + 2a_2 b_2 + \cdots + 2a_n b_n)x + (b_1^2 + b_2^2 + \cdots b_n^2)$

Rewrite as $\mathbf{Ax^2} + \mathbf{Bx} + \mathbf{C}$

Take derivatives to check convexity

$g'(x) = 2\mathbf{Ax} + \mathbf{B}$

$g''(x) = 2\mathbf{A} \geq \mathbf{0}$ since it's a sum of squares, $\therefore g$ is convex

Also, since $g(x) \geq 0$ it cannot cross the x-axis so there are no real roots or the only root is x = 0, therefore $\mathbf{B}^2 - 4\mathbf{AC} \leq \mathbf{0}$ (quadratic formula)

Rewrite as $\mathbf{B}^2 \leq 4\mathbf{AC}$

Substitute values: $(2a_1 b_1 + \cdots + 2a_n b_n)^2 \leq 4(a_1^2 + a_2^2 + \cdots + a_n^2)\ (b_1^2 + b_2^2 + \cdots b_n^2)$

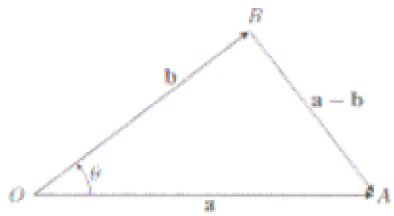
Factor out 2 in left hand side: $[2(a_1 b_1 + \cdots + a_n b_n)]^2$

Take it out of the square: $4(a_1 b_1 + \cdots + a_n b_n)^2$

Cancel the 4 and apply def of dot product and length:

$(a \cdot b)^2 \leq (a_1^2 + a_2^2 + \cdots + a_n^2)\ (b_1^2 + b_2^2 + \cdots b_n^2)$ //

Orthogonality - $\mathbf{a} \perp \mathbf{b} \Leftrightarrow \mathbf{a}\cdot\mathbf{b} = 0$

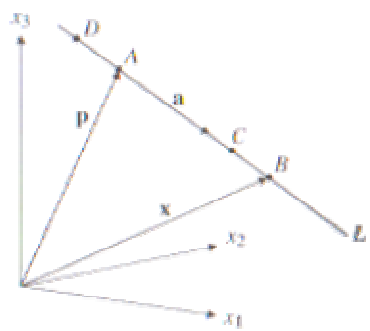


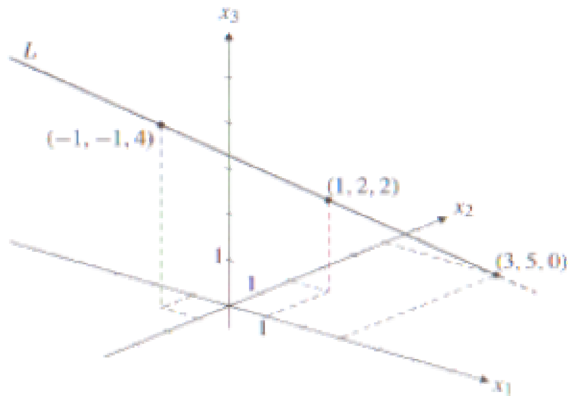
$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\cdot\|\mathbf{b}\|}$$, if $\theta = 90^{\circ}$, $\cos\theta = 0$ so $\mathbf{a}\cdot\mathbf{b} = 0$

Right angle means:
$\|\mathbf{a} - \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$ (Pythagorean Theorem)
$(\mathbf{a} - \mathbf{b})\cdot(\mathbf{a} - \mathbf{b}) = \mathbf{a}\cdot\mathbf{a} + \mathbf{b}\cdot\mathbf{b}$
$\mathbf{a}\cdot\mathbf{a} - \mathbf{a}\cdot\mathbf{b} - \mathbf{b}\cdot\mathbf{a} + \mathbf{b}\cdot\mathbf{b} = \mathbf{a}\cdot\mathbf{a} + \mathbf{b}\cdot\mathbf{b}$
$-2(\mathbf{a}\cdot\mathbf{b}) = 0$
$\mathbf{a}\cdot\mathbf{b} = 0$

Two vectors determine a line: $\mathbf{x} = (1-\lambda)\mathbf{a} + \lambda\mathbf{b}$

In 2 dimensions: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (1-\lambda)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \lambda\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$





Hyperplanes - a plane is a linear combination of 3 points; vector $\mathbf{p}$ is orthogonal to a plane if it is orthogonal to all lines on the plane; if $\mathbf{a}$ and $\mathbf{x}$ define points on the plane, then

$$\mathbf{p}\cdot(\mathbf{x} - \mathbf{a}) = p_1(x_1 - a_1) + p_2(x_1 - a_2) + \dots + p_n(x_n - a_n) = 0$$

Matrix Operations

Equal - $\mathbf{A} = \mathbf{B} \Leftrightarrow \mathbf{A}$ and $\mathbf{B}$ are m x n matrices with $a_{ij} = b_{ij}\ \forall\ i = 1, 2,\dots, m;\ j = 1, 2,\dots, n$

Sum - $\mathbf{A} + \mathbf{B} = (a_{ij})_{m \times n} + (b_{ij})_{m \times n} = (a_{ij} + b_{ij})_{m \times n}$ (add corresponding elements of m x n matrices)

<u>Rules for Matrix Addition</u>

$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$	(associative)
$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$	(cumulative)
$\mathbf{A} + \mathbf{0} = \mathbf{A}$	(identity)
$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$	
$(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$	(distributive over scalar addition)
$\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$	(distributive over scalar multiplication)

Scalar Multiplication - $\alpha\mathbf{A} = (\alpha a_{ij})_{m \times n}$ (multiply each element of $\mathbf{A}$ by α)

Product - $\mathbf{A} = (a_{ij})_{m \times n}$ and $\mathbf{B} = (b_{ij})_{n \times p}$, $\mathbf{C} = \mathbf{AB}$ is $m \times p$ matrix $\mathbf{C} = (c_{ij})_{m \times p}$, with

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}$$

Basically taking dot product of each row of $\mathbf{A}$ with corresponding column in $\mathbf{B}$:

Example:

$$\begin{pmatrix} 0 & 1 & 2 \\ \boxed{2 \quad 3 \quad 1} \\ 4 & -1 & 6 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 \\ 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ \boxed{8} & 5 \\ 5 & 14 \end{pmatrix}$$

$c_{21} = 2 \cdot 3 + 3 \cdot 1 + 1 \cdot (-1) = 6 + 3 - 1 = 8$

Rules for Matrix Multiplication

$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$	(associative)
$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$	(left distributive)
$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$	(right distributive)

Cautions

$\mathbf{AB} \neq \mathbf{BA}$	(not commutative; **BA** may not even exist)
$\mathbf{A}(\mathbf{B} + \mathbf{C}) \neq (\mathbf{B} + \mathbf{C})\mathbf{A}$	(not commutative)

$\mathbf{AB} = \mathbf{0} \not\Rightarrow \mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$

$\mathbf{AB} = \mathbf{AC}$ and $\mathbf{A} \neq \mathbf{0} \not\Rightarrow \mathbf{B} = \mathbf{C}$

Identity Matrix - of order n is n x n matrix having ones along the main diagonal and zeros elsewhere

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n \times n}$$

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

Transpose - If A is m x n matrix, transpose of A is defined as n x m matrix whose first column is the first row of A, whose second column is the second row of A, etc:

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

Rules for Transposition

$(\mathbf{A}')' = \mathbf{A}$

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

$(\alpha\mathbf{A})' = \alpha\mathbf{A}'$

$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

Symmetric Matrix = $\mathbf{A} = \mathbf{A}'$

Orthogonal Matrix = $\mathbf{A}'\mathbf{A} = \mathbf{I}n$

Determinants - denoted by det(A) or |A|; has to be square matrix

2nd Order

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

3rd Order ("easy" formula; doesn't work for larger matrices)

$$|\mathbf{A}| = \left| \begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \right. =$$

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}$$

Cofactors

1) Pick a column or row (which has most number of zeros)... e.g., row 1
2) For each element, take determinant of matrix made by deleting the corresponding row and column ($\mathbf{M}_{ij}$)
3) Multiply the determinant by -1 to the $i + j$ power and by the corresponding element a_{ij}

$$(-1)^{i+j}|\mathbf{M}_{ij}|a_{ij} \qquad \{(-1)^{i+j}|\mathbf{M}_{ij}| \text{ is the cofactor }\}$$

Example using row 1

$$(-1)^{1+1}a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{1+2}a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + (-1)^{1+3}a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Example:

$$\begin{vmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 2 & 1 & 0 \\ 1 & 2 & 1 & 1 & 0 \end{vmatrix} = -1\begin{vmatrix} 1 & 2 & 1 & 2 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 1 & 1 \end{vmatrix} = -1\left[2\begin{vmatrix} 1 & 2 & 2 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \end{vmatrix} - 1\begin{vmatrix} 1 & 2 & 1 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \end{vmatrix}\right] = -1[2(0) - 1(0)] = 0$$

<u>Rules for Determinants</u>

If all elements in a row (or column) of $\mathbf{A}$ are 0, then $|\mathbf{A}| = 0$

If all elements in a row (or column) are multiplied by α, determinant is multiplied by α

Value of determinant is unchanged if a multiple of one row (or column) is added to a different row (or column)

If two rows (or two columns) are interchanged, determinant changes sign, but absolute value is unchanged

If two rows (or columns) are proportional, then $|\mathbf{A}| = 0$

$|\mathbf{A}'| = |\mathbf{A}|$

$|\mathbf{AB}| = |\mathbf{A}|\cdot|\mathbf{B}|$

$|\alpha\mathbf{A}| = \alpha^n|\mathbf{A}|$ ($\mathbf{A}$ is n x n matrix)

<u>Cautions</u>

$|\mathbf{A} + \mathbf{B}| \neq |\mathbf{A}| + |\mathbf{B}|$

Cramer's Rule - solving a system of equations

$$x_1 = \frac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{21}a_{12}} = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{|\mathbf{A}|} \qquad x_2 = \frac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{21}a_{12}} = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{|\mathbf{A}|}$$

To find x1, replace A's first column with b; take determinant and divide by |A|

To find x2, replace A's second column with b; take determinant and divide by |A|

Example:

2x1 + 4x2 = 7

2x1 - 2x2 = -2

$$|\mathbf{A}| = \begin{vmatrix} 2 & 4 \\ 2 & -2 \end{vmatrix} = -4 - 8 = -12$$

$$x_1 = \frac{\begin{vmatrix} 7 & 4 \\ -2 & -2 \end{vmatrix}}{|\mathbf{A}|} = \frac{-14+8}{-12} = \frac{1}{2}$$

$$x_2 = \frac{\begin{vmatrix} 2 & 7 \\ 2 & -2 \end{vmatrix}}{|\mathbf{A}|} = \frac{-4-14}{-12} = \frac{3}{2}$$

Inverse - denoted by $\mathbf{A}^{-1}$; has to be square matrix

$\mathbf{AX} = \mathbf{XA} = \mathbf{I}_n$

Singular - matrix with no inverse

Nonsingular - matrix has an inverse $\Leftrightarrow |\mathbf{A}| \neq 0$

<u>Rules for Inverses</u>

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

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

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

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

Example:

$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, find $\mathbf{A}^{-1} = \begin{pmatrix} w & x \\ y & z \end{pmatrix}$.... $\mathbf{AA}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\mathbf{AA}^{-1} = \begin{pmatrix} aw+by & ax+bz \\ cw+dy & cx+dz \end{pmatrix}$... four equations with four unknowns

Use Cramer's rule:

$$w = \frac{\begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{d}{ad-bc}, \quad y = \frac{\begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{-c}{ad-bc}$$

$$x = \frac{\begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{-b}{ad-bc}, \quad z = \frac{\begin{vmatrix} a & 0 \\ c & 1 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} = \frac{a}{ad-bc}$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{|\mathbf{A}|}\mathrm{Adj}(\mathbf{A})$$

Adjoint - $\mathrm{Adj}(\mathbf{A})$; transpose of the matrix made up of all the cofactors

"Easy" Way - make a new matrix with the original next to an identity matrix; use the following operations to get an identity matrix on the left; the result on the right is the inverse:

Types of Operations

1) switch rows

2) multiply row by scalar

3) add multiples of a row to other rows

Example

$$\left(\begin{matrix}2 & 3 & 5\\4 & 1 & 6\\3 & 1 & 7\end{matrix}\right\|\left.\begin{matrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{matrix}\right)$$

Multiply first row by 1/2: $\left(\begin{matrix}1 & 3/2 & 5/2\\4 & 1 & 6\\3 & 1 & 7\end{matrix}\right\|\left.\begin{matrix}1/2 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{matrix}\right)$

Add -4 times first row to second row: $\left(\begin{matrix}1 & 3/2 & 5/2\\0 & -5 & -4\\3 & 1 & 7\end{matrix}\right\|\left.\begin{matrix}1/2 & 0 & 0\\-2 & 1 & 0\\0 & 0 & 1\end{matrix}\right)$

Add -3 times first row to third row: $\left(\begin{matrix}1 & 3/2 & 5/2\\0 & -5 & -4\\0 & -7/2 & -1/2\end{matrix}\right\|\left.\begin{matrix}1/2 & 0 & 0\\-2 & 1 & 0\\-3/2 & 0 & 1\end{matrix}\right)$

Multiply second row by -5: $\left(\begin{matrix}1 & 3/2 & 5/2\\0 & 1 & 4/5\\0 & -7/2 & -1/2\end{matrix}\right\|\left.\begin{matrix}1/2 & 0 & 0\\2/5 & -1/5 & 0\\-3/2 & 0 & 1\end{matrix}\right)$

Add -3/2 times second row to first row: $\left(\begin{matrix}1 & 0 & 13/10\\0 & 1 & 4/5\\0 & -7/2 & -1/2\end{matrix}\right\|\left.\begin{matrix}-1/10 & 3/10 & 0\\2/5 & -1/5 & 0\\-3/2 & 0 & 1\end{matrix}\right)$

Add 7/2 times second row to first row: $\left(\begin{matrix}1 & 0 & 13/10\\0 & 1 & 4/5\\0 & 0 & 23/10\end{matrix}\right\|\left.\begin{matrix}-1/10 & 3/10 & 0\\2/5 & -1/5 & 0\\-1/10 & -7/10 & 1\end{matrix}\right)$

Multiply third row by 10/23: $\left(\begin{matrix}1 & 0 & 13/10\\0 & 1 & 4/5\\0 & 0 & 1\end{matrix}\right\|\left.\begin{matrix}-1/10 & 3/10 & 0\\2/5 & -1/5 & 0\\-1/23 & -7/23 & 10/23\end{matrix}\right)$

Add -13/10 times third row to first row: $\left(\begin{matrix}1 & 0 & 0\\0 & 1 & 4/5\\0 & 0 & 1\end{matrix}\right\|\left.\begin{matrix}-1/23 & 16/23 & -13/23\\2/5 & -1/5 & 0\\-1/23 & -7/23 & 10/23\end{matrix}\right)$

Add -4/5 times third row to the second row: $\left(\begin{matrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{matrix}\right\|\left.\begin{matrix}-1/23 & 16/23 & -13/23\\10/23 & 1/23 & -8/23\\-1/23 & -7/23 & 10/23\end{matrix}\right)$

Systems of Equations

Coefficient Matrix - m x n matrix of all the coefficients of a linear system of m equations with n unknowns; from page 1, $\mathbf{A}$ is the coefficient matrix to the linear system above

RHS Vector - $\mathbf{b}$ is a column vector containing the b_js on the right hand side of a linear system

Can rewrite system of linear equations from page 1 as $\mathbf{Ax} = \mathbf{b}$

Solve with inverse: $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

Linearly Independent

$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (all $k \times 1$ vectors) are

linearly dependent if $\exists\ \lambda_1, \lambda_2, \ldots, \lambda_n \neq (0,0,\ldots,0)$ such that $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \lambda_n\mathbf{x}_n = \mathbf{0}_{k\times1}$

linearly independent if $\nexists\ \lambda_1, \lambda_2, \ldots, \lambda_n \neq (0,0,\ldots,0)$ such that $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \lambda_n\mathbf{x}_n = \mathbf{0}_{k\times1}$

How to check, arrange each vector as a column (or row) in a matrix, $\mathbf{X}$, if $|\mathbf{X}| = \mathbf{0}$, they're linearly dependent

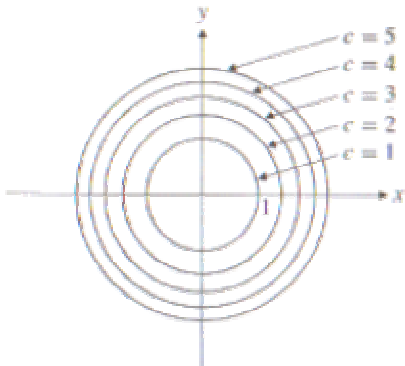
Partial Derivatives

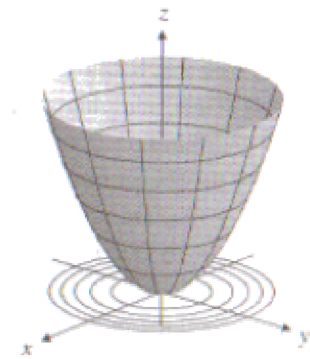
Function of Several Variables

Function - rule that assigns a specified number $f(\mathbf{x})$ to each n vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ in the domain D

Example: $f(x,y) = x^2 + y^2$

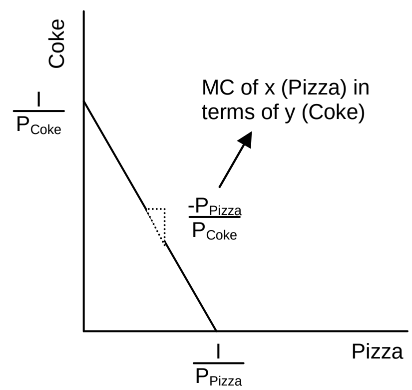
3d plot:

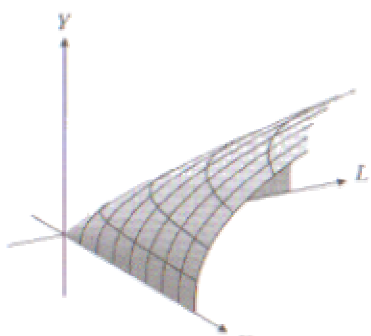


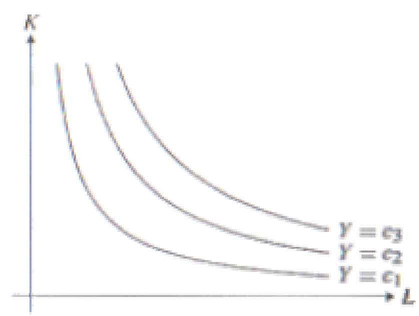


2d elevation plot - plots all combinations of x and y that give a constant value for $f(x,y)$

Example: Budget Constraint, Coke & Pizza







Partial Derivatives

One variable:

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

Two variables:

$$\frac{\partial f(x,y)}{\partial x} = \lim_{h \to 0} \frac{f(x+h,y) - f(x,y)}{h}$$ (partial of f with respect to x; hold y constant)

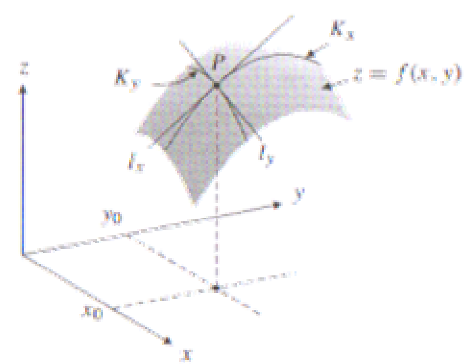
Interpretation - partial derivative is slope of tangent to function in plane where you're taking the partial (e.g., if you're in 3 dimensions, partial wrt x gives you tangent to curve in plane where y is held constant; this plane with be parallel to the xz plane); all partials determine a plane that is tangent to the function

Example - Cobb-Douglass production function of capital, K, and labor, L

$$f(K,L) = A \cdot K^a L^b \quad (K > 0, L > 0)$$

$$\frac{\partial f(K,L)}{\partial K} = A \cdot aK^{a-1}L^b$$

$$\frac{\partial f(K,L)}{\partial L} = A \cdot K^a bL^{b-1}$$



Example:

$$f(x,y) = x^3 + 2y^2$$

$$\frac{\partial f(x,y)}{\partial x} = 3x^2$$

$$\frac{\partial f(x,y)}{\partial y} = 4y$$

Higher-Order Partial Derivatives

z = f(x,y)

First order: $\frac{\partial f(x,y)}{\partial x}$

Second order: $f_{xx} = \frac{\partial^2 f(x,y)}{\partial x^2}$, $f_{yx} = \frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f(x,y)}{\partial y}\right)$, also have f_{xy} and f_{yy}

Example: $f(x,y) = x^2 e^y$

$$\frac{\partial f(x,y)}{\partial x} = 2xe^y, \quad \frac{\partial f(x,y)}{\partial y} = x^2 e^y$$

$$\frac{\partial^2 f(x,y)}{\partial x^2} = 2e^y, \quad \frac{\partial^2 f(x,y)}{\partial y^2} = x^2 e^y, \quad \frac{\partial f(x,y)}{\partial x \partial y} = \frac{\partial f(x,y)}{\partial y \partial x} = 2xe^y$$

Note: Usually $f_{xy} = f_{xy}$, but not always

Example: $w = f(x,y,z) = x^4 + 2y^3 + xz^2$

$$\frac{\partial w}{\partial x} = 4x^3 + z^2$$

$$\frac{\partial^2 w}{\partial z \partial x} = 2z, \quad \frac{\partial^2 w}{\partial x^2} = 12x^2$$

$$\frac{\partial^3 w}{\partial x \partial z \partial x} = 0, \quad \frac{\partial^2 w}{\partial z \partial x^2} = 0$$

Young's Theorem - function that has nth order continuous derivatives, if any two derivatives involve differentiating with respect to each of the variables the same number of times, then the derivatives are necessarily equal (i.e., the order of the differentiation doesn't matter)

Example where Young's doesn't work:

$$f(x,y) = \begin{cases} \dfrac{x^3 y - xy^3}{x^2 + y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

If you hold x = 0, f(x,y) = 0 (regardless of what happens to y); also f(x,y) = 0 when y = 0 (regardless of x)

$$\lim_{y \to 0} \frac{f(0,y) - f(0,0)}{y} = \frac{\partial f}{\partial y}(0,0) = 0$$

$$\lim_{x \to 0} \frac{f(x,0) - f(0,0)}{x} = \frac{\partial f}{\partial x}(0,0) = 0$$

$$\frac{\partial f}{\partial x} = \frac{(3x^2 y - y^3)(x^2 + y^2) - (x^3 y - xy^3)(2x)}{(x^2 + y^2)^2}$$

$$\frac{\partial f(0,y)}{\partial x} = \frac{-y^5}{y^4} = -y$$

$$\frac{\partial f}{\partial y} = \frac{(x^3 - 3y^2)(x^2 + y^2) - (x^3 y - xy^3)(2y)}{(x^2 + y^2)^2}$$

$$\frac{\partial f(x,0)}{\partial y} = \frac{x^5}{x^4} = x$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial f_x(0,y)}{\partial y} = \lim_{y \to 0} \frac{\dfrac{\partial f(0,y)}{\partial x} - \dfrac{\partial f(0,0)}{\partial x}}{y} = \frac{-y - 0}{y} = -1$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial f_y(x,0)}{\partial x} = \lim_{x \to 0} \frac{\dfrac{\partial f(x,0)}{\partial y} - \dfrac{\partial f(0,0)}{\partial y}}{x} = \frac{x - 0}{x} = 1$$

Why are these different? Look at second order partial where y = x

$$\frac{\partial f}{\partial y} = \frac{x^3 - 3xy^2}{x^2 + y^2} - \frac{2x^3 y^2 - 2xy^4}{(x^2 + y^2)^2}$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{(3x^2 - 3y^2)(x^2 + y^2) - (x^3 - 3xy^2)2x}{(x^2 + y^2)^2} +$$

$$\frac{(6x^2 y^2 - 2y^4)(x^2 + y^2)^2 - (2x^3 y^2 - 2xy^4)2(x^2 + y^2)2x}{(x^2 + y^2)^4}$$

$$\frac{\partial^2 f}{\partial x \partial y}(x,x) = \frac{-2x(x^3 - 3x^3)}{(x^2 + x^2)^2} - \frac{6x^4 - 2x^4}{(x^2 + x^2)^2} = \frac{-2x^4 + 6x^4 - 6x^4 + 2x^4}{(x^2 + x^2)^2} = 0$$

But we just showed that the partial at (0,0) is 1; that means the function is not continuous at (0,0)

Gradient - vector of all first order partial derivatives of a function evaluated as some point

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$ evaluated at $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$

Hessian Matrix - all combinations of second order partial derivatives

$$\mathbf{H} = f''(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) = \begin{pmatrix} f''_{11}(\mathbf{x}) & f''_{12}(\mathbf{x}) & \cdots & f''_{1n}(\mathbf{x}) \\ f''_{21}(\mathbf{x}) & f''_{22}(\mathbf{x}) & \cdots & f''_{2n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f''_{n1}(\mathbf{x}) & f''_{n2}(\mathbf{x}) & \cdots & f''_{nn}(\mathbf{x}) \end{pmatrix}$$

Young's Theorem says this is usually a symmetric matrix

Example: Cobb-Douglas function: $f(K,L) = AK^a L^b$

$$\frac{\partial f}{\partial K} = AaK^{a-1}L^b$$

$$\frac{\partial f}{\partial L} = AK^a bL^{b-1}$$

$$\frac{\partial^2 f}{\partial K \partial L} = AabK^{a-1}L^{b-1}, \quad \frac{\partial^2 f}{\partial L \partial K} = AabK^{a-1}L^{b-1}$$

$$\frac{\partial^2 f}{\partial K^2} = Aa(a-1)K^{a-2}L^b, \quad \frac{\partial^2 f}{\partial L^2} = Ab(b-1)K^a L^{b-2}$$

$$\mathbf{H} = \begin{pmatrix} Aa(a-1)K^{a-2}L^b & AabK^{a-1}L^{b-1} \\ AabK^{a-1}L^{b-1} & Ab(b-1)K^a L^{b-2} \end{pmatrix}$$

Aside: for Cobb-Douglas production function:

a + b = 1 constant returns to scale

a + b > 1 increasing returns to scale

a + b < 1 decreasing returns to scale

Quadratic Forms

Functions with variables to the second power

Example: $f(x,y) = ax^2 + 2bxy + cy^2$

In matrix notation: $\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by & bx+cy \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2$

$$\frac{\partial f}{\partial x} = 2ax + 2by$$

$$\frac{\partial f}{\partial y} = 2bx + 2cy$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2b, \quad \frac{\partial^2 f}{\partial x \partial y} = 2b, \quad \frac{\partial^2 f}{\partial x^2} = 2a, \quad \frac{\partial^2 f}{\partial y^2} = 2c$$ so $\mathbf{H} = \begin{pmatrix} 2a & 2b \\ 2b & 2c \end{pmatrix}$

General Case:

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Positive Definite - $f(x,y) > 0 \ \forall \ (x,y) \neq (0,0)$
Positive Semidefinite - $f(x,y) \geq 0 \ \forall \ (x,y) \neq (0,0)$
Negative Definite - $f(x,y) < 0 \ \forall \ (x,y) \neq (0,0)$
Negative Semidefinite - $f(x,y) \leq 0 \ \forall \ (x,y) \neq (0,0)$

$f(x,y)$ is positive definite iff $a > 0$, $c > 0$, and $ac - b^2 = |\mathbf{A}| > 0$

Proof:

Let $x = 1$ and $y = 0$

$f(x,y) = a > 0$

Let $x = 0$ and $y = 1$

$f(x,y) = c > 0$

$f(-b/a,1) = a\dfrac{b^2}{a^2} + 2b\dfrac{-b}{a}1 + c(1) = \dfrac{ac-b^2}{a} > 0$ and since $a > 0$ then $ac - b^2 > 0$

That proves positive definite leads to $a > 0$, $c > 0$, and $ac - b^2 = |\mathbf{A}| > 0$

Now suppose $a > 0$, $c > 0$, and $ac - b^2 = |\mathbf{A}| > 0$

Factor out a and add/subtract b/ay:

$$f(x,y) = a\left(x^2 + \frac{2b}{a}xy + \frac{c}{a}y^2\right) = a\left(x^2 + \frac{2b}{a}xy + \frac{b^2}{a^2}y^2 - \frac{b^2}{a^2}y^2 + \frac{c}{a}y^2\right)$$

$f(x,y) = a\left(x + \dfrac{b}{a}y\right)^2 + \left(\dfrac{ac-b^2}{a}\right)y^2 > 0$ so $f(x,y)$ is positive definite

$f(x,y)$ is positive semidefinite iff $a \geq 0$, $c \geq 0$, and $ac - b^2 = |\mathbf{A}| \geq 0$
$f(x,y)$ is negative definite iff $a < 0$, $c < 0$, and $ac - b^2 = |\mathbf{A}| > 0$
$f(x,y)$ is negative semidefinite iff $a \leq 0$, $c \leq 0$, and $ac - b^2 = |\mathbf{A}| \geq 0$

Example: $f(x,y) = x^2 + y^2$ ... positive definite

Example: $f(x,y) = (x-y)^2$ ... positive semidefinite

Example: $f(x,y) = -(x-y)^2$ ... negative semidefinite

Example: $f(x,y) = x^2 - y^2$ ... indefinite

Chain Rule

If $x = g(t)$ and $y = h(t)$ so $f(x,y) = f(g(t),h(t))$

Total Derivative of f wrt t: $\dfrac{\partial f}{\partial t} = \dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial t} + \dfrac{\partial f}{\partial y}\dfrac{\partial y}{\partial t}$

General Case: $f(x_1, x_2, \ldots, x_n)$, $x_i = h_i(t_1, t_2, \ldots, t_m)$

$$\frac{\partial f}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j}$$

Example: $f(x_0 + th, y_0 + tk)$

$$\frac{\partial f}{\partial t} = f_x(x_0 + th, y_0 + tk)h + f_y(x_0 + th, y_0 + tk)k$$

Total Differentials & Taylor Approximations

NOTE: $F_{12}(x,y) = \dfrac{\partial^2 F(x,y)}{\partial y \partial x}$

Implicit Functions

Given $F(x,y) = c$ (level curve), what happens to y as we change x... if $y = f(x)$, we look at new function $g(x) = F(x, f(x)) = c$

Take derivative of both sides: $F_1(x, f(x)) + F_2(x, f(x)) \cdot \dfrac{\partial f(x)}{\partial x} = 0$

Solve for $\dfrac{\partial y}{\partial x} = \dfrac{\partial f(x)}{\partial x} = \dfrac{-F_1(x, f(x))}{F_2(x, f(x))} = \dfrac{-F_1(x,y)}{F_2(x,y)}$

Example: $P_x x + P_y y = I \quad \Rightarrow \quad y = \dfrac{I - P_x x}{P_y}$

$$\frac{\partial y}{\partial x} = \frac{-P_x}{Py}$$

Example: $F(x,y) = x^3 + x^2 y - 2y^2 - 10y = 0$

$\dfrac{\partial y}{\partial x}$ at $(x,y) = (2,1)$

$F_1(x,y) = 3x^2 + 2xy \quad \Rightarrow \quad F_1(2,1) = 3(2^2) + 2(2)(1) = 16$

$F_2(x,y) = x^2 - 4y - 10 \quad \Rightarrow \quad F_2(2,1) = 2^2 - 4(1) - 10 = -10$

$\dfrac{\partial y}{\partial x}(2,1) = \dfrac{-16}{-10} = \dfrac{8}{5}$

Example: $x^2 + e^{xy} = 0$

Can compute the implicit derivative, but it won't matter

Problem... there's no (x,y) that solves the equation

Other cases that doesn't work: $F_2(x,y) = 0$ (i.e., vertical lines)

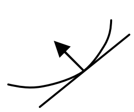
Directional Derivative

From h to k: $D_{h,k}$ where $x = x_0 + h$ and $y = y_0 + k$

Approximates change in function as function of t, as $g(t) = f(x_0 + th, y_0 + tk)$

Gradient Vector

$\nabla F = \begin{pmatrix} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} \end{pmatrix}$



Two properties:

Perpendicular to tangent

Parallel to direction of maximum increase in function

Now, $D = F_1(x-x_0) + F_2(y-y_0) = 0$ can be rewritten as $\nabla F \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} = 0$ or

$\nabla F \cdot \begin{pmatrix} x-x_0 & y-y_0 \end{pmatrix}$

You can use $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos\theta$ with $\mathbf{a} = \nabla F$ and $\mathbf{b} =$ any vector; a special case has

$D = \|\nabla F\| \cdot \|\mathbf{z}\|$, where $\mathbf{z}$ is a vector pointing to the maximum increase in the function (i.e., parallel to the gradient so $\cos 0^0 = 1$, NOTE: $\cos 0^0 = 1$ is the maximum value for $\cos$ so this is the maximum increase in the function).

Second Derivative

$$\frac{\partial y}{\partial x} = \frac{-F_1(x,y)}{F_2(x,y)} = \frac{-G(x,y)}{H(x,y)}$$

$$\frac{\partial^2 y}{\partial x^2} = -\frac{G'H - GH'}{H^2}$$

$$G' = \frac{\partial G(x, f(x))}{\partial x} = \frac{\partial F_1(x,y)}{\partial x} = F_{11}(x,y) + F_{12}(x,y)\frac{\partial y}{\partial x} = F_{11} + F_{12}\frac{-F_1}{F_2}$$

$$H' = \frac{\partial H(x, f(x))}{\partial x} = \frac{\partial F_2(x,y)}{\partial x} = F_{21}(x,y)) + F_{22}(x,y)\frac{\partial y}{\partial x} = F_{21} + F_{22}\frac{-F_1}{F_2}$$

$$\frac{\partial^2 y}{\partial x^2} = -\frac{\left(F_{11} + F_{12}\frac{-F_1}{F_2}\right)F_2 - F_1\left(F_{21} + F_{22}\frac{-F_1}{F_2}\right)}{(F_2)^2}$$

$$\frac{-1}{(F_2)^3}\left[F_{11}F_2^2 - 2F_{12}F_1F_2 + F_{22}F_1^2\right]$$

$$\frac{1}{F_2^3}\begin{vmatrix} 0 & F_1 & F_2 \\ F_1 & F_{11} & F_{12} \\ F_2 & F_{21} & F_{22} \end{vmatrix}$$ ("bordered" Hessian)

Linear Approximation

Approximation: $f(x,y) \approx f(x_0,y_0) + f_x(x-x_0) + f_y(y-y_0)$

Plane through a point (x_0, y_0, z_0): $z - z_0 = A(x-x_0) + B(y-y_0)$

If $z = f(x,y)$, plane tangent to $f(x,y)$ at (x_0, y_0, z_0): $z - z_0 = z_x(x-x_0) + z_y(y-y_0)$

$z_x = \frac{\partial z}{\partial x}$ and $z_y = \frac{\partial z}{\partial y}$

In 2d say $x = x_0 + dx$ x or $y = y_0 + dy$; now say $z = z_0 + dz$ so $dz = z - z_0 = z_x dx + z_y dy$,

similar to saying $df(x,y) = f_x dx + f_y dy$

Rules:

$d(af + bg) = a \cdot df + b \cdot dg$ (a & b scalars & f & g functions)

$d(fg) = g \cdot df + f \cdot dg$

$d\left(\frac{f}{g}\right) = \frac{g \cdot df - f \cdot dg}{g^2}$

$z = g(f()) \Rightarrow dz = g' df$

Proof of second rule:

$d(f(x,y)g(x,y)) = [f_x(x,y)g(x,y) + f(x,y)g'(x,y)]dx + [f_y g + f g_y]dy$

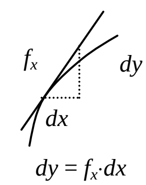
$g[f_x dx + f_y dy] + f[g_x dx + g_y dy] = g \cdot df + f \cdot dg$

Doesn't matter if x and y are functions of other variables:

$z = F(x,y)$ where $x = f(t,s)$ and $y = g(t,s)$

$z = F(f(t,s), g(t,s))$

$dz = z_t \cdot dt + z_s ds$, $z_t = F_x \cdot x_t + F_y \cdot y_t$ and $z_s = F_x \cdot x_s + F_y \cdot y_s$



$dz = (F_x \cdot x_s + F_y \cdot y_t)dt + (F_x \cdot x_s + F_y \cdot y_s)ds$

$dz = F_x(x_t dt + x_s ds) + F_y(y_t dt + y_s ds) = F_x dx + F_y dy$

General case: $F(x_1, x_2, \ldots, x_n)$; $dF = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \cdot dx_i$ where $dx_i = x_i - x_i^0$ (old - new)

Example: $z = xy^2 + x^3$

$dz = z_x dx + z_y dy = (y^2 + 3x^2)dx + 2xy \cdot dy$

If x changes from 3 to 5 and y changes from 2 to 7 (really too big for this technique)

$dx = 2$; $dy = 5$

$dz \approx (2^2 + 3(3^2))(2) + 2(3)(2)(5)$

Taylor Series with Several Variables

Move from $f(x_1^0, x_2^0)$ to $f(x_1, x_2)$; know value of x_1 and x_2, but not $f(x_1, x_2)$

Can approximate it.

Can rewrite $x_1 = x_1^0 + th_1$ and $x_2 = x_2^0 + th_2$ so now $f(x_1, y_1) = y(t) = f(x_1^0 + th_1, x_2^0 + th_2)$ which is a function of 1 variable (t) and can use Taylor Series

$y(t) \approx y(0) + y'(0)t + \frac{y'(0)}{2}t^2 + \cdots + \frac{y^{(n)}(0)}{n!}t^n$ or

$y(t) = y(0) + y'(0)t + \frac{y'(0)}{2}t^2 + \cdots + \frac{y^{(n)}(t^*)}{n!}t^n$, where $0 \le t^* \le t$

How do you get y' wrt t? $y'(t) = f_{x1}(x_1^0 + th_1, x_2^0 + th_2) \cdot h_1 + f_{x2}(x_1^0 + th_1, x_2^0 + th_2) \cdot h_2$

so $y'(0) = f_{x1}(x_1^0, x_2^0) \cdot h_1 + f_{x2}(x_1^0, x_2^0) \cdot h_2$

Next is y'': $y''(0) = \sum_{i=1}^{2}\sum_{j=1}^{2} f_{ij}(x_1^0, x_2^0)h_i h_j$

Can write second order Taylor Series using matrix notation with the Hessian matrix:

$f(x_0 + h) = f(x_0) + f'(\mathbf{x_0}) \cdot \mathbf{h} + 2\mathbf{h}'\mathbf{H}\mathbf{h}$

Example: $f(x,y) = 3x^2 + 5xy + 6y^2 + 10$

Evaluate Talor series approximation for $(x_0, y_0) = (1,1)$

$f(2,2) \approx f(1,1) + f_1(1,1)\cdot(x - x_0) + f_2(1,1)\cdot(y - y_0) + \tfrac{1}{2} f_{11}(1,1)\cdot(x - x_0) + \tfrac{1}{2} f_{22}(1,1)\cdot(y - y_0) +$

$f_{12}(1,1)\cdot(x - x_0)(y - y_0)$ (higher order derivatives are all zero so this is exact)

$f_1 = 6x + 5y \qquad f_{11} = 6 \qquad f_{12} = 5$

$f_2 = 5x + 12y \qquad f_{22} = 12$

$f(2,2) \approx 24 + 11(1) + 17(1) + \tfrac{1}{2}6(1) + \tfrac{1}{2}12(1) + 5 = 66$

Check approximation: $3(2^2) + 5(2)(2) + 6(2^2) + 10 = 66$

Taylor Series with 2 variables and n = 3

$$f(x,y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) +$$

$$\frac{1}{2!}\left[f_{xx}(x_0, y_0)(x - x_0)^2 + f_{yy}(x_0, y_0)(y - y_0)^2 + 2 f_{xy}(x_0, y_0)(x - x_0)(y - y_0)\right] +$$

$$\frac{1}{3!}\left[f_{xxx}(x_0, y_0)(x - x_0)^3 + f_{yyy}(x_0, y_0)(y - y_0)^3 + 3 f_{xxy}(x_0, y_0)(x - x_0)^2 (y - y_0) +\right.$$

$$\left. 3 f_{xyy}(x_0, y_0)(x - x_0)(y - y_0)^2 + f_{yyy}(x_0, y_0)(y - y_0)^3\right]$$

Optimization with Several Variables

Extreme Value Theorem (Existence Theorem)

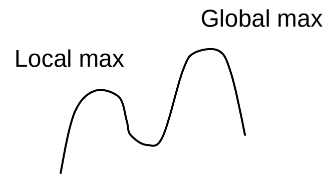
Any continuous function attains a maximum and a minimum in a compact set (i.e., closed and bounded set), there exists a maximum $\mathbf{c}$ and a minimum $\mathbf{d}$; sufficient, but not necessary condition

Maximum: $f(\mathbf{x}) \le f(\mathbf{c})$; Minimum: $f(\mathbf{x}) \ge f(\mathbf{d})$ (<u>Extreme Points</u>)

First-Order Conditions

For interior point $\mathbf{c} = (c_1, c_2, \dots, c_n)$ to be a max or min of $f(\mathbf{c})$, necessary conditoin is:

$\dfrac{\partial f(\mathbf{c})}{\partial c_i} = 0 \;\; \forall \; i = 1,2,\dots,n$ (<u>Stationary Points</u>)



Condition is also necessary for local max or min

Aside: Will also have directional derivative $D_{hk} f(\mathbf{c}) = \sum_{i=1}^{n} f_i(c) \cdot h = 0$

Theorem

f defined over $S \in R^n$
$F(x)$ defined in range of *f*
$g(x)$ defined on S (domain of *f*)
$g(x) = F(f(x))$

a) if F is increasing and $\mathbf{c}$ maximizes (minimizes) *f* over S, then $\mathbf{c}$ also maximizes (minimizes) g over S

b) if F is <u>strictly</u> increasing then $\mathbf{c}$ maximizes (minimizes) *f* over S <u>iff</u> $\mathbf{c}$ maximizes (minimizes) g over S

Example: $f(K, L) = AK^a L^b \quad (K > 0, L > 0)$

Let $F(u) = \ln u$ (strictly increasing)

$g(K, L) = \ln A + a \ln K + b \ln L$

Find K & L to max g, and they'll also max *f*

Proof of a:

Assume $\mathbf{c}$ miximizes *f*

$f(\mathbf{c}) \ge f(\mathbf{x}) \;\; \forall \; \mathbf{x} \in S$

$F(f(\mathbf{c})) \ge F(f(\mathbf{x})) \; \forall \; \mathbf{x} \in S$ because F is increasing

Proof of b:

Assume $\mathbf{c}$ maximizes g

$F(f(\mathbf{c})) > F(f(\mathbf{x})) \;\; \forall \; \mathbf{x} \in S$

$f(\mathbf{c}) \ge f(\mathbf{x}) \;\; \forall \; \mathbf{x} \in S$

Finding Max or Min

of a differentiable function *f* defined on a compact set

1) Find all stationary points
2) Find largest & smallest values of function on the boundary

Example: if domain is $x \in$ [0,3] and $y \in$ [0,3], need to look all around square determined by x & y at their limits

3) Compute values of function for all points in steps 1 & 2
4) Compare values of the function
5) Largest value is the max; smallest is the min

Using First Derivative for Max/Min

With single variable, looked at stationary point c, if $f'(x) > 0$ for $x < c$ and $f'(x) < 0$ for $x > c$, then c is a max

Doesn't work with several variables:

$f(x, y) = x^2 + 2bxy + y^2$

Stationary points:

$f_x = 2x + 2by = 0$

$f_y = 2bx + 2y = 0$

(0,0) is a stationary point

Is is a min?

Fix $y = 0$ and look at $g(x) = f(x,0) = x^2$ ... has min as $x = 0$

Same for $x = 0$

But if $b = -3/2$, $f(x, y) = x^2 - 3xy + y^2$ so $f(x, x) = -x^2 < 0$ (when $x \neq 0$)

Second Order Conditions

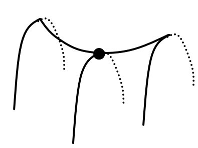
Let $f(x, y)$ be continuous function with continuous first and second order derivatives in a doman S, and (x_0, y_0) be an interior stationary point

$A = f_{11}(x_0, y_0)$ (i.e., second derivative wrt x evaluated at (x_0, y_0))

$B = f_{12}(x_0, y_0) = f_{21}(x_0, y_0)$ (because f is continuous)

$C = f_{22}(x_0, y_0)$

Saddle point:
Max in one axis, but in in another



1) $A < 0$ and $AC - B^2 > 0$, then (x_0, y_0) is local max
2) $A > 0$ and $AC - B^2 > 0$, then (x_0, y_0) is local min
3) $AC - B^2 < 0$, then (x_0, y_0) is a saddle point
4) $AC - B^2 = 0$, then (x_0, y_0) is a min, max, or saddle point

NOTE: $AC - B^2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix}$ (the Hessian)

Example: $f(x, y) = -x^3 + xy + y^2 + x$

Necessary first order Conditions:

$f_1(x, y) = -3x^2 + y + 1 = 0$

$f_2(x, y) = x + 2y = 0$

Plug in $x = -2y$ into first eqn to get $12y^2 + y + 1 = 0$ so $y = 1/3$ or $-1/4$

Plug back in to $x = -2y$ to get two stationary points: $(-2/3, 1/3)$ and $(1/2, -1/4)$

Now look at second order conditions:

$f_{11}(x,y) = -6x$

$f_{12}(x,y) = 1$

$f_{22}(x,y) = 2$

Compute Hessian at the two points

$(-2/3, 1/3)$ $\mathbf{H} = \begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} = 7 > 0 \Rightarrow$ Minimum

$(1/2, -1/4)$ $\mathbf{H} = \begin{vmatrix} -3 & 1 \\ 1 & 2 \end{vmatrix} = -7 < 0 \Rightarrow$ Saddle Point

Example: Three functions (a) $z = -x^4 - y^4$, (b) $z = x^4 + y^4$, (c) $z = x^3 + y^3$

(a) Stationary points:

$z_1 = -4x^3 = 0$

$z_2 = -4y^3 = 0 \Rightarrow$ (0,0) is stationary point

$A = z_{11} = -12x^2$, $B = z_{12} = 0$, $C = -12y^2 = 0$

So at (0,0), $\mathbf{H} = 0$

(0,0) can be either min, max or saddle point

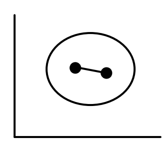
Look at function... $\forall x \neq 0, \forall y \neq 0 \Rightarrow z(x,y) < z(0,0) = 0$ so it's a max

(b) & (c) are similar, i.e., (0,0) is stationary oint with $\mathbf{H} = 0$

For (b) $\forall x \neq 0, \forall y \neq 0 \Rightarrow z(x,y) > z(0,0) = 0$ so it's a min

For (c) $z(1,1) = 2 > z(0,0) = 0$ and $z(-1,-1) = -2 < z(0,0) = 0$ so it's a saddle point

Convex Sets



$S \in R^n$ and $\mathbf{x},\mathbf{y} \in S$ so $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$

S is convex set if $z = (1 - \lambda)x + \lambda y \in S \ \forall \ \lambda \in [0,1]$

(i.e., line segment connecting x and y resides in S)

Example: budget set: $Pxx + pyy \leq I$ ($x \geq 0$, and $y \geq 0$)

Example: Cobb-Douglas production or utility functions: $f(K,L) = AK^aL^b$ (K,L ≥ 0, a > 0, b ≤ 1)

Level curve: $f(K,L) = AK^aL^b = c$

So $K = \left(\frac{c}{A}\right)^{1/a} L^{-b/a}$

$$\frac{\partial K}{\partial L} = \left(\frac{c}{A}\right)^{1/a} \frac{-b}{a} L^{(-b/a)-1}$$

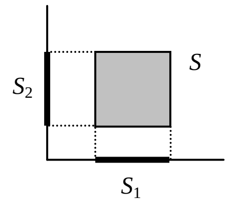
$\frac{\partial^2 K}{\partial L^2} = \left(\frac{c}{A}\right)^{1/a} \left(\frac{-b}{a}\right)\left(\frac{-b}{a} - 1\right) L^{(-b/a)-2} > 0$ so it's a convex set

Facts of Convex Sets:

1) If S1, S2,..., Sn are convext sets, $S_1 \cap S_2 \cap \ldots \cap S_n$ is a convex set

2) If S2, S2 are convext sets: $S = \{(x,y) : x \in S_1, y \in S_2\}$ is a convex set

(Cartesian Product: $S = S_1 \times S_2$



Convex Function

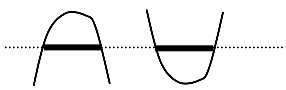
f(x) defined on convex set is

Concave if $f\left[(1-\lambda)\mathbf{x}+\lambda\mathbf{y}\right] \geq (1-\lambda)f(\mathbf{x})+\lambda f(\mathbf{y}) \;\; \forall \mathbf{x},\mathbf{y} \in S, \lambda \in (0,1)$

Convex if $f\left[(1-\lambda)\mathbf{x}+\lambda\mathbf{y}\right] \leq (1-\lambda)f(\mathbf{x})+\lambda f(\mathbf{y}) \;\; \forall \mathbf{x},\mathbf{y} \in S, \lambda \in (0,1)$

Strictly if using > and <

A linear function is both concave and convex, but is neither strictly concave nor strictly convex



Formally

f is concave $M_f = \{(\mathbf{x}, y) : \mathbf{x} \in S$ and $f(\mathbf{x}) \geq y\}$ is a convex set

f is concave $M_f = \{(\mathbf{x}, y) : \mathbf{x} \in S$ and $f(\mathbf{x}) \leq y\}$ is a convex set

Jenson's Inequality

Discrete and continuous verisons

A function *f* of *n* variables is concave on a convex set $S \in R^n$ <u>iff</u> $\forall$ $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in$ S and $\lambda_i \geq 0$ with

$$\sum_{i=1}^{n} \lambda_i = 1: \;\; f(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \lambda_n\mathbf{x}_n) \geq \lambda_1 f(\mathbf{x}_1) + \lambda_2 f(\mathbf{x}_2) + \cdots + \lambda_n f(\mathbf{x}_n)$$

Continuous Functions

Any function obtained by adding, substracting, multiplying, dividing or composing (function of a funciton) some continuous functoins is also a continuous function

(We've been assuming this, but hadn't written formally yet)

Last Day!!!

Theorem

$f(\mathbf{x})$, $g(\mathbf{x})$ defined on S (convex set) $\in R^n$

a) if f, g are concave and $a \geq 0$, $b \geq 0 \Rightarrow G(\mathbf{x}) = af(\mathbf{x}) + bg(\mathbf{x})$ is also concave

b) if f, g are convex and $a \geq 0$, $b \geq 0 \Rightarrow G(\mathbf{x}) = af(\mathbf{x}) + bg(\mathbf{x})$ is also concave

c) if $F(\mathbf{x})$ is concave and increasing & $f(\mathbf{x})$ is concave $\Rightarrow U = F(f(\mathbf{x}))$ is concave

d) if $F(\mathbf{x})$ is convex and increasing & $f(\mathbf{x})$ is convex $\Rightarrow U = F(f(\mathbf{x}))$ is convex

e) if f, g are concave $\Rightarrow H(\mathbf{x}) = \min(f, g)$ is concave

f) if f, g are convex $\Rightarrow H(\mathbf{x}) = \max(f, g)$ is convex

g) for a & b, add a + b > 0, & f, g <u>strictly</u>, concave/convex $\Rightarrow G(\mathbf{x})$ is strictly concave/convex

Proof of a:

Pick $\mathbf{x}, \mathbf{y} \in S$

$$G[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}] = af[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}] + bg[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}] \geq$$
$$a[(1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})] + b[(1-\lambda)g(\mathbf{x}) + \lambda g(\mathbf{y})] =$$
$$(1-\lambda)[af(\mathbf{x}) + bg(\mathbf{x})] + \lambda[af(\mathbf{y}) + bg(\mathbf{y})] = (1-\lambda)G(\mathbf{x}) + \lambda G(\mathbf{y})$$

Proof of c:

$$U[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}] = F[f[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}]]$$

Since f is concave: $f[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}] \geq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$

Since F is increasing: $F[f[(1-\lambda)\mathbf{x} + \lambda\mathbf{y}]] \geq F[(1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})]$

Since F is concave: $F[(1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})] \geq (1-\lambda)F(f(\mathbf{x})) + \lambda F(f(\mathbf{y}))$

By def: $(1-\lambda)F(f(\mathbf{x})) + \lambda F(f(\mathbf{y})) = (1-\lambda)U(\mathbf{x}) + \lambda U(\mathbf{y})$

So U is concave

Gradient Refresher

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$
$$\mathbf{x} - \mathbf{x}_0 = \begin{pmatrix} x_1 - x_1^0 & x_2 - x_2^0 & \cdots & x_n - x_n^0 \end{pmatrix}$$

Theorem

Let f defined on open convex set S

a) f is concave $\Leftrightarrow$ $f(\mathbf{x}) - f(\mathbf{x_0}) \leq \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - x_i^0) = \nabla F \cdot (\mathbf{x} - \mathbf{x_0})$

b) f is strictly concave $\Leftrightarrow$ $f(\mathbf{x}) - f(\mathbf{x_0}) < \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - x_i^0)$

c) f is convex $\Leftrightarrow$ $f(\mathbf{x}) - f(\mathbf{x_0}) \geq \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - x_i^0)$

d) f is strictly convex $\Leftrightarrow$ $f(\mathbf{x}) - f(\mathbf{x_0}) > \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - x_i^0)$

Proof of a:

Suppose $f(\mathbf{x})$ is concave

$$f((1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}) \geq (1-\lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}) =$$

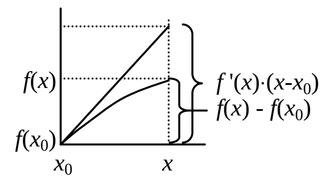
$$f(\mathbf{x}_0) - \lambda f(\mathbf{x}_0) + \lambda f(\mathbf{x}) = f(\mathbf{x}_0) + \lambda(f(\mathbf{x}) - f(\mathbf{x}_0))$$

$$f(\mathbf{x}) - f(\mathbf{x}_0) \leq \frac{f((1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}) - f(\mathbf{x}_0)}{\lambda} = \frac{f(\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0) - f(\mathbf{x}_0)}{\lambda}$$

Let $g(\lambda) = f(\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0))$

So rhs of inequality is now: $\frac{g(\lambda) - g(0)}{\lambda}$

$$g'(0) = \lim_{\lambda \to 0} \frac{g(\lambda) - g(0)}{\lambda}$$



$$g'(\lambda) = \sum_{i=1}^{n} f_i^{'}(\lambda) \cdot (x_i - x_i^0) = \nabla f \cdot (\mathbf{x} - \mathbf{x}_0)$$

Suppose $f(\mathbf{x}) - f(\mathbf{x}_0) \leq \nabla f \cdot (\mathbf{x} - \mathbf{x}_0)$

For brevity let $\mathbf{z} = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}$ be a point between $\mathbf{x}$ and $\mathbf{x}_0$

$f(\mathbf{x}_0) - f(\mathbf{z}) \leq \nabla f(\mathbf{z}) \cdot (\mathbf{x}_0 - \mathbf{z})$ (tangent going towards $\mathbf{x}_0$)

$f(\mathbf{x}) - f(\mathbf{z}) \leq \nabla f(\mathbf{z}) \cdot (\mathbf{x} - \mathbf{z})$ (tangent going towards $\mathbf{x}$)

Multiply first ineq by (1 - λ) and second ineq by λ and add them together

$$(1-\lambda) f(\mathbf{x}_0) - (1-\lambda) f(\mathbf{z}) + \lambda f(\mathbf{x}) - \lambda f(\mathbf{z}) \leq \nabla f(\mathbf{z})\left[(1-\lambda)\mathbf{x}_0 - (1-\lambda)\mathbf{z} + \lambda\mathbf{x} - \lambda\mathbf{z}\right]$$

$$(1-\lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}) - f(\mathbf{z}) \leq \nabla f(\mathbf{z})\left[(1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x} - \mathbf{z}\right]$$

$$(1-\lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}) - f(\mathbf{z}) \leq \nabla f(\mathbf{z})[\mathbf{0}] = 0$$

$(1-\lambda) f(\mathbf{x}_0) + \lambda f(\mathbf{x}) \leq f(\mathbf{z}) = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}$ so $f(\mathbf{x})$ is concave

Theorem

Let f defined on open convex set $S \in R^n$ with $f \in C^2$ (i.e., continuous first and second order derivatives); $\mathbf{x}_0$ is an interior point of S

a) If f is concave, $\mathbf{x}_0$ is max $\Leftrightarrow$ $\mathbf{x}_0$ is stationary point (all first order derivatives = 0)

b) If f is convex, $\mathbf{x}_0$ is min $\Leftrightarrow$ $\mathbf{x}_0$ is stationary point (all first order derivatives = 0)

Proof of a:

$$\nabla f(\mathbf{x}_0) = \mathbf{0}$$

$$f(\mathbf{x}) - f(\mathbf{x}_0) \leq \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$$

$f(\mathbf{x}) \leq f(\mathbf{x}_0)$ so $\mathbf{x}_0$ is max

Theorem

$f(x, y) \in C^2$ and domain S is open and convex

a) If f is concave $\Leftrightarrow$ $f_{xx}'' \leq 0$, $f_{yy}'' \leq 0$, $\begin{vmatrix} f_{xx}'' & f_{xy}'' \\ f_{yx}'' & f_{yy}'' \end{vmatrix} \geq 0$

b) If f is convex $\Leftrightarrow$ $f_{xx}'' \geq 0$, $f_{yy}'' \geq 0$, $\begin{vmatrix} f_{xx}'' & f_{xy}'' \\ f_{yx}'' & f_{yy}'' \end{vmatrix} \geq 0$

c) Strictly convex/concave... remove = from ≥ and ≤

Theorem (Sufficient Conditions for Global Extreme Points)

If this happens, you have a max or min

$f(x, y) \in C^2$ and domain S is convex with interior, stationary point (x_0, y_0)

a) $\forall (x, y) \in S$ $f_{xx}'' \leq 0$, $f_{yy}'' \leq 0$, $\begin{vmatrix} f_{xx}'' & f_{xy}'' \\ f_{yx}'' & f_{yy}'' \end{vmatrix} \geq 0$ $\Leftrightarrow$ (x_0, y_0) is global maximum

a) $\forall (x, y) \in S$ $f_{xx}'' \geq 0$, $f_{yy}'' \geq 0$, $\begin{vmatrix} f_{xx}'' & f_{xy}'' \\ f_{yx}'' & f_{yy}'' \end{vmatrix} \geq 0$ $\Leftrightarrow$ (x_0, y_0) is global minimum

Theorem

Function defined on a convex set is both concave and convex iff it is linear

(i.e., has form: $f(\mathbf{x}) = \mathbf{ax} + b = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + b$)