

Master Program in *Data Science and Business Informatics*

Statistics for Data Science

Lesson 06 - Recalls on calculus

Salvatore Ruggieri

Department of Computer Science
University of Pisa, Italy
salvatore.ruggieri@unipi.it

*J. Ward, J. Abdey. **Mathematics and Statistics**. University of London, 2013. Chapters 1-8 of Part 1.*

- Errata-corrige at pag. 30: $\frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + c \cdot b}{b \cdot d}$ and $\frac{a}{b} - \frac{c}{d} = \frac{a \cdot d - c \cdot b}{b \cdot d}$

Sets and functions

- Numerical sets

- ▶ $\mathbb{N} = \{0, 1, 2, \dots\}$

[Natural numbers]

- ▶ $\mathbb{Z} = \mathbb{N} \cup \{-1, -2, \dots\}$

[Integers]

- ▶ $\mathbb{Q} = \{m/n \mid m, n \in \mathbb{Z}, n \neq 0\}$

[Rationals]

- ▶ $\mathbb{R} = \{ \text{fractional numbers with possibly infinitely many digits} \} \supseteq \mathbb{Q}$

[Reals]

- ▶ $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$

[Irrationals]

- y such that $y \cdot y = 2$ belongs to \mathbb{I}

- Functions

- ▶ $\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$

[Cartesian product]

- ▶ $f : \mathbb{R} \rightarrow \mathbb{R}$ is a subset $f \subseteq \mathbb{R} \times \mathbb{R}$ such that $(x, y_0), (x, y_1) \in f$ implies $y_0 = y_1$

[Functions]

- usually written $f(x) = y$ for $(x, y) \in f$

- $f(x) = v$ for all x

[Constant functions]

- $f(x) = a \cdot x + b$ for fixed a, b

[Linear functions]

- $f(x) = a \cdot x^2 + b$ for fixed a, b

[Quadratic functions]

- $f(x) = \sum_{i=0}^n a_i \cdot x^i$ for fixed a_0, \dots, a_n

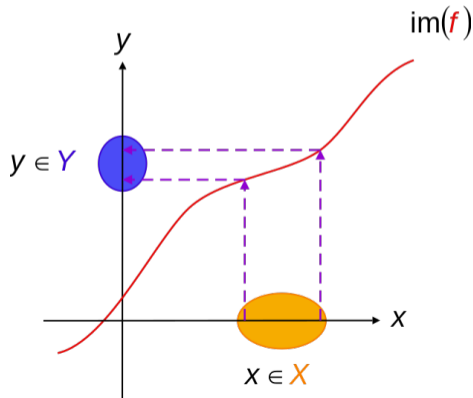
[Polinomials]

See R script

Functions

- $dom(f) = \{x \in \mathbb{R} \mid \exists y \in \mathbb{R}.(x, y) \in f\}$
- $im(f) = \{y \in \mathbb{R} \mid \exists x \in \mathbb{R}.(x, y) \in f\}$
- $f^{-1} = \{(y, x) \mid (x, y) \in f\}$
 - ▶ f^{-1} is a function iff f is injective
 - ▶ $f^{-1}(y) = x$ iff $f(x) = y$
 - ▶ $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$
- Examples
 - ▶ $\sqrt{y} = x$ iff $x^2 = y$ over $x \geq 0$
 - ▶ $\sqrt[n]{y} = x$ iff $x^n = y$ over $x \geq 0$ [positive root]

[Domain or Support]
[Co-domain or Image]
[Inverse function, also f^{inv}]



Powers and logarithms

Power laws

The power laws state that

$$a^n \cdot a^m = a^{n+m} \quad \frac{a^n}{a^m} = a^{n-m} \quad (a^n)^m = a^{nm}$$

provided that both sides of these expressions exist. In particular, we have

$$a^0 = 1 \quad \text{and} \quad a^{-n} = \frac{1}{a^n}.$$

If it exists, we also define the *positive* n th root of a , written $\sqrt[n]{a}$, to be $a^{\frac{1}{n}}$.

- $\log_a(y) = x$ iff $a^x = y$ for $a \neq 1, x > 0$
- for $n/m \in \mathbb{Q}$: $a^{n/m} \stackrel{\text{def}}{=} (a^n)^{1/m}$
- what is a^x for $x \in \mathbb{I}$?

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots = \sum_{n \geq 0} \frac{x^n}{n!}$$

and $a^x = (e^{\log_e(a)})^x = e^{x \cdot \log_e(a)}$

- $X \sim \text{Poi}(\mu), \quad \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} \cdot e^{\mu} = 1$

[Logarithms]

See R script

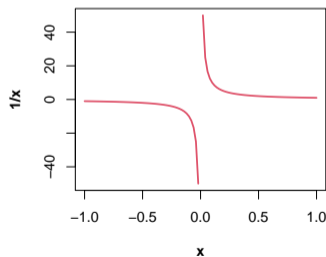
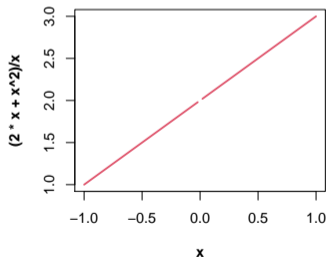
Limits

For a function $f()$, and $a \in \mathbb{R} \cup \{-\infty, \infty\}$

$$\lim_{x \rightarrow a} f(x) = L \quad \text{or} \quad f(x) \rightarrow L \text{ as } x \rightarrow a$$

if $f(x)$ can be made as close to L as desired, by making x close enough, but not equal, to a .

- Example: $\lim_{x \rightarrow 0} \frac{2 \cdot x + x^2}{x} = 2$
- A function $f()$ is called *continuous* at c , if $\lim_{x \rightarrow c} f(x) = f(c)$



- The limit may not exist, e.g., $\lim_{x \rightarrow 0} 1/x$

Gradient and derivatives

- The **gradient** of a straight line is a measure of how 'steep' the line is.

$$y = a \cdot x + b$$

a is the gradient and b the intercept (at $x = 0$)

- For $y = f(x) = x^2$?

- ▶ Tangent at $x = a$ is $y = m \cdot x + b$

- ▶ $m = \frac{f(a+\delta) - f(a)}{\delta} = \frac{2 \cdot a \cdot \delta + \delta^2}{\delta} \rightarrow 2 \cdot a$ for $\delta \rightarrow 0$

- ▶ $b = 2 \cdot a - a^2$ because $m \cdot a + b = a^2$

- More in general?

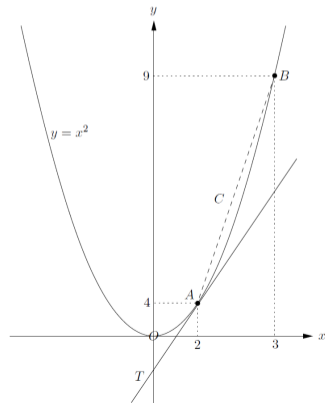
- ▶ For $y = f(x)$, $m = f'(x)$

- ▶ $f'(x) = \lim_{\delta \rightarrow 0} \frac{f(x+\delta) - f(x)}{\delta}$ is called the **derivative** of $f()$,

- ▶ $f'(x)$ also written $\frac{\delta f}{\delta x}$ or $\frac{df}{dx}$

- ▶ Not all functions are differentiable!

See **R script** or **this Colab Notebook**



Derivatives

Standard derivatives

- If k is a constant, then $f(x) = k$ gives $f'(x) = 0$.
- If $k \neq 0$ is a constant, then $f(x) = x^k$ gives $f'(x) = kx^{k-1}$.
- $f(x) = e^x$ gives $f'(x) = e^x$.
- $f(x) = \ln x$ gives $f'(x) = \frac{1}{x}$.

- Constant multiple rule:

$$\frac{d}{dx}[k \cdot f(x)] = k \cdot \frac{df}{dx}(x)$$

- Sum rule:

$$\frac{d}{dx}[f(x) + g(x)] = \frac{df}{dx}(x) + \frac{dg}{dx}(x)$$

Derivatives

- Product rule:

$$\frac{d}{dx}[f(x) \cdot g(x)] = \frac{df}{dx}(x) \cdot g(x) + f(x) \cdot \frac{dg}{dx}(x)$$

- Quotient rule:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \left[\frac{df}{dx}(x) \cdot g(x) - f(x) \cdot \frac{dg}{dx}(x)\right] \cdot \frac{1}{g(x)^2}$$

- Chain rule:

$$\frac{d}{dx}[f(g(x))] = \frac{df}{dg}(g(x)) \cdot \frac{dg}{dx}(x)$$

- $\frac{d}{dx}e^{-x} = \dots$

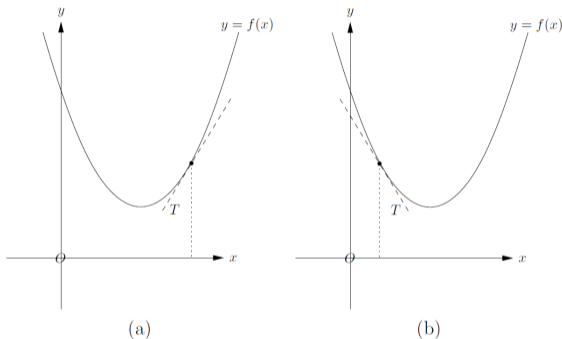
- Inverse rule:

$$\frac{d}{dx}[f^{-1}(x)] = \frac{1}{\frac{df}{dx}(f^{-1}(x))}$$

- $\frac{d}{dx}\log x = \dots$

See R script or this Colab Notebook

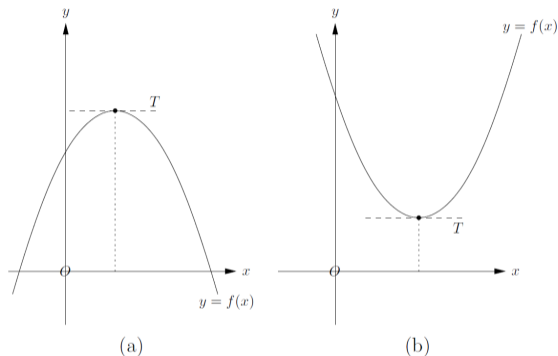
Optimization



- $f'(x) > 0$ implies $f()$ is increasing at x
- $f'(x) < 0$ implies $f()$ is decreasing at x
- $f'(x) = 0$ we cannot say

[Stationary point]

Optimization - second derivatives



- $f''(x) < 0$ implies $f(x)$ is a maximum
- $f''(x) > 0$ implies $f(x)$ is a minimum
- $f''(x) = 0$ we cannot say

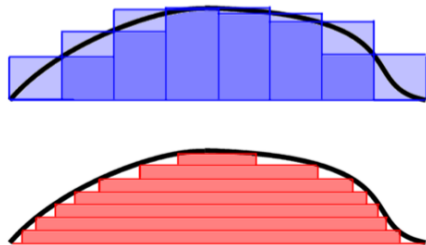
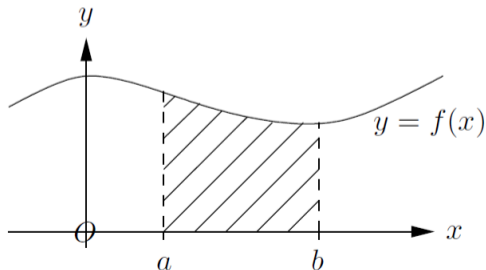
[Maximum, minimum, or point of inflection]

See this Colab Notebook

Integration

- Given $f(x)$, what is $F(x)$ such that $f(x) = \frac{d}{dx}F(x)$? i.e, such that $F'(x) = f(x)$
- Quick answer: $F(x) = \int_{-\infty}^x f(t)dt$
 - ▶ Integration is the inverse of differentiation
- Geometrical definition of integrals:
 - ▶ $\int_a^b f(x)dx$ is the area below $f(x)$
 - ▶ defined as approximation of domain partitioning (Riemann–Darboux integrals) or image partitioning (Lebesgue integrals)

[Fundamental theorem of calculus]



Key concepts in integration

If $F(x)$ is a function whose derivative is the function $f(x)$, then we have

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

where c is an arbitrary constant. In particular, we call the

- function, $f(x)$, the *integrand* as it is what we are integrating,
- function, $F(x)$, an *antiderivative* as its derivative is $f(x)$,
- constant, c , a *constant of integration* which is completely arbitrary,[†] and
- integral, $\int f(x) dx$, an *indefinite integral* since, in the result, c is arbitrary.

- Definite integrals over an interval $[a, b]$:

$$\int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a)$$

Standard integrals

- If $k \neq -1$ is a constant, then $\int x^k dx = \frac{x^{k+1}}{k+1} + c$.

In particular, if $k = 0$, we have $\int 1 dx = \int x^0 dx = x + c$.

- $\int x^{-1} dx = \ln|x| + c$.
- $\int e^x dx = e^x + c$.

- Constant multiple rule:

$$\int [k \cdot f(x)] dx = k \cdot \int f(x) dx$$

- Sum rule:

$$\int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx$$

See R script

Integration by parts

- From the product rule of derivatives:

$$\frac{d}{dx}[f(x) \cdot g(x)] = \frac{df}{dx}(x) \cdot g(x) + f(x) \cdot \frac{dg}{dx}(x)$$

- take the inverse of both sides:

$$f(x) \cdot g(x) = \int f'(x) \cdot g(x) dx + \int f(x) \cdot g'(x) dx$$

- and then:

$$\int f(x) \cdot g'(x) dx = f(x) \cdot g(x) - \int f'(x) \cdot g(x) dx$$

- $\int \lambda x e^{-\lambda x} dx = \dots = -e^{-\lambda x}(x + 1/\lambda)$
 - ▶ consider $f(x) = x$ and $g'(x) = \lambda e^{-\lambda x}$
 - ▶ $g(x) = -e^{-\lambda x}$ and $f'(x) = 1$

Integration by change of variable

- Change of variable rule:

$$\int f(y)dy =_{y=g(x)} \int f(g(x))g'(x)dx$$

- $\int \frac{\log x}{x} dx = \int y dy = y^2/2$ for $y = \log x$ hence, $\int \frac{\log x}{x} dx = (\log x)^2/2$
 - ▶ consider $f(y) = y$ and $g(x) = \log x$

Functions of two or more variables

- Symmetry of second derivatives

[Schwarz's theorem]

$$\frac{d}{dx} \frac{d}{dy} f(x, y) = \frac{d}{dy} \frac{d}{dx} f(x, y)$$

- Leibniz integral rule

$$\frac{d}{dx} \int_a^b f(x, y) dy = \int_a^b \frac{d}{dx} f(x, y) dy$$

- Gradient (pronounced "del")

[direction and rate of fastest increase]

$$\nabla f(x, y) = \begin{pmatrix} \frac{d}{dx} f(x, y) \\ \frac{d}{dy} f(x, y) \end{pmatrix}$$

- Hessian matrix (2×2 case):

[Generalize the second derivative test for max/min]

$$\mathbf{H}_2(x, y) = \begin{pmatrix} \frac{d}{dx} \frac{d}{dx} f(x, y) & \frac{d}{dx} \frac{d}{dy} f(x, y) \\ \frac{d}{dy} \frac{d}{dx} f(x, y) & \frac{d}{dy} \frac{d}{dy} f(x, y) \end{pmatrix}$$

Feynman's trick

$$F(t) = \int_0^{\infty} e^{-tx} dx = \left[-\frac{e^{-tx}}{t} \right]_0^{\infty} = \frac{1}{t}$$

- using Leibniz integral rule

$$\frac{d}{dt} F(t) = \frac{d}{dt} \int_0^{\infty} e^{-tx} dx = \int_0^{\infty} \frac{d}{dt} e^{-tx} dx = - \int_0^{\infty} x e^{-tx} dx = -\frac{1}{t^2}$$

- Taking further derivatives yields:

$$\int_0^{\infty} x^n e^{-tx} dx = -\frac{n!}{t^{n+1}}$$

- and for $t = 1$:

$$n! = \int_0^{\infty} x^n e^{-x} dx$$

[Euler's $\Gamma(n)$]