
Introduzione alla programmazione per le scuole superiori

basics

07 dic 2024

I	Introduzione alla programmazione	3
1	Introduzione alla programmazione - in Python	5
2	Variabili, tipi e funzioni elementari built-in	7
2.1	Commenti	7
2.2	Tipi built-in	8
3	Controllo del flusso	11
3.1	Alternativa	11
3.2	Iterazione	12
4	Funzioni	15
4.1	Default value of function arguments	16
5	Classi	17
6	Librerie	19
II	Introduzione al calcolo scientifico	21
7	Introduzione al calcolo scientifico	23
8	Sistemi lineari	25
8.1	Sistemi lineari quadrati con matrici piene	25
8.2	Matrici sparse	28
9	Equazioni algebriche non lineari	31
9.1	Equazioni non lineari	31
9.2	Sistemi di equazioni non lineari	34
10	Approssimazione di funzioni	37
10.1	Interpolazione	37
10.2	Regressione	37
11	Derivate di funzioni	39
11.1	Differenze finite	39

12 Integrali	43
12.1 Integrazione di Newton-Cotes	43
12.2 Integrazione di Gauss	45
13 Equazioni differenziali ordinarie	49
13.1 Problemi di Cauchy ai valori iniziali	49
13.2 Problemi al contorno	50
14 Ottimizzazione	51
 III Statistica	 53
15 Metodi per la statistica	55
16 Statistica descrittiva	57
16.1 Dimensione dei dati	58
16.2 Rappresentazione grafica	58
16.3 Variabili 1-dimensionali	58
16.4 Variabili 2-dimensionali	58
16.5 Variabili di grandi dimensioni	58
17 Introduzione alla probabilità	59
17.1 Variabili casuali	59
17.2 Processi casuali	72
18 Statistica inferenziale	75
18.1 Stima	76
18.2 Test di verifica delle ipotesi	77
19 Dimostrazioni	101
19.1 Distribuzione gaussiana come limite della distribuzione t -Student	101
19.2 Somma/differenza di variabili gaussiane indipendenti	102
19.3 Distribuzione della statistica $t \sim t_{N-1}$ nel t -test per campioni singoli	102
19.4 Indipendenza di media campionaria e varianza campionaria per variabili gaussiane iid	103
19.5 $\sum_{n=1}^N \left(\frac{X_n - \bar{X}}{\sigma^2} \right)^2 \sim \chi_{N-1}^2$ per variabili gaussiane iid	103
 IV Introduzione ai metodi in statistica e AI	 105
20 Introduzione ai metodi in AI	107
21 Supervised Learning	109
22 Unsupervised Learning	111
22.1 PCA	111
22.2 ICA e PCA	114
23 Reinforcement Learning	123
 V Supporto tecnico	 125
24 Supporto tecnico	127
Proof Index	129

Questo libro fa parte del materiale pensato per le scuole superiori, nell'ambito del progetto **basics-books**. E' disponibile in versione in .pdf.

Obiettivi generali. Questo lavoro punta ad essere un'opera di *formazione alla tirchieria, prigrizia e onestà*, almeno in ambito informatico. I più benevoli potranno condividere questo obiettivo, riassumibile come il naturale desiderio di *non voler buttare nel WC soldi, tempo, e pazienza*.

Obiettivi generali

Oltre alle nozioni minime, l'obiettivo principale di questo libro è la formazione a:

- **indipendenza** in ambito informatico: evitare di pagare per qualcosa che non serve; evitare di pagare per qualcosa di inutile o dannoso in presenza di alternative libere; meglio dedicare le risorse a ciò che vale la pena pagare
- **ordine**: i moderni strumenti informatici permettono di lavorare in maniera ordinata, risparmiando tempo, soldi e pazienza
- **trasparenza e onestà**: anche se spesso in maniera non lineare, la conoscenza procede seguendo il metodo scientifico: i risultati mostrati e le tesi proposte devono essere supportate da dati e logica; i dati e le analisi svolte per poter produrre risultati devono essere disponibili, controllabili e soggetti a critica. Tutto il resto, almeno qui, almeno nell'ambito della conoscenza che procede con il metodo scientifico, sarà considerata confusione nella migliore delle ipotesi o direttamente *merda*.

Questo stesso libro è scritto seguendo questi criteri: oltre al dispositivo elettronico usato per consultare il materiale (online o offline, una volta scaricato), non è necessaria la spesa per nessun altro dispositivo o infrastruttura informatica; il materiale è sviluppato localmente, ospitato e disponibile su [Github](https://github.com/Basics2022/bbooks-programming-hs) all'indirizzo <https://github.com/Basics2022/bbooks-programming-hs>.

necessità di una connessione internet, se non si porta il progetto su un sistema locale, con tutti gli strumenti necessari - non tanti, e standard, ma comunque devono esserci «per funzionare»

Metodo.

- **Impostazione degli strumenti necessari. todo...**
- **Linguaggio di programmazione.** In questa introduzione si sceglie di usare **Python** come linguaggio di programmazione. Un approccio più generale all'informatica e alla programmazione prevederebbe l'utilizzo di altri linguaggio di programmazione (come C). Considerata la **diffusione** di Python, la quantità di **librerie disponibili** (con eventuali binding a librerie sviluppate usando linguaggi di programmazione più efficienti) e strumenti per la **programmazione collaborativa** e remota (**Colab, Jupyter,...**), Python risulta comunque una buona scelta per un corso improntato alla presentazione delle basi di programmazione dirette a un'applicazione abbastanza immediata.
Verranno posta attenzione sulla sintassi «particolare» di Python quando si discosta maggiormente dagli altri linguaggi di programmazione.
- **Argomenti. todo...**

Parte I

Introduzione alla programmazione

Introduzione alla programmazione - in Python

L'introduzione alla programmazione in Python userà le risorse messe a disposizione da [Google](#) con il progetto [Jupyter](#) per lo sviluppo di codice open-source, con open-standard e servizi interattivi utilizzabili su diversi dispositivi usando diversi linguaggi di programmazione, come [Python](#), [Julia](#) o [R](#)

Oltre al dispositivo elettronico utilizzato per consultare al materiale, non è necessario nessun altro dispositivo informatico: un account Google personale permette l'accesso libero ai servizi base di cloud computing di [Colab](#)

Variabili, tipi e funzioni elementari built-in

- tipi
- variabili
 - ...
 - by-reference o by-value

2.1 Commenti

```
# I commenti permettono di aggiungere brevi descrizione al codice

# In Python, è possibile aggiungere commenti al codice con il carattere #: tutto_
↳quello
# che viene dopo il carattere # su una riga è considerato un commento, e non codice da
# eseguire

# E' buona regola aggiungere qualche commento al codice, e sarebbe bene iniziare a_
↳farlo
# in lingua inglese:
# - il codice non è auto-esplicativo!
# - il codice potrebbe essere usato da altri in giro per il mondo, ed è più probabile_
↳che
#   si conosca l'inglese invece dell'italiano

# So let's switch to English for scripts, both for comments and "for variable names"

# Comments are not documentation! todo Add some paragraph about documentation!
```

2.2 Tipi built-in

- numero: intero, reale, complesso
- booleano
- stringa
- bytes
- lista
- tupla
- insieme
- dizionario, dict

2.2.1 Numeri

```
"""
Numbers

in Python, variables are not declared. Thus, a number variable is not defined
as an integer, a real or a complex variable, but its type is inferred by its
initialization
"""

# Numbers
a_int = 1
a_real = 1.
a_complex = 1.+ 0.j

# Strings
a_str = '1.0'

print(f"type(a_int)      : {type(a_int)}")
print(f"type(a_real)    : {type(a_real)}")
print(f"type(a_complex): {type(a_complex)}")
print(f"type(a_str)     : {type(a_str)}")
```

```
type(a_int)      : <class 'int'>
type(a_real)     : <class 'float'>
type(a_complex): <class 'complex'>
type(a_str)      : <class 'str'>
```

2.2.2 Booleani - logici

2.2.3 Stringhe

```
"""  
Strings  
  
strings are character  
"""
```

```
'\nStrings\n\nstrings are character \n'
```

2.2.4 Liste, tuple e insiemi

2.2.5 Dizionari

Controllo del flusso

Nei paradigmi di *programmazione imperativa* (**todo** fare riferimento ai paradigmi di programmazione. Ha senso questa distinzione?), vengono usate delle strutture di controllo del flusso di esecuzione di un programma.

Si possono distinguere due categorie delle strutture di controllo:

- condizionale ed alternativa: if, if-then, if-then-else
- iterazione: for, while, ...

3.1 Alternativa

3.1.1 if-then statement

```
""" if-then example """
# Try this script changing the user input

# User input
a = 2          # a is initialize as an integer
# a = 2.1      # if a is initialized as a real, a % 2 perform automatic casting,
↳int(a) % 2    # uncomment previous line and try!

word = 'odd'
if ( a % 2 == 0 ):    # automatic casting into an int -> a % 2 = floor(a) % 2
    word = 'even'

print(f"User input ({a}) is an {word} number")
```

```
User input = 2 is an even number
```

3.1.2 if-then-else statement

```
""" if-then-else example"""
# Try this script changing the user input

# User input
a = 15

if ( a % 3 == 1 ):      # First condition
    reminder = 1
elif ( a % 3 == 2 ):   # Other condition
    reminder = 2
else:                  # All the other conditions
    reminder = 0

print(f"User input ({a}). {a} % 3 = {reminder}")
```

```
User input (15.7). 15.7 % 3 = 0
```

3.2 Iterazione

3.2.1 for loop

```
""" for loop examples

Loops over:
- elements of a list
- elements in range
- keys, values of a dict
- ...
"""

# Loop over elements of a list
seq = ['a', 3, 4. , {'key': 'value'}]

print("\nLoop over elements of the list: seq = {seq}")
for el in seq:
    print(f"element {el} has type {type(el)}")

# Loop over elements of a tuple
# ...

# Loop over elements of a range
n_el = 5
range_el = range(5)
print("\nLoop over elements of the list, seq = {seq}")
print(f"range({n_el}) has type: {type(range_el)}")
print(f"range({n_el}): {range_el}")

for i in range_el:
    print(i)
```

(continues on next page)

(continua dalla pagina precedente)

```
# Loop over keys, values of a dict
d = {'a': 1., 'b': 6, 'c': {'c1': 1, 'c2': True}}
print(f"\nLoop over elements of the dict, d = {d}")

for i,k in d.items():
    print(i, k)
```

```
Loop over elements of the list: seq = {seq}
element a has type <class 'str'>
element 3 has type <class 'int'>
element 4.0 has type <class 'float'>
element {'key': 'value'} has type <class 'dict'>
```

```
Loop over elements of the list, seq = {seq}
range(5) has type: <class 'range'>
range(5): range(0, 5)
0
1
2
3
4
```

```
Loop over elements of the dict, d = {'a': 1.0, 'b': 6, 'c': {'c1': 1, 'c2': True}}
a 1.0
b 6
c {'c1': 1, 'c2': True}
```

3.2.2 while loop

```
""" while loop example """

a = 3

while ( a < 5 ):
    a += 1
    print(f">> in while loop, a: {a}")

print(f"after while loop, a: {a}")
```

```
>> in while loop, a: 4
>> in while loop, a: 5
after while loop, a: 5
```

3.2.3 altri cicli

todo

```
#> Define a function to tell if a number is positive or not
def is_positive(x):
    """ Function returning True if x>0, False if x<=0 """
    return x > 0

#> User input to test the function: tuple of numbers
n_tuple = [ -1, 2., .003, 1./3., -2.2, 0, -7.4 ]

#> Test function on all the elements in the user-defined tuple
for n in n_tuple:
    if ( is_positive(n) ):
        string = ' '
    else:
        string = 'not '

    print(f'{n} is '+string+'positive')
```

```
-1 is not positive
2.0 is positive
0.003 is positive
0.3333333333333333 is positive
-2.2 is not positive
0 is not positive
-7.4 is not positive
```

4.1 Default value of function arguments

```
#> Define a user function to tell is the first argument is greater than the second.
# If no second argument is given, it's set = 0 (default value, defined in the
↪function)
def is_greater_than(x, y=0):
    """ """
    return is_positive(x-y)

a, b = -2, -3
print(f"Is {a} greater than {b}? is_greater_than({a}, {b}):"+str(is_greater_than(a,
↪b)))
print(f"Is {a} greater than 0 ? is_greater_than({a},    ):"+str(is_greater_than(a)))
```

```
Is -2 greater than -3? is_greater_than(-2, -3):True
Is -2 greater than 0 ? is_greater_than(-2,    ):False
```

CAPITOLO 5

Classi

- classi, metodi, oggetti
- ereditarietà, overloading

CAPITOLO 6

Librerie

- Definizione del concetto
- librerie «standard», scritte da qualcun'altro
- scrivere una libreria

Parte II

Introduzione al calcolo scientifico

Introduzione al calcolo scientifico

In questa introduzione al calcolo numerico, vengono presentati alcuni algoritmi. Dove sensato, viene implementata la versione elementare di alcuni di questi algoritmi. Per usi non didattici, e quando possibile, si raccomanda l'uso di algoritmi implementati in librerie disponibili, per questioni di tempo ed efficienza: è lavoro già fatto, da persone che lo sanno fare meglio di noi, controllato, migliorato nel corso degli anni, ottimizzato per ogni sistema e spesso in linguaggi di programmazione diversi da Python, come C o Fortran.

In questa introduzione viene fatto affidamento e uso di alcune librerie disponibili per Python:

- librerie con algoritmi e strumenti matematici per il calcolo numerico: [NumPy](#), [SciPy](#),...
- librerie per la creazione di grafici: [Matplotlib](#), [Plotly](#),...
- librerie per l'analisi dati e la statistica: [pandas](#),...
- librerie per il machine learning: [sci-kit](#), [PyTorch](#),...
- ...

Introduzione al calcolo numerico

- Equazioni lineari
- Equazioni non lineari
- Approssimazione di funzioni
- Derivate
- Integrali
- Equazioni differenziali ordinarie:
 - problema di Cauchy ai valori iniziali
 - problema ai valori al contorno
- Ottimizzazione, vincolata e non

Metodi per la statistica

Introduzione al machine learning

La soluzione di sistemi lineari è un problema che compare in molte altre applicazioni di calcolo numerico.

Formalismo matriciale. Con il formalismo matriciale, un sistema di equazioni lineari può essere scritto come

$$\mathbf{Ax} = \mathbf{b}$$

Classificazione. In generale, i sistemi di equazioni lineari possono essere classificati:

- in base al numero di incognite n_u ed equazioni indipendenti n_e : $n_e = n_u$ sistemi determinati con un'unica soluzione; $n_e > n_u$ sistemi sovradeterminati: con nessuna soluzione in generale; $n_e < n_u$ sistemi indeterminati, con infinite soluzioni in generale
- in base alla «struttura» del sistema:
 - diagonale, tridiagonale, ...
- in base al numero di coefficienti non-nulli della matrice \mathbf{A} : sistemi con matrice \mathbf{A} piena o **sparsa**; questa distinzione non è netta, ma il più delle volte risulta chiara dalla particolare applicazione/metodo.

Algoritmi. Esistono due grandi classi di metodi/algoritmi per la soluzione di sistemi lineari:

- i **metodi diretti**, che si basano su una fattorizzazione della matrice
- i **metodi indiretti**, che si basano sul calcolo di prodotti matrice-vettore

8.1 Sistemi lineari quadrati con matrici piene

In questa sezione si discute la soluzione di sistemi lineari quadrati con matrici piene con le funzioni disponibili nella libreria NumPy.

8.1.1 Esempio 1. Sistema quadrato determinato

Il sistema lineare

$$\begin{cases} x_1 + 2x_2 = 0 \\ x_1 + x_3 = -1 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

può essere riscritto con il formalismo matriciale nella forma $\mathbf{Ax} = \mathbf{b}$,

$$\underbrace{\begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}}_{\mathbf{b}}$$

e risolto grazie alla funzione `solve(A, b)` della libreria `numpy.linalg`.

```
"""
Linear systems with full square non-singular matrices
"""

import numpy as np

A = np.array([[1., 2., 0.], [1., 0., 1.], [1., 1., 1.]])
b = np.array([0., -1., 1.])

x = np.linalg.solve(A, b)

print(f"Sol, x: {x}")
print(f"Proof : Ax = {A @ x}")    # Check that Ax = b
print(f"      b = {b}")
```

```
Sol, x: [-4.  2.  3.]
Proof : Ax = [ 0. -1.  1.]
           b = [ 0. -1.  1.]
```

8.1.2 Esempio 2. Sistemi quadrati non determinati

I sistemi lineari

$$\begin{cases} x_1 + 2x_2 = 1 \\ x_1 + x_3 = -1 \\ 2x_1 + 2x_2 + x_3 = 1 \end{cases}, \quad \begin{cases} x_1 + 2x_2 = 0 \\ x_1 + x_3 = -1 \\ 2x_1 + 2x_2 + x_3 = 1 \end{cases}$$

sono due sistemi quadrati non determinati. Il primo sistema non ha soluzioni, mentre il secondo ne ha infinite della forma

$$(x_1, x_2, x_3) = (-2, 1, 1) + \alpha(2, -1, -2), \quad \alpha \in \mathbb{R}.$$

Dopo aver riscritto i sistemi lineari con il formalismo matriciale, si può provare a risolverli usando la funzione `solve(A, b)` della libreria `numpy.linalg`. In entrambi i casi, la funzione `solve(A, b)` restituisce un errore, segnalando che la matrice del sistema lineare è singolare, definizione equivalente di sistemi non determinati.

- **todo** dare interpretazione geometrica, fare grafico?
- **todo** spiegare motivo?

- *Esistono algoritmi che trovano almeno una soluzione nel caso in cui ne esistano infinite?*: discutere gli algoritmi implementati nella funzione `numpy.linalg.solve()` e rimandare alla documentazione della libreria; discutere altri algoritmi che rendono possibile trovare una soluzione
- *Esistono algoritmi che trovano una soluzione approssimata nel caso in cui non ne esistano?*: **minimi quadrati**, minimizzano l'errore, dare un'interpretazione geometrica

```
"""
Linear systems with full square singular matrices
"""

import numpy as np

A = np.array([[1., 2., 0.], [1., 0., 1.], [2., 2., 1.]])
b = np.array([1., -1., 1.])
# b = np.array([0., -1., 1.])

x = np.linalg.solve(A, b)

print(f"Sol, x: {x}")
print(f"Proof : Ax = {A @ x}")    # Check that Ax = b
print(f"      b = {b}")
```

```
-----
LinAlgError                                Traceback (most recent call last)
Cell In[2], line 11
      8 b = np.array([1., -1., 1.])
      9 # b = np.array([0., -1., 1.])
--> 11 x = np.linalg.solve(A, b)
      13 print(f"Sol, x: {x}")
      14 print(f"Proof : Ax = {A @ x}")    # Check that Ax = b

File <__array_function__ internals>:200, in solve(*args, **kwargs)

File ~/.local/lib/python3.8/site-packages/numpy/linalg/linalg.py:386, in solve(a, b)
    384 signature = 'DD->D' if isComplexType(t) else 'dd->d'
    385 extobj = get_linalg_error_extobj(_raise_linalgerror_singular)
--> 386 r = gufunc(a, b, signature=signature, extobj=extobj)
    388 return wrap(r.astype(result_t, copy=False))

File ~/.local/lib/python3.8/site-packages/numpy/linalg/linalg.py:89, in _raise_linalgerror_singular(err, flag)
    88 def _raise_linalgerror_singular(err, flag):
--> 89     raise LinAlgError("Singular matrix")

LinAlgError: Singular matrix
```

8.2 Matrici sparse

Una matrice sparsa ha un elevato numero di elementi nulli. Una matrice sparsa viene definita in maniera efficiente salvando in memoria solo gli elementi non nulli (**limiti di memoria**); gli algoritmi per le matrici sparse risultano spesso efficienti perché evitano un molte operazioni che darebbero risultati parziali nulli (**velocità**).

- **todo** dire due parole sui formati
- **todo** fare esempio di calcolo del prodotto matrice vettore per matrici sparse

8.2.1 Esempio 1 - Matrice di rigidezza di elementi finiti

Il sistema lineare

$$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

è descritto da una matrice, $N = 5$, $N \times N = 25$, che ha $N + (N - 1) + (N - 1) = 13$ elementi non nulli. Il rapporto tra il numero di elementi non nulli e il numero di elementi totali è $\frac{3N-2}{N^2} \sim \frac{3}{N}$. Al crescere della dimensione del problema, la matrice **A** diventa sempre più sparsa e diventa sempre più conveniente definirla come matrice sparsa, ed usare gli algoritmi pensati per questo tipo di matrici.

```
"""
Linear systems with square non-singular matrices, in sparse format
"""

from scipy import sparse

# Printout level: the higher the number, the more verbose the script
printout_level = 1

n_nodes = 5
i_nodes = list(np.arange(5))

# Build sparse stiffness matrix, I: row indices, J: col indices, E: matrix elems
I = np.array(i_nodes+i_nodes[:-1]+i_nodes[ 1:])
J = np.array(i_nodes+i_nodes[ 1:]+i_nodes[:-1])
E = np.array(n_nodes*[2]+(n_nodes-1)*[-1]+(n_nodes-1)*[-1])

A = sparse.coo_array((E, (I,J))).tocsr()

if ( printout_level > 50 ): # print matrix in sparse format
    print(f" I: {I}\n J: {J}\n E: {E}")
    print(f" A:\n {A}")

if ( printout_level > 60 ): # convert and print matrix in full format
    print(f" A.todense(): {A.todense()}")

# RHS
b = np.array(5*[1])

# Solve linear system
x = sparse.linalg.spsolve(A, b)
```

(continues on next page)

(continua dalla pagina precedente)

```
print(f"Sol, x: {x}")
```

```
Sol, x: [2.5 4.  4.5 4.  2.5]
```

Equazioni algebriche non lineari

Questa sezione si occupa della soluzione delle equazioni algebriche non lineari, distinguendo le equazioni non lineari con una sola incognita $x \in \mathbb{R}$

$$f(x) = 0 ,$$

e i sistemi di equazioni non lineari con un numero di incognite pari al numero di equazione,

$$\mathbf{f}(\mathbf{x}) = \mathbf{0} .$$

9.1 Equazioni non lineari

Vengono presentati i metodi di bisezione e di Newton per la soluzione di un'equazione non lineare,

$$f(x) = 0 ,$$

e applicati alla soluzione del problema con $f(x) = e^x + x$, la cui derivata è nota e immediata da calcolare $f'(x) = e^x + 1$. L'espressione della derivata verrà utilizzata nel metodo di Newton.

```
""" Import libraries """
```

```
import numpy as np  
from time import time
```

```
"""  
Example.  $f(x) = e^x + x$   
"""
```

```
# Function  $f$  and its derivative  
f = lambda x: np.exp(x) + x  
df = lambda x: np.exp(x) + 1
```

9.1.1 Metodo di bisezione

Il metodo di bisezione per la ricerca degli zeri di una funzione continua $F(x)$ si basa sul teorema dei valori intermedi per le funzioni continue.

Dati due numeri reali a, b tali che $f(a)f(b) < 0$, allora esiste un punto $c \in (a, b)$ tale che $f(c) = 0$.

```
"""
Define bisection_method_scalar() function to solve nonlinear scalar equations with
↪ bisection method
"""

def bisection_method_scalar(f, a, b, tol=1e-6, max_niter=100):
    """ Function implementing the bisection method for scalar equations """

    niter = 0

    if ( not f(a) * f(b) < 0 ):
        print("Bisection algorithm can't start, f(a)f(b)>= 0")
    else:
        x = .5 * (a+b)
        fx = f(x)
        while ( np.abs(fx) > tol and niter < max_niter ):

            if ( f(x) * f(a) <= 0 ): # new range [a,c]
                b = x
            else: # new range [a,b]
                a = x

            # Update solution and residual
            x = .5 * (a+b)
            fx = f(x)

            # Update n.iter
            niter += 1

    return x, np.abs(fx), niter, max_niter
```

```
""" Use bisection_method_scalar() function to solve the example """

# Find 2 values so that $f(a) f(b) < 0$
a, b = -2., 0.

t1 = time()
x, res, niter, max_niter = bisection_method_scalar(f, a, b,)

print("Bisection method summary: ")
if ( niter < max_niter ):
    print(f"Convergence reached")
    print(f"Sol, x = {x}")
else:
    print(f"max n.iter reached without convergence")

print(f"residual      : {f(x)}")
print(f"n. iterations: {niter}")
print(f"elapsed time : {time()-t1}")
```

```

Bisection method summary:
Convergence reached
Sol, x = -0.567143440246582
residual      : -2.348157265297246e-07
n. iterations: 20
elapsed time  : 0.0006182193756103516

```

9.1.2 Metodo di Newton

Per trovare la soluzione del problema non lineare

$$f(x) = 0,$$

il metodo di Newton sfrutta l'espansione in serie troncata al primo grado della funzione $f(x)$, per scrivere

$$0 = f(x^n + \Delta x) \approx f(x^n) + f'(x^n)\Delta x$$

e ottenere l'incremento della soluzione Δx come soluzione del sistema lineare

$$f'(x^n)\Delta x = -f(x^n)$$

e aggiornare la soluzione $x^{n+1} = x^n + \Delta x$.

```

"""
Define newton_method_scalar() function to solve nonlinear scalar equations with Newton
↪ 's method
"""

def newton_method_scalar(f, df, x=.0, tol=1e-6, max_niter=100):
    """ Function implementing Newton's method for scalar equations """

    res = f(x)
    niter = 0

    # Newton algorithm
    while ( np.abs(res) > tol and niter < max_niter ):
        # Solve linear approximation step, and update solution
        dx = - res / df(x)
        x += dx

        #> Evaluate new residual and n. of iter
        res = f(x)
        niter += 1

    return x, res, niter, max_niter

```

```

""" Use newton_method_scalar() function to solve the example """

# import numpy as np    # already imported

# Parameters of the Newton method, for stopping criteria
# tol = 1e-6            # tolerance on the residual |f(x)| < tol
# max_niter = 10        # max n. of iterations          niter > max_niter
x0 = -1.

```

(continues on next page)

(continua dalla pagina precedente)

```

t1 = time()
x, res, niter, max_niter = newton_method_scalar(f, df, x=x0)

print("Newton's method summary: ")
if ( niter < max_niter ):
    print("Convergence reached")
    print(f"Sol, x = {x}")
else:
    print(f"max n.iter reached without convergence")

print(f"residual      : {f(x)}")
print(f"n. iterations: {niter}")
print(f"elapsed time : {time()-t1}")

```

```

Newton's method summary:
Convergence reached
Sol, x = -0.567143285989123
residual      : 6.927808993140161e-09
n. iterations: 3
elapsed time : 0.0005326271057128906

```

9.2 Sistemi di equazioni non lineari

9.2.1 Metodo di Newton

Il metodo di Newton sfrutta l'espansione lineare della funzione $\mathbf{f}(\mathbf{x})$ nell'intorno di un valore \mathbf{x} ,

$$\mathbf{0} = \mathbf{f}(\mathbf{x} + \mathbf{h}) \simeq \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x}) \mathbf{h}$$

per costruire un metodo iterativo composto da due passi a ogni iterazione:

- ricerca dell'incremento:

$$\mathbf{f}'(\mathbf{x}^n) \mathbf{h}^{(n)} = -\mathbf{f}(\mathbf{x}^{(n)})$$

- aggiornamento della soluzione

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \mathbf{h}^{(n)} .$$

Esempio. Il metodo di Newton viene applicato al sistema non lineare

$$\begin{cases} x_0 - x_1 = 0 \\ -x_0^2 + x_1 = -1 \end{cases}$$

che può essere scritto con il formalismo matriciale come

$$\mathbf{0} = \mathbf{f}(\mathbf{x}) = \begin{bmatrix} x_0 - x_1 \\ -x_0^2 + x_1 - 1 \end{bmatrix} .$$

La derivata della funzione $\mathbf{f}(\mathbf{x})$, rispetto alla variabile indipendente \mathbf{x} ,

$$\mathbf{f}'(\mathbf{x}) = \begin{bmatrix} 1 & -1 \\ -2x_0 & 1 \end{bmatrix}$$

può essere rappresentata come una matrice che ha come elemento alla riga i e alla colonna j la derivata della funzione $f_i(\mathbf{x})$ rispetto alla variabile x_j , $[f']_{ij} = \frac{\partial f_i}{\partial x_j}$, così che l'approssimazione al primo ordine dell'incremento della funzione può essere scritto come

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{h} + o(\|\mathbf{h}\|) .$$

```
"""
Example. f(x) = np.array([ x[0]      - x[1]
                          -x[0]**2 + x[1] - 1 ])
"""

# Function f and its derivative
f = lambda x: np.array([ x[0] - x[1], -x[0]**2 + x[1] + 1])
df = lambda x: np.array([[1, -1], [-2*x[0], 1]])
```

```
"""
Define newton_method_scalar() function to solve nonlinear systems of equations with
↳Newton's method
"""

def newton_method_system(f, df, x=.0, tol=1e-6, max_niter=100):
    """ Function implementing Newton's method for systems of equations """

    res = f(x)
    niter = 0

    # Newton algorithm
    while ( np.linalg.norm(res) > tol and niter < max_niter ):
        # Solve linear approximation step, and update solution
        dx = - np.linalg.solve(df(x), res)
        x += dx

        #> Evaluate new residual and n. of iter
        res = f(x)
        niter += 1

    return x, res, niter, max_niter
```

```
""" Use newton_method_scalar() function to solve the example """

# import numpy as np    # already imported

# Parameters of the Newton method, for stopping criteria
# tol = 1e-6            # tolerance on the residual |f(x)| < tol
# max_niter = 10        # max n. of iterations          niter > max_niter
x0 = np.array([ -1. , 1 ])

t1 = time()
x, res, niter, max_niter = newton_method_system(f, df, x=x0)

print("Newton's method summary: ")
if ( niter < max_niter ):
    print(f"Convergence reached")
    print(f"Sol, x = {x}")
else:
    print(f"max n.iter reached without convergence")
```

(continues on next page)

(continua dalla pagina precedente)

```
print(f"residual      : {f(x)}")
print(f"n. iterations: {niter}")
print(f"elapsed time : {time()-t1}")
```

```
Newton's method summary:
Convergence reached
Sol, x = [-0.61803399 -0.61803399]
residual      : [ 0.00000000e+00 -2.10942375e-13]
n. iterations: 4
elapsed time  : 0.0006396770477294922
```

- **todo** L'algoritmo di Newton trova solo una soluzione del problema. Cercare le altre soluzioni cambiando il tentativo iniziale.
- **todo** ... altro?

CAPITOLO 10

Approssimazione di funzioni

10.1 Interpolazione

10.2 Regressione

11.1 Differenze finite

Il calcolo della derivata di una funzione $f(x)$ derivabile in un punto x_0 può essere svolto utilizzando l'espansione locale in serie di Taylor di una funzione.

11.1.1 Derivata prima

Usando le espansioni

$$\begin{aligned}f(x+h) &= f(x) + f'(x)h + f''(x)\frac{h^2}{2} + f'''(x)\frac{h^3}{3!} + o(h^3) \\f(x-h) &= f(x) - f'(x)h + f''(x)\frac{h^2}{2} - f'''(x)\frac{h^3}{3!} + o(h^3)\end{aligned}$$

si possono ricavare:

- gli schemi del **primo ordine**

$$\begin{aligned}f'(x) &= \frac{f(x+h) - f(x)}{h} + O(h) \\f'(x) &= \frac{f(x) - f(x-h)}{h} + O(h)\end{aligned}$$

- lo schema **centrato del secondo ordine**

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

- lo schema non centrato del secondo ordine:

$$f'(x) = \frac{-3f(x) + 4f(x+h) - f(x+2h)}{2h} + O(h^2).$$

Dimostrazione

Usando le espansioni in serie,

$$f(x+h) = f(x) + f'(x)h + f''(x)\frac{h^2}{2} + f'''(x)\frac{h^3}{3!} + o(h^3)$$

$$f(x+2h) = f(x) + f'(x)2h + 2f''(x)h^2 + f'''(x)\frac{4}{3}h^3 + o(h^3)$$

si cerca una coppia di coefficienti della combinazione lineare $a_1 f(x+h) + a_2 f(x+2h)$ che annullano il termine di secondo grado.

In particolare è facile dimostrare che una di queste scelte è $\alpha_1 = 4, \alpha_2 = -1$, e la combinazione lineare per questi valori diventa,

$$4f(x+h) - f(x+2h) = 3f(x) + 2hf'(x) + O(h^3).$$

A questo punto è semplice isolare $f'(x)$ per trovare lo schema numerico desiderato,

$$f'(x) = \frac{-3f(x) + 4f(x+h) - f(x+2h)}{2h} + O(h^2).$$

- ...
- schemi di ordine superiore...

11.1.2 Derivata seconda

Usando le stesse espansioni in serie, si può ottenere uno schema del secondo ordine per la derivata seconda

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2)$$

Dimostrazione

Usando le espansioni in serie

$$f(x+h) = f(x) + f'(x)h + f''(x)\frac{h^2}{2} + f'''(x)\frac{h^3}{3!} + f^{iv}(x)\frac{h^4}{4!} + O(h^5)$$

$$f(x-h) = f(x) - f'(x)h + f''(x)\frac{h^2}{2} - f'''(x)\frac{h^3}{3!} + f^{iv}(x)\frac{h^4}{4!} + O(h^5)$$

si può notare che nella somma che compare al numeratore dello schema numerico si annullano sia il termine di primo grado sia il termine di terzo grado,

$$f(x+h) - 2f(x) + f(x-h) = f''(x)h^2 + O(h^4),$$

e quindi lo schema numerico proposto è del secondo ordine,

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2).$$

```
""" Numerical schemes for derivatives """
```

```
df_first_order_left    = lambda f, x, h: (f(x) - f(x-h)) / h
df_first_order_right   = lambda f, x, h: (f(x+h) - f(x)) / h
df_second_order_center = lambda f, x, h: (f(x+h) - f(x-h)) / ( 2. * h )
df_second_order_left   = lambda f, x, h: ( - 3*f(x) - 4*f(x-h) + f(x-2*h)) / ( 2. * h )
df_second_order_right  = lambda f, x, h: ( - 3*f(x) + 4*f(x+h) - f(x+2*h)) / ( 2. * h )

df_fun_dict = {
    '1_left' : df_first_order_left,
    '1_right': df_first_order_right,
    '2_center': df_second_order_center,
    '2_left' : df_second_order_left,
    '2_right': df_second_order_right
}
```

```
""" Example """
```

```
import numpy as np
```

```
f = lambda x: 2. * np.exp(- x**2)
```

```
x, h = 1., .01
```

```
for df_label, df_fun in df_fun_dict.items():
    print(f"Scheme: {df_label.ljust(10)}, value: {df_fun(f,x,h)}" )
```

```
Scheme: 1_left      , value: -1.4788256893130014
Scheme: 1_right     , value: -1.4641117378933477
Scheme: 2_center    , value: -1.4714687136031745
Scheme: 2_left      , value: -1.4716195509633268
Scheme: 2_right     , value: -1.4716121945026028
```



```
""" Import libraries """
import numpy as np
```

12.1 Integrazione di Newton-Cotes

Valutazioni di integrali definiti di funzioni $f : [a, b] \in \mathbb{R} \rightarrow \mathbb{R}$, continue sull'intervallo. Si veda l'*esempio* per intuire la necessità della continuità della funzione.

- Formula del punto medio
- Formula del trapezio

12.1.1 Metodo di integrazione del punto medio

Il metodo di integrazione del punto medio ricorda la definizione dell'integrale di Riemann (**todo** aggiungere link). Data la funzione $f : [a, b] \in \mathbb{R} \rightarrow \mathbb{R}$, e una partizione $P = \{a = x_0 < x_1 < \dots < x_n = b\}$ dell'intervallo $[a, b]$, l'integrale viene calcolato come la somma dell'area dei rettangoli elementari di lati $\Delta x_k := x_k - x_{k-1}$ e $f(\xi_k)$, con $\xi_k \in [x_{k-1}, x_k]$,

$$\int_{x=a}^b f(x) dx \simeq \sum_{k=1:n} f(\xi_k) \Delta x_k .$$

```
""" Mid-point method """

def integral_rect(f, a, b, n):
    """
    Inputs:
    - f: function
    - a, b: extreme points of the range
```

(continues on next page)

(continua dalla pagina precedente)

```

- n: n. of intervals of the range
"""
# Partition of the range [a,b]
# uniform partition here; refined algorithms may use non-uniform partitions
xv = a + ( b - a ) * np.arange(n+1) / n;  xv[-1] = b
xc = .5 * ( xv[1:] + xv[:-1] )
dx = xv[1:] - xv[:-1]

return np.sum( dx * f(xc) )

```

12.1.2 Metodo di integrazione del trapezio

Data la funzione $f : [a, b] \in \mathbb{R} \rightarrow \mathbb{R}$, e una partizione $P = \{a = x_0 < x_1 < \dots < x_n = b\}$ dell'intervallo $[a, b]$, l'integrale viene calcolato come la somma dell'area dei trapezi rettangoli elementari con basi $f(x_{k-1})$, $f(x_k)$ e altezza $\Delta x_k := x_k - x_{k-1}$,

$$\int_{x=a}^b f(x)dx \simeq \sum_{k=1:n} \frac{1}{2} (f(x_{k-1}) + f(x_k)) \Delta x_k.$$

```

""" Trapezoid method """

def integral_trapz(f, a, b, n):
    """
    Inputs:
    - f: function
    - a, b: extreme points of the range
    - n: n. of intervals of the range
    """
    # Partition of the range [a,b]
    # uniform partition here; refined algorithms may use non-uniform partitions
    xv = a + ( b - a ) * np.arange(n+1) / n;  xv[-1] = b
    dx = xv[1:] - xv[:-1]

    return np.sum( .5 * dx * ( f(xv[1:]) + f(xv[:-1]) ) )

```

12.1.3 Esempi

Esempio 1.

Si valuta l'integrale

$$\int_{x=0}^1 x^2 dx = \frac{1}{3},$$

con i metodi del punto medio e del trapezio. La funzione $f(x) = x^2$ è continua ovunque e quindi è continua nell'intervallo $[0, 1]$.


```
f = lambda x: x**2

a, b = 0., 1.
n = 10

# Evaluate integral with the rect and trapz methods
val_rect = integral_rect(f, a, b, n)
val_trapz = integral_trapz(f, a, b, n)

print(f"Value of the integral, x \in [{a}, {b}] with {n} intervals")
print(f"- mid-point method: {val_rect}")
print(f"- trapezoid method: {val_trapz}")
```

```
Value of the integral, x \in [0.0, 1.0] with 10 intervals
- mid-point method: 0.3325
- trapezoid method: 0.33499999999999996
```

Esempio 2. Necessità della continuità della funzione.

12.2 Integrazione di Gauss

L'integrazione di Gauss permette di calcolare in **maniera esatta** l'integrale di una **funzione polinomiale** $p_n(x)$ su un intervallo $[a, b]$, come somma pesata della funzione valutata in alcuni punti dell'intervallo,

$$\int_a^b p^{(n)}(x) dx = \sum_g w_g f(x_g),$$

per un numero di nodi di Gauss, n_{Gauss} che soddisfa la relazione

$$n < 2n_{Gauss} - 1.$$

Per motivi di generalizzazione dell'algoritmo, nella definizione dei **pesi** w_i e dei **nodi di Gauss** x_i , l'integrale viene riportato all'integrale su un intervallo di riferimento, tramite una trasformazione di coordinate. Per domini 1D, l'intervallo di riferimento per la quadratura di Gauss è l'intervallo $\xi = [-1, 1]$ e il cambio di variabili è

$$x = \frac{a+b}{2} + \frac{b-a}{2} \xi,$$

così che l'integrale originale può essere scritto come

$$\begin{aligned} \int_{x=a}^b p^{(n)}(x) dx &= \int_{\xi=-1}^1 p^{(n)}(x(\xi)) \frac{dx}{d\xi} d\xi = \\ &= \frac{b-a}{2} \int_{\xi=-1}^1 p^{(n)}(x(\xi)) d\xi = \frac{b-a}{2} \sum_g w_g p^{(n)}(x(\xi_g)) \end{aligned}$$

```
""" Gauss integration """

# Dict of Gauss weights and nodes on the reference interval [-1,1] for 1D integration
gauss_nw = { 1: {'nodes' : [ 0. ],
                 'weights': [ 2. ]},
            2: {'nodes' : [ -1./np.sqrt(3.), 1./np.sqrt(3.) ],
```

(continues on next page)

(continua dalla pagina precedente)

```

        'weights': [ 1., 1.]},
3: {'nodes' : [ 0., -np.sqrt(3./5.), np.sqrt(3./5.) ],
    'weights': [ 8./9., 5./9., 5./9. ]},
}

def gauss_int_1(f, n, a=-1, b=1, gauss_nw=gauss_nw):
    """
    Integral of the function f(x)
    over a physical domain, x \in [a, b],
    using n number of Gauss nodes
    """
    # Cap n.nodes to the max n.nodes stored in gauss_nw dict
    n = np.min([n, np.max(list(gauss_nw.keys()))])

    # Gauss nodes (from reference to actual domain) and weights
    xg = .5 * ( a + b + ( b - a ) * np.array(gauss_nw[n]['nodes']) )
    wg = np.array(gauss_nw[n]['weights'])

    return .5 * (b-a) * np.sum( wg * f(xg) )

def gauss_int_n(f, n_gauss, a, b, n_elems):
    """
    Integral of the function f(x)
    over a physical domain, x \in [a, b], (uniformly) splitted in n_elems
    using n number of Gauss nodes per each elem
    """
    # Partition of the interval [a,b]
    xv = a + ( b - a ) * np.arange(n_elems+1) / n_elems; xv[-1] = b

    return np.sum([ gauss_int_1(f, n_gauss, xv[i], xv[i+1]) for i in np.arange(n_elems) ])

```

12.2.1 Esempio 1 - integrazione di Gauss.

Si valuta l'integrale

$$\int_{x=0}^1 x^2 dx = \frac{1}{3},$$

con il metodo di integrazione di Gauss. Si chiede di:

- osservare che l'integrazione è esatta, a meno degli arrotondamenti dovuti all'aritmetica finita, poiché la funzione integranda è di grado 2, e il numero di nodi di Gauss è $n_{Gauss} = 3$; questo è vero indipendentemente dal numero di sotto-intervalli;
- cambiare le funzioni e il numero di nodi usati nelle funzioni per l'integrazione di Gauss per verificare che l'integrazione è esatta per funzioni polinomiali di grado $2n_{Gauss} - 1$.

```

""" Evaluate integrals using Gauss integration """

f = lambda x: x**2
a, b = 0., 1.

# Gauss nodes, and n.of elements

```

(continues on next page)

(continua dalla pagina precedente)

```
n_gauss = 3
n_elems = 10

val_1 = gauss_int_1(f, n_gauss, 0., 1.)
val_n = gauss_int_n(f, n_gauss, 0., 1., n_elems)

print(f"Value of the integral, x \in [{a}, {b}] using Gauss integration methods with
↳{n_gauss} nodes per elem")
print(f"- using one elem: {val_1}")
print(f"- using {n_elems} elems: {val_n}")
```

```
Value of the integral, x \in [0.0, 1.0] using Gauss integration methods with 3↳
↳nodes per elem
- using one elem: 0.3333333333333337
- using 10 elems: 0.3333333333333337
```


13.1 Problemi di Cauchy ai valori iniziali

13.1.1 Approccio a un problema di Cauchy di ordine n

Un problema di Cauchy di ordine n

$$\begin{cases} F(y^{(n)}(x), y^{(n-1)}(x), \dots, y'(x), y(x), x) = 0 \\ y(x_0) = y^0 \\ y'(x_0) = y'^0 \\ \dots \\ y^{(n-1)}(x_0) = y^{(n-1),0} \end{cases}$$

con funzione incognita $y(x) : D \in \mathbb{R} \rightarrow \mathbb{R}$, può essere riscritto come un problema di «ordine 1» per la funzione incognita $\mathbf{z}(x) : D \in \mathbb{R} \rightarrow \mathbb{R}^n$, definita come

$$\mathbf{z}(x) = (z_0(x), z_1(x), \dots, z_{n-1}(x)) := (y(x), y'(x), \dots, y^{(n-1)}(x)) .$$

Esplicitando le relazioni tra le componenti di $\mathbf{z}(x)$ e le derivate della funzione $y(x)$, $z_k(x) = y^{(k)}(x) = y^{(k-1)'}(x) = z'_{k-1}(x)$, il problema di Cauchy può essere riformulato come

$$\begin{cases} z'_0 - z_1 = 0 \\ z'_1 - z_2 = 0 \\ \dots \\ z'_{n-2} - z_{n-1} = 0 \\ F(z'_{n-1}(x), z_{n-1}(x), \dots, z_1(x), z_0(x)) = 0 , \end{cases} \quad \text{i.c.} \quad \begin{cases} z_0(x_0) = y^0 \\ z_1(x_0) = y'^0 \\ \dots \\ z_{n-1}(x_0) = y^{(n-1),0} \end{cases}$$

che può essere riscritto con il formalismo vettoriale come

$$\begin{cases} \mathbf{F}(\mathbf{z}'(x), \mathbf{z}) = \mathbf{0} \\ \mathbf{z}(x_0) = \mathbf{z}_0 \end{cases}$$

13.1.2 Caratteristiche (cenni)

- accuratezza, consistenza, convergenza
- stabilità: 0-, A- condizionata e incondizionata

13.1.3 Schemi numerici

Schemi numerici a un passo

Schemi numerici multi-step

13.2 Problemi al contorno

13.2.1 Differenze finite

13.2.2 Elementi finiti

13.2.3 Volumi finiti

CAPITOLO 14

Ottimizzazione

Le tecniche di ottimizzazione sono alla base di molti metodi di interesse, dall'approssimazione di funzioni, alla regolazione e controllo, agli algoritmi usati in intelligenza artificiale

Parte III

Statistica

Metodi per la statistica

La statistica si occupa dello studio dei fenomeni in condizioni di incertezza o di non determinismo, determinate da una conoscenza incompleta del modello.

La statistica è applicabile in **diversi (tutti?) gli ambiti**: dalla scienze naturali, alle scienze sociali, ... Secondo uno studio, i 12 metodi statistici (cos'è un metodo statistico?) più usati sono: ANOVA, χ^2 -test, t -test, regressione lineare, coefficiente di correlazione di Pearson, Mann-Whitney U-test, Kruskal-Wallis test, Shannon's diversity index, Turkey's range test, cluster analysis, Spearman's rank correlation coefficient, and PCA.

La statistica è uno strumento fondamentale del **metodo scientifico**, e si occupa dell'intero processo di:

- progetto dell'esperimento
- svolgimento esperimento e raccolta dati
- analisi dati

Statistica descrittiva

La statistica descrittiva si occupa di descrivere e riassumere le informazioni contenute nei dati disponibili, con lo scopo di:

- comunicare il maggior numero di informazioni rilevanti
- nella maniera più semplice e sintetica possibile

Questo, in accordo con la *percezione umana* **todo**, avviene tramite indicazioni di:

- **posizione**, come la media, la mediana, la moda, o loro variazioni sul tema
- **dispersione**, come la deviazione standard, la varianza, lo scarto interquartile,...
- **forma**, come skewness o curtosi, o momenti di ordine superiore
- **correlazione** o **dipendenza** **todo** *correlazione non è dipendenza, fare un riferimento; è necessario dedurre una dipendenza? E' tra i compiti della statistica descrittiva?*

Oggi sono disponibili librerie e strumenti informatici gratuiti e liberi per l'analisi e la rappresentazione dei dati. Ad esempio, native o compatibili con Python,

- pandas, e geoapndas per dati geografici
- [plotly](#) come libreria grafica (con dash come ambiente per lo sviluppo di semplici app interattive)
- ...

Per una galleria di esempi, si rimanda agli esempi e alle risorse messe a disposizione direttamente dagli sviluppatori, come ad esempio i [grafici elementari](#), o i grafici frequentemente usati in [statistica](#), e un ["introduzione"](#) con le impostazioni di base. Per un livello di dettaglio maggiore, è spesso possibile consultare i file sorgenti con l'implementazione degli strumenti resi disponibili su repository pubblici: nel caso di plotly è possibile consultare questi file nel [repository pubblico](#).

16.1 Dimensione dei dati

- variabili 1-dimensionali
- variabili 2-dimensionali o di dimensione piccola
- variabili di grandi dimensioni: possono essere necessarie tecniche di riduzione dei dati per far emergere le proprietà significative contenute nei dati raccolti

16.2 Rappresentazione grafica

EDA: exploratory data analysis. Può essere utilizzata in applicazioni di statistica descrittiva o nelle prime fasi di applicazioni di statistica inferenziale, «per vedere che faccia hanno i dati raccolti» e per fare una prima selezione dei metodi da applicare nelle fasi successive dell'indagine.

Attualmente, sono disponibili i mezzi informatici e un gran numero di librerie per l'analisi dati, dall'elaborazione alla rappresentazione.

16.3 Variabili 1-dimensionali

16.4 Variabili 2-dimensionali

16.5 Variabili di grandi dimensioni

Introduzione alla probabilità

In questa sezione vengono descritti alcuni argomenti di calcolo di probabilità.

Variabili casuali. Vengono definite le variabili casuali e viene mostrato come possano essere uno strumento per rappresentare fenomeni che non è possibile descrivere accuratamente con equazioni deterministiche. Vengono discusse le variabili casuali a valori discreti e valori continui. Vengono poi mostrate alcune trasformazioni e combinazioni di variabili casuali che permettono poi di introdurre alcune distribuzioni di probabilità comuni. Vengono infine presentati alcuni risultati su un insieme di variabili indipendenti identicamente distribuite, che possono essere riassunti nella *legge dei grandi numeri* e nel *teorema del limite centrale*; questi risultati verranno poi utilizzati nel campionamento di una popolazione e nella stima delle sue caratteristiche a partire dal campione osservato, una delle principali attività di statistica inferenziale.

Processi casuali. I processi casuali vengono definiti come variabili casuali che dipendono da una variabile indipendente, t , in molti casi variabile che rappresenta il tempo.

17.1 Variabili casuali

Qualitativamente, una variabile casuale è una grandezza che può assumere valori diversi come risultato di un fenomeno aleatorio, un fenomeno che non è possibile descrivere accuratamente con delle equazioni deterministiche. **todo** fare esempi: lancio del dado e estrema dipendenza dalle condizioni iniziali, soprattutto con urto; aggiungere simulazione, già presente sul drive

Contenuti. *Definizione di variabile casuale*

Variabili casuali discrete

Variabili casuali continue

Trasformazione di funzioni di probabilità

Funzioni di probabilità di r.v. discrete comuni

Funzioni di probabilità di r.v. continue comuni

Variabili casuali indipendenti e identicamente distribuite

Campionamento (**todo** qui o nella parte di statistica inferenziale, stima?)

17.1.1 Definizione di variabile casuale

Qualitativamente, una variabile casuale è una grandezza che può assumere valori diversi come risultato di un fenomeno aleatorio, un fenomeno che non è possibile descrivere accuratamente con delle equazioni deterministiche. Per fare esempi: lancio del dado e estrema dipendenza dalle condizioni iniziali, soprattutto con urto; aggiungere simulazione, già presente sul drive

Una definizione più rigorosa prevede i concetti di spazio di probabilità e σ -algebra. Senza pretendere di fornire una descrizione dettagliata e rigorosa di questi oggetti matematici, non si vuole rinunciare a dare la definizione rigorosa di variabile casuale, fornendo un'interpretazione degli oggetti matematici coinvolti.

Definizione

Dato uno *spazio di probabilità* $(\Omega, \mathcal{F}, \nu)$ e uno *spazio misurabile* (E, \mathcal{E}) , una **variabile aleatoria** è una funzione misurabile $X : \Omega \rightarrow E$.

Ora, cerchiamo di interpretare questa definizione criptica. Iniziamo dalle cose semplici:

- Ω ed E sono due *insiemi*: Ω viene definito *insieme degli eventi* o *spazio campionario*, E è l'insieme dei **valori** che può assumere la variabile casuale.

Se l'insieme E dei possibili valori della variabile casuale è un insieme di elementi discreti, si definisce la variabile casuale **discreta**. Se l'insieme E è invece un insieme di valori continui, come ad esempio un intervallo $I \subset \mathbb{R}$, la variabile viene definita **continua**. Le variabili casuali possono essere **numeriche** o **categoriali**: le variabili casuali numeriche hanno valori numerici, quelle categoriali possono essere identificate da etichette: ad esempio, una variabile X che rappresenta il mezzo preferito di locomozione con valori nell'insieme $\{\text{piedi, bici, treno, auto, altro}\}$ è una variabile categoriale. Le variabili categoriali possono essere o non essere ordinabili; se non sono numeriche, non è possibile compiere operazioni numeriche su di esse.

Continuiamo con gli altri oggetti meno immediati:

- lo spazio di probabilità $(\Omega, \mathcal{F}, \nu)$ può essere pensato come formato da due parti, lo spazio misurabile (Ω, \mathcal{F}) e la misura di probabilità ν ;
- nella definizione ora sono coinvolti due spazi misurabili, (Ω, \mathcal{F}) , (E, \mathcal{E}) , che a loro volta sono composti da due oggetti:
 - un insieme di elementi;
 - una σ -algebra: per quanto interessa a noi, una σ -algebra applicata agli elementi di un insieme offre la possibilità di applicare le operazioni degli insiemi (unione, intersezione,...) agli elementi dell'insieme
- la misura di probabilità ν è una funzione compatibile con gli spazi misurabili coinvolti nella definizione, che traduce in matematica i concetti della probabilità come, ad esempio:
 - $P(\omega \in \Omega) = 1$,
 - $P(\omega \in A \vee \omega \in B) \leq P(\omega \in A) + P(\omega \in B)$, e l'uguaglianza vale se A, B disgiunti
 - ...

Spieghiamoci peggio con due esempi.

todo. L'evento è sempre osservabile? O si può osservare solo il valore assunto dalla variabile casuale?

Esempio 1. Lancio dado non truccato, con variabile casuale corrispondente al valore della faccia

Viene lanciato un dado a 6 facce non truccato. Questo corrisponde a un insieme degli eventi costituito da i 6 eventi distinti $\Omega = \{\text{faccia 1}, \dots, \text{faccia 6}\}$, e ognuno di questi 6 eventi ha una misura di probabilità uguale a $\nu(\text{faccia 1}) = \dots = \nu(\text{faccia 6}) = \frac{1}{6}$. Si sceglie come variabile casuale la funzione che associa al numero letto sulla faccia lo stesso numero, $X : \Omega \rightarrow E$, $X(\text{faccia } n) = n$. Si vogliono ora analizzare le probabilità associate ai seguenti eventi:

- Ω_1 : eventi che portano a risultato 3:

$$P(X(\omega \in \Omega_1) = 3) = P(\omega \in \Omega_1) = \nu(\text{faccia 3}) = \frac{1}{6}$$

- Ω_2 : eventi che portano a risultato pari,

$$\begin{aligned} P(\omega \in \Omega_2) &= P(X(\omega \in \Omega_2) = 2 \vee X(\omega \in \Omega_2) = 4 \vee X(\omega \in \Omega_2) = 6) = \\ &= \nu(\text{faccia 2}) + \nu(\text{faccia 4}) + \nu(\text{faccia 6}) = \frac{1}{2} \end{aligned}$$

- Ω_3 : eventi che portano a risultato dispari
- Ω_4 : eventi che portano a risultato inferiore a 4
- Ω_5 : eventi che portano a un risultato compreso tra 1 e 6 inclusi
- Ω_6 : eventi che verificano gli eventi Ω_2 o gli eventi Ω_4
- Ω_7 : eventi che verificano gli eventi Ω_2 e gli eventi Ω_4

Esempio 2. Lancio dado non truccato, con variabile casuale diverso dal valore della faccia

Si studino ora i casi in cui la variabile casuale associa all'evento:

1. il valore della faccia diviso 2 e aumentato di 1
2. il massimo divisore diverso del numero sulla faccia, diverso dal numero stesso

17.1.2 Variabili casuali discrete

Funzione di probabilità

La funzione di probabilità, $p_X(x)$, o funzione di massa, o *densità di probabilità* con un abuso per analogia con le variabili continue, è la funzione che associa la probabilità al valore x della variabile casuale X . Data la variabile casuale $X : \Omega \rightarrow E$, la probabilità del valore discreto $x \in E$ è data dalla somma delle probabilità degli eventi disgiunti $\Omega_x := \{\omega \in \Omega \mid X(\omega) = x\}$ che producono il valore x ,

$$p_X(x) = \sum_{\omega_i \in \Omega_x} \nu(\omega_i).$$

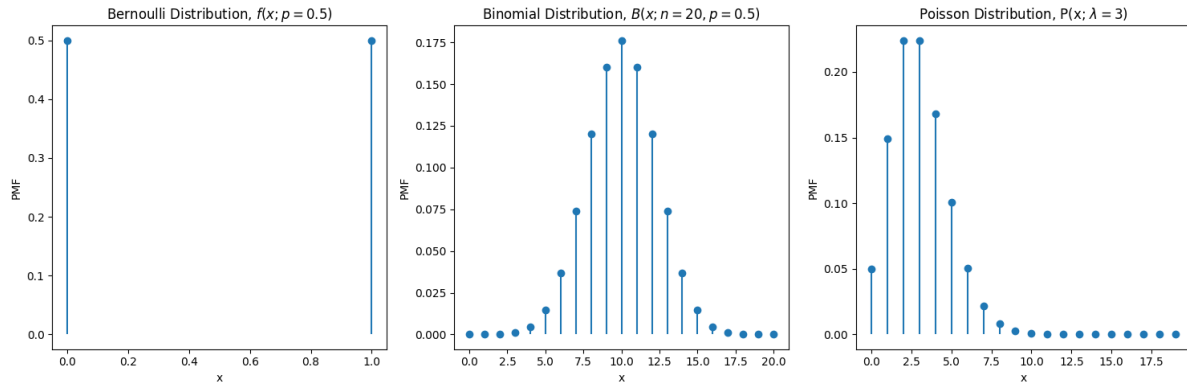
Proprietà.

- $p_X \geq 0$
- $\sum_{x_k \in E} p_X(x_k) = 1$

E da queste due proprietà segue che la probabilità del singolo valore è non superiore a 1, $p(x_k) \leq 1$.

Esempi di variabili discrete

Si anticipano qui alcuni *esempi di variabili casuali discrete*, rappresentandone la loro distribuzione di probabilità: la distribuzione di Bernoulli, la distribuzione binomiale, e la distribuzione di Poisson. La distribuzione di *Bernoulli* può rappresentare la probabilità che il lancio di una moneta dia testa:0 o croce:1, la distribuzione *binomiale* può rappresentare la probabilità di ottenere x volte croce in n lanci; la distribuzione di *Poisson* viene usata per rappresentare eventi discreti rari, dei quali si conosce la media λ .



Momenti

Valore medio. Il valore medio di una variabile casuale è la media dei valori possibili $x_k \in E$ pesata della probabilità $p(x_k)$,

$$\begin{aligned}\mu_X = E[X] &:= \sum_{x_k \in E} p(x_k) x_k = \\ &= \sum_{x_k \in E} \sum_{\omega_i \in \Omega_x} \nu(\omega_i) X(\omega_i) = \\ &= \sum_{\omega_i \in \Omega} \nu(\omega_i) X(\omega_i) .\end{aligned}$$

equivalente alla media dei valori della variabile $X(\omega_i)$ pesata sulle probabilità degli eventi $\omega_i \in \Omega$. La media di una variabile casuale X viene comunemente indicata con μ_X , e così verrà fatto in seguito per motivi di sintesi.

Varianza. La varianza è una misura della dispersione dei valori di una variabile casuale attorno al suo valore medio, ed è definita come la media pesata degli scarti quadratici,

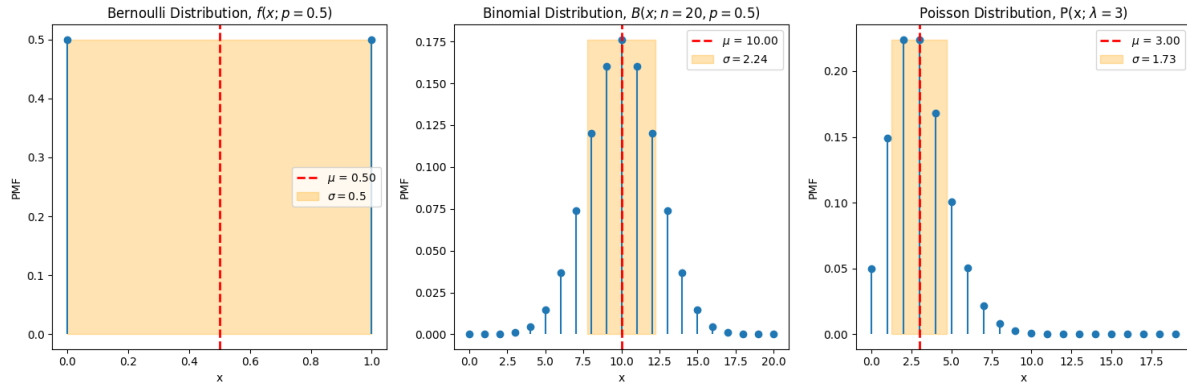
$$\sigma_X^2 = E[(X - \mu_X)^2] = \sum_k p(x_k) (x_k - \mu_X)^2 .$$

Analogamente a quanto fatto per il valore medio, anche questa media può essere riferita sia ai valori sia agli eventi.

La **variazione standard** è definita come la radice quadrata della varianza. A differenza della varianza, è una grandezza *omogenea* (!) alla variabile casuale: ad esempio, la media di una misura di lunghezza può essere 2 m con deviazione standard 0.1 m , mentre la varianza è 0.01 m^2 .

Di seguito vengono mostrati il valore medio e la deviazione standard - rappresentata qui come un intervallo centrato attorno alla media, anche se questa rappresentazione può essere ingannevole (!) per distribuzioni di probabilità non simmetriche - delle variabili casuali presentate in precedenza. In seguito, verrà introdotto il concetto di *intervallo di confidenza*. **todo riferimento?**

Oss. Nonostante siano rappresentati in tutto il grafico, si deve pensare alla media come un valore sull'asse x dei valori della variabile casuale, e la deviazione standard come un intervallo sullo stesso asse x , ricordando le osservazioni fatte in precedenza sulla rappresentazione della deviazione standard.



Variabili casuali multi-dimensionali

- **Probabilità congiunta.** $p(x, y)$
- **Probabilità condizionale.** $p(x|y)$
- **Probabilità marginale.** $p(x)$

Probabilità congiunta e teorema di Bayes

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x) \quad (17.1)$$

Teorema di Bayes

Dalla formula (17.1) della probabilità congiunta è immediato ricavare l'espressione del teorema di Bayes

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} .$$

Variabili indipendenti

Due variabili casuali sono indipendenti se la probabilità condizionale di una variabile coincide con la sua probabilità marginale,

$$p(x|y) = p(x) ,$$

così che la probabilità congiunta di due variabili casuali indipendenti è il prodotto delle probabilità marginali,

$$p(x, y) = p(x)p(y) .$$

Covarianza

$$\sigma_{ij}^2 = E[(X_i - \mu_i)(X_j - \mu_j)] = R_{ij} - \mu_i \mu_j$$

Correlazione

$$\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{E[(X - \mu_X)^2]^{1/2} E[(Y - \mu_Y)^2]^{1/2}} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

17.1.3 Variabili casuali continue**17.1.4 Trasformazione di probabilità di densità****Variabile singola****Cambio di variabile****Traslazione**

Data la variabile casuale X con distribuzione di probabilità $f_X(x)$ nota, la variabile casuale

$$Y = X + a$$

ha funzione di probabilità

$$f_Y(x + a) = f_X(x) .$$

Tra la media e la varianza delle due variabili casuali valgono le relazioni

$$\mu_Y = \mu_X + a \quad , \quad \sigma_Y^2 = \sigma_X^2 .$$

La dimostrazione è lasciata come esercizio.

Dimostrazione.

La probabilità che la variabile $Y = X + a$ abbia valore $y = x + a \in [y_0, y_0 + dy]$, è uguale alla probabilità che l'evento X abbia valore $x \in [x_0, x_0 + dx]$, con $x_0 = y_0 - a$ e $dy = dx$,

$$\begin{aligned} P(Y \in [y_0, y_0 + dy]) &= P(X + a \in [y_0, y_0 + dy]) = \\ &= P(X \in [y_0 - a, y_0 - a + dy]) = \\ &= P(X \in [x_0, x_0 + dx]) \end{aligned}$$

Per la definizione di densità di probabilità,

$$\begin{aligned}
 P(Y \in [y_0, y_0 + dy]) &= \int_{y=y_0}^{y_0+dy} f_Y(y) dy = & y_0 &= x_0 + a \\
 &= \int_{y=x_0+a}^{x_0+a+dx} f_Y(y) dy = & x &= y - a, \quad dy = dx \\
 &= \int_{x=x_0}^{x_0+dx} f_Y(x-a) dx \\
 P(X \in [x_0, x_0 + dx]) &= \int_{x=x_0}^{x_0+dx} f_X(x) dx
 \end{aligned}$$

dal confronto tra i due termini, e dall'arbitrarietà di x_0 e dx , si ottiene $f_X(x) = f_Y(x-a)$.

Scalatura

Data la variabile casuale X con distribuzione di probabilità $f_X(x)$ nota, la variabile casuale

$$Y = \frac{X}{a}$$

ha funzione di probabilità

$$f_Y\left(\frac{x}{a}\right) = a f_X(x)$$

Tra la media e la varianza delle due variabili casuali valgono le relazioni

$$\mu_Y = \frac{\mu_X}{a}, \quad \sigma_Y^2 = \frac{\sigma_X^2}{a^2}.$$

La dimostrazione è lasciata come esercizio.

Dimostrazione.

La probabilità che la variabile $Y = \frac{X}{a}$ abbia valore $y = \frac{x}{a} \in [y_0, y_0 + dy]$, è uguale alla probabilità che l'evento X abbia valore $x \in [x_0, x_0 + dx]$, con $x_0 = ay_0$ e $dy = adx$,

$$\begin{aligned}
 P(Y \in [y_0, y_0 + dy]) &= P\left(\frac{X}{a} \in [y_0, y_0 + dy]\right) = \\
 &= P(X \in [ay_0, ay_0 + ady]) = \\
 &= P(X \in [x_0, x_0 + dx])
 \end{aligned}$$

Per la definizione di densità di probabilità,

$$\begin{aligned}
 P(Y \in [y_0, y_0 + dy]) &= \int_{y=y_0}^{y_0+dy} f_Y(y) dy = & y_0 &= \frac{x_0}{a} \\
 &= \int_{y=\frac{x_0}{a}}^{\frac{x_0}{a}+\frac{dx}{a}} f_Y(y) dy = & y &= \frac{x}{a}, \quad dy = \frac{dx}{a} \\
 &= \frac{1}{a} \int_{x=x_0}^{x_0+dx} f_Y\left(\frac{x}{a}\right) dx \\
 P(X \in [x_0, x_0 + dx]) &= \int_{x=x_0}^{x_0+dx} f_X(x) dx
 \end{aligned}$$

dal confronto tra i due termini, e dall'arbitrarietà di x_0 e dx , si ottiene $f_X(x) = \frac{1}{a} f_Y\left(\frac{x}{a}\right)$.

Trasformazione affine

Combinando la traslazione e la scalatura, si ottiene la più generale trasformazione affine

$$X = \frac{Y - y_0}{a},$$

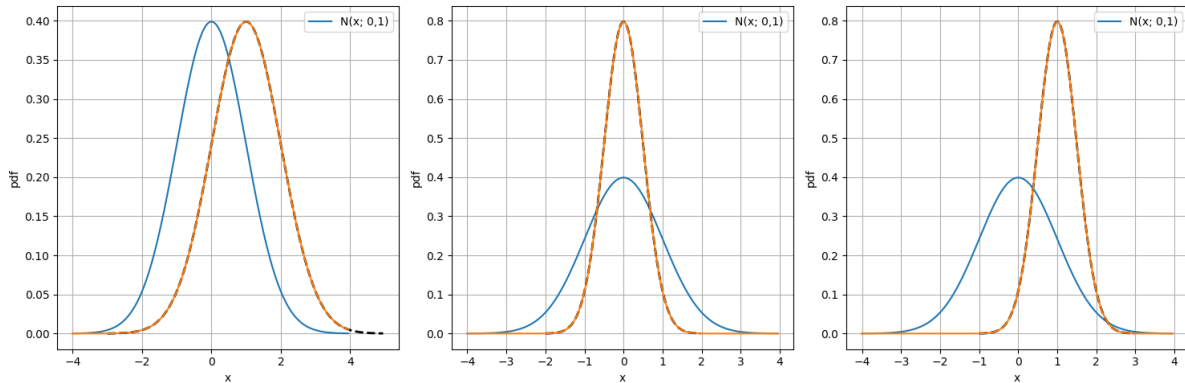
usata spesso come **scalatura** delle variabili nel *preprocessing*. **todo** riferimento a scalatura di variabili su valori significativi del problema, per bilanciare il problema e lavorare con variabili non-dimensionali significate con lo stesso ordine di grandezza (!!!)

Tra la media e la varianza delle due variabili casuali valgono le relazioni

$$\mu_X = \frac{\mu_Y - y_0}{a}, \quad \sigma_X^2 = \frac{\sigma_Y^2}{a^2}.$$

La dimostrazione è lasciata come esercizio.

Le tre trasformazioni discusse vengono applicate alla *variabile normale* $X \sim N(x; \mu = 0, \sigma^2 = 1)$. In particolare si mostrano gli effetti delle trasformazioni: 1. $X = Y + 1$; 2. $X = 2Y$; 3. $X = 2(Y - 1)$



Multi-variabile

Cambio di variabili

Combinazione di variabili

Somma

Date due variabili casuali X, Y , con probabilità congiunta $p_{XY}(x, y)$, la loro somma $Z = X + Y$ è una variabile casuale dipendente con distribuzione di probabilità

$$p(z) = \int_y p_{XY}(z - y, y) dy = \int_x p_{XY}(x, z - x) dx.$$

Se le due variabili sono tra di loro *statisticamente indipendenti*, la densità di probabilità congiunta è uguale al prodotto delle densità di probabilità delle singole variabili, $p_{XY}(x, y) = p_X(x)p_Y(y)$, e quindi la densità di probabilità della somma è uguale alla **convoluzione** tra le densità di probabilità delle due variabili,

$$p(z) = \int_y p_X(z - y) p_Y(y) dy.$$

Valore atteso

Il **valore atteso** della somma è quindi

$$E[Z] = \int_z z p(z) dz = \int_{y,z} z p_{XY}(z - y, y) dy dz = \int_{x,y} (x + y) p_{XY}(x, y) dx dy ,$$

e, nel caso in cui le due variabili siano tra di loro *statisticamente indipendenti*,

$$\begin{aligned} E[Z] &= \int_{x,y} (x + y) p_X(x) p_Y(y) dx dy = \\ &= \int_{x,y} x p_X(x) p_Y(y) dx dy + \int_{x,y} y p_X(x) p_Y(y) dx dy = \\ &= \int_x x p_X(x) dx + \int_y y p_Y(y) dy = E[X] + E[Y] . \end{aligned}$$

Varianza

La **varianza** della somma è

$$\sigma_Z^2 = E[(Z - E[Z])^2] = E[Z^2] - E[Z]^2 = \int_z z^2 p(z) dz - E[Z]^2$$

e nel caso le due variabili siano tra di loro *statisticamente indipendenti*,

$$\begin{aligned} \sigma_Z^2 &= \int_z z^2 p(z) dz - E[Z]^2 = \\ &= \int_{x,y} (x + y)^2 p_X(x) p_Y(y) dx dy - (E[X] + E[Y])^2 = \\ &= \int_x x^2 p_X(x) dx + \int_y y^2 p_Y(y) dy + 2E[X]E[Y] - (E[X]^2 + 2E[X]E[Y] + E[Y]^2) = \\ &= \int_x x^2 p_X(x) dx - E[X]^2 + \int_y y^2 p_Y(y) dy - E[Y]^2 = \\ &= \sigma_X^2 + \sigma_Y^2 . \end{aligned}$$

Prodotto

Date due variabili casuali X, Y con probabilità congiunta $p_{XY}(x, y)$, il loro prodotto $Z = X \cdot Y$ è una variabile casuale dipendente con distribuzione di probabilità

$$p(z) = \int_y p_{XY}\left(\frac{z}{y}, y\right) dy = \int_x p_{XY}\left(x, \frac{z}{x}\right) dx$$

Se le due variabili sono *statisticamente indipendenti*

$$p(z) = \int_x p_X(x) p_Y\left(\frac{z}{x}\right) dx$$

(**todo** ha qualche uso questa formula? Si ritrova la convoluzione con il cambio di variabili $z = e^v, x = e^u$)

Valore atteso

Il **valore atteso** del prodotto è

$$E[Z] = \int_z p(z) dz = \int_{y,z} z p_{XY} \left(\frac{z}{y}, y \right) dy dz = \int_{x,y} xy p_{XY}(x, y) dx dy ,$$

e, nel caso in cui le due variabili siano tra di loro *statisticamente indipendenti*,

$$\begin{aligned} E[Z] &= \int_{x,y} xy p_X(x) p_Y(y) dx dy = \\ &= \int_x x p_X(x) dx \cdot \int_y y p_Y(y) dy = E[X] \cdot E[Y] . \end{aligned}$$

Varianza

La **varianza** del prodotto è

$$\sigma_Z^2 = \dots$$

e nel caso le due variabili siano tra di loro *statisticamente indipendenti*,

$$\begin{aligned} \sigma_Z^2 &= \int_{x,y} x^2 y^2 p_X(x) p_Y(y) dx dy - (E[X] E[Y])^2 = \\ &= R_X^2 R_Y^2 - E[X]^2 E[Y]^2 = \\ &= (\sigma_X^2 + E[X]^2) (\sigma_Y^2 + E[Y]^2) - E[X]^2 E[Y]^2 = \\ &= \sigma_X^2 \sigma_Y^2 + \sigma_X^2 E[Y]^2 + \sigma_Y^2 E[X]^2 \end{aligned}$$

In termini di correlazione, $R_X^2 = \sigma_X^2 + E[X]^2$,

$$R_Z^2 = R_X^2 R_Y^2 .$$

17.1.5 Esempi di funzioni di probabilità discreta

Bernoulli

Binomiale

Poisson

17.1.6 Esempi di funzioni di probabilità continua

Distribuzione gaussiana o normale, \mathcal{N}

La funzione densità di probabilità di una variabile casuale gaussiana $X \sim \mathcal{N}(\mu, \sigma^2)$ con valore atteso $\mathbb{E}[X] = \mu$ e varianza $\mathbb{E}[(X - \mathbb{E}[X])^2]$ è

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

Distribuzione chi-quadrato, χ_N^2

Date N variabili casuali $\{X_n\}_{n=1:N}$ iid con distribuzione normale $X_n \sim \mathcal{N}(0, 1)$, la somma dei loro quadrati,

$$\chi_n^2 = \sum_{k=1}^n X_k^2,$$

è una variabile casuale con densità di probabilità χ_N^2 , con N definito come il numero di gradi di libertà. La distribuzione χ_n^2 ha una funzione densità di probabilità

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \propto x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$$

Distribuzione t -Student

La t di Student è la distribuzione di probabilità che governa il rapporto tra due variabili casuali indipendenti,

$$t_\nu = \frac{Z}{\sqrt{\frac{K}{\nu}}},$$

con il numeratore con distribuzione normale, $Z \sim \mathcal{N}(0, 1)$, e il denominatore con distribuzione chi quadrato, $K \sim \chi_n^2$. La pdf è

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \propto \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Proprietà. Per $N \rightarrow +\infty$, la distribuzione t_N tende alla distribuzione gaussiana $\mathcal{N}(0, 1)$.

17.1.7 Variabili casuali indipendenti e identicamente distribuite, iid.

Due variabili stocastiche X, Y sono

- **statisticamente indipendenti** se la loro probabilità congiunta è uguale al prodotto delle probabilità marginali,

$$p_{XY}(x, y) = p_X(x) p_Y(y),$$

- **identicamente distribuite** se hanno la stessa densità di probabilità,

$$p_X(x) = p_Y(x)$$

Teoremi sulle variabili casuali i.i.d.

Dato un insieme di N variabili casuali iid $\{X_n\}_{n=1:N}$, con valore atteso $\mathbb{E}[X_n] = \mu$ e varianza $\mathbb{E}[(X_n - \mu)^2] = \sigma$, allora la sua media campionaria

$$\bar{X}_N = \frac{1}{N} \sum_{n=1}^N X_n$$

- per il **teorema dei grandi numeri**, converge al valore atteso μ della distribuzione di probabilità.
- per il **teorema del limite centrale**, è una variabile casuale che converge in distribuzione a una variabile casuale gaussiana con valore atteso μ e varianza $\frac{\sigma^2}{n}$,

$$\bar{X}_N \rightarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right), \quad \text{as } N \rightarrow \infty$$

Convergenza in statistica, todo

- **quasi certamente.** Esempio: teorema dei grandi numeri in forma forte. Il limite della media campionaria è diverso da una variabile casuale μ (**todo controllare!**) solo nel caso di eventi di probabilità nulla,

$$P\left(\lim_{N \rightarrow \infty} \bar{X}_N = \mu\right) = 1.$$

- **in probabilità.** Esempio: teorema dei grandi numeri in forma debole. Per ogni valore di $\varepsilon > 0$,

$$P\left(\left|\lim_{N \rightarrow \infty} \bar{X}_N - \mu\right| < \varepsilon\right) = 1.$$

- **convergenza in distribuzione.** Esempio: teorema del limite centrale...

17.1.8 Campionamento

17.1.9 Media campionaria

La media campionaria di N campioni indipendenti identicamente distribuiti,

$$\bar{X}_N = \frac{1}{N} \sum_{n=1}^N X_n,$$

è una variabile casuale, che può essere usata come **stimatore senza bias della media** della popolazione. Uno **stimatore senza bias della varianza** è definito

$$\hat{\sigma}_N^2 = \frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X}_N)^2.$$

17.1.10 Dimensione del campione, teorema del limite centrale e distribuzione *t*-Student

La media di N variabili iid con media μ e varianza σ^2 è una variabile casuale, la cui distribuzione di probabilità tende alla distribuzione normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$ per $N \rightarrow \infty$.

Per un numero di campioni ridotti, nel caso la popolazione sia formata da variabili iid con distribuzione gaussiana, la distribuzione della variabile

$$T_N = \frac{\bar{X} - \hat{X}}{\frac{\hat{\sigma}}{\sqrt{N}}},$$

costruita con gli stimatori non-biased della media e della varianza, è una variabile casuale con distribuzione ***t*-Student** con $N - 1$ gradi di libertà. Al tendere di $N \rightarrow \infty$, gli stimatori senza bias tendono ai valori veri delle statistiche della popolazione, e la variabile

$$\frac{\bar{X} - \hat{X}}{\frac{\hat{\sigma}}{\sqrt{N}}} \sim \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}},$$

tende alla distribuzione normale $\mathcal{N}(0, 1)$

```
#

import numpy as np
import scipy as sp
import matplotlib.pyplot as plt

def sample_to_pdf(x, bins=100, density=True):
    """ evaluate pdf, from histogram, and normalizing with uniform integration rule """
    hist, bin_edges = np.histogram(x, bins=bins, density=True) # get data from
    # histogram
    bin_cen = 0.5 * ( bin_edges[:-1] + bin_edges[1:] )           # bin center
    d_bin = bin_edges[1:] - bin_edges[:-1]                       # bin width
    pdf = hist / np.sum(hist * d_bin)                             # normalization to get
    # int pdf = 1
    return pdf, bin_cen
```

```
#> Population size
pop_size = 1000000

#> Random number generators
pop_distribution = 'normal'

if ( pop_distribution == 'uniform' ):
    rng = np.random.default_rng().uniform # Uniform distribution
    rng_params = { 'low': -1. , 'high': 1., 'size': pop_size }
else: # 'normal by default'
    rng = np.random.default_rng().normal # Normal distribution
    rng_params = { 'loc': 0. , 'scale': 1., 'size': pop_size }

#> Generate population
pop = rng(**rng_params)
```

```
sample_size = [ 2, 5, 10, 30 ]
n_samples = 10000

sample_avgs = []
sample_vars = []
for s_s in sample_size:
    sample_avg = []
    sample_var = []
    for i_s in np.arange(n_samples):
        sample_avg += [ np.mean(pop[i_s*s_s:(i_s+1)*s_s]) ]
        sample_var += [ np.var(pop[i_s*s_s:(i_s+1)*s_s], ddof=1) ]

    sample_avgs += [ sample_avg ]
    sample_vars += [ sample_var ]

# print(sample_avgs)

plt.figure()
for is_s in np.arange(len(sample_size)):
    ns = sample_size[is_s]
    var = np.array( sample_vars[is_s] ) / ns
    scaled = sample_avgs[is_s]/var**.5
```

(continues on next page)

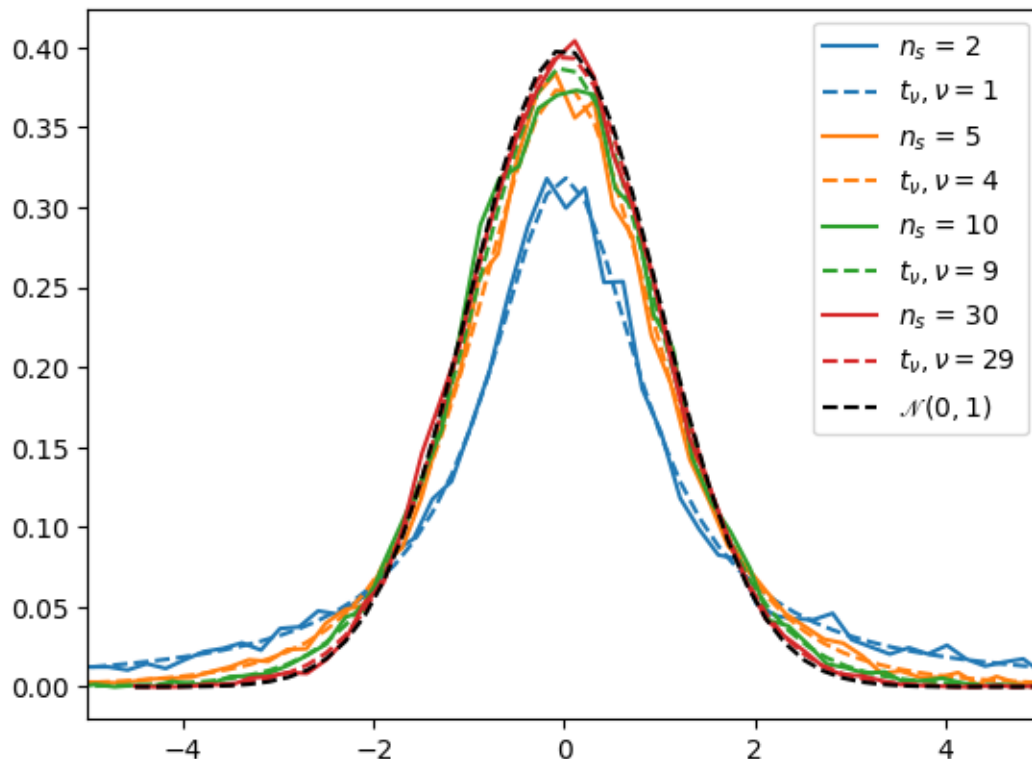
(continua dalla pagina precedente)

```

bin_width = .2; n_bins = int((np.max(scaled)-np.min(scaled))/bin_width)
f, x = sample_to_pdf(scaled, bins=n_bins)
plt.plot(x,f, color=plt.cm.tab10(is_s), label=f"$n_s$ = {ns}")
plt.plot(x, sp.stats.t.pdf(x, ns-1, loc=0, scale=1), '--', color=plt.cm.tab10(is_
↪s), label=f"$t_{\\nu}$, $\\nu$={ns-1}")
plt.plot(x, sp.stats.norm.pdf(x, loc=0, scale=1), '--', color='black', label="$\\
↪mathscr{N}(0,1)$")
plt.legend()
plt.xlim(-5,5)

```

(-5.0, 5.0)



17.2 Processi casuali

Un processo casuale può essere definito come una variabile casuale che dipende dalla variabile tempo t , $X(t)$.

todo è probabile che i dettagli di questa sezione richiedano una conoscenza abbastanza approfondita della materia, e di alcuni strumenti matematici come le trasformate

todo

- definizione
- statistiche (in tempo e in frequenza) e definizioni (stazionarietà, ergodicità,...)
- esempi:
 - discreti: Markov

- continui (e discretizzati al computer):
 - * white noise, Wiener
 - * random walk, Brown
 - * applicazioni:
 - diffusione
 - white noise nei sistemi dinamici (risposte a forzanti stocastiche)

Statistica inferenziale

La statistica inferenziale si occupa di stabilire le caratteristiche di una popolazione dall'osservazione di un suo campione, usando gli strumenti del calcolo delle probabilità.

Tra le applicazioni della statistica inferenziale si possono distinguere due applicazioni principali:

- **stima** di parametri di una popolazione, noto un campione di essa:
 - **valori puntuali**, come ad esempio la media o la varianza di una popolazione
 - **intervalli**, come ad esempio gli intervalli di confidenza
- **verifica ipotesi**. La verifica delle ipotesi è un'attività intimamente connessa al metodo sperimentale e alla formazione della conoscenza nella storia umana secondo criteri logici. Nonostante sia poco sensata una classificazione rigida delle attività che coinvolgono la verifica delle ipotesi (seguendo la documentazione di `scipy.stats`), si possono distinguere
 - test su **campioni singoli** o **coppie di campioni**: *t*-test, Wilcoxon,...
 - test di **correlazione**: LS, Pearson *r*-test, Spearman *r*-test,
 - test di **indipendenza**, verifica se i campioni provengono dalla stessa popolazione (e quindi dalla stessa distribuzione), o da popolazioni con distribuzioni con proprietà in comune: ANOVA, Kruskal-Wallis, Tukey
 - **ri-campionamento** e metodi MC: per valutare *p*-value e intervalli di confidenza test di validità di ipotesi
 - *ipotesi multiple*

Esistono (almeno) due approcci alla statistica inferenziale:

- l'approccio classico **frequentista**: ... **todo dire due parole**
- l'approccio **bayesiano**: utilizza le osservazioni per trovare una stima migliore (*a posteriori*, dopo aver osservato il campione) di modello di partenza del sistema (*a priori*, da conoscenze precedenti, da stime sensate,...), grazie all'uso del *teorema di Bayes*.

Questo capitolo si concentra principalmente sull'approccio classico frequentista; *dell'approccio bayesiano, ci si limita a dire che usa delle conoscenze pregresse sul fenomeno e che è stato reso praticabile dai recenti sviluppi nelle capacità di calcolo a disposizione*. **todo aggiungere sezione sull'approccio bayesiano?** **todo esempio con variabili discrete di dimensione limitata, senza introdurre i problemi sulle scelte di priori-posteriori**

todo Scelta modello? Serve una discuterne? Rimandato a ML?

18.1 Stima

Una delle attività principali della statistica inferenziale è la stima di parametri o valori di una popolazione X della quale è disponibile solo un campione $\mathbf{X} = \{X_n\}_{n=1:N}$. E' possibile stimare valori puntuali, come la media o la varianza, o intervalli, come intervalli di confidenza. **todo** come stimare gli intervalli di confidenza? Serve conoscenza o ipotesi sulla distribuzione? Gaussiana per un numero sufficiente di campioni, per teorema del limite centrale?

18.1.1 Stimatori

Uno stimatore $\hat{\theta}(\mathbf{X})$ è una statistica, funzione dei dati del campione osservato $\mathbf{X} = \{X_n\}_{n=1:N}$, che viene usata per dedurre il valore di un parametro della distribuzione di probabilità della popolazione, $p(X|\theta)$, funzione del parametro.

Bias. Il bias di uno stimatore è la differenza tra il valore atteso dello stimatore $\mathbb{E}[\hat{\theta}]$ e il valore del parametro θ ,

$$B(\hat{\theta}) := \mathbb{E}[\hat{\theta}] - \theta .$$

Media e varianza campionaria senza bias

Dato un campione $\mathbf{X} = \{X_n\}_{n=1:N}$ di N osservazioni indipendenti estratto da una popolazione, X , la media campionaria \bar{X} e la varianza campionaria corretta S^2 ,

$$\bar{X} := \frac{1}{N} \sum_{n=1}^N X_n \quad , \quad S^2 := \frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})^2 ,$$

sono stimatori senza bias della media, $\hat{\mu}$, e della varianza della variabile osservata della popolazione, $\hat{\sigma}^2$. Le singole osservazioni possono essere dei dati generati in maniera indipendente con uguale distribuzione di probabilità.

todo oss. Dalle dimostrazioni, sembra che l'identica distribuzione non sia strettamente necessaria, ma che siano richieste: indipendenza (e non correlazione) delle variabili che producono l'osservazione; variabili con stesso valore di media e di varianza, indipendentemente dalla «forma» della distribuzione di probabilità.

La media campionaria è stimatore senza bias della media della popolazione, $\mathbb{E}[\bar{X}] = \mu_X$

Dato un campione di n variabili indipendenti $\{X_n\}_{n=1:N}$ osservate in una popolazione con media $\mu = \mathbb{E}[X]$, e varianza $\sigma^2 = \mathbb{E}[(X - \mu)^2]$, allora la media campionaria \bar{X} è uno stimatore senza bias $\hat{\mu}$ della media $\mu = \mathbb{E}[X]$ della popolazione. Il valore atteso della media campionaria coincide con la media della popolazione,

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{N} \sum_{n=1}^N X_n\right] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[X_n] = \frac{1}{N} N \mu = \mu .$$

e quindi il bias di questo stimatore è nullo, $B(\hat{\mu}) = \mathbb{E}[\hat{\mu}] - \mu = 0$.

La varianza campionaria è stimatore senza bias della varianza della popolazione, $\mathbb{E}[S^2] = \sigma_X^2$

Dato un campione di n variabili indipendenti $\{X_n\}_{n=1}^N$ osservate in una popolazione con media $\mu = \mathbb{E}[X]$, e varianza $\sigma^2 = \mathbb{E}[(X - \mu)^2]$, allora la varianza campionaria corretta S^2 è uno stimatore senza bias $\hat{\sigma}^2$ della varianza $\sigma^2 = \mathbb{E}[(X - \mu)^2]$ della popolazione. Per dimostrare questa affermazione, si usa la proprietà della covarianza

$$\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathbb{E}[X_i X_j] - \mu_i \mathbb{E}[X_j] - \mu_j \mathbb{E}[X_i] + \mu_i \mu_j = \mathbb{E}[X_i X_j] - \mu_i \mu_j,$$

che, nel caso di variabili indipendenti non correlate con uguale varianza, si riduce a

$$\mathbb{E}[X_i X_j] = \sigma^2 \delta_{ij} + \mu_i \mu_j$$

Il valore atteso della varianza campionaria corretta coincide con la varianza della popolazione,

$$\begin{aligned} (N-1)\mathbb{E}[S^2] &= \mathbb{E}\left[\sum_{n=1}^N (X_n - \bar{X})^2\right] = \\ &= \mathbb{E}\left[\sum_{n=1}^N \left(X_n - \frac{1}{N} \sum_{m=1}^N X_m\right)^2\right] = \\ &= \mathbb{E}\left[\sum_{n=1}^N X_n^2 + \sum_{m=1}^N \left(-\frac{2}{N} X_n X_m + \frac{2}{N^2} \sum_{p \neq m} X_m X_p + \frac{X_m^2}{N^2}\right)\right] = \\ &= \sum_{n=1}^N \mathbb{E}[X_n^2] + \sum_{n,m=1}^N \left(-\frac{2}{N} \mathbb{E}[X_n X_m] + \frac{1}{N^2} \mathbb{E}[X_m^2]\right) - \frac{2}{N^2} \sum_{n,m=1}^N \sum_{p>m} \mathbb{E}[X_m X_p] = \\ &= \sum_{n=1}^N (\sigma^2 + \mu^2) + \sum_{n,m=1}^N \left(-\frac{2}{N} (\sigma^2 \delta_{nm} + \mu^2) + \frac{1}{N^2} (\sigma^2 + \mu^2)\right) + \frac{2}{N^2} \sum_{n,m=1}^N \sum_{p>m} (\sigma^2 \delta_{mp} + \mu^2) = \\ &= N(\sigma^2 + \mu^2) - \frac{2}{N} (N\sigma^2 + N^2\mu^2) + \frac{1}{N^2} N^2(\sigma^2 + \mu^2) + \frac{2}{N^2} \frac{N^2(N-1)}{2} \mu^2 = \\ &= \sigma^2(N-1) + \mu^2(N-2N+1+N-1) \\ &= \sigma^2(N-1). \end{aligned}$$

e quindi, dividendo per il fattore $N-1$ entrambi i termini, segue la dimostrazione che la varianza campionaria corretta per la varianza della popolazione è uno stimatore con bias nullo, $B(\hat{\sigma}^2) = \mathbb{E}[\hat{\sigma}^2] - \sigma^2 = 0$.

18.2 Test di verifica delle ipotesi

Popper. Filosofia della scienza e principio di falsificabilità.

Fisher. Viene formulata un'ipotesi falsificabile, chiamata **ipotesi nulla**, H_0 , che viene ritenuta vera fino a prova contraria. Il test di verifica delle ipotesi ha l'obiettivo di verificare se i dati disponibili, riassunti in una variabile test, possono smentire l'ipotesi con una probabilità desiderata.

Test di verifica di Fisher

Il test di verifica di un'ipotesi di Fisher si ispira al principio di falsificabilità di Popper, e può essere riassunto nei seguenti passaggi:

1. formulazione di un'ipotesi falsificabile, definita **ipotesi nulla**, H_0 , da verificare e che viene ritenuta vera fino a prova contraria
2. scelta di una **statistica test**, x , o variabile esplicativa, un parametro disponibile rilevante per il fenomeno indagato e di cui è nota - o approssimabile, sotto ipotesi ragionevoli sul fenomeno - la distribuzione di probabilità condizionata alla validità dell'ipotesi nulla,

$$p(x|H_0) .$$

3. scelta del test statistico (es. una coda o due code,...), in base anche alle caratteristiche del fenomeno indagato
4. scelta del livello di **significatività**, α , del test; il valore del livello di significatività traduce «il livello di evidenza richiesto» per falsificare l'ipotesi e - insieme alle caratteristiche del fenomeno e del test - determina le *regioni di rifiuto*, $A_r(\alpha)$, e di *accettazione dell'ipotesi*, $A_a(\alpha)$, gli intervalli di valori della statistica test che determinano se l'ipotesi è stata falsificata o meno;
5. raccolta dati e calcolo statistica test sul campione, x_s
6. confronto del valore calcolato della statistica test con gli intervalli di rifiuto e accettazione della variabile soggetta all'ipotesi nulla, e verdetto sull'ipotesi,

$$\begin{array}{ll} x_s \in A_r(\alpha) & \rightarrow \text{il test falsifica } H_0, \text{ che deve essere scartata} \\ x_s \in A_a(\alpha) & \rightarrow \text{il test non falsifica } H_0 \end{array}$$

All'aumentare del livello di significatività del test richiesto, si riduce la regione di falsificazione e quindi diventano più stringenti i requisiti sul test per scartare l'ipotesi di partenza.

Risultati ed errori. Un test di ipotesi può portare a due tipi di errore:

- **tipo I, falso positivo:** rifiuto di un'ipotesi H_0 valida
- **tipo II, falso negativo:** mancato rifiuto di un'ipotesi H_0 non valida

Example 18.2.1 (Giudizio di un imputato)

In un tribunale, si vuole decidere se l'imputato è innocente. Si considera come ipotesi nulla H_0 : «l'imputato è innocente». Le indagini svolte cercano di valutare lo stato dell'imputato e portano alla decisione di condanna o assoluzione. Un errore di **tipo I** corrisponde alla condanna di un imputato innocente; un errore di tipo II corrisponde all'assoluzione di un imputato colpevole.

Example 18.2.2 (Test medico)

todo Definire *potenza*, β , di un test in presenza di un'ipotesi alternativa. Fare esempio con falsi positivi, e affidabilità tamponi Covid.

- H_0 nulla vera: α : probabilità di rifiutare H_0 ; $1 - \alpha$: probabilità di non rifiutare H_0
 - H_1 alternativa vera: $1 - \beta$: probabilità di rifiutare H_0 ; β : probabilità di non rifiutare H_0
-

Usando questo approccio, è possibile distinguere diversi test a seconda dei dati disponibili (numero di campioni, dimensione dei campioni, distribuzione attesa delle osservazioni,...), della variabile test utilizzata, e dell'obiettivo del test. Una classificazione rigida dettagliata rischierebbe di risultare poco pratica e imperfetta.

Usi frequenti.

- Test su campioni singoli o coppie di campioni
- Test di correlazione tra campioni diversi
- Test di indipendenza tra campioni, intesa come verifica della possibilità che i campioni disponibili provengano dalla stessa popolazione o da popolazioni con distribuzioni simili

18.2.1 Test su campioni singoli o coppie di campioni

Test più comuni

- t -test
- χ^2 -test
- Z -test
- Wilcoxon
- ...

Student t -test

I t -test sono dei test statistici di posizione in cui la statistica test segue una distribuzione t di Student sotto l'ipotesi nulla H_0 ,

$$p(t|H_0) \sim t_\nu,$$

dove il numero di gradi di libertà ν della distribuzione t_ν dipende dal metodo considerato.

Questi test sono «esatti» se la popolazione ha distribuzione normale. **todo** in caso contrario, la statistica test costruita non segue una distribuzione t -Student. Bisogna valutare l'attendibilità del test in questo caso; per distribuzioni «sufficientemente simili a quella gaussiana» potrebbero essere validi; il teorema del limite centrale aiuta l'ipotesi di gaussianità; in caso di distribuzioni generiche, meglio affidarsi ad altri test - non parametrici

Test per un campione - valore medio

E' un test di posizione/stima della media della popolazione della quale è noto un campione di N osservazioni, $\mathbf{X} = \{X_n\}_{n=1:N}$.

- l'ipotesi nulla H_0 dichiara che la popolazione ha media μ_0 ;
- la statistica test T ,

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{N}}}$$

è costruita con la *media e la varianza campionaria*,

$$\bar{X} = \frac{1}{N} \sum_{n=1}^N X_n, \quad S^2 = \frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})^2.$$

- quindi, una volta definito il livello di significatività del test, α , si controlla se la statistica test t valutata con il campione a disposizione cade nella regione di rifiuto dell'ipotesi della distribuzione t_{N-1} o meno, per determinare se si deve considerare H_0 falsificata dal test o meno

Ipotesi. *Come dimostrato in appendice*, se i campioni sono variabili i.i.d. gaussiane, con media μ_0 sotto l'ipotesi nulla H_0 , $X_n \sim \mathcal{N}(\mu_0, \sigma^2)$, allora la statistica test segue una distribuzione t -Student con $N - 1$ gradi di libertà,

$$T \sim t_{N-1}.$$

todo cosa succede quando la popolazione non ha distribuzione gaussiana? Qual è la robustezza del metodo alla caduta di questa ipotesi? Esempio già disponibile con campionamento variabile con distribuzione uniforme; altri metodi, non parametric?

Test per una coppia di campioni con stessa varianza - valore medio

Un t -test per una coppia di campioni $\{X_{1,n}\}_{n=1:N}$, $\{X_{2,n}\}_{n=1:N}$ presi da due popolazioni X_1 , X_2 di cui si ipotizza la stessa varianza σ^2 , si riduce al t -test per campione singolo per la variabile casuale $X_2 - X_1$.

Variabili gaussiane. Nel caso in cui le variabili casuali X_1 , X_2 abbiano distribuzione gaussiana con la stessa varianza, $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma^2)$, la variabile $X_2 - X_1$ è gaussiana con distribuzione \rightarrow

$$X_2 - X_1 \sim \mathcal{N}(\mu_2 - \mu_1, 2\sigma^2)$$

todo

Altri t -test

todo Se campioni di dimensione differente o con varianze differenti, *Welch t -test*...

Z -test

Test identico al t -test in cui la statistica test z segue una distribuzione gaussiana sotto l'ipotesi nulla. Per il legame esistente tra la distribuzione t_ν e \mathcal{N} , si può interpretare lo Z -test come limite del t -test per un numero di campioni sufficientemente grande.

χ^2 -test di Pearson

Il test χ^2 di Pearson si applica a dati **categoriali**, per confrontare le **frequenze** delle K categorie per diversi insiemi. Può essere applicato:

- a un singolo campione, confrontando le frequenze osservate $\{O_k\}_{k=1:K}$ con le frequenze attese $\{E_k\}_{k=1:K}$, che costituiscono l'ipotesi nulla H_0
- a più campioni (**todo link a test di indipendenza statistica**) per stabilire se provengono dalla stessa popolazione, o comunque da popolazioni con la stessa funzione di probabilità.

Il test χ^2 a un campione viene utilizzato per confrontare le frequenze osservate $\{O_k\}_{k=1:K}$ e attese $\{E_k\}_{k=1:K}$ di un insieme di possibili eventi, $\{X_n\}_{n=1:N}$.

- l'**ipotesi nulla** H_0 afferma che il fenomeno osservato segue la distribuzione di probabilità descritta dalle frequenze attese, $\{E_k\}$. *Supponendo che i dati campionati siano iid*, l'ipotesi nulla afferma che il fenomeno dal quale è estratto il campione segue una distribuzione di probabilità categoriale, $\text{Categorical}(E_1, \dots, E_K)$
- La statistica test è

$$X^2 := \sum_{k=1}^K \frac{(O_k - E_k)^2}{E_k},$$

Nel caso in cui le frequenze attese sono «sufficientemente grandi» e quelle osservate O_k abbiano una distribuzione gaussiana, allora X^2 segue una distribuzione χ^2_{K-1} .

... **todo**

Esempio: dado truccato

Ci vengono dati due dadi a 6 facce e ci viene chiesto di verificare con $n_{rolls} = 50$ lanci se questi dadi possono essere considerati truccati. Noi non lo sappiamo, ma il primo dado non è truccato, mentre il secondo dado è sbilanciato in modo tale che la possibilità che esca la faccia 3 è doppia rispetto a tutte le altre,

$$X_1 \sim \text{Categorical}\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$$

$$X_2 \sim \text{Categorical}\left(\frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}\right)$$

Ipotesi nulla. Per ogni esperimento, l'ipotesi nulla H_0 è che il dado non sia truccato e che ogni faccia abbia una probabilità $\frac{1}{6}$. La distribuzione di probabilità attesa è

$$X|H_0 \sim \text{Categorical}\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right).$$

Statistica test. Assumendo che la differenza tra le frequenze osservate e quelle attese abbiano una distribuzione gaussiana $\sim \mathcal{N}(0, E^2)$, la statistica test ha una distribuzione

$$X^2 = \sum_{k=1}^K \frac{(O_k - E_k)^2}{E_k} \sim \chi^2_{N-1}.$$

todo link al motivo per cui i dof sono $N - 1$: sostanzialmente poiché la pdf è determinata da un solo parametro, un solo vincolo, che riduce di 1 il numero di dof

Livello di significatività. Si sceglie un livello di significatività $\alpha = 0.05$, cioè solo il 5

Se si vuole essere più conservativi e rigettare l'ipotesi nulla «solo per dadi palesemente truccati», si può ridurre il livello di significatività del test. Ad esempio, scegliendo $\alpha = 0.01$ si identificano come truccati solo i risultati estremi che hanno probabilità cumulata di verificarsi dell'1% sotto l'ipotesi nulla (1% di falsi positivi).

```
# Show influence of alpha on rejection set

print(f"\nRejection and acceptance regions for significance, alpha = {alpha}")
print(f"- Not rejection for X2 < {X2_rej_lim:.4f}")
print(f"- Rejection      for X2 > {X2_rej_lim:.4f}")

print("\nRejection limits as a function of significance level")
alpha_l = [ .05, .01, .002, .0004 ]
for al in alpha_l:
    print(f"X_rej_lim( alpha={al:0.3f} ) = {chi2.ppf(1-al, n_dofs):.4f}")

# Plot probability density, f(x), cumulative probability, F(x), and significance_
↪ alpha(x)=1-F(x)
x_plot = np.arange(0., 25., .05)
fig, axes = plt.subplots(3, 1, figsize=(5,8))
axes[0].plot(x_plot, chi2.pdf(x_plot, n_dofs) )
axes[0].set_xlim([x_plot[0], x_plot[-1]])
axes[0].set_title(f"$\chi^2_{n\_dofs}$")
axes[0].set_ylabel(f"pdf")
```

(continues on next page)

(continua dalla pagina precedente)

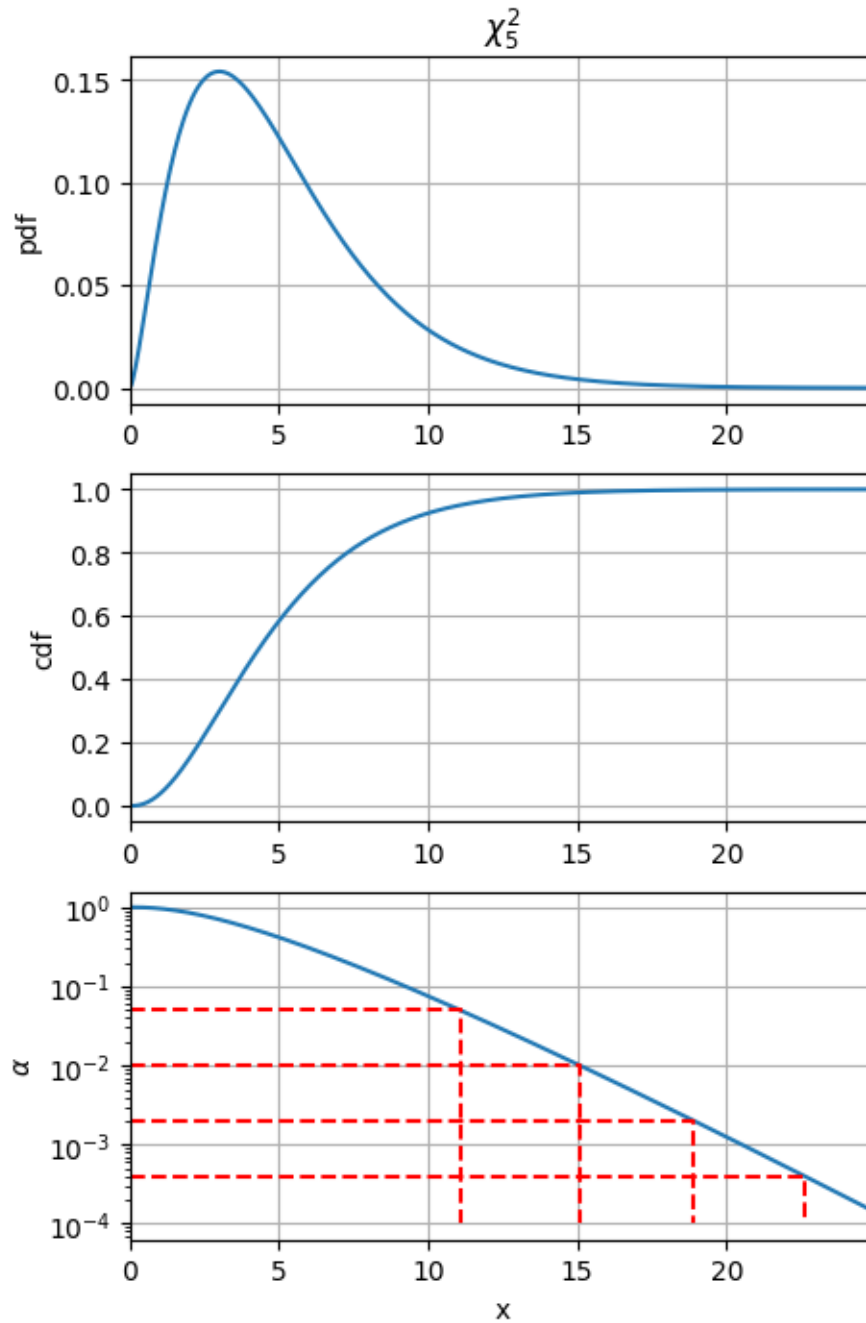
```
axes[0].grid()
axes[1].plot(x_plot, chi2.cdf(x_plot, n_dofs) )
axes[1].set_xlim([x_plot[0], x_plot[-1]])
axes[1].set_ylabel(f"cdf")
axes[1].grid()
axes[2].semilogy(x_plot, 1-chi2.cdf(x_plot, n_dofs) )
axes[2].set_xlim([x_plot[0], x_plot[-1]]); axes[2].set_xlabel('x')
# axes[2].fill_between(x_plot, 1-chi2.cdf(x_plot, n_dofs))
axes[2].set_ylabel(f"$\\alpha$")
for al in alpha_l:
    xl = chi2.ppf(1-al, n_dofs)
    axes[2].plot([0., xl, xl], [al, al, 1e-4], '--', color='red')
axes[2].grid()
```

Rejection and acceptance regions for significance, $\alpha = 0.05$

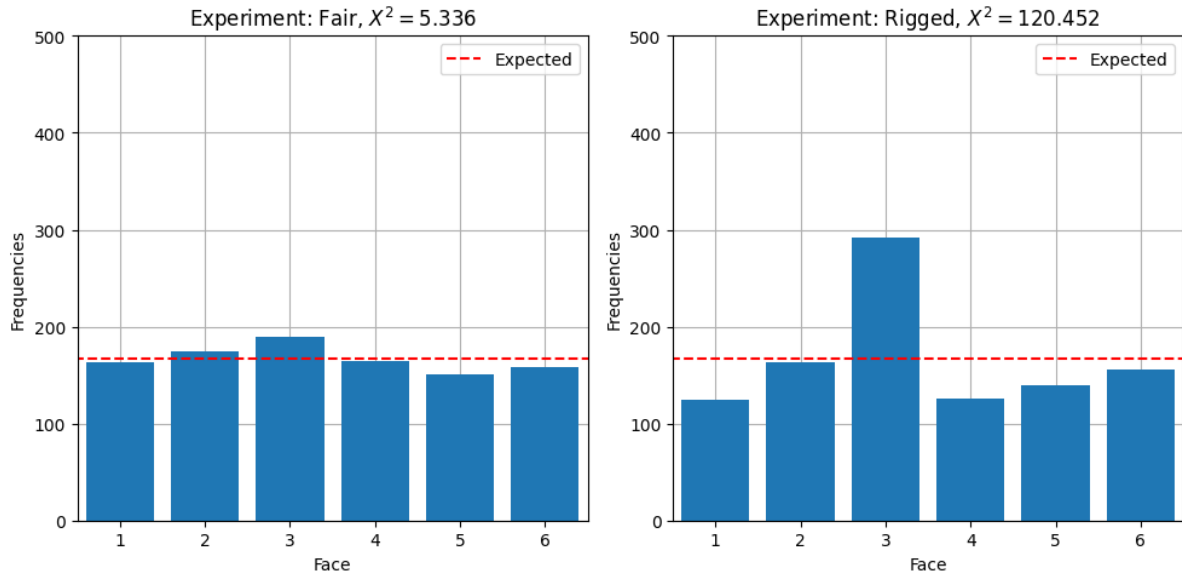
- Not rejection for $X_2 < 11.0705$
- Rejection for $X_2 > 11.0705$

Rejection limits as a function of significance level

$X_{\text{rej_lim}}(\alpha=0.050)$	= 11.0705
$X_{\text{rej_lim}}(\alpha=0.010)$	= 15.0863
$X_{\text{rej_lim}}(\alpha=0.002)$	= 18.9074
$X_{\text{rej_lim}}(\alpha=0.000)$	= 22.6141



Vengono svolti gli esperimenti, raccolti i dati, calcolata la statistica test sui campioni e questa viene utilizzata per trarre le conclusioni sull'ipotesi nulla.



```
Experiment: Fair, nrolls: 1000
- probability (unknown in real exp.) : [0.1667 0.1667 0.1667 0.1667 0.1667 0.1667]
- Observed events                     : [163. 174. 189. 165. 151. 158.]
- Test statistics, X2                  : 5.3360
>> H0 not rejected (5.3360 < X_lim=11.0705): dice is fair

Experiment: Rigged, nrolls: 1000
- probability (unknown in real exp.) : [0.1429 0.1429 0.2857 0.1429 0.1429 0.1429]
- Observed events                     : [124. 163. 292. 126. 139. 156.]
- Test statistics, X2                  : 120.4520
>> H0 rejected (120.4520 > X_lim=11.0705): dice is rigged
```

Dipendenza del risultato del test dal numero di prove. Si analizzano i risultati in funzione del numero di lanci fatti in ogni test. Si osserva che:

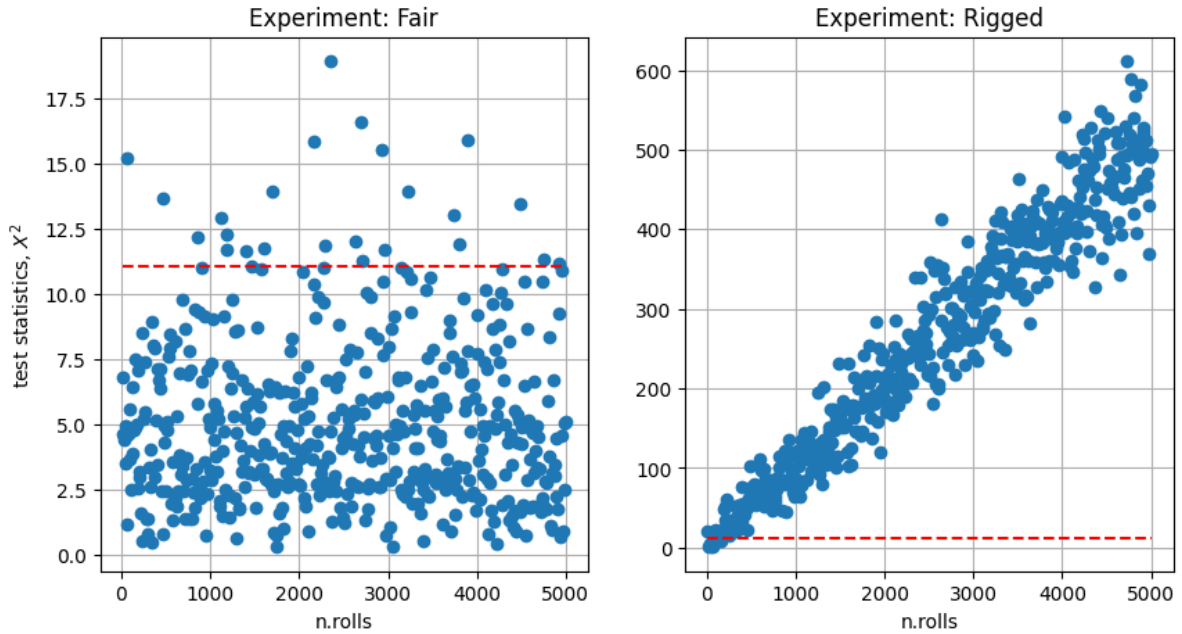
- nel caso di dado non truccato la statistica test X^2 non dimostra una convergenza (almeno fino a $n_{rolls} = 5000$) ma rimane limitata
- nel caso di dado truccato, la statistica test X^2 cresce circa linearmente con il numero di lanci n_{rolls}

Nei 500 test svolti, il test ha dato risultati falso positivi nel 4.8

Il progetto di un esperimento prevede la determinazione del numero di lanci necessario (senza andare troppo oltre! Tempo, soldi e pazienza sono una risorse finite!) per distinguere con la sufficiente accuratezza un dado truccato.

N. of rejections of H0, over 500 tests

```
- Experiment: Fair      ,      24,  4.8% (false positive)
- Experiment: Rigged    ,     493, 98.6%
```

Wilcoxon

...todo

18.2.2 Test di correlazione

Test più comuni

- regressione lineare
- correlazione di Pearson
- correlazione di Spearman

Regressione lineare

- regressione lineare
- regressione lineare generalizzata
- ...

Metodo dei minimi quadrati per una regressione lineare

Dati due campioni accoppiati $\{X_n\}_{n=1:N}$, $\{Y_n\}_{n=1:N}$ campionati dalle popolazioni X , Y , si calcolano i coefficienti $\theta = (a, b)$ dell'approssimazione lineare $y(x|\theta) = ax + b$ che rendono minimo l'errore quadratico medio

$$e(\theta) = \frac{1}{2N} \sum_{n=1}^N (y(x_n|\theta) - y_n)^2 = \frac{1}{2N} \sum_{n=1}^N (ax_n + b - y_n)^2.$$

La funzione è quadratica in a , b e semi-definita positiva. Si cercano i valori che rendono minimo l'errore ponendo uguale a 0 le derivate parziali dell'errore rispetto ai parametri

$$\begin{aligned} 0 = \frac{\partial e}{\partial a} &= \frac{1}{N} \sum_{n=1}^N x_n (ax_n + b - y_n) = \frac{1}{N} \sum_{n=1}^N x_n^2 a + \frac{1}{N} \sum_{n=1}^N x_n b - \frac{1}{N} \sum_{n=1}^N x_n y_n \\ 0 = \frac{\partial e}{\partial b} &= \frac{1}{N} \sum_{n=1}^N (ax_n + b - y_n) = \frac{1}{N} \sum_{n=1}^N x_n a + b - \frac{1}{N} \sum_{n=1}^N y_n \end{aligned}$$

Queste due equazioni formano un sistema lineare di due equazioni in due incognite, che può essere riscritto con il formalismo matriciale

$$\begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n^2 & \frac{1}{N} \sum_{n=1}^N x_n \\ \frac{1}{N} \sum_{n=1}^N x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n y_n \\ \frac{1}{N} \sum_{n=1}^N y_n \end{bmatrix}.$$

Il valore dei coefficienti ottimi $\theta^* = (a, b)$ che minimizzano l'errore quadratico medio si trova come soluzione del sistema lineare, l'errore quadratico medio minimo dovuto all'approssimazione viene calcolato in seguito, $e(\theta^*)$. La bontà dell'approssimazione lineare può essere valutata in termini dell'errore quadratico medio.

Regressione e SL. Una delle applicazioni fondamentali in ML, in particolare in SL, è la regressione o l'approssimazione di funzioni; scelta del modello...

Metodo ed errore

Il sistema lineare può essere riscritto come

$$\begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n^2 & \bar{x} \\ \bar{x} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n y_n \\ \bar{y} \end{bmatrix},$$

e la soluzione

$$\begin{aligned} \begin{bmatrix} a \\ b \end{bmatrix} &= \frac{1}{\frac{1}{N} \sum_{n=1}^N x_n^2 - \bar{x}^2} \begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & \frac{1}{N} \sum_{n=1}^N x_n^2 \end{bmatrix} \begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n y_n \\ \bar{y} \end{bmatrix} = \\ &= \frac{1}{\frac{1}{N} \sum_{n=1}^N x_n^2 - \bar{x}^2} \begin{bmatrix} \frac{1}{N} \sum_{n=1}^N x_n y_n - \bar{x} \bar{y} \\ -\bar{x} \sum_{n=1}^N x_n y_n + \bar{y} \sum_{n=1}^N x_n^2 \end{bmatrix} = \\ &= \frac{1}{R_x - \bar{x}^2} \begin{bmatrix} R_{xy} - \bar{x} \bar{y} \\ -\bar{x} R_{xy} + \bar{y} R_x \end{bmatrix} \end{aligned}$$

L'errore minimo diventa quindi

$$\begin{aligned} e &= \frac{1}{2N} \sum_{m=1}^N (ax_m + b - y_m)^2 = \\ &= \frac{1}{2N} \sum_{m=1}^N (a^2 x_m^2 + b^2 + y_m^2 + 2abx_m - 2by_m - 2ax_m y_m) \\ 2e &= \frac{1}{N} \left[a^2 \sum_{m=1}^N x_m^2 + Nb^2 + \sum_{m=1}^N y_m^2 + 2ab\bar{x} - 2b\bar{y} - 2a \sum_{m=1}^N x_m y_m \right] = \\ &= a^2 R_x + b^2 + R_y + 2ab\bar{x} - 2b\bar{y} - 2a R_{xy} \\ 2e &= \left(\frac{R_{xy} - \bar{x} \bar{y}}{R_x - \bar{x}^2} \right)^2 R_x + \left(\frac{-\bar{x} R_{xy} + \bar{y} R_x}{R_x - \bar{x}^2} \right)^2 + R_y + 2 \left(\frac{R_{xy} - \bar{x} \bar{y}}{R_x - \bar{x}^2} \right) \left(\frac{-\bar{x} R_{xy} + \bar{y} R_x}{R_x - \bar{x}^2} \right) \bar{x} - 2 \left(\frac{-\bar{x} R_{xy} + \bar{y} R_x}{R_x - \bar{x}^2} \right) \bar{y} - 2 \left(\frac{R_{xy} - \bar{x} \bar{y}}{R_x - \bar{x}^2} \right) \bar{y} \\ 2e &= \frac{1}{D^2} [R_{xy}^2 R_x - 2xy R_{xy} R_x + x^2 y^2 R_x + \dots] \end{aligned}$$

Correlazione di Pearson

Il **coefficiente di correlazione** tra due variabili casuali X, Y è definito come il rapporto tra la loro covarianza e il prodotto delle loro deviazioni standard,

$$\rho_{XY} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]} \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Coefficiente di correlazione per campioni. Usando le statistiche campionarie per i campioni a disposizione delle due popolazioni,

$$\begin{aligned} r_{xy} &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}} = \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_i - \bar{y}}{S_y} \right) = \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \end{aligned}$$

Coefficiente di correlazione in statistica inferenziale. I metodi di statistica basati sul coefficiente di correlazione di Pearson hanno due obiettivi:

1. valutare l'ipotesi nulla H_0 che afferma che l'assenza di correlazione tra le due popolazioni, $\rho = 0$, sulla base del valore del coefficiente r calcolato sui campioni
2. fornire un intervallo di confidenza che contenga ρ

Alcuni metodi per ottenere questi due obiettivi sono:

- permutazione
- bootstrap
- errore standard, assumendo una relazione lineare tra x e y , con l'aggiunta di un errore gaussiano ε , $y = ax + b + \varepsilon$,
 $\sigma_r = \sqrt{\frac{1-r^2}{n-2}}$
- t -test sulla variabile $t = \frac{r}{\sigma_r}$
- ...

Correlazione di Spearman

- per ranghi
- per relazioni non lineari ma monotone ... **todo**

Esempio

Si valutano i metodi di correlazione per 4 coppie di campioni, 2 che mostrano una buona correlazione e 2 con correlazione scarsa

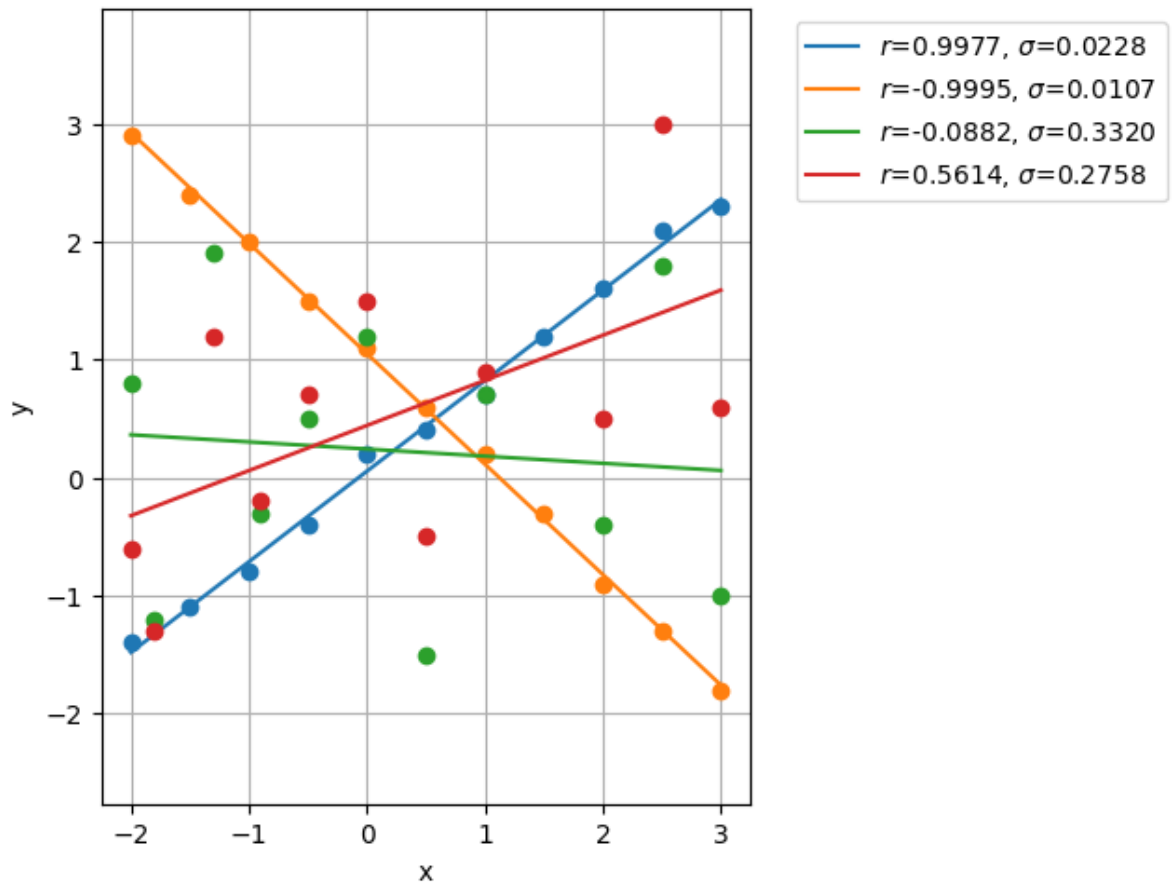
```
> Pair 0
Regression coeffs y = 0.7673 x + 0.0527, with error 0.0831
Pearson correlation: 0.9977, sigma: 0.0228
> Pair 1
```

(continues on next page)

(continua dalla pagina precedente)

```
Regression coeffs y = -0.9364 x + 1.0500, with error 0.0477
Pearson correlation: -0.9995, sigma: 0.0107
> Pair 2
Regression coeffs y = -0.0604 x + 0.2410, with error 1.1254
Pearson correlation: -0.0882, sigma: 0.3320
> Pair 3
Regression coeffs y = 0.3824 x + 0.4404, with error 0.9296
Pearson correlation: 0.5614, sigma: 0.2758
```

```
Text(44.22222222222214, 0.5, 'y')
```



18.2.3 Test di indipendenza

Test più comuni

- t -test per campioni indipendenti e Welch t -test
- ANOVA
- Tukey test
- Mann-Whitney U -test
- Kruskal-Wallis H -test

***t*-test e Welch *t*-test**

Il *t*-test di Welch è un test di posizione per stimare se due popolazioni di cui sono noti due campioni hanno la stessa media. E' una modifica del *t*-test, e si dimostra un criterio più affidabile quando i due campioni hanno varianza diversa e dimensione diversa.

Ipotesi. Il test assume che la media dei campioni abbia una distribuzione normale.

...**todo**

ANOVA (Analysis of variance)

- ...insieme di metodi statistici...
- differenza tra due o più campioni
- i metodi più semplici valutano se le medie di 2 o più popolazioni hanno valori uguali
- 3 classi: fixed-effect, random-effect, mixed-effect
- *F*-test
- single/multiple factors
- caso particolare di regressione lineare generalizzata

...**todo**

Tukey test

...**todo**

Mann-Whitney *U*-test

Il test di Mann-Whitney è un test statistico non parametrico usato per verificare se le popolazioni X, Y delle quali sono disponibili due campioni $\{X_n\}_{n=1:n_x}, \{Y_m\}_{m=1:n_y}$, sono identiche.

... **todo**

Kruskal-Wallis *H*-test

...**todo**

18.2.4 Esempi

...

Test di verifica d'ipotesi - Fisher

Il test di verifica di un'ipotesi di Fisher si ispira al principio di falsificabilità di Popper, e può essere riassunto nei seguenti passaggi:

1. formulazione di un'ipotesi falsificabile, definita **ipotesi nulla** H_0 , da verificare e che viene ritenuta vera fino a prova contraria
2. scelta di una variabile esplicativa, o **statistica test**, x , un parametro disponibile, rilevante per il fenomeno indagato e di cui è nota - o approssimabile, sotto ipotesi ragionevoli sul fenomeno - la distribuzione di probabilità, $p(x|H_0)$
3. scelta del **test statistico** (es. una coda o due code,...), in base anche alle caratteristiche del fenomeno indagato
4. scelta del **livello di significatività**, α , del test; il valore del livello di significatività traduce «il livello di evidenza richiesto» per falsificare l'ipotesi e - insieme alle caratteristiche del fenomeno e del test - determina le *regioni di rifiuto e di accettazione dell'ipotesi*, gli intervalli di valori della statistica test x che determinano se l'ipotesi è stata falsificata o meno;
5. raccolta dati e **calcolo statistica test sul campione**
6. **confronto** del valore calcolato della statistica test con gli intervalli di rifiuto e accettazione della variabile soggetta all'ipotesi nulla, e **verdetto sull'ipotesi**

Esempio: moneta truccata o no?

Dati i risultati di n lanci di una moneta, si vuole stabilire con una certa probabilità se la moneta è truccata o meno.

Il lancio di una moneta viene modellato come una variabile casuale X di Bernoulli, con due possibili uscite testa, $X = 0$, o croce, $X = 1$. La forma generale della distribuzione di probabilità di una variabile casuale di Bernoulli $B(p)$ è

$$p(X) = \begin{cases} p & , \quad X = H : \text{Head} \\ 1 - p & , \quad X = T : \text{Tail} \end{cases} ,$$

essendo $p \in [0, 1]$ la probabilità associata al valore $X = H$ e $1 - p$ quella associata al valore $X = T$.

Per la verifica dell'ipotesi si organizza una campagna sperimentale di n lanci e si sceglie come **statistica test** x il numero di volte che il risultato del lancio è testa, $X = H$. Per una variabile casuale con distribuzione di probabilità di Bernoulli $B(p)$, il numero x di risultati $X = H$ in n ripetizioni indipendenti dell'evento è a sua volta una variabile casuale, con distribuzione di probabilità binomiale $\mathcal{B}(n, p)$

$$p_n(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Import librerie

```
# import libraries
%reset -f
import numpy as np
import scipy as sp

import matplotlib.pyplot as plt
```

Campione generato dal processo casuale

Viene lanciata una moneta n_{flips} volte. Dati

```
#> Random process of interest, unknown except for the sample.
p_head = .5
p_tail = 1. - p_head

# Coin flip as a Bernoulli probability with outcomes: a = [0, 1] with prob p = [p_
  ↪head, p_tail]
flip_rng = np.random.default_rng(42).choice
n_flips = 30
flip_params = { 'a': [0,1], 'p': [p_head, p_tail], 'size': n_flips }

#> Running coin flip experiment
ov = flip_rng(**flip_params)

coin_dict = [ 'head', 'tail' ]
# print([ coin_dict[o] for o in ov])
print("\n> Experiment")
print(f"n.samples: {n_flips}")
print(f"heads:tails {np.sum(ov)}:{n_flips - np.sum(ov)}")
print()
```

```
> Experiment
n.samples: 30
heads:tails 17:13
```

```
# ! Approach for low-dimensional problems, with no memory or performance issues

#> Null hypothesis, H0: the coin is fair
# The outcome of a single flip is a r.v. distributed like a Bernoulli variable with p_
  ↪head = .5
# Every flip is statistically independent from the other
ixv = np.arange(n_flips+1)
xv = ixv / n_flips

#> Test statistics: average n.of heads in flip_n samples
# H0 hypothesis implies that the outcome of flip_n is a r.v. with Binomial pdf
x_H0_fv = sp.stats.binom.pmf(ixv, n_flips, p_head)

#> Test characteristics: symmetric
test_type = 'value' # 'symmetric', 'right', 'left', 'value'

#> Significance level, alpha = .05 ("default")
alpha = .05
```

```
#> Acceptance and rejection regions, for discrete pdf
# Starting from the value of the test statistics x_max = max(x_H0_fv), expand

def find_acceptance_region(p, alpha, test_type='value'):
    """ Find acceptance region for a discrete pdf, supposed to be unimodal """
    ix_max = np.argmax(p)
    nx = len(p)
    threshold = 1. - alpha
```

(continues on next page)

(continua dalla pagina precedente)

```
# Initialization
p_acc, ixl, ixr = p[ix_max], ix_max, ix_max

# if ( test_type == 'value' ):
while ( p_acc < threshold ):
    if ( p[ixl-1] >= p[ixr+1] ):
        ixl -= 1; p_acc += p[ixl]
    else:
        ixr += 1; p_acc += p[ixr]

# else:
return ixl, ixr
```

```
#> Compute acceptance region
ixl, ixr = find_acceptance_region(x_H0_fv, alpha,)

#> Evaluate test statistics on the samples
ixs = np.sum(ov)
xs = ixs / n_flips
```

```
# Compare test statistics on the samples with the acceptance region
if ( ixl <= ixs <= ixr ):
    print("H0 accepted")
else:
    print("H0 rejected")

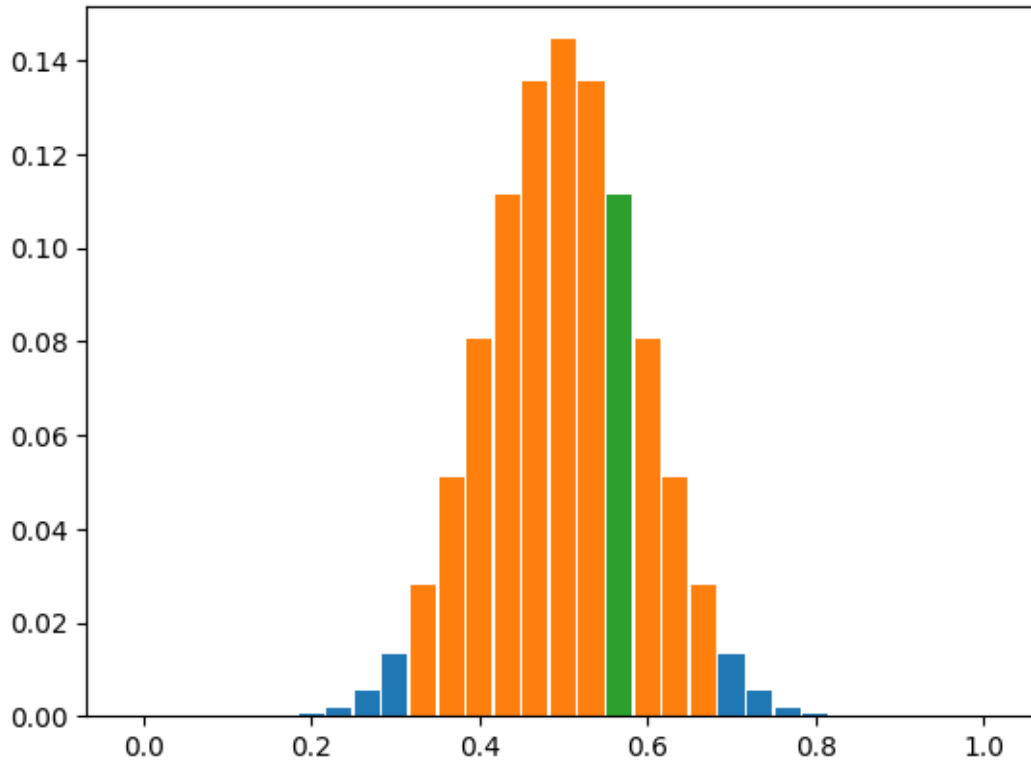
ix_acc = np.arange(ixl, ixr+1)
x_acc = xv[ix_acc]

dx_bar = (xv[1] - xv[0])*0.90
plt.figure()
plt.bar( xv, x_H0_fv, width=dx_bar)
plt.bar(x_acc, x_H0_fv[ix_acc], width=dx_bar)
plt.bar( xs, x_H0_fv[ixs], width=dx_bar)

# print(xv)
```

H0 accepted

<BarContainer object of 1 artists>



Moneta truccata

Si vuole valutare se una moneta è truccata. L'ipotesi nulla H_0 da falsificare è «la moneta non è truccata». Si sceglie come statistica test la percentuale di risultati «testa» sul numero di lanci.

L'esperimento viene condotto per un numero incrementale di esperimenti.

Import librerie e funzioni utili

Import librerie

Reset delle variabili e import librerie.

Funzione per lo svolgimento di un esperimento

La funzione per lo svolgimento di un esperimento qui definita prende come argomenti un generatore di numeri casuali e i suoi parametri, e restituisce i campioni prodotti secondo la distribuzione data. Vengono definiti dei parametri di default, corrispondenti a una distribuzione di Bernoulli uniforme

Funzione che restituisce la pdf $p(x|H_0)$

Funzioni per la ricerca degli intervalli di accettazione dell'ipotesi

Vengono qui definite due funzioni per la ricerca degli intervalli di accettazione, dato il livello di significatività del test richiesto. La prima funzione ricerca un unico intervallo di accettazione riferito a un unico livello di significatività; la seconda funzione ricerca tanti intervalli di accettazione quanti sono i livelli di significatività cercati: ad esempio, si possono cercare simultaneamente i livelli di significatività associati a σ , 2σ , 3σ ...

Esperimento

Livelli di accettazione

Svolgimento esperimento

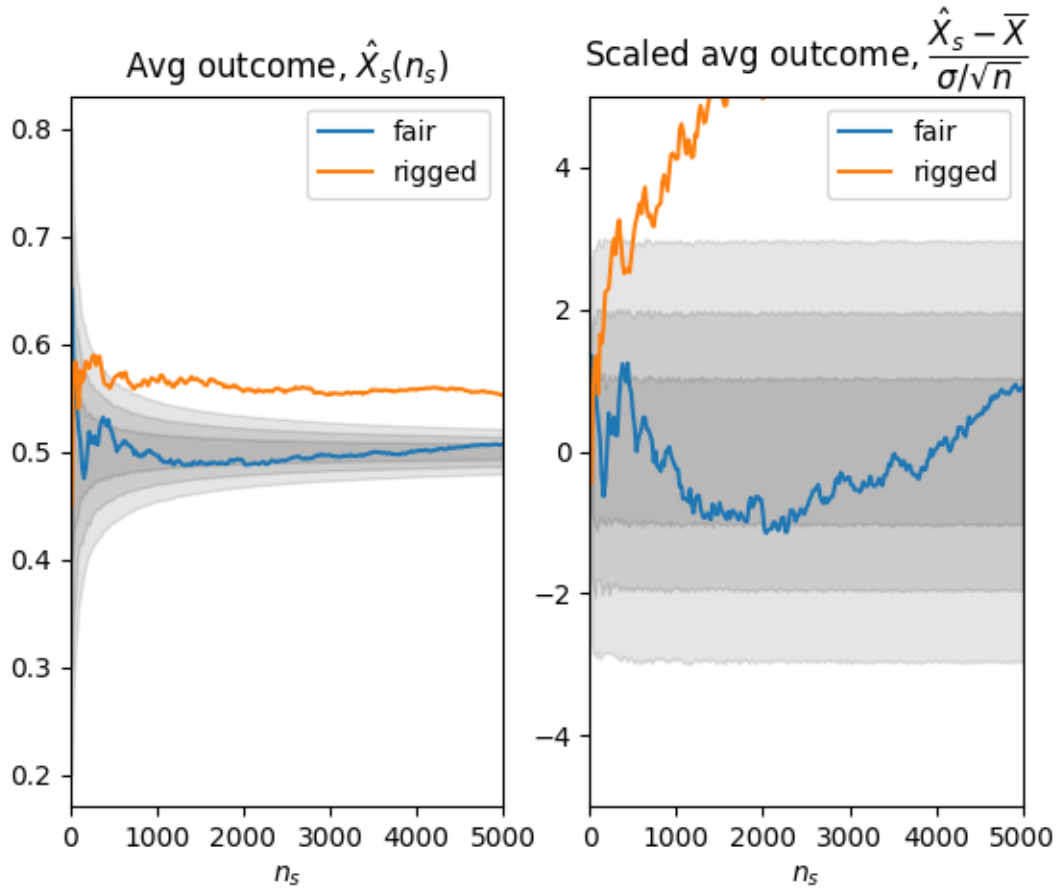
L'esperimento viene svolto usando due monete, la cui natura è a priori incognita, che possono essere modellate con variabili casuali di Bernoulli: una moneta non truccata produce testa o croce con la probabilità uniforme 0.5; una moneta truccata che produce testa o croce con probabilità $0.45 - 0.55$

```
/home/davide/.local/lib/python3.8/site-packages/scipy/stats/_discrete_distns.  
↳py:77: RuntimeWarning: divide by zero encountered in _binom_pdf  
    return _boost._binom_pdf(x, n, p)
```

Grafici

Il valore medio delle volte che è uscita croce viene usata come statistica test. Il valore della statistica test per gli esperimenti condotti con le due monete viene confrontato con la funzione di distribuzione $p(x|H_0)$ per una moneta non truccata. Viene mostrato l'andamento delle statistiche test e degli intervalli di accettazione in funzione del numero di lanci della moneta. Vengono mostrati i dati raccolti e successivamente depurati della media attesa e scalati per la varianza della media campionaria $\frac{\sigma}{\sqrt{n_s}}$, per mettere in evidenza i livelli di significatività $[0.3, 0.05, 0.003]$

```
<matplotlib.legend.Legend at 0x7fd49b8aed60>
```



Gambler rabbit

t -test

Un t -test è un test di verifica delle ipotesi nel quale la **statistica test** ha una distribuzione **t di Student** sotto l'ipotesi nulla H_0 ,

$$p(x|H_0) \sim t_\nu$$

t -test per un campione

Il t -test per un campione è un test di verifica del valore di un parametro, per la media di una variabile casuale sotto un'opportuna ipotesi nulla H_0 . Il parametro t ,

$$t = \frac{\hat{x} - \mu_0}{\frac{s}{\sqrt{n}}},$$

confronta la distanza della media \bar{x} del campione dal valore medio μ_0 della distribuzione $p(x|H_0)$, con la deviazione standard del campione s opportunamente scalata della grandezza del campione n .

Se le osservazioni della variabile casuale sono indipendenti tra di loro, allora t è una variabile casuale che tende a una variabile normale $\mathcal{N}(0, 1)$ per il teorema del limite centrale **todo link a una sezione sul campionamento**

```

# Example

import numpy as np
from scipy.stats import ttest_1samp, t

import matplotlib.pyplot as plt

# Test statistics has expected value mu0 under H0 hypothesis
mu0 = 86.          # Expected value of the test statistics under H0
sigma = 0.05       # Significance level

# Sample data: test scores from two classes
sample = [88, 92, 85, 91, 87]
n_s = len(sample)
mu_s = np.sum(sample)/n_s
s_s = ( np.sum((sample-mu_s)**2)/(n_s-1) )**.5

print("Sample statistics:")
print(f" sample avg      : {mu_s}")
print(f" sample std.dev: {s_s}")

# Perform independent t-test
t_statistic, p_value = ttest_1samp(sample, mu0)

print(f"\n1-sample t-test")
print(f" t-Statistic: {t_statistic}")
print(f" p-Value      : {p_value}")

# Interpretation
if p_value < sigma:
    print("\n> Reject the null hypothesis, H0: the means of the two classes are_
    ↪significantly different.")
else:
    print("\n> Fail to reject the null hypothesis, H0: no significant difference in_
    ↪means between the classes.")

x_plot = np.arange(80, 100, .01)
t_pdf_plot = t.pdf(x_plot, n_s, mu_s, s_s)

plt.figure()
plt.plot(x_plot, t_pdf_plot, color=plt.cm.tab10(0))
plt.plot(sample, np.zeros(n_s), 'o', color=plt.cm.tab10(0))
plt.plot(mu_s, 0., 'x', color=plt.cm.tab10(0), markersize=10, markeredgewidth=2)
plt.plot(mu0 , 0., 'x', color='black'          , markersize=10, markeredgewidth=2)

```

```

Sample statistics:
sample avg      : 88.6
sample std.dev: 2.8809720581775866

1-sample t-test
t-Statistic: 2.017991366836461
p-Value      : 0.11375780482862627

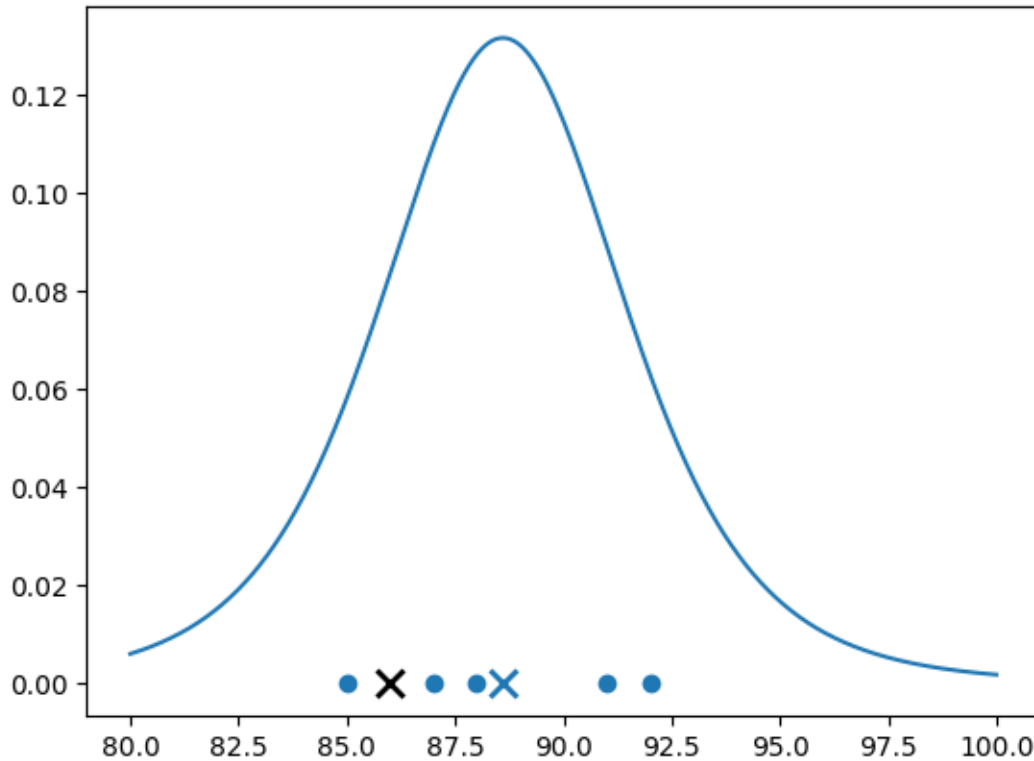
```

(continues on next page)

(continua dalla pagina precedente)

```
> Fail to reject the null hypothesis, H0: no significant difference in means
  ↳ between the classes.
```

```
[<matplotlib.lines.Line2D at 0x7f954c6f0a60>]
```



t-test per due campioni

- test con varianza uguale (o simile), *t*-test/test con varianza diversa, *Welch*
- campioni indipendenti/campioni dipendenti (paired samples)

t-test per due variabili indipendenti, con varianza simile e stesso numero di osservazioni

Nell'ipotesi che i campioni siano ottenuti da due variabili indipendenti, i risultati ottenuti nella sezione sulla *combinazione di variabili casuali*

- il **valore atteso** della somma/differenza di variabili casuali indipendenti è uguale alla somma/differenza dei valori attesi delle singole variabili
- la **varianza** della somma/differenza di variabili casuali indipendenti è uguale alla somma delle varianze delle singole variabili.

Il *t*-test per due campioni equivale al *t*-test per un campione uguale alla differenza delle osservazioni nei due campioni,

$$z_i = x_i - y_i .$$

```

# Example

import numpy as np
from scipy.stats import ttest_ind

# Significance level
sigma = 0.05

# Sample data: test scores from two classes
class_A = [88, 92, 85, 91, 87]
class_B = [78, 85, 80, 83, 79]

nA = len(class_A); mu_A = np.sum(class_A)/nA; sA = ( np.sum((class_A-mu_A)**2) / (nA-1) )  

↪ **.5
nB = len(class_B); mu_B = np.sum(class_B)/nB; sB = ( np.sum((class_B-mu_B)**2) / (nB-1) )  

↪ **.5

print("Sample statistics:")
print("Sample A")
print(f" sample avg      : {mu_A}")
print(f" sample std.dev: {sA}")
print("Sample B")
print(f" sample avg      : {mu_B}")
print(f" sample std.dev: {sB}")

# Perform independent t-test
t_statistic, p_value = ttest_ind(class_A, class_B)

print(f"\ns-sample t-test")
print(f" t-Statistic: {t_statistic}")
print(f" p-Value      : {p_value}")

print((mu_A-mu_B)/((sA**2 + sB**2)/nA)**.5)

# Interpretation
if p_value < sigma:
    print("\nReject the null hypothesis: The means of the two classes are_  

↪ significantly different.")
else:
    print("\nFail to reject the null hypothesis: No significant difference in means_  

↪ between the classes.")

x_plot = np.arange(70, 100, .01)
t_pdf_A = t.pdf(x_plot, nA, mu_A, sA)
t_pdf_B = t.pdf(x_plot, nB, mu_B, sB)

plt.figure()
plt.plot(x_plot, t_pdf_A, color=plt.cm.tab10(0))
plt.plot(x_plot, t_pdf_B, color='orange')
plt.plot(class_A, np.zeros(nA), 'o', color=plt.cm.tab10(0), markersize=10)
plt.plot(class_B, np.zeros(nB), 'o', color='orange' )
plt.plot(mu_A, 0., 'x', color=plt.cm.tab10(0), markersize=10, markeredgewidth=2)
plt.plot(mu_B, 0., 'x', color='orange' , markersize=10, markeredgewidth=2)

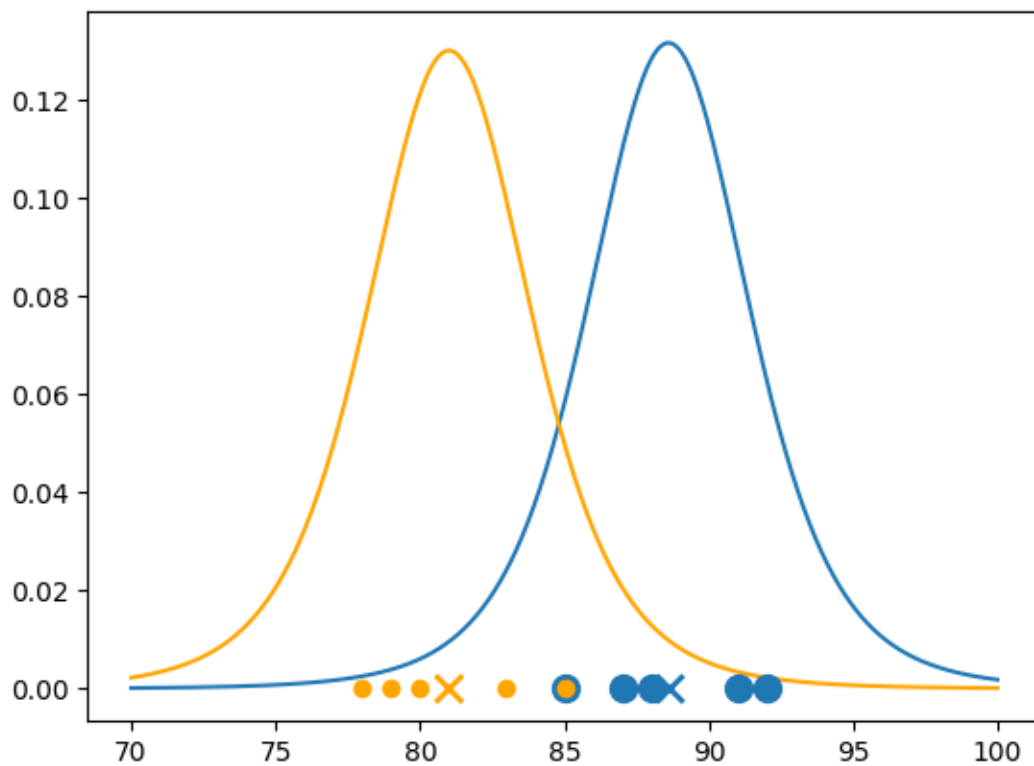
```

```
Sample statistics:
Sample A
  sample avg      : 88.6
  sample std.dev: 2.8809720581775866
Sample B
  sample avg      : 81.0
  sample std.dev: 2.9154759474226504

s-sample t-test
t-Statistic: 4.146139914483853
p-Value      : 0.0032260379191180397
4.146139914483853

Reject the null hypothesis: The means of the two classes are significantly
different.
```

```
[<matplotlib.lines.Line2D at 0x7f954c612f40>]
```



#

Regressione lineare

Nella sua forma elementare, la regressione lineare è un'approssimazione (stimatore) lineare

$$\hat{Y}(X|\theta) = aX + b,$$

tra una coppia di campioni $\{X_n\}_{n=1:N}, \{Y_n\}_{n=1:N}$ estratti dalle popolazioni X, Y , i cui parametri $\theta = (a, b)$ rendono minimo l'errore quadratico medio

$$e := \sum_{n=1}^N (\hat{y}(X_n) - Y_n)^2 = \sum_{n=1}^N (a X_n + b - Y_n)^2,$$

il cui valore permette di stimare la significatività dell'approssimazione lineare.

Algoritmo

L'algoritmo base consiste nella:

- normalizzazione dei campioni:
 - vengono calcolate le medie e le varianze campionarie

$$\begin{aligned} \mu_x &= \frac{1}{N} \sum_{n=1}^N X_n, & S_x^2 &= \frac{1}{N-1} (X_n - \mu_x)^2 \\ \mu_y &= \frac{1}{N} \sum_{n=1}^N Y_n, & S_y^2 &= \frac{1}{N-1} (Y_n - \mu_y)^2 \end{aligned}$$

- vengono definiti i campioni normalizzati

$$x_n = \frac{X_n - \mu_x}{S_x}, \quad y_n = \frac{Y_n - \mu_y}{S_y}$$

che hanno media nulla e varianza campionaria unitaria.

- calcolo dei valori ottimi dei parametri $\theta = (a, b)$ del modello applicato ai dati scalati,

$$\hat{y}(x|\theta) = ax + b$$

che rendono minimo l'errore quadratico,

$$e := \sum_{n=1}^N (\hat{y}(x_n) - y_n)^2 = \sum_{n=1}^N (a x_n + b - y_n)^2.$$

L'errore quadratico è una funzione definita positiva con un unico minimo in corrispondenza del sistema lineare formato dalle condizioni di derivate parziali nulle,

$$\begin{cases} 0 &= \frac{\partial e}{\partial a} = 2 \sum_{n=1}^N (a x_n + b - y_n) x_n \\ 0 &= \frac{\partial e}{\partial b} = 2 \sum_{n=1}^N (a x_n + b - y_n) \end{cases}$$

che ha forma diagonale e, introducendo il coefficiente di correlazione dei campioni,

$$r_{XY}^2 = \frac{S_X^2 Y}{S_X S_Y}, \quad S_{XY}^2 = \frac{1}{N-1} \sum_{n=1}^N (X_n - \mu_X)(Y_n - \mu_Y),$$

e usando il formalismo matriciale può essere riscritto come,

$$\begin{bmatrix} N-1 & 0 \\ 0 & N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (N-1)r_{XY} \\ 0 \end{bmatrix},$$

la cui soluzione è $a^* = r_{XY}$, $b^* = 0$. La forma della regressione lineare in termini delle variabili normalizzate è

$$\hat{y}(x) = r_{XY}x.$$

L'errore quadratico minimo nelle variabili normalizzate vale

$$\begin{aligned} e^* &= \sum_{n=1}^N (a^*x_n + b^* - y_n)^2 = \\ &= \sum_{n=1}^N (r_{XY}x_n - y_n)^2 = \\ &= r_{XY}^2(N-1) - 2(N-1)r_{XY}r_{XY} + (N-1) = \\ &= (N-1)(1 - r_{XY}^2). \end{aligned}$$

Il modello nelle variabili originali diventa

$$\begin{aligned} \hat{Y}(X) &= \mu_Y + S_Y r_{XY} \frac{X - \mu_X}{S_X} = \\ &= r_{XY} \frac{S_Y}{S_X} X + \mu_Y - \mu_X r_{XY} \frac{S_Y}{S_X} = \\ &= \frac{S_{XY}^2}{S_X^2} X + \mu_Y - \mu_X \frac{S_{XY}^2}{S_X^2} \end{aligned}$$

Regressione lineare come MLE

La regressione lineare può essere interpretata come risultato di un metodo di Maximum Likelihood Estimation, supponendo che ogni osservazione y_n sia il risultato del modello lineare $ax_n + b$ con l'aggiunta di un errore ε_n ,

$$y_n = ax_n + b + \varepsilon_n,$$

di variabili gaussiane indipendenti non correlate, identicamente distribuite a media nulla e varianza σ^2 ,

$$\varepsilon_n \sim \mathcal{N}(0, \sigma^2) \sim \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right]$$

$$\mathbb{E}[\varepsilon_m \varepsilon_n] = \sigma^2 \delta_{mn}.$$

Assumendo le x_n , a , b osservazioni e parametri deterministici, senza incertezza, segue che

$$y_n \sim \mathcal{N}(ax_n + b, \sigma^2).$$

I parametri ottimi del modello a , b , σ^2 rendono massima la probabilità,

$$p(\mathbf{x}, \mathbf{y}|\theta) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(ax_n + b - y_n)^2}{2\sigma^2}\right],$$

nel caso in cui le osservazioni siano indipendenti e quindi valga $p(\mathbf{x}, \mathbf{y}|\theta) = \prod_{n=1}^N p(x_n, y_n|\theta)$, o il suo logaritmo

$$\ln p(\mathbf{x}, \mathbf{y}|\theta) = -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{n=1}^N \frac{(ax_n + b - y_n)^2}{2\sigma^2}.$$

L'annullamento delle derivate parziali produce il sistema lineare

$$\begin{cases} 0 = \frac{\partial p}{\partial a} = \sum_{n=1}^N (ax_n + b - y_n)x_n \\ 0 = \frac{\partial p}{\partial b} = \sum_{n=1}^N (ax_n + b - y_n) \\ 0 = \frac{\partial p}{\partial \sigma^2} = -\frac{N}{2} \frac{1}{\sigma^2} + \sum_{n=1}^N \frac{(ax_n + b - y_n)^2}{2\sigma^4} \end{cases}$$

la cui soluzione, ipotizzando di aver normalizzato i campioni sulle medie e varianze campionarie,

$$a^* = r_{XY} \quad , \quad b^* = 0 \quad , \quad \sigma^{2*} = \frac{1}{N} \sum_{n=1}^N (ax_n + b - y_n)^2 = \frac{e^*}{N} = \frac{N-1}{N} (1 - r_{XY}^2)$$

Statistiche χ^2 e t

Siano y_n delle variabili gaussiane con varianza σ^2 e media $ax_n + b$, con $b = 0$ e

$$\begin{aligned} a = r &= \frac{S_{XY}^2}{S_X S_Y} = \\ &= \frac{\frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})(Y_n - \bar{Y})}{\sqrt{\frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})^2} \sqrt{\frac{1}{N-1} \sum_{n=1}^N (Y_n - \bar{Y})^2}} = \\ &= \frac{1}{N-1} \sum_{n=1}^N \left(\frac{X_n - \bar{X}}{S_X} \right) \left(\frac{Y_n - \bar{Y}}{S_Y} \right) = \\ &= \frac{1}{N-1} \sum_{n=1}^N x_n y_n \end{aligned}$$

$$\mathbb{E}[a] = \frac{1}{N-1} \sum_{m=1}^N x_m \mathbb{E}[y_m]$$

$$\varepsilon_n := y_n - ax_n \sim \mathcal{N}(0, \sigma^2) .$$

$$0 = \mathbb{E}[\varepsilon_n] = \mathbb{E}[y_n] - \mathbb{E}[a] x_n$$

$$\begin{aligned} \sigma^2 \delta_{mn} &= \mathbb{E}[\varepsilon_m \varepsilon_n] = \\ &= \mathbb{E}[(y_m - ax_m)(y_n - ax_n)] = \\ &= \mathbb{E}[y_m y_n] - x_m \mathbb{E}[ay_n] - y_m \mathbb{E}[ax_n] + x_n x_m \mathbb{E}[a^2] = \\ &= \mathbb{E}[y_m y_n] = \mathbb{E}[(\bar{y}_m + \delta y_m)(\bar{y}_n + \delta y_n)] = \\ &= \bar{y}_m \bar{y}_n + \mathbb{E}[\delta y_m \delta y_n] = \\ &= \bar{y}_m \bar{y}_n + \sigma^2 \delta_{mn} \end{aligned}$$

$$\begin{aligned}
\mathbb{E} \left[\sum_{n=1}^N (y_n - ax_n)^2 \right] &= \mathbb{E} \left[\sum_{n=1}^N \left(y_n - \frac{1}{N-1} \sum_{m=1}^N y_m x_m x_n \right)^2 \right] = \\
&= \sum_{n=1}^N \mathbb{E} \left[y_n^2 - \frac{2}{N-1} \sum_{m=1}^N y_n y_m x_m x_n + \frac{2}{(N-1)^2} \sum_{m=1}^N \sum_{p>m}^N y_m y_p x_m x_p x_n^2 + \frac{1}{(N-1)^2} \sum_{m=1}^N y_m^2 x_m^2 x_n^2 \right] = \\
&= \sum_{n=1}^N (\bar{y}_n^2 + \sigma^2) - \frac{2}{N-1} \sum_{n,m=1}^N (\bar{y}_m \bar{y}_n + \sigma^2 \delta_{mn}) x_m x_n + \frac{2}{(N-1)^2} \sum_{n,m=1}^N \sum_{p>m}^N (\bar{y}_m \bar{y}_p + \sigma^2 \delta_{mp}) x_m x_p x_n^2 + \frac{1}{(N-1)^2} \sum_{m,n=1}^N y_m^2 x_m^2 x_n^2 = \\
&= \bar{a}^2 \sum_{n=1}^N x_n^2 + N\sigma^2 - \frac{2\bar{a}^2}{N-1} \sum_{m,n=1}^N x_m^2 x_n^2 - \frac{2}{N-1} \sigma^2 \sum_{n=1}^N x_n^2 + \\
&\quad + \frac{2}{(N-1)^2} \sum_{n,m=1}^N \sum_{p>m}^N x_m^2 x_p^2 x_n^2 + \frac{\bar{a}^2}{(N-1)^2} \sum_{m,n=1}^N x_m^4 x_n^2 + \frac{1}{(N-1)^2} \sigma^2 \sum_{m,n=1}^N x_m^2 x_n^2 = \\
&= \bar{a}^2(N-1) + N\sigma^2 - 2\bar{a}^2(N-1) - 2\sigma^2 + 2 \sum_{p \dots} \dots + \dots \bar{a}^2 \sum_{\dots} x_m^4 + \sigma^2
\end{aligned}$$

ANOVA - Analysis of variance

Metodo

Esempio

```

# ANOVA - Analysis of variance

import numpy as np
from scipy.stats import f_oneway

# Sample data: test scores from three teaching methods
method_A = [88, 92, 85, 91, 87]
method_B = [78, 85, 80, 83, 79]
method_C = [94, 89, 91, 96, 90]

# Perform one-way ANOVA
f_statistic, p_value = f_oneway(method_A, method_B, method_C)

# Display results
print(f"F-Statistic: {f_statistic}")
print(f"P-Value: {p_value}")

# Interpretation
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: There is a significant difference between the
    ↪ groups.")
else:
    print("Fail to reject the null hypothesis: No significant difference between the
    ↪ groups.")

```

```

F-Statistic: 18.806324110671927
P-Value: 0.00020023178541394456
Reject the null hypothesis: There is a significant difference between the groups.

```


Questa sezione contiene le dimostrazioni usate all'interno del bbook. Le dimostrazioni vengono fornite per completezza, ma vengono raccolte qui per non appesantire inutilmente le altre sezioni. Alcune di esse possono richiedere qualche strumento che va oltre gli strumenti del valore atteso degli strumenti messi a disposizione dalla preparazione fornita da una scuola superiore.

La distinzione da una raccolta di appunti personali potrebbe non essere così definita.

19.1 Distribuzione gaussiana come limite della distribuzione t -Student

Per $N \rightarrow +\infty$, la distribuzione t_N tende alla distribuzione gaussiana $\mathcal{N}(0, 1)$.

Dimostrazione

Ricordandosi la definizione di e^x come limite della successione $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$, e usando l'approssimazione asintotica per $n \rightarrow \infty$ della funzione Gamma $\Gamma(n + \alpha) \sim \Gamma(n)n^\alpha$, si può dimostrare che

$$\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} = \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}} \left[\left(1 + \frac{x^2}{n}\right)^n\right]^{-\frac{1}{2}} \rightarrow 1 \cdot e^{-\frac{x^2}{2}}$$

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n} \Gamma\left(\frac{n}{2}\right)} \sim \frac{\Gamma\left(\frac{n}{2}\right) \left(\frac{n}{2}\right)^{\frac{1}{2}}}{\sqrt{n} \Gamma\left(\frac{n}{2}\right)} = \frac{1}{\sqrt{2}},$$

e quindi la funzione t -Student tende alla distribuzione normale con valore atteso 0 e varianza unitaria,

$$f(x) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

19.2 Somma/differenza di variabili gaussiane indipendenti

Con la stessa varianza

todo

Con varianza diversa

todo

19.3 Distribuzione della statistica $t \sim t_{N-1}$ nel t -test per campioni singoli

Dimostrazione

Dati N campioni $\{X_n\}_{n=1:N}$ iid con distribuzione gaussiana $X_n \sim \mathcal{N}(\mu, \sigma^2)$ allora la statistica t dei t -test per singolo campione o per coppie di campioni, costruita con la media campionaria \bar{X} e la varianza campionaria S segue una distribuzione t -Student con $N - 1$ gradi di libertà,

$$t := \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}} \sim t_{N-1}.$$

Questa affermazione risulta vera poiché la statistica test t può essere riportata alla definizione di una distribuzione t_ν -Student,

$$t_\nu \sim T = \frac{Z}{\sqrt{\frac{V}{\nu}}},$$

con $Z \sim \mathcal{N}(0, 1)$ e $V \sim \chi_\nu^2$, con $\nu = N - 1$. Infatti:

- la media campionaria è gaussiana con varianza $\frac{\sigma^2}{N}$. Si definisce quindi la variabile Z con distribuzione normale di media nulla e varianza unitaria, $Z := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \sim \mathcal{N}(0, 1)$
- la somma $V := \sum_{n=1}^N \left(\frac{X_n - \bar{X}}{\sigma} \right)^2$ è una somma di N variabili normali, di cui solo $N - 1$ sono indipendenti (poiché è stata usata la media campionaria e non il valore medio «vero» della distribuzione), $V \sim \chi_{N-1}^2$ (*dimostrazione*). La variabile V può essere espressa in termini della varianza campionaria, $V = (N - 1) \frac{S^2}{\sigma^2}$
- per definizione della distribuzione di t -Student, $\frac{Z}{\sqrt{V/(N-1)}} \sim \chi_{N-1}^2$
- questo rapporto coincide alla statistica test, infatti:

$$t_{N-1} \sim \frac{Z}{\sqrt{\frac{V}{N-1}}} = \frac{\frac{(\bar{X} - \mu)\sqrt{N}}{\sigma}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}} =: t.$$

19.4 Indipendenza di media campionaria e varianza campionaria per variabili gaussiane iid

Dimostrazione

todo

$$19.5 \sum_{n=1}^N \left(\frac{X_n - \bar{X}}{\sigma^2} \right)^2 \sim \chi_{N-1}^2 \text{ per variabili gaussiane iid}$$

Dimostrazione

todo

Parte IV

Introduzione ai metodi in statistica e AI

Introduzione ai metodi in AI

In questa sezione si descriveranno alcuni metodi utilizzati nelle applicazioni di **machine learning**. I contenuti sono organizzati in 3 sezioni, seguendo una delle più comuni classificazioni in machine learning.

Approccio. Alla completezza e al rigore dello sviluppo teorico dei metodi, viene preferita una descrizione dei metodi tramite esempi e applicazioni.

Benché questo possa essere un approccio ad alto rischio di creazione di utenti acritici di strumenti che non comprendono, la speranza è di mitigare questo rischio con continui moniti a prestare attenzione e un supporto del lettore da parte di gente con un minimo di esperienza.

Argomenti. In questa sezione verranno presentati alcuni metodi e applicazioni del **machine learning**, che possono essere classificate in tre grandi classi di apprendimento:

- SL, supervised learning: regression and classification
- UL, unsupervised learning: clustering
- ML, machine learning: control

Dopo aver presentato le tecniche classiche, verranno introdotte le reti neurali, gli algoritmi fondamentali che hanno permesso un uso pratico ed efficiente di reti profonde, e alcune architetture fondamentali di reti neurali.

Le applicazioni verranno affrontate inizialmente con approcci classici, affidandoci alla libreria [sci-kit](#), e successivamente con tecniche «deep» grazie alla libreria [PyTorch](#).

CAPITOLO 21

Supervised Learning

Unsupervised Learning

Le principali attività dell'apprendimento non supervisionato sono:

- il clustering: raggruppamento dei dati in classi più o meno omogenee
- la riduzione delle dimensioni: sintesi delle dinamiche/informazioni «principali» contenute nei dati. Una caratteristica che permette di distinguere gli algoritmi è la definizione di «principali»; così ad esempio:
 - PCA (o la cara vecchia SVD): seleziona le dinamiche «più energetiche» tra di loro ortogonali contenute in un segnale
 - ICA: seleziona le dinamiche «maggiormente indipendenti» contenute nel segnale

Un confronto tra ICA e PCA può chiarire il significato delle diverse definizioni e le differenze nei risultati prodotti.

22.1 PCA

22.1.1 Esempio: compressione di immagini

Un'immagine in scala di grigio può essere rappresentata con una matrice \mathbf{A} , i cui elementi rappresentano il valore di grigio di ogni pixel.

```
(-0.5, 511.5, 511.5, -0.5)
```

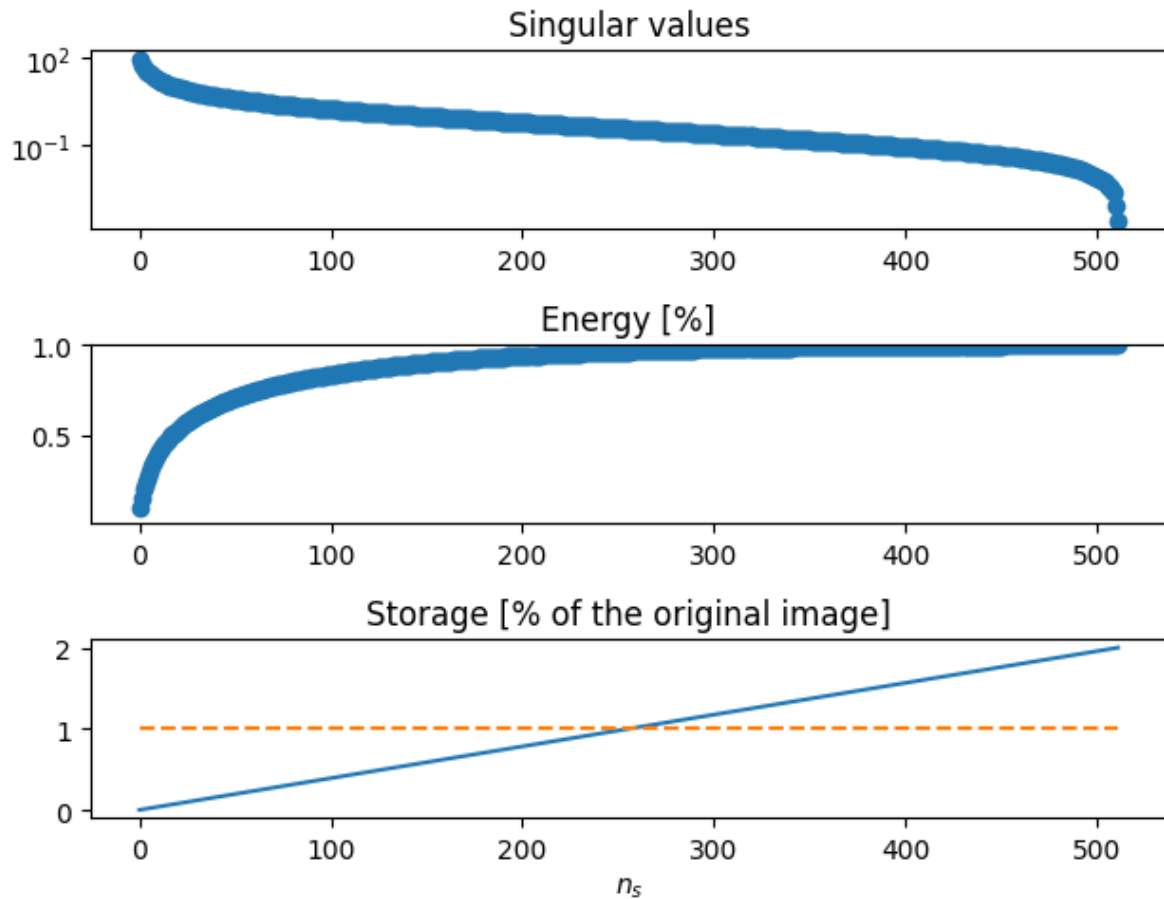


Preprocessing e SVD

Viene rimosso il valore medio dall'immagine, prima di valutare la SVD della matrice A

Risultati della SVD

I valori singolari rappresentano il contenuto energetico dei singoli modi. La somma cumulativa dei valori singolari rappresenta il contenuto energetico della



Ricostruzione dell'immagine compressa

```
Image reconstruction with 100 singular values:
> Energy error [% original]: 0.1647581016503445
> Storage      [% original]: 0.3870964050292969
```

Compressed image with 100 s.v.



Difference



22.2 ICA e PCA

22.2.1 Riferimenti

scikit

- <https://scikit-learn.org/dev/modules/generated/sklearn.decomposition.FastICA.html>

Carten Klein

- <https://github.com/akcarsten/>
- https://github.com/akcarsten/Independent_Component_Analysis

22.2.2 FastICA on 2D point clouds

Esempio disponibile sul sito di [scikit-learn](https://scikit-learn.org/).

Librerie e funzioni di comodo

Import librerie

```
%reset -f

import numpy as np
from sklearn.decomposition import PCA, FastICA

import matplotlib.pyplot as plt
```

Funzioni utili

```
### utils

def plot_samples(S, axis_list=None):
    """ util function for plotting scatter plots """
    plt.scatter(
        S[:,0], S[:,1], s=2, marker="o", zorder=10, color="steelblue", alpha=0.5
    )
    if axis_list is not None:
        for axis, color, label in axis_list:
            x_axis, y_axis = axis / axis.std()
            plt.quiver(
                (0, 0),
                (0, 0),
                x_axis,
                y_axis,
                zorder=11,
                width=0.01,
                scale=6,
                color=color,
                label=label,
```

(continues on next page)

(continua dalla pagina precedente)

```

    )

plt.hlines(0, -3, 3, color="black", linewidth=0.5)
plt.vlines(0, -3, 3, color="black", linewidth=0.5)
plt.xlim(-3, 3)
plt.ylim(-3, 3)
plt.gca().set_aspect("equal")
plt.xlabel("x")
plt.ylabel("y")

```

Generazione campione

La formulazione più comune dei metodi usa l'espressione

$$\mathbf{X} = \mathbf{A} \mathbf{S},$$

che lega osservazioni X e segnali S tramite la matrice di mixing A . Per le strutture dati di Python, è conveniente (**todo provare! O è dovuto solo alla forma mentis di chi usa Python?**) scrivere la relazione trasposta,

$$\mathbf{X}^T = \mathbf{S}^T \mathbf{A}^T.$$

```

#> Sample data generation
# Initialize a default random number generator
rng = np.random.default_rng(42)
# rng = np.random.RandomState(42)
# np.random.RandomState() is deprecated! Use random.default_rng(),
# but FastICA needs a np.random.RandomState as an optional input random_state

# Create data: non-isotropic mixing of 2 t-Student variables:
# 1. Create signals S, as 2 t-Student distributions
# 1. sampling from a t-Student distribution with degrees of freedom df
# matrix S has dimensions (n_rows, n_cols) = (n_samples, n_dim), interpreting
# each row as a sample, and each row as a dimension of the data
df, n_samples, n_dims = 1.5, 10000, 2
S = rng.standard_t(df, size=(n_samples, n_dims))

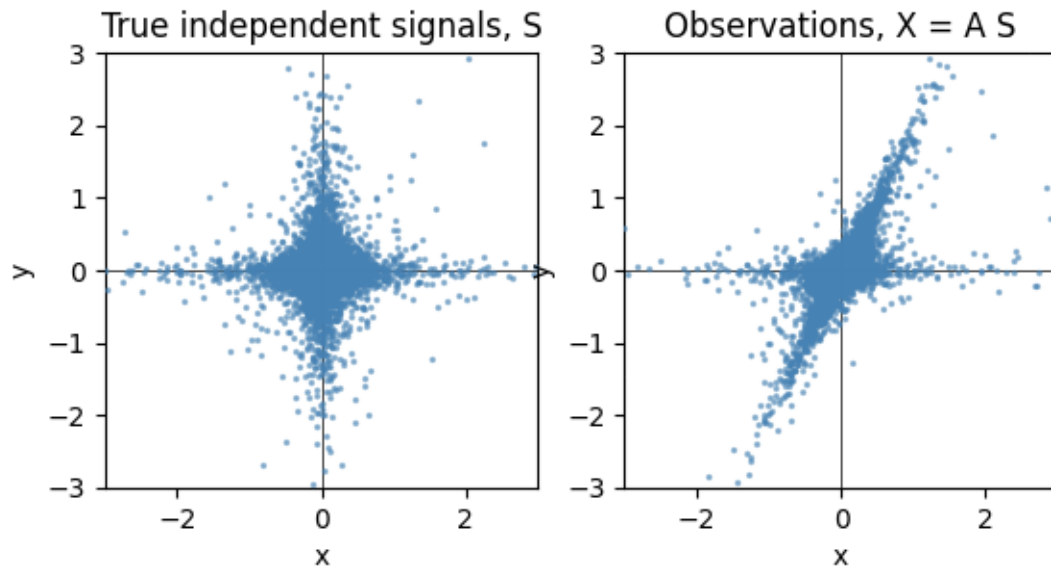
# 2. scale one component
S[0,:] *= 2.0
plt.subplot(1, 2, 1)
plot_samples(S / S.std())
plt.title("True independent signals, S")

# 3. Mix components
A = np.array([[1, 1], [0, 2]]) # Mixing matrix
X = np.dot(S, A.T)             # Generate observations

plt.subplot(1, 2, 2)
plot_samples(X / X.std())
plt.title("Observations, X = A S")

```

```
Text(0.5, 1.0, 'Observations, X = A S')
```



ICA e PCA

```
#> PCA
pca = PCA()
S_pca_ = pca.fit(X).transform(X)

#> ICA
ica = FastICA(random_state=3, whiten="arbitrary-variance")
# ica = FastICA(random_state=np.random.RandomState(42), whiten="arbitrary-variance")
S_ica_ = ica.fit(X).transform(X) # Estimate the sources
```

Risultati

```
#> Print results:
# Normalization: principal and independent components are usually defined up to a
# multiplicative factor;
# PCA and ICA provides information about the "shape" of the main components in a
# signal; usually, it's a
# good practice to have normalized info/results, that contains only shape info and no
# other arbitrary (non)-info
# - PCA results are usually normalized
# - ICA results seem to be not normalized by the algorithm; easy to perform
# normalization, as done below,
# to have unit-norm vectors to be easily compared with PCA.
# If normalization done outside functions, remember to normalize both signals and
# mixing matrix

print("\nPCA, principal components (rows of the matrix)")
print(pca.components_)
print("\nICA, independent components (rows of the matrix)")
print(ica.mixing_.T) # non-normalized
```

(continues on next page)

(continua dalla pagina precedente)

```
# print(ica.mixing_.T / np.linalg.norm(ica.mixing_.T,axis=1)[: , np.newaxis]) #_
↪normalized
print()
```

```
PCA, principal components (rows of the matrix)
[[-0.48647086 -0.8736968 ]
 [ 0.8736968  -0.48647086]]

ICA. indepndent components (rows of the matrix)
[[-9.46908181e+02 -1.89462935e+03]
 [ 6.85936578e+02 -8.07940837e-01]]
```

```
#> Plots
plt.figure()
plt.subplot(2, 2, 1)
plot_samples(S / S.std())
plt.title("True independent signals, S")

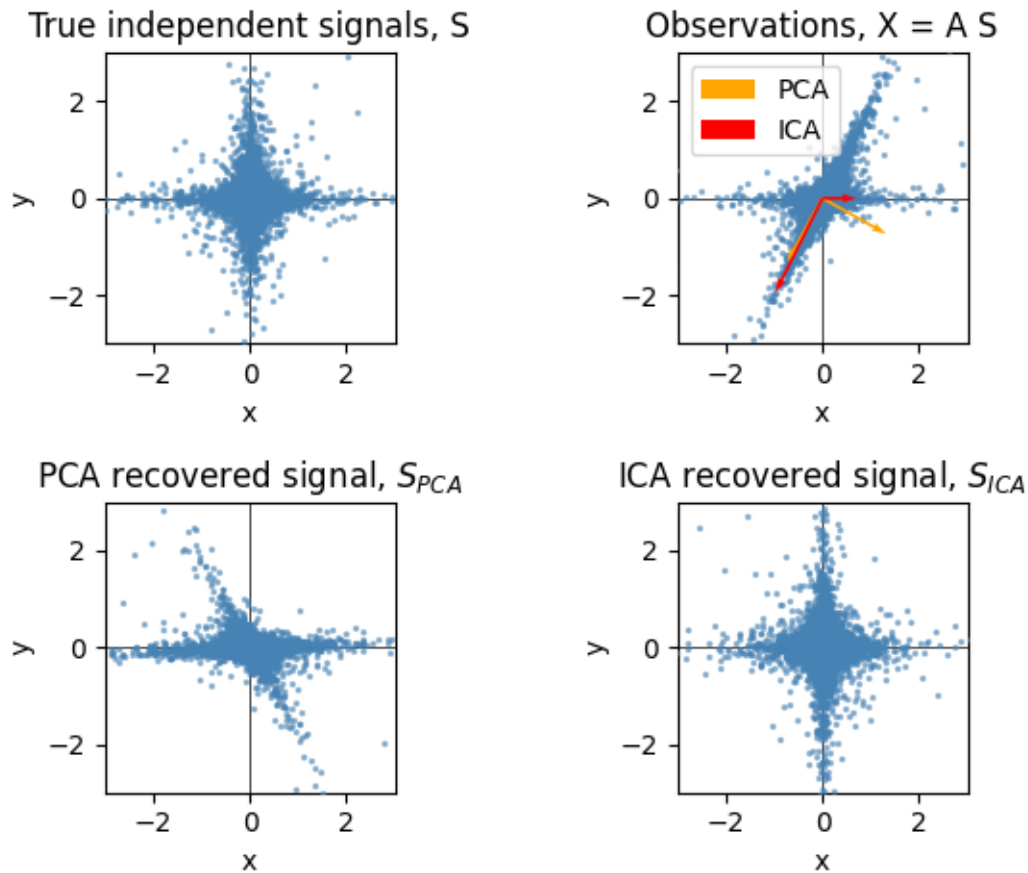
axis_list = [(pca.components_.T, "orange", "PCA"), (ica.mixing_, "red", "ICA")]
plt.subplot(2, 2, 2)
plot_samples(X / np.std(X), axis_list=axis_list)
legend = plt.legend(loc="upper left")
legend.set_zorder(100)

plt.title("Observations, X = A S")

plt.subplot(2, 2, 3)
plot_samples(S_pca_ / np.std(S_pca_))
plt.title("PCA recovered signal, $S_{PCA}$")

plt.subplot(2, 2, 4)
plot_samples(S_ica_ / np.std(S_ica_))
plt.title("ICA recovered signal, $S_{ICA}$")

plt.subplots_adjust(0.09, 0.04, 0.94, 0.94, 0.26, 0.36)
plt.tight_layout()
plt.show()
```



22.2.3 ICA e PCA: segnali in tempo

22.2.4 ICA e PCA - dettagli

Prendendo spunto da quanto fatto da [A.K.Carsten](#) si discutono alcuni dettagli dell'algoritmo ICA,¹ applicandolo a un problema di riconoscimento delle componenti indipendenti di segnali in tempo.

In particolare, si vuole discutere **todo**:

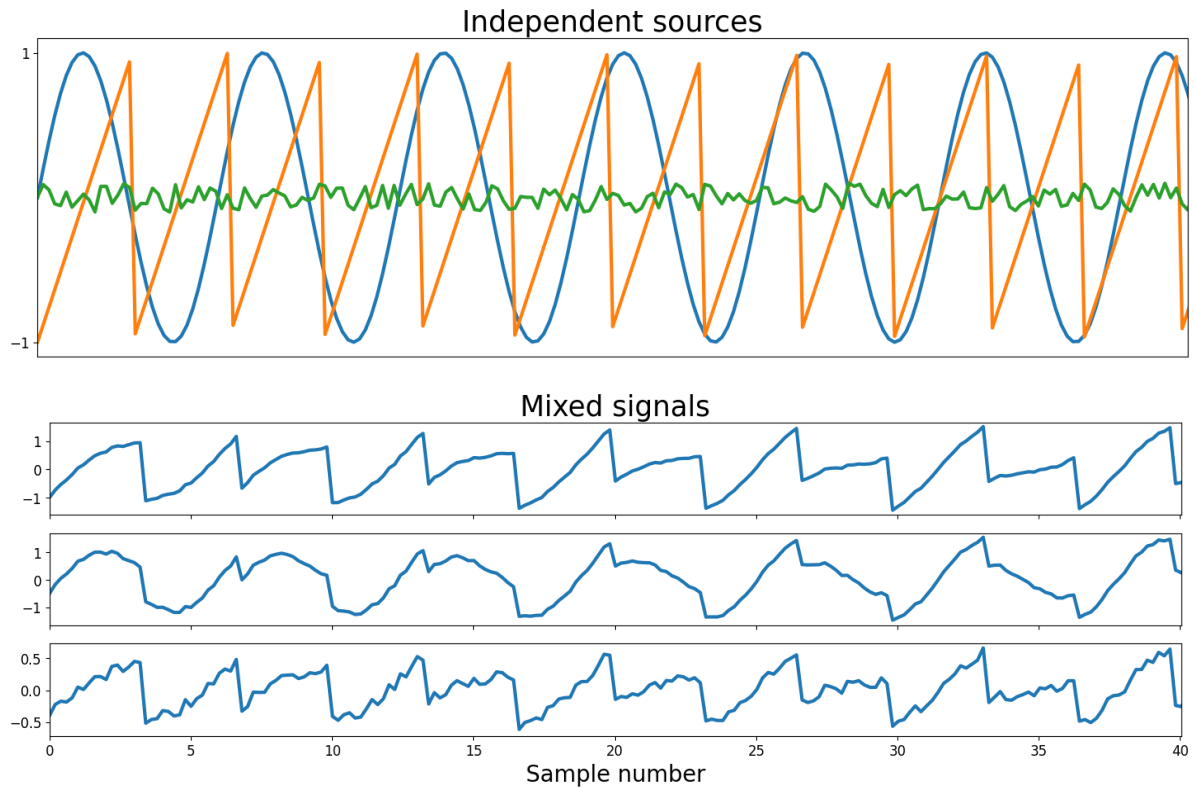
- i dettagli dell'algoritmo, prestando attenzione a
 - la relazione «non-gaussianità» \sim «indipendenza»
 - la misura di non gaussianità tramite neg-entropia, e le sue approssimazioni
 - l'espressione dell'iterazione di Newton nell'ottimizzazione della neg-entropia, che dà vita a un problema di punto fisso
- la necessità della non-gaussianità dei segnali

¹ A.Hyvarinen, E.Oja «Independent component analysis: algorithms and applications»

Librerie e funzioni utili

Generazione segnale

Viene generato un campione test con 200 campioni di 3 osservazioni, come risultato del mix di 3 segnali indipendenti, da ricostruire con l'algoritmo ICA.



Algoritmo

Pre-processing

L'osservazione \mathbf{X} vengono **depurate dalla media** e viene usata una trasformazione per combinare le osservazioni originali e ottenere un nuovo segnale \mathbf{X}_w con **componenti non correlate**. Il procedimento viene illustrato per i dati organizzati nella matrice,

$$\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_s] ,$$

in cui la colonna j -esima contiene le n_x osservazioni relative all'indice j (istante di tempo, paziente,...), mentre la riga i -esima contiene l'osservazione della quantità identificata dall'indice i per ogni valore di j (istante di tempo, paziente,...).

Teoria

Vengono svolte due operazioni:

- rimozione della media dalle osservazioni di ogni quantità, quindi rimozione del valore medio di ogni riga, $X_{ij} \leftarrow X_{ij} - \frac{1}{n_s} \sum_{j=1}^{n_s} X_{ij}$, in modo da avere i nuovi segnali a media nulla;
- ricerca di una trasformazione di coordinate (combinazione delle osservazioni) che renda la nuove componenti non correlate. Una stima senza bias della correlazione è

$$\hat{\mathbf{R}} = \frac{1}{n_s - 1} \mathbf{X} \mathbf{X}^*.$$

In generale, questa è una matrice piena di dimensioni (n_x, n_x) . Si vuole cercare la trasformazione di coordinate $\mathbf{X}_w = \mathbf{W}_w \mathbf{X}$, che garantisce che la nuove osservazioni siano non correlate con varianza unitaria (come *white noise*, e da qui il nome *whitening*), cioè

$$\mathbf{I} = \hat{\mathbf{R}}_w = \frac{1}{n_s - 1} \mathbf{X}_w \mathbf{X}_w^* = \frac{1}{n_s - 1} \mathbf{W}_w \mathbf{X} \mathbf{X}^* \mathbf{W}_w^* = \mathbf{W}_w \hat{\mathbf{R}} \mathbf{W}_w^*$$

È possibile trovare la matrice desiderata \mathbf{W}_w usando una tecnica di scomposizione della matrice di correlazione $\hat{\mathbf{R}}$, simmetrica (semi)definita positiva, come ad esempio la

- scomposizione agli autovalori, $\hat{\mathbf{R}} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^{-1}$, con $\mathbf{\Lambda}$ diagonale; sfruttando le proprietà delle matrici sdP, si può scrivere $\hat{\mathbf{R}} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^*$, con la matrice \mathbf{E} ortogonale tale che $\mathbf{E} \mathbf{E}^* = \mathbf{I}$.
- la SVD, $\hat{\mathbf{R}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$, con $\mathbf{\Sigma}$ diagonale; nel caso di matrice di partenza sdP, si può scrivere $\hat{\mathbf{R}} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^*$, con la matrice \mathbf{U} ortogonale tale che $\mathbf{U} \mathbf{U}^* = \mathbf{I}$.

Sfruttando quindi la proprietà delle scomposizioni presentate sopra, osservando l'analogia del risultato delle due scomposizioni nel caso di matrice sdP, e definendo $\mathbf{\Sigma}^{1/2}$ la matrice diagonale con elementi le radici quadrate di $\mathbf{\Sigma}$, si può scrivere

$$\begin{aligned} \mathbf{I} &= \mathbf{W}_w \hat{\mathbf{R}} \mathbf{W}_w^* = \\ &= \mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{\Sigma}^{1/2} \mathbf{U}^* \mathbf{W}_w^* = \\ &= \mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^* \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^* \mathbf{W}_w^* = \\ &= (\mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^*) (\mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^*)^*, \end{aligned}$$

e quindi (**todo** giustificare la comparsa di $\mathbf{U}^* \mathbf{U}$ per ottenere una matrice quadrata, giustificare la scelta di $\mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^* = \mathbf{I}$)

$$\mathbf{W}_w \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{U}^* = \mathbf{I} \quad \rightarrow \quad \mathbf{W}_w = \mathbf{U}^* \mathbf{\Sigma}^{-1/2} \mathbf{U}$$

FastICA

Viene qui illustrato l'algoritmo FastICA^{Pag. 118, 1}

todo

```
def fastIca(signals, alpha = 1, thresh=1e-8, iterations=5000):
    m, n = signals.shape

    # Initialize random weights
    W = np.random.rand(m, m)
```

(continues on next page)

(continua dalla pagina precedente)

```

for c in range(m):
    w = W[c, :].copy().reshape(m, 1)
    w = w / np.sqrt((w ** 2).sum())

    i = 0
    lim = 100
    while ((lim > thresh) & (i < iterations)):

        # Dot product of weight and signal
        ws = np.dot(w.T, signals)

        # Pass w*s into contrast function g
        wg = np.tanh(ws * alpha).T

        # Pass w*s into g prime
        wg_ = (1 - np.square(np.tanh(ws))) * alpha

        # Update weights
        wNew = (signals * wg.T).mean(axis=1) - wg_.mean() * w.squeeze()

        # Decorrelate weights
        wNew = wNew - np.dot(np.dot(wNew, W[:c].T), W[:c])
        wNew = wNew / np.sqrt((wNew ** 2).sum())

        # Calculate limit condition
        lim = np.abs(np.abs((wNew * w).sum()) - 1)

        # Update weights
        w = wNew

        # Update counter
        i += 1

    W[c, :] = w.T
return W

```

Applicazione del metodo al segnale

Il metodo viene applicato al problema considerato. In questo caso (segnali non gaussiani,...) il metodo riesce a ricostruire bene (**todo** è un bene «a occhio», qualitativo. *Quantificare!*) i 3 segnali indipendenti di partenza **a meno di un fattore moltiplicativo**, arbitrarietà propria del metodo.

todo Dire qualcosa sulle componenti principali indipendenti, discutendo le componenti della matrice **W**

```

#> Preprocessing
# Center signals
Xc, meanX = center(X)

# Whiten mixed signals
Xw, whiteM = whiten(Xc)

```

```

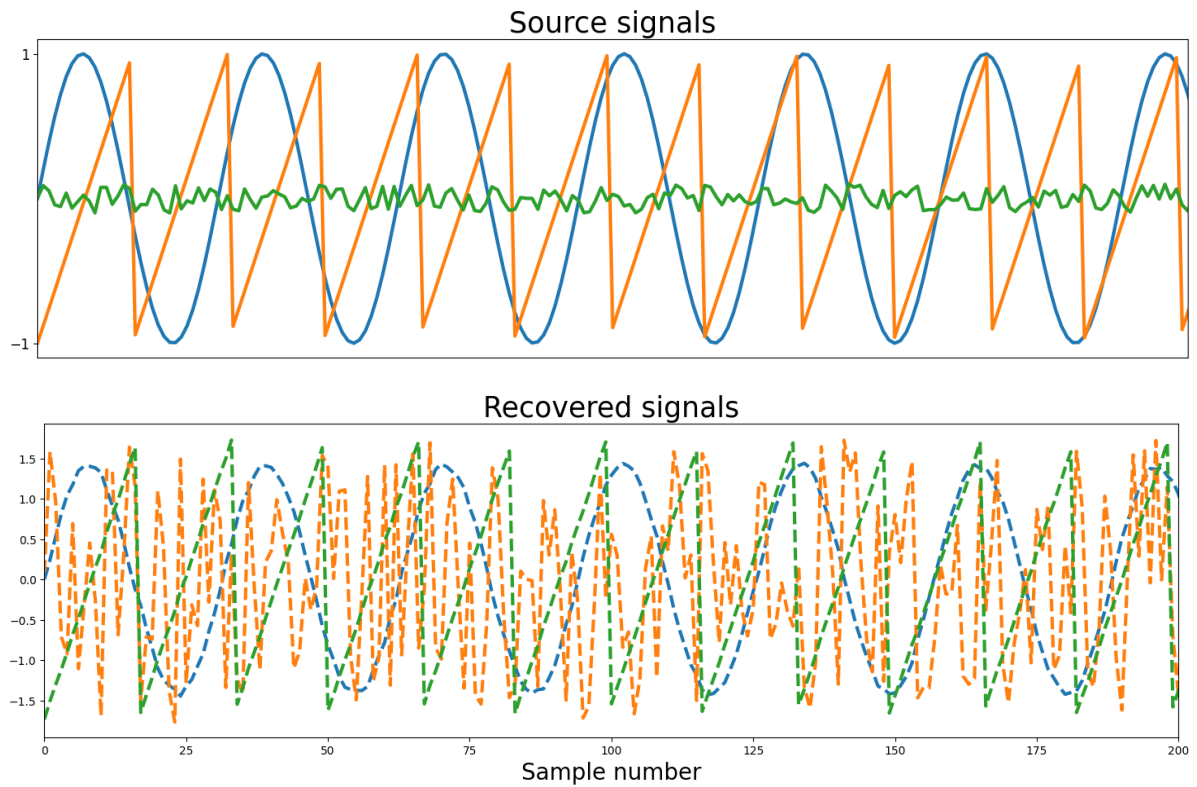
#> FastICA
W = fastIca(Xw, alpha=1)

```

(continues on next page)

(continua dalla pagina precedente)

```
#> Find unmixed signals using  
unMixed = Xw.T.dot(W.T)
```



```
#> Comparison of matrices involved, A, whiteM, W  
# todo...
```

CAPITOLO 23

Reinforcement Learning

Parte V

Supporto tecnico

CAPITOLO 24

Supporto tecnico

- git
- Servizi Google
- TeX

example-0

example-0 (*ch/statistics/hp-test*), [78](#)

example-1

example-1 (*ch/statistics/hp-test*), [78](#)