Οπτικοποίηση δεδομένων

Οπτικοποίηση δεδομένων

Το γεγονός ότι οι μελέτες διαφόρων φυσικών διεργασιών γίνεται με πειραματικές διατάξεις και επομένως μετρήσεις διαφόρων φυσικών μεγεθών, σημαίνει ότι θα πρέπει να είμαστε σε θέση να μελετήσουμε συσχετίσεις μεταξύ των μεγεθών αλλά και τα χαρακτηριστικά του κάθε μεγέθους που μετρήθηκε.

Θα πρέπει επομένως να μπορούμε να οπτικοποιήσουμε με ένα περισσότερο γραφικό τρόπο το τι έχουμε μετρήσει

Για το σκοπό αυτό χρησιμοποιούμε γραφικά πακέτα που μας προσφέρει η PYTHON

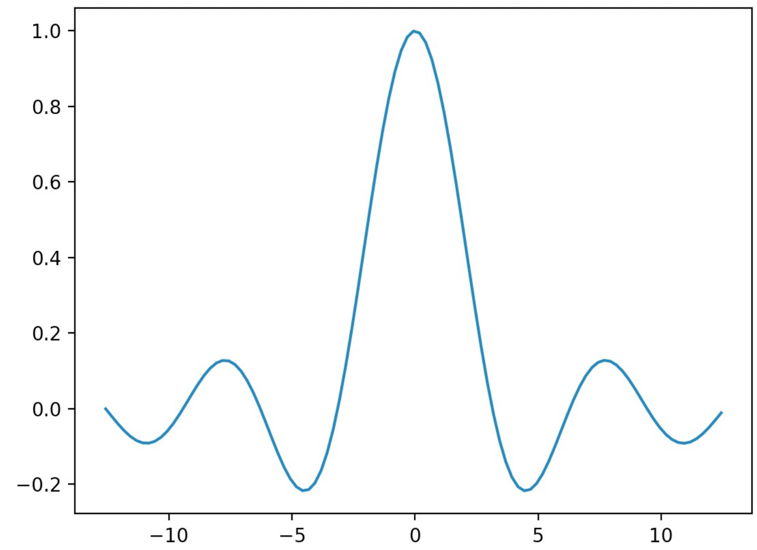
Συγκεκριμένα, θα χρησιμοποιήσουμε την βιβλιοθήκη **matplotlib**

Η εντολή που θα πρέπει να δώσουμε είναι: **import matplotlib.pyplot as plt**

Matplotlib - https://matplotlib.org/stable/index.html

Η πλέον συνηθισμένη εντολή για να κάνουμε το γράφημα μιας συνάρτησης είναι να δημιουργήσουμε μια λίστα (πίνακα) με τιμές της ανεξάρτητης μεταβλητής και να υπολογίσουμε τις αντίστοιχες τιμές της συνάρτησης

```
>>>>import numpy as np
>>>>import matplotlib.pyplot as plt
>>>>def sinc(x):
>>>>   return np.sin(x)/x
>>>>val = -4*np.pi
>>>>x = []
>>>>while val <= 4*np.pi:
>>>>    x += [val]
>>>>    val = val + 0.25
>>>>sincx = list(map(sinc,x))
>>>>plt.figure()
>>>>plt.plot(x,sincx)
>>>>plt.show()
```



Η μέθοδος **figure()** χρειάζεται για να δημιουργήσουμε έναν καμβά για να σχεδιαστεί το γράφημα

Η μέθοδος **plot()** θα κάνει ένα γράφημα 2-Δ αν δοθεί ένα σετ ζευγών (x,y).

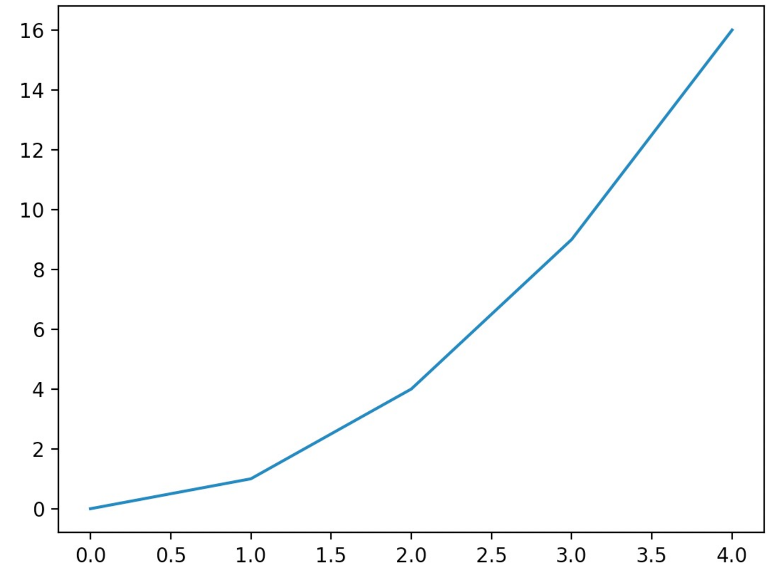
Η μέθοδος **show()** χρειάζεται για να δούμε στο τέλος όλα τα γραφήματα

Η εντολή plot και σχετιζόμενες εντολές

Η μέθοδος pyplot θα κάνει ένα 2-Δ γράφημα αν δοθεί ένα σετ ζευγών (x,y).

Για παράδειγμα θα έχουμε:

```
import numpy as np, matplotlib.pyplot as plt
dx = 1
xmin =0; xmax = 4
x = xmin
xval=[]; yval = []
while x <= xmax :
    xval +=[x]
    yval +=[x*x]
    x += dx
plt.plot(xval, yval)
plt.show()
```



Η εντολή plot και σχετιζόμενες επιλογές

Η εντολή plot δέχεται και επιπλέον προαιρετικά ορίσματα που βοηθούν να αλλάξουμε το χρώμα, το είδος του συμβόλου (marker) ή της γραμμής της γραφικής

Η σύνταξη είναι: plt.plot(x, y, ’**CS**’)

C: επιλογή χρώματος

[**k** = μαύρο, **g** = πράσινο, **b** = μπλε, **r** = κόκκινο,
c = κυανό, **y** = κίτρινο, **m** = magenta, **w** = άσπρο]

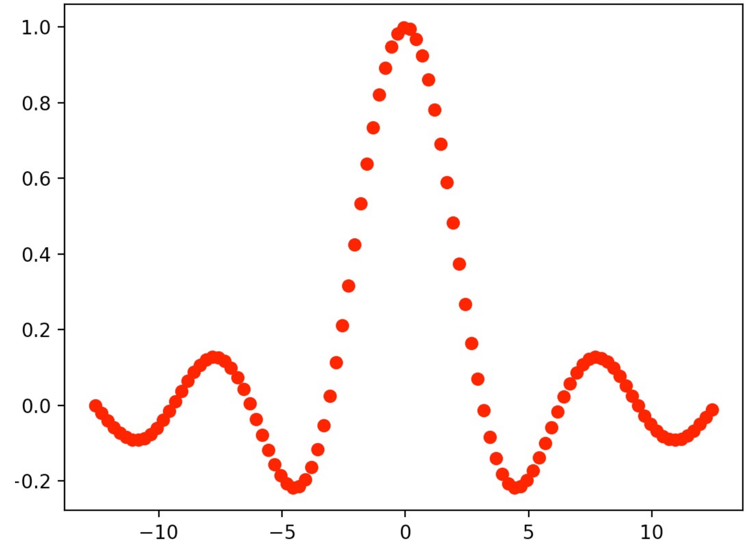
S: επιλογή marker

[**o**=κύκλοι, **s** = τετράγωνο, **.** = τελεία, **v** = τρίγωνο κάτω
^ = τρίγωνο πάνω, **<** = τρίγωνο αριστερά, **>** = τρίγωνο δεξιά
D = διαμάντι, **d** = λεπτό διαμάντι, ***** = αστεράκι, **x** = x, **+** = +]

S: επιλογή είδους γραμμής καμπύλης

[**-** = συνεχής, **--** = διακεκομμένη,
-. = -. **:** = διακεκομμένη με ….

```
>>>>plt.clf()   #clear figure
>>>>plt.figure()
>>>>plt.plot(x,sincx,'ro')
>>>>plt.show()
```



Η εντολή plot και σχετιζόμενες εντολές

Έστω για παράδειγμα θέλουμε το γράφημα του συνημίτονου από -2 έως 1.

Θα πρέπει πρώτα να διακριτοποιήσουμε τις τιμές κατά μήκος του x-άξονα και να εκτιμήσουμε την τιμή της συνάρτησης για κάθε τιμή του x.

```
import numpy as np; import matplotlib.pyplot as plt
xvals = [x/100 for x in range(-200,101)]
yvals = [np.cos(x/100) for x in range (-200,101)]
plt.plot(xvals,yvals)
plt.show()
```

Από το -2. έως το 1 με βήμα 0.01

Έστω ότι θέλουμε να συγκρίνουμε το αποτέλεσμα αυτό με το αποτέλεσμα που θα είχαμε αν χρησιμοποιούσαμε το ανάπτυγμα Taylor του συνημίτονου και κρατούσαμε μόνο τους δύο πρώτους όρους. Δηλαδή cos(x) = 1 – 0.5 * x * x

```
newyvals = [(1-0.5*x/100*x/100) for x in range(-200,101)]
plt.plot(xvals,newyvals, ‘r--’)
plt.title(‘Example plots’)
plt.xlabel(‘Input’)
plt.ylabel(‘Function values’)
plt.show()
```

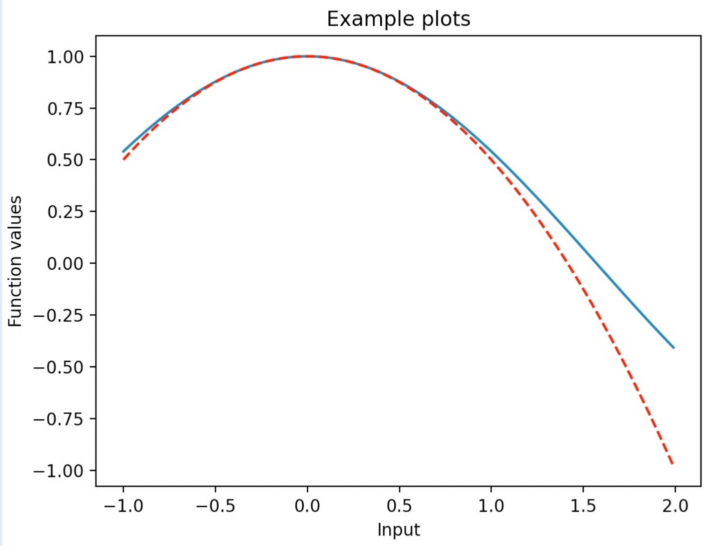
Το 3^{o} όρισμα προσδιορίζει να γίνει το γράφημα με κόκκινη (**r**) διακεκομμένη γραμμή (**--**)

Σε όλα τα γραφήματα θα πρέπει να υπάρχουν κατάλληλα ονοματισμένοι άξονες. Αυτό γίνεται με τις εντολές **title**, **xlabel** και **ylabel** που δίνουν τίτλο και ονόματα στους x, y- άξονες

Η εντολή plot και σχετιζόμενες εντολές

Αν θέλουμε και τα δύο γραφήματα μαζί, θα πρέπει να δώσουμε την εντολή show() στο τέλος αφού έχουμε κάνει όλα τα γραφήματα που θέλουμε:

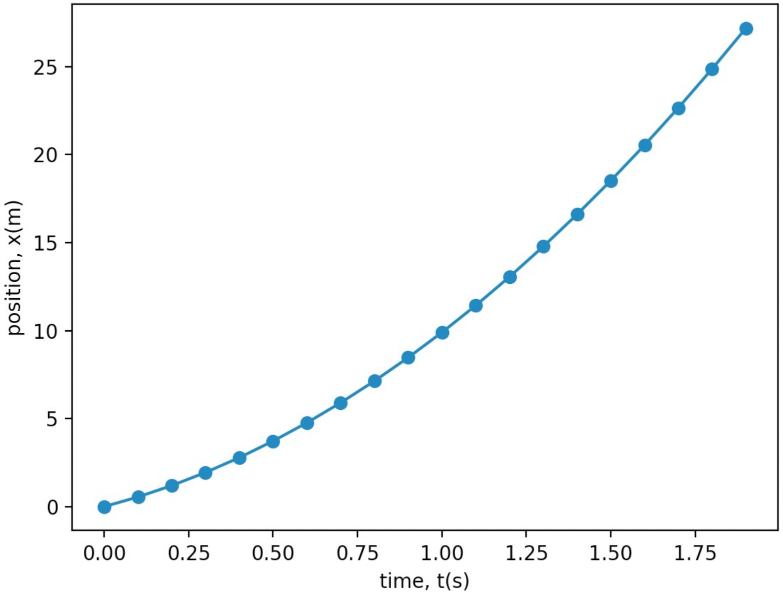
```
import numpy as np
import matplotlib.pyplot as plt
xvals = [x/100 for x in range(-100,201)]
yvals = [np.cos(x/100) for x in range (-100,201)]
plt.plot(xvals,yvals)
#plt.show()

newyvals = [(1-0.5*x*x/(100*100)) for x in range(-100,201)]
plt.plot(xvals,newyvals, ‘r--’)
plt.title(‘Example plots’)
plt.xlabel(‘Input’)
plt.ylabel(‘Function values’)
plt.show()
```



Παράδειγμα

Έστω ότι θέλουμε να δούμε την κίνηση με σταθερή επιτάχυνση

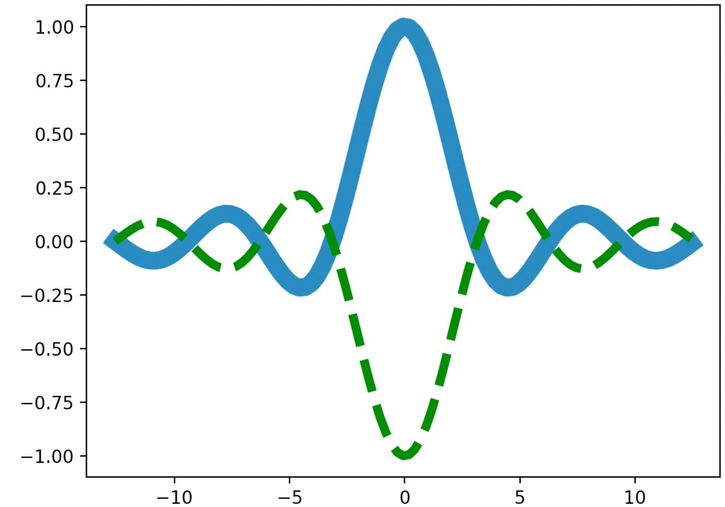
```
import numpy as np
import matplotlib.pyplot as plt

# Xrisi tis sunartisis split() gia na dwsoume polles times
# Tha prpei na dwthoun xwris "," alla me kena anamesa
# An thelame na dwthoun me "," tha grafame split(",")

a,v0 = input("Enter the acceleration and initial velocity ").split()
a = float(a)            # Prosoxi to a kai v0 poy diavazontai apo tin
v0= float(v0)           # entoli input einai strings. Metatropi se arithmous

t, dt, n = 0.0, 0.1, 20  # initial time, dt xroniko vima, and time intervals

ta, xa = [],[]          # Dimiourgia 2 listwn gia tis xronikes sti
                        # kai antistoixes theseis toy swmatos

for i in range(n):
    ta.append(t)
    xa.append(v0*t + a*t*t/2.0)
    t = t + dt

plt.figure()           # Dimiourgia neou parathyrou grafimatos

plt.plot(ta,xa, '-o')   # Kane to grafima me ta simeia me kuklous

plt.xlabel("time, t(s)")
plt.ylabel("position, x(m)")
plt.show()
```



Η εντολή plot και σχετιζόμενες επιλογές

Μπορούμε να έχουμε επιπλέον επιλογές ως προς το πάχος της γραμμής (**lw**) ή το πόσο διαφανής ή συμπαγής (**alpha**) είναι (alpha=1 συμπαγής, 0=διαφανής)

```
>>>>plt.clf()   #clear figure
>>>>plt.figure()
>>>>plt.plot(x,sincx,lw=10,alpha=0.5)
>>>>plt.plot(x,-sincx,'g--',lw=5)
>>>>plt.show()
```

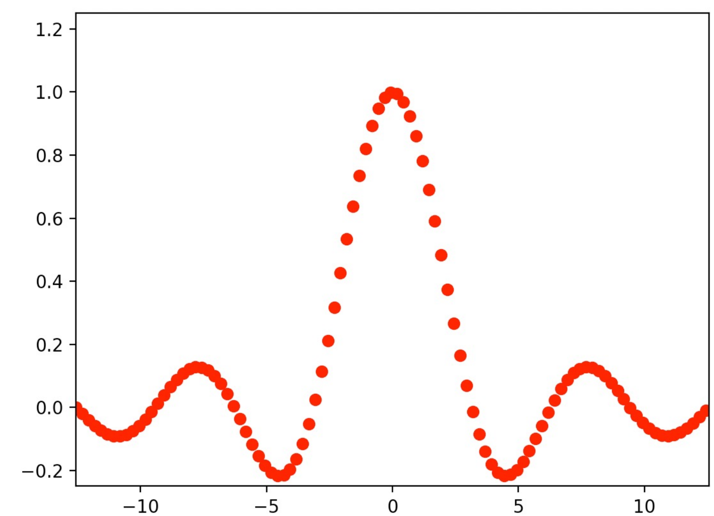


Μπορούμε να δώσουμε το εύρος της κλίμακας του x και y άξονα

plt.xlim([x_min, x_max]) ή
plt.xlim((x_min, x_max)) ή
plt.xlim(xmax=x_max, xmin=x_min)

plt.ylim([ymin,ymax])

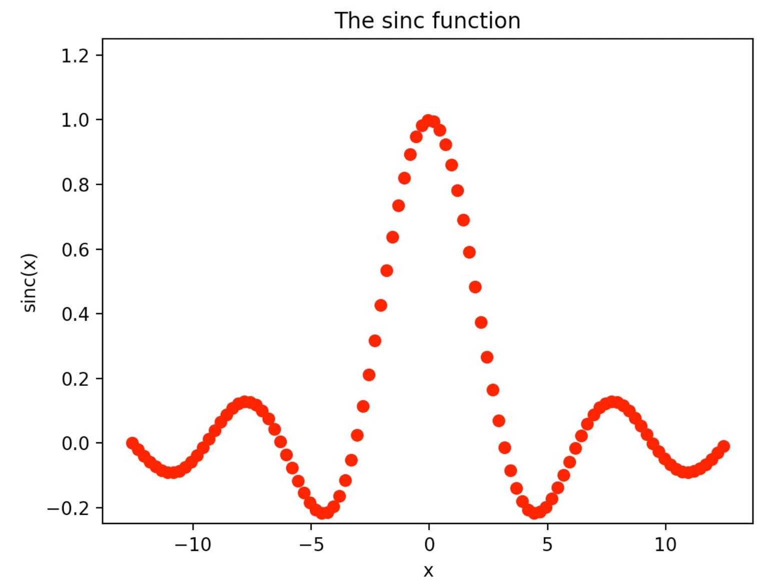
```
>>>>plt.clf()   #clear figure
>>>>plt.figure()
>>>>plt.plot(x,sincx,'ro')
>>>>plt.xlim(-4*np.pi, 4*np.pi)
>>>>plt.ylim(-0.25,1.25)
>>>>plt.show()
```



Η εντολή plot και σχετιζόμενες εντολές

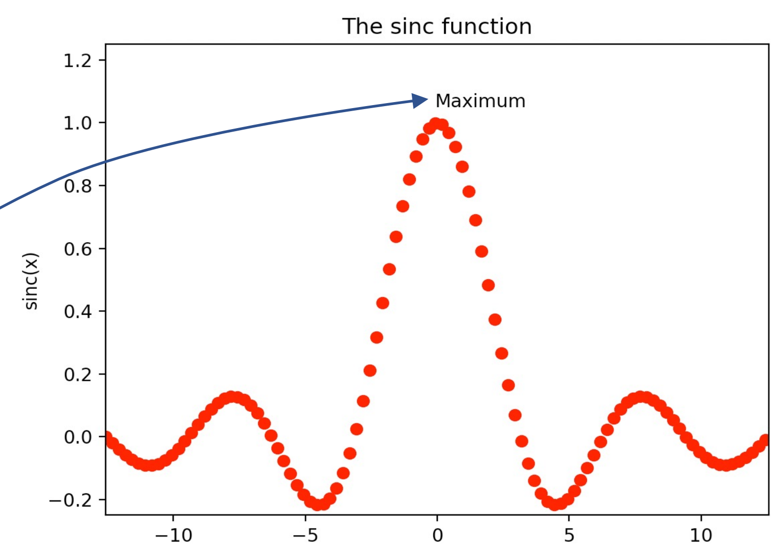
Σε όλα τα γραφήματα θα πρέπει να υπάρχουν κατάλληλα ονοματισμένοι άξονες. Αυτό γίνεται με τις εντολές **title**, **xlabel** και **ylabel** που δίνουν τίτλο και ονόματα στους x, y- άξονες

```
>>>>plt.clf()   #clear figure
>>>>plt.figure()
>>>>plt.plot(x,sincx,'ro')
>>>>plt.xlim(-4*np.pi, 4*np.pi)
>>>>plt.ylim(-0.25,1.25)
>>>>plt.title('The sinc function')
>>>>plt.xlabel('x')
>>>>plt.ylabel('sinc(x)')
>>>>plt.show()
```



Πολλές φορές είναι χρήσιμο να έχουμε text στο γράφημα για να επισημάνουμε κάτι.
Αυτό επιτυγχάνεται με την εντολή **plt.text(x,y,s)** όπου x, y οι συντεταγμένες και s string

```
>>>>plt.text(-0.1, 1.1,'Maximum')
```



Πολλαπλές καμπύλες στο ίδιο γράφημα

Μπορούμε να κάνουμε πολλές καμπύλες στο ίδιο γράφημα:

```
import numpy as np
import matplotlib.pyplot as plt

xv=[x*0.02 for x in range(51)]
y1=[np.exp(x*0.02) for x in range(51)]
y2=[(x*0.02)*(x*0.02) for x in range(51)]
plot.plt(xv,y1,xv,y2)
```

Εισαγωγή πολλαπλών ζευγών (x, y1) και (x, y2)

Η python θα διαλέξει διαφορετικά χρώματα μόνο της

Αποθήκευση γραφημάτων

Μπορείτε να αποθηκεύσετε τα γραφήματα σε pdf format με την εντολή:

plt.savefig(“myfigures.pdf”)

Κάνοντας τα plots ευπαρουσίαστα

Τίτλος:

Μπορούμε να δώσουμε τίτλο στο γράφημα με την εντολή:

plt.title(“Title”, **size**=24, **weight**=‘bold’)

Ονομασία αξόνων:

plt.xlabel(“speed”’)

plt.ylabel(“kinetic energy”’)

Αλλαγή του στυλ της γραμμής:

Μπορούμε να αλλάξουμε το στυλ της γραμμής αφού σχεδιάσουμε τη γραμμή:

plt.setp (**lines[0]**, **linestyle**=‘--’, **linewidth**=3, **color**=‘r’)

Αλλαγή της πρώτης γραμμής (lines[0]) σε --, με πάχος 3 και χρώματος κόκκινο

Εισαγωγή legend: (περιγραφή μιας γραμμής για παράδειγμα**)**

plt.plot(xvalue, yvalue, **label**=”Energy 1”)

plt.plot(xvalue, yvalue*yvalue, **label**=”Energy 2”)

plt.legend() εμφανίζει το legend

Εισαγωγή σφαλμάτων:

Μπορούμε να κάνουμε ένα γράφημα με σφάλματα (error bars)

plt.errorbar(xvalue, yvalue, yerr=y_errors, xerr=x_errors)

Ευπαρουσίαστα γραφήματα

Μπορούμε να αλλάξουμε τον y-άξονα από γραμμικό σε ημιλογαριθμικό άξονα:

plt.semilogy

Αν επιθυμούμε ημιλογαριθμικό τον x-άξονα τότε μπορούμε να γράψουμε:

plt.semilogx

Αν επιθυμούμε λογαριθμικούς και τους δύο άξονες τότε δίνουμε την εντολή:

plt.loglog

Πολλαπλά γραφήματα στον ίδιο κανβά

Πολλές περιπτώσεις χρειάζεται να έχουμε πολλά διαφορετικά γραφήματα στον ίδιο κανβά ακόμα και γραφήματα με διαφορετικούς άξονες

Κάθε φορά που δίνουμε την εντολή plt.figure() δημιουργείται ένας νέος κανβάς που έχει ένα νούμερο (1,2,3, κλπ)

Μπορούμε να δώσουμε νούμερο σε κάποιο figure και να προσδιορίσουμε το μέγεθός του:

plt.figure(num=3, figsize=(7,5)) Το μέγεθος είναι (x=7)x (y=5) ίντσες

Με βάση τον αριθμό του figure επιλέγουμε ποιο μπορεί να είναι ενεργό για γράφημα:

Αν θέλουμε να δημιουργήσουμε πολλαπλά γραφήματα μέσα στο ίδιο figure, μπορούμε να χωρίσουμε τον κανβά σε πολλά τμήματα χρησιμοποιώντας την εντολή subplot()

plt.subplot(nrows, ncols, curplot)

Η μέθοδος δέχεται 3 ορίσματα, το 1° δηλώνει τον αριθμό των γραμμών, το 2° τον αριθμό των στηλών και το 3° ποια θέση από τις nrows*ncols είναι ενεργός

```
>>>plt.subplot(2,3,1)
```
2 γραμμές, 3 στήλες

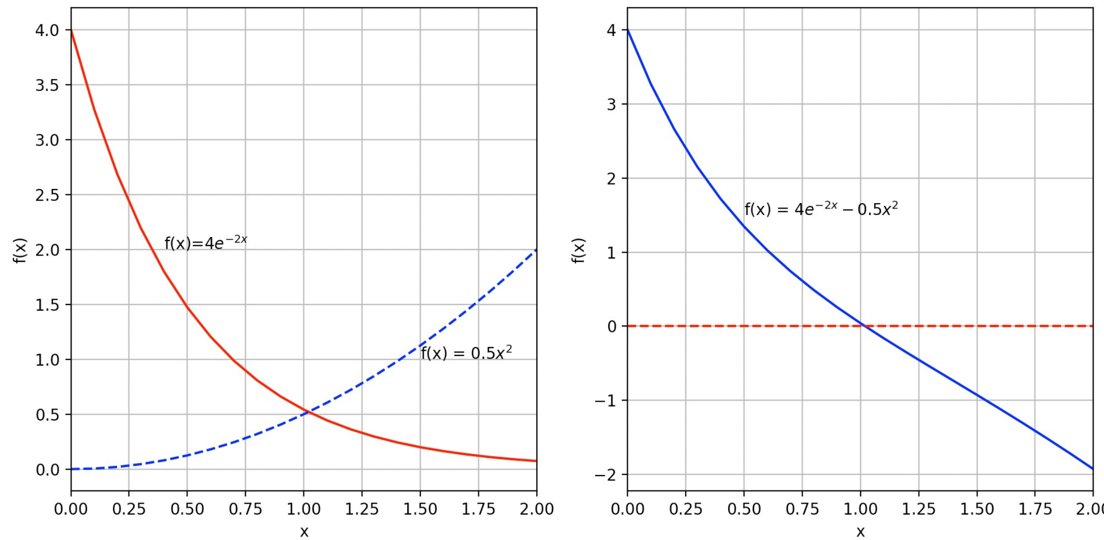
ncols =3

nrows = 2			
	1	2	3
	4	5	6

Τα νούμερα μπορούν να δοθούν ως ένα νούμερο πχ. 231 αντί για 2,3,2

Παράδειγμα

```
import numpy as np; import matplotlib.pyplot as plt
def func1(x):
   return 4*np.exp(-2*x)
def func2(x):
   return 0.5*x**2
def func_dif(x):
   return 4*np.exp(-2*x)-0.5*x**2
x=[x*0.1 for x in rage(21)]
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.plot(x,list(map(func1,x)),’r-’)
plt.plot(x,list(map(func2,x)),’b--’)
plt.text(0.4,2.0,r’f(x)=$4e^{-2x}$’)
plt.text(1.5,1.0,r'f(x) = $0.5x^2$’)
plt.xlabel('x’)
plt.ylabel('f(x)’)
plt.xlim(0,2)
plt.grid(True)
plt.subplot(1,2,2)
plt.plot(x,list(map(func_dif,x)),’b-’)
plt.text(0.5,1.5,r'f(x) = $4e^{-2x} - 0.5x{^2}$’)
plt.xlabel('x’)
plt.ylabel('f(x)’)
plt.xlim(0,2)
plt.grid(True)
plt.axhline(y=0,color='red',linestyle='--’)
plt.tight_layout()
plt.show()
```



Το $ χρησιμοποιείται για εισαγωγή special συμβόλων όπως superscripts(^), supscripts(_). Θα πρέπει το super(sup) script να περικλείεται σε {}. Πρέπει να κλείνει με ανάλογο $

εισαγωγή οριζόντιας γραμμής στο y=0

Ημιλογαριθμικό χαρτί

Πολλές φορές τα δεδομένα μας δεν περιγράφονται από μια απλή ευθεία αλλά ο νόμος της φυσικής που περιγράφει το φαινόμενο έχει μια εκθετική μορφή

Για παράδειγμα η ραδιενεργός διάσπαση κάποιων ραδιοϊσοτόπων.
Η διάσπαση ακολουθεί εκθετική μορφή σύμφωνα με τη σχέση: $A(t) = A_0 e^{-t/\tau}$

όπου A_0 είναι η ενεργότητα τη στιγμή t = 0 και τ η σταθερά διάσπασης.

Μπορούμε και πάλι να χρησιμοποιήσουμε την ίδια τεχνική για να βρούμε τη σταθερά διάσπασης και την αρχική ενεργότητα του δείγματος. Αρκεί να γράψουμε την παραπάνω εξίσωση σε γραμμική μορφή.

Η μετατροπή της εκθετικής εξίσωσης σε γραμμική γίνεται εύκολα λογαριθμίζοντας την εξίσωση:

$$\ln[A(t)] = \ln[A_0 e^{-t/\tau}] = \ln[A_0] - \frac{t}{\tau} = \ln[A_0] + \lambda t$$

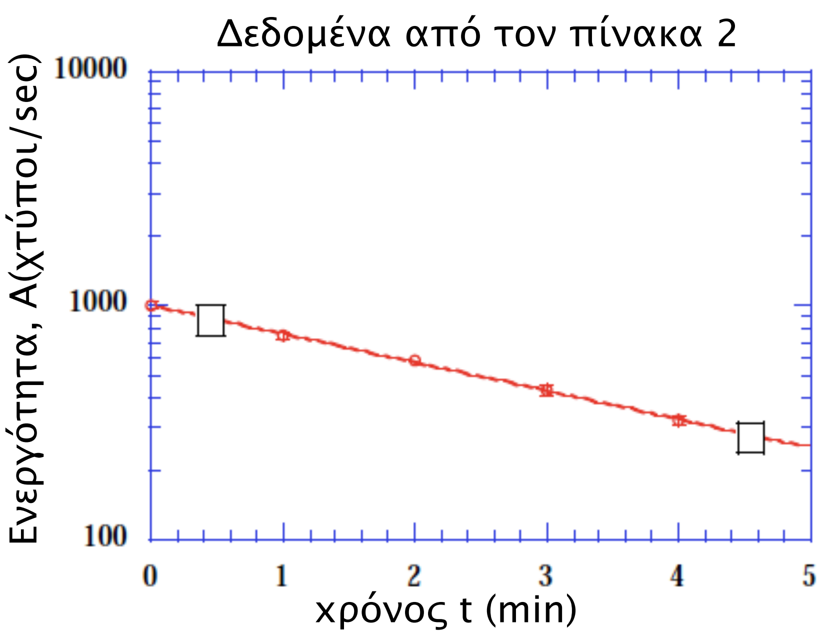
Ο όρος $\ln[A_0]$ αντιπροσωπεύει τον σταθερό όρο της εξίσωσης της ευθείας ενώ ο όρος $\lambda = -\frac{1}{\tau}$ την κλίση της. Ο όρος *t* αποτελεί την ανεξάρτητη μεταβλητή ενώ ο λογάριθμος της ενεργότητας, $\ln[A(t)]$, την εξαρτώμενη μεταβλητή.

Ημιλογαριθμικό χαρτί

Για τη γραφική παράσταση θα μπορούσαμε να υπολογίσουμε τους λογαρίθμους και να χρησιμοποιήσουμε χιλιοστομετρικό χαρτί ή να χρησιμοποιήσουμε το λεγόμενο ημιλογαριθμικό χαρτί.

Η χρήση του ημιλογαριθμικού χαρτιού διευκολύνει στην περίπτωση αυτή γιατί μπορούμε να θέσουμε τις μετρήσεις απευθείας στο γράφημα χωρίς επιπλέον υπολογισμούς. Το χαρτί περιέχει τον κάθετο άξονα με τέτοιο τρόπο ώστε οι κατακόρυφες αποστάσεις να είναι ανάλογες των επιθυμητών λογαρίθμων.

Ο παρακάτω πίνακας περιέχει τα δεδομένα της ραδιενεργούς διάσπασης

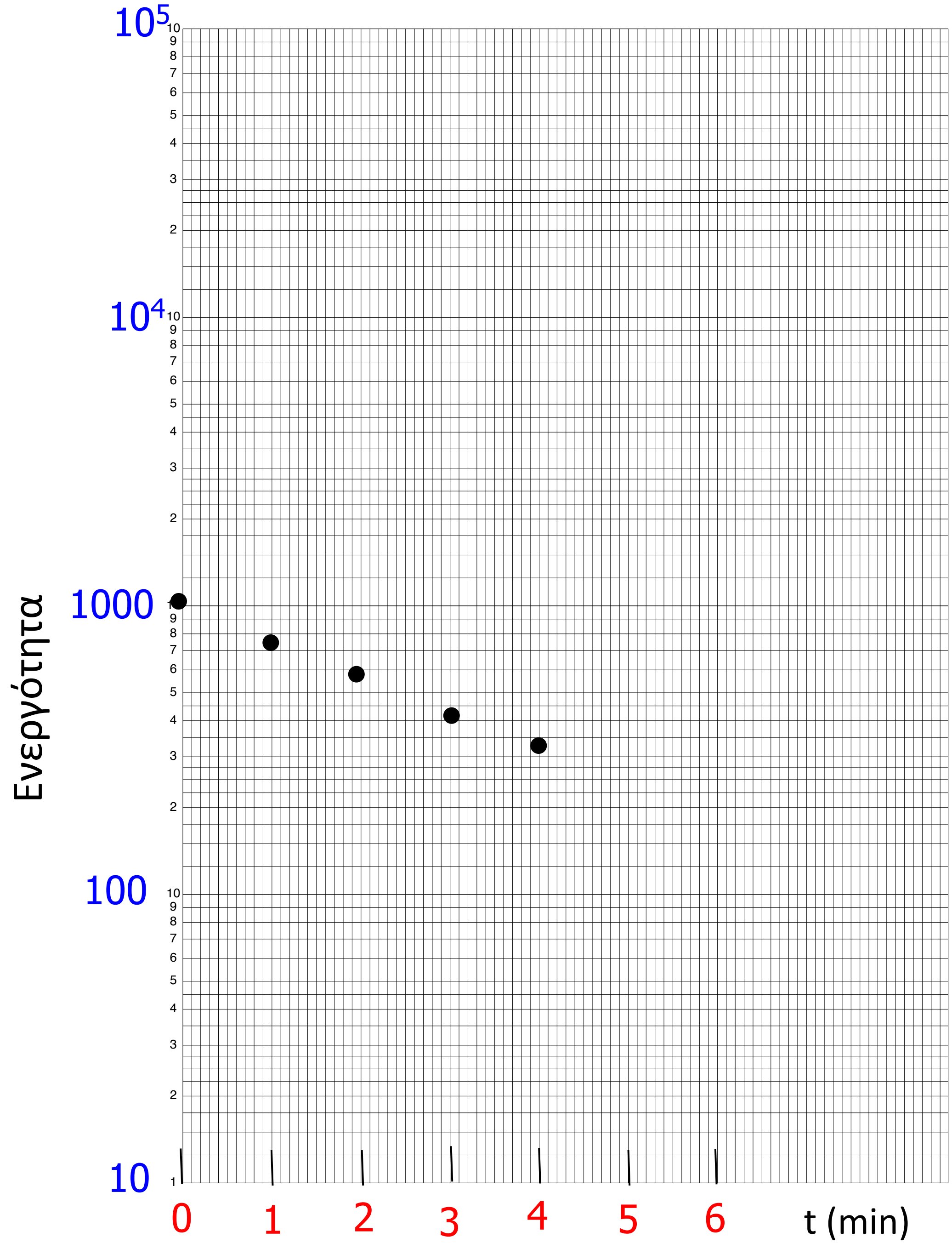


Πίνακας 2. Ενεργότητα ως προς χρόνο για το ραδιενεργό δείγμα μας

Χρόνος (min)	Ενεργότητα A και ΔA (χτύποι/s)
0	1010±31
1	740±28
2	585±24
3	435±21
4	325±18

Ημιλογαριθμικό χαρτί

Χρόνος (min)	Ενεργότητα Α και ΔΑ (χτύποι/s)
0	1010±31
1	740±28
2	585±24
3	435±21
4	325±18



Ημιλογαριθμικό χαρτί

Χρειάζεται ωστόσο κάποια προσοχή, γιατί τα δεδομένα δίνονται αυτόματα σε λογαριθμική κλίμακα, αλλά οι λογάριθμοι δεν υπολογίστηκαν. Απλά το χαρτί παρέχει τον κατάλληλο μετασχηματισμό.

Από την εξίσωση: $lny = lnA_0 - \frac{t}{\tau}$ έχουμε: $lny_1 - \ln y_2 = -\frac{1}{\tau}(t_1 - t_2) = \lambda(t_1 - t_2)$

Για να υπολογίσουμε την κλίση στο ημιλογαριθμικό χαρτί δεν μπορούμε να πάρουμε τις τιμές από το χαρτί αλλά πρέπει να υπολογίσουμε τους λογαρίθμους.

Για παράδειγμα: $-\frac{1}{\tau} = \frac{\ln(740/sec) - \ln(325/sec)}{1-4} = -0.274/\min \Rightarrow \tau = 3.6496$

Η τεταγμένη θα δίνεται από: $lny = lnA_0 - \frac{t}{\tau} \Rightarrow ln(\mathrm{A}_0) = \ln(y) + \frac{t}{\tau} \Rightarrow \mathrm{A}_0 = e^{[\ln(y)+t/\tau]}$

Επομένως για $y = 585$ και $t = 2min$ θα έχουμε $\mathrm{A}_0 = \mathrm{e}^{[\ln(585)+2(0.274)]}$

$$\Rightarrow \mathrm{A}_0 = \mathrm{e}^{6.9196} \Rightarrow \mathrm{A}_0 = 1011.92$$

Βρίσκουμε ότι η τεταγμένη είναι όντως η τιμή που έχουμε για *t*=0. και επομένως μπορούμε να την διαβάσουμε απευθείας από το γράφημα.
Για την κλίση ωστόσο θα πρέπει να υπολογίσουμε τους κατάλληλους λογαρίθμους

Log-Log χαρτί (λογαριθμικό – λογαριθμικό)

Πολλές φορές μπορούμε να βρούμε την συναρτησιακή εξάρτηση ενός φυσικού μεγέθους y από το ανεξάρτητο μέγεθος x, θεωρώντας το λογάριθμο των δεδομένων που μετράμε.

Αν το εξαρτώμενο μέγεθος y είναι ανάλογο κάποιας δύναμης του ανεξάρτητου μεγέθους x, τότε το γράφημα του y ως προς x, σχεδιαζόμενο σε ένα λογαριθμικό-λογαριθμικό χαρτί θα είναι ευθεία, η κλίση της οποίας θα είναι ίση με τον εκθέτη στον οποίο είναι υψωμένο το ανεξάρτητο μέγεθος.

Για παράδειγμα, έστω ότι μετρούμε κάποια δεδομένα τα οποία κατανέμονται σύμφωνα με την εξίσωση $y = x^n$

Θεωρώντας το λογάριθμο θα έχουμε: $\log(y) = \log(x^n) \Rightarrow \log(y) = n\log(x)$

Προφανώς από ένα γράφημα με λογαριθμικούς άξονες μπορούμε να βρούμε αμέσως την κλίση

Αν είχαμε περισσότερο πολύπλοκη μορφή:

$$y = Ax^n \Rightarrow \log(y) = \log(Ax^n) = \log(A) + n\log(x)$$

Έχει διαφορά στο γράφημα αν θεωρήσουμε Log_{10} ή ln ($\log_e$) σχέσεις?

Log-Log χαρτί (λογαριθμικό – λογαριθμικό)

X	Y
5	10
15	90
30	360
50	1000
95	3610

Για να βρούμε την κλίση της ευθείας:

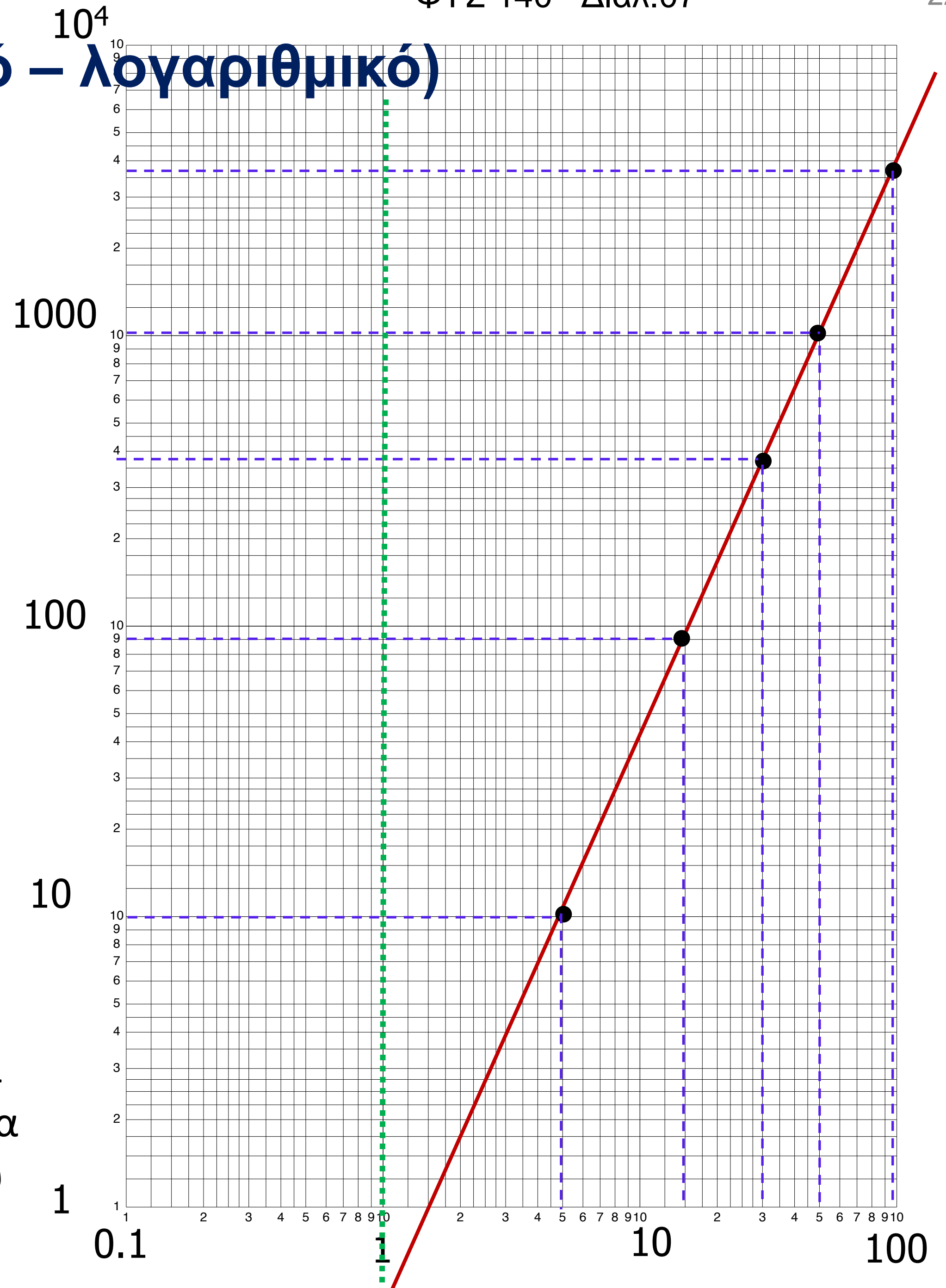
$\log(y) = \log(A) + n log(x)$

χρειαζόμαστε δύο σημεία (x_1,y_1) και (x_2,y_2):

$$n = \frac{\log(y_1) - \log(y_2)}{\log(x_1) - \log(x_2)}$$

$$\text{π.χ. } n = \frac{\log(10) - \log(360)}{\log(5) - \log(30)} = 2$$

Η τεταγμένη της παραπάνω ευθείας βρίσκεται θέτοντας log(x)=0, δηλαδή x=1 και μπορεί να διαβαστεί από το γράφημα αν φέρουμε την κάθετο στο x=1 εκεί που τέμνει την ευθεία



Log-Log χαρτί (λογαριθμικό – λογαριθμικό)

Θα μπορούσαμε να βρούμε την τεταγμένη με βάση την εξίσωση της ευθείας και τις συντεταγμένες ενός σημείου.

Για παράδειγμα έστω το σημείο (30,360):

$$\log(y) = \log(A) + n log(x) \Rightarrow \log(A) = \log(y) - n log(x) \Rightarrow \log(A) = \log(360) - 2 log(30)$$

$$\Rightarrow A = \mathrm{e}^{[\log(360) - 2\log(30)]} \Rightarrow A = \mathrm{e}^{-0.91629} \Rightarrow A = 0.4$$

Το αποτέλεσμα συνάδει με την τιμή που αντιστοιχεί στην τομή της ευθείας *x*=1 και της ευθείας που περνά από τα σημεία στο προηγούμενο γράφημα.

Αν ξέραμε την κλίση και την τεταγμένη της ευθείας $\log(y) = \log(A) + n log(x)$ για να την κατασκευάσουμε σε λογαριθμικό χαρτί, θα πρέπει για τυχαία *x* να εφαρμόσουμε την εξίσωση της ευθείας και να βρούμε τα αντίστοιχα *y.*

Θα είχαμε επομένως: $\log(y) = \log(A) + n log(x) = \log(Ax^n) \Rightarrow y = Ax^n$

Για το παράδειγμά μας, για *n*=2 και *A*=0.4, θεωρώντας ότι *x* = 20 θα πάρουμε:

$y = 0.4 \times 20^2 = 160$ τιμή που συνάδει με την προηγούμενη γραφική

Ιστογράμματα

Τα ιστογράμματα αποτελούν έναν εύχρηστο οπτικό τρόπο για να εξαγάγουμε την κατανομή που ακολουθούν μια σειρά από μεγάλο αριθμό μετρήσεων ενός μεγέθους αλλά παράλληλα δίνουν την δυνατότητα για εύκολη στατιστική ανάλυση μετρήσεων.

Αν έχουμε μια σειρά από μεγάλο αριθμό μετρήσεων ενός μεγέθους, τότε μπορούμε να ταξινομήσουμε τις μετρήσεις και να βρούμε τη συχνότητα εμφάνισης κάθε τιμής.

Δεν κρατάμε κάθε μέτρηση ξεχωριστά αλλά χωρίζουμε το εύρος των τιμών σε κατάλληλα υποδιαστήματα και αθροίζουμε τις τιμές που πέφτουν σε κάθε υποδιάστημα. Με τον τρόπο αυτό παίρνουμε την κατανομή της συχνότητας εμφάνισης κάθε τιμής.

Για παράδειγμα, έστω ότι μετρήσαμε κάποιο μέγεθος 30 φορές και βρήκαμε:

8.16	8.10	8.12	8.12	8.16	8.16
8.14	8.18	8.12	8.10	8.21	8.14
8.12	8.18	8.17	8.14	8.14	8.17
8.16	8.18	8.06	8.09	8.16	8.18
8.18	8.24	8.10	8.16	8.13	8.21

Οι τιμές των μετρήσεων έχουν συχνότητα:

Τιμή	#	Τιμή	#	Τιμή	#
8.06	1	8.13	1	8.20	0
8.07	0	8.14	4	8.21	2
8.08	0	8.15	0	8.22	0
8.09	1	8.16	6	8.23	0
8.10	3	8.17	2	8.24	1
8.11	0	8.18	5		
8.12	4	8.19	0		

Δηλαδή κάθε τιμή x_i εμφανίζεται n_i φορές

Η μέση τιμή μπορεί να γραφεί ως:

$$\boxed{\overline{x} = \frac{\sum x_i}{N} = \frac{\sum x_k n_k}{\sum n_k}}$$

Ιστογράμματα

Ομαδοποιώντας τις μετρήσεις, βρίσκουμε ευκολότερα πόσο συνεισφέρει κάθε τιμή στο άθροισμα της μέσης τιμής: $\sum x_k n_k$

Στο παράδειγμα, όλες οι τιμές βρίσκονται στο διάστημα 8.00 - 8.30 και άρα η πραγματική τιμή θα βρίσκεται στο διάστημα αυτό. Εφόσον οι τιμές 8.12 - 8.18 είναι οι πιο συχνές περιμένουμε ότι και η πραγματική τιμή θα είναι στο υποδιάστημα αυτό.

Όντως η μέση τιμή είναι 8.15.

Συνήθως χωρίζουμε το διάστημα σε ίσα υποδιαστήματα κατάλληλου εύρους ώστε στο καθένα να συμπεριλαμβάνεται μεγάλος αριθμός μετρήσεων και βρίσκουμε το ποσοστό των μετρήσεων σε κάθε υποδιάστημα,

$F_k = \frac{n_k}{N}$ **Συχνότητα εμφάνισης** μιας μέτρησης x_k. $\boxed{\bar{x} = \sum_k F_k x_k}$

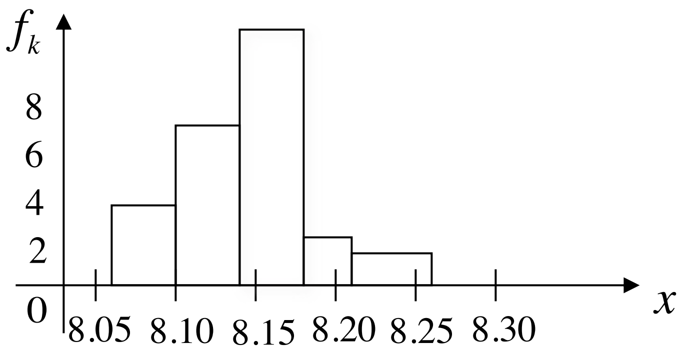
$f_k = \frac{F_k}{\Delta x_k}$ **Πυκνότητα Πιθανότητας** (probability density) μιας μεταβλητής x να βρεθεί στο k-διάστημα Δx

το ποσοστό δηλαδή των μετρήσεων στο διάστημα Δx_k

Ιστογράμματα

Το ιστόγραμμα αποτελεί τη γραφική παράσταση της f_k ως προς το x

Η ολική επιφάνεια των ορθογωνίων θα είναι ίση με την μονάδα: $\sum F_k = 1 = \sum f_k \Delta x_k$



Στην περίπτωση αυτή χρησιμοποιούμε Δx = 0.04 cm

Αν ο αριθμός των μετρήσεων γίνει αρκετά μεγάλος τότε το ιστόγραμμα παίρνει τη μορφή της κατανομής από την οποία προέρχονται οι μετρήσεις.

Η καμπύλη αποτελεί την οριακή κατανομή.

Κατασκευή ιστογράμματος με matplotlib

Για να κατασκευάσουμε ένα ιστόγραμμα στην python με την matplotlib, χρειάζεται να δώσουμε την εντολή:

content**, **binvalues**, **interm **=** ***plt.hist*(x, bins=τιμή, range=(τιμές))**

- **Τα ορίσματα της συνάρτησης *hist*:**

x: *array* [*data*] ή *list* από *arrays* [*data1, data2,...*] με τις τιμές της μεταβλητής που θέλουμε να κάνουμε το γράφημα (όπως και στην *plt.plot*)

bins: ακέραιος αριθμός που δηλώνει τον αριθμό των υποδιαστημάτων ίσου εύρους στα οποία είναι χωρισμένος ο *x*-άξονας

range *[optional]*: *tuple* που περιέχει την τιμή του *xmin* και *xmax.*

Αν δεν προσδιοριστεί, τότε αυτόματα ως *range*=(*x.min*(), *x.max*()) λαμβάνεται η ελάχιστη και μέγιστη τιμή των στοιχείων του array *x.*

Κατασκευή ιστογράμματος με matplotlib

***content*, *binvalues*, *interm* = *plt.hist*(x, bins=τιμή, range=(τιμές))**

➢ **Τι επιστρέφει η συνάρτηση *hist*:** ένα tuple με τα ακόλουθα στοιχεία

content: array μεγέθους N=bins με στοιχεία τη συχνότητα εμφάνισης των δεδομένων που δόθηκαν για κάθε bin. Αν δόθηκε *list* με *arrays* τότε το *content* περιέχει μια *list* με *arrays* με τη συχνότητα εμφάνισης για κάθε ξεχωριστό *array*.

binvalues: array μεγέθους N=bins+1 με τις τιμές του *x*-άξονα που αντιστοιχούν στα υποδιαστήματα που ορίστηκαν. Ο πίνακας περιέχει την αριστερή τιμή του κάθε υποδιαστήματος και την δεξιά τιμή του τελευταίου υποδιαστήματος.

***interm**:* Μια λίστα από ενδιάμεσους υπολογισμούς που χρησιμοποιηθήκαν στην κατασκευή του ιστογράμματος

Ιστογράμματα – Επιπλέον επιλογές για ορίσματα εισόδου

facecolor *[optional]* = *'b'* το χρώμα του ιστογράμματος (bleu)

density *[optional]* = *True/None[default]*

True: επιστρέφει την πυκνότητα πιθανότητας για κάθε bin: $\frac{N_i}{N_{tot}\Delta x}$.

Το εμβαδό της επιφάνειας κάτω από το ιστόγραμμα (το ολοκλήρωμα δηλαδή) θα ισούται με 1

Η επιλογή *Density* είναι παρόμοια με την επιλογή ***Normed*** η οποία επιστρέφει απλά ως αποτέλεσμα N_i/N_{tot}

histtype: {*'bar','barstacked','step','stepfilled'*} *[optional]*

bar: ραβδόμορφο ιστόγραμμα

barstacked: ραβδόμορφα ιστογράμματα το ένα πάνω στο άλλο αν δοθεί list από data arrays ως x

step: δημιουργεί γράφημα ως καμπύλη και η περιοχή στο εσωτερικό δεν χρωματίζεται

stepfilled: δημιουργεί γράφημα ως καμπύλη και η περιοχή στο εσωτερικό χρωματίζεται

Παράδειγμα

```
import numpy as np
import matplotlib.pyplot as plt
from random import random, seed, gauss
seed(12345678)
mu, sigma = 100, 15
x=[]
'''
```

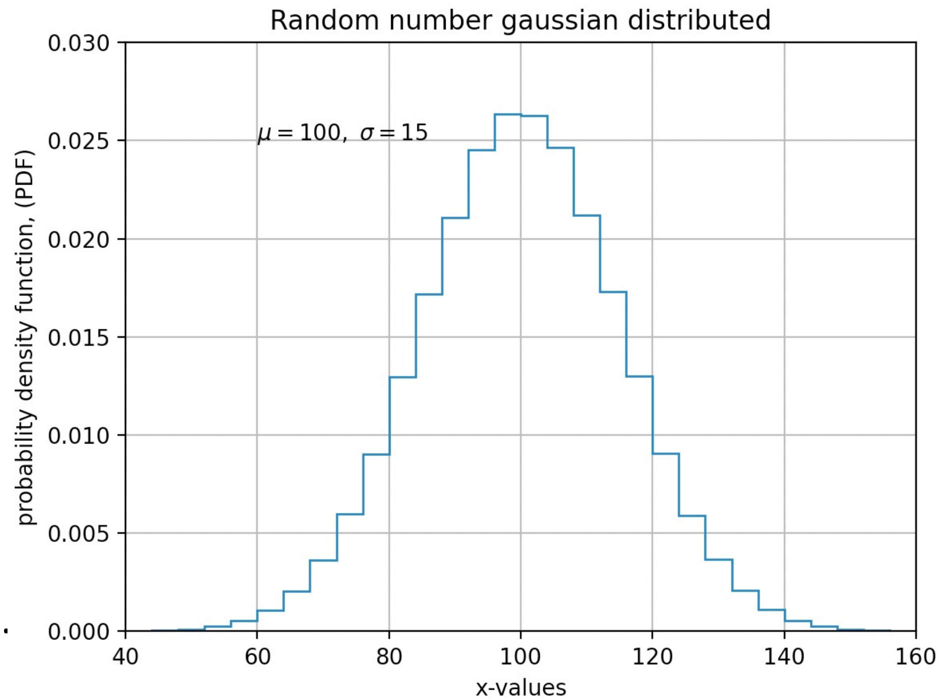
Δημιουργία 1M τυχαίων αριθμών κατανεμημένους σύμφωνα με την Gauss κατανομή με μέση τιμή μ = 100 και σ = 15

```
'''
for i in range(1000000): #
    x.append( gauss(mu,sigma ))
```

\# Δημιουργία ιστογράμματος με 50 bins στο *x* [0-200].

\# *Δx = 4* = (xmax-xmin)/bins

```
cont,binval,intr=plt.hist(x, bins=50, range=(0.,200.),
                          density=True, histtype='step')
plt.xlabel('x-values')
plt.ylabel('probability density function, (PDF)')
plt.title('Random number gaussian distributed")
plt.text(60, 0.025, r'$\mu=100, \ \sigma = 15$')
plt.xlim(40,160)
plt.ylim(0., 0.03)
plt.grid(True)
plt.show()
```



Ιστογράμματα - Οριακή κατανομή

Η οριακή κατανομή είναι μια θεωρητική καμπύλη που δεν μπορεί να προκύψει ποτέ από τις μετρήσεις.

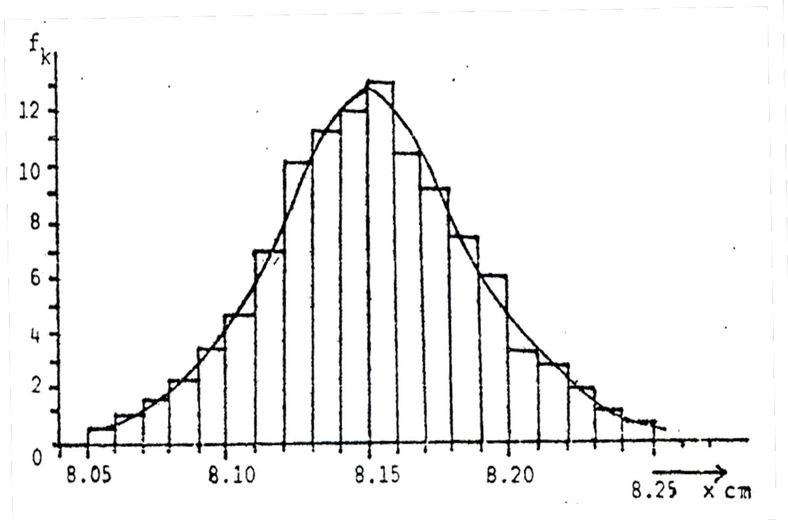
Το εμβαδό κάθε παρ/μου του ιστογράμματος είναι ισοδύναμο με τον αριθμό μετρήσεων που περιλαμβάνονται στο διάστημα αυτό.

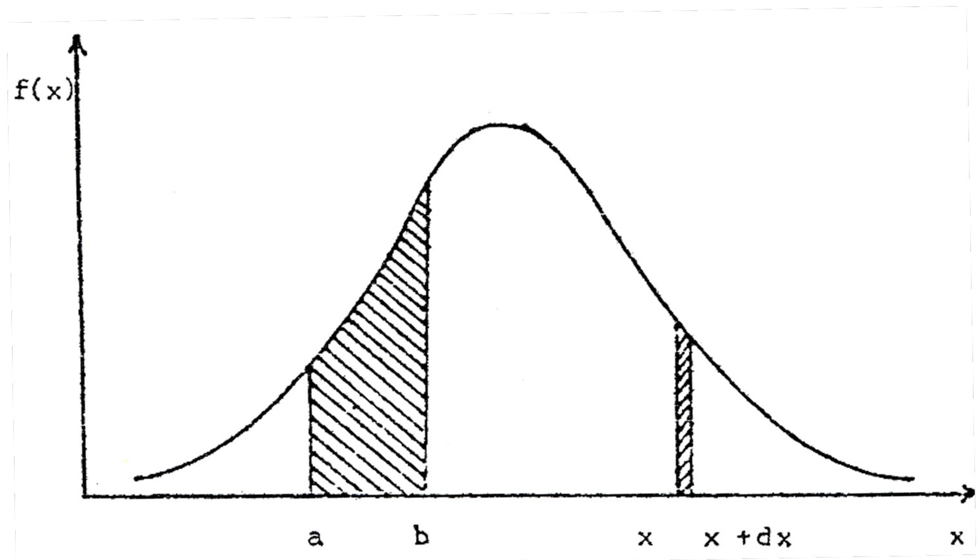
$f(x)dx$ πιθανότητα μια μέτρηση να περιλαμβάνεται μεταξύ x και x+Δx

$\int_a^b f(x)dx$ ποσοστό μετρήσεων μεταξύ x=α και x=b, η πιθανότητα δηλαδή μια μέτρηση να είναι στο διάστημα α,b

$\int_{-\infty}^{+\infty} f(x)dx = 1$ Η πιθανότητα μια μέτρηση να βρίσκεται μεταξύ -∞ και +∞

Λέμε τότε ότι η $f(x)$ είναι κανονικοποιημένη





Ιστογράμματα - Οριακή κατανομή πολλών μετρήσεων

Αν ξέρουμε την οριακή κατανομή μπορούμε να υπολογίσουμε τη μέση τιμή

$$\bar{x} = \sum x_k F_k \xrightarrow[\lim \Delta x_k \to 0]{} \bar{x} = \int_{-\infty}^{+\infty} x f(x)\, dx \quad \text{και ανάλογα} \quad \sigma_x^2 = \int_{-\infty}^{+\infty} \left(x - \bar{x}\right)^2 f(x)\, dx$$

Αν οι μέτρησεις επηρεάζονται από πολλές πηγές μικρών τυχαίων σφαλμάτων τότε η οριακή κατανομή προσεγγιζει αυτή της κατανομής Gauss

Για πολύ μεγάλο αριθμό μετρήσεων και για τυχαία σφάλματα οι τιμές που αποκλίνουν πολύ από τη μέση τιμή κατανέμονται συμμετρικά ως προς τη μέση τιμή, τότε στην άθροιση της εύρεσης της μέσης τιμής, αυτές οι τιμές αλληλοαναιρούνται και η τελική τιμή πλησιάζει την πιστή τιμή του μεγέθους

$$f(x) = G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \Rightarrow \int_{-\infty}^{+\infty} G(x)\, dx = 1$$

Η πιθανότητα να πάρουμε μια τιμή x στο διάστημα [x_1, x_1+dx) είναι: $P(x_1) \propto \sigma^{-1} e^{-(x_1-\mu)^2/2\sigma^2}$

Η πιθανότητα να πάρουμε μια τιμή x στο διάστημα [x_2,x_2+dx) είναι: $P(x_2) \propto \sigma^{-1} e^{-(x_2-\mu)^2/2\sigma^2}$

Αν συνεχίσουμε για όλες τις τιμές x_N τότε η πιθανότητα για την x_N: $P(x_N) \propto \sigma^{-1} e^{-(x_N-\mu)^2/2\sigma^2}$

Μέση τιμή ως η καλύτερη εκτίμηση της πραγματικής τιμής

Η πιθανότητα να παρατηρήσουμε το συγκεκριμένο δείγμα των N μετρήσεων είναι το γινόμενο των πιθανοτήτων:

$$P(x_1, x_2, \cdots, x_N) = P(x_1) \times P(x_1) \times \cdots \times P(x_N) \;\Rightarrow P(x_1, x_2, \cdots, x_N) \propto \frac{1}{\sigma^{\mathrm{N}}} e^{-\sum (x_i - \mu)^2 / 2\sigma^2}$$

Δεδομένων N παρατηρούμενων τιμών x_1, x_2, …, x_N η καλύτερη εκτίμηση της πραγματικής τιμής είναι αυτή που παίρνουμε αν μεγιστοποιήσουμε την πιθανότητα

Αυτό θα συμβεί όταν το άθροισμα στον εκθέτη είναι ελάχιστο $\sum (x_i - \mu)^2 / 2\sigma^2$

Επομένως θα πρέπει: $$\frac{d\left(\sum (x_i - \mu)^2 / 2\sigma^2\right)}{d\mu} = 0 \Rightarrow \sum (x_i - \mu) / \sigma^2 = 0 \Rightarrow \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

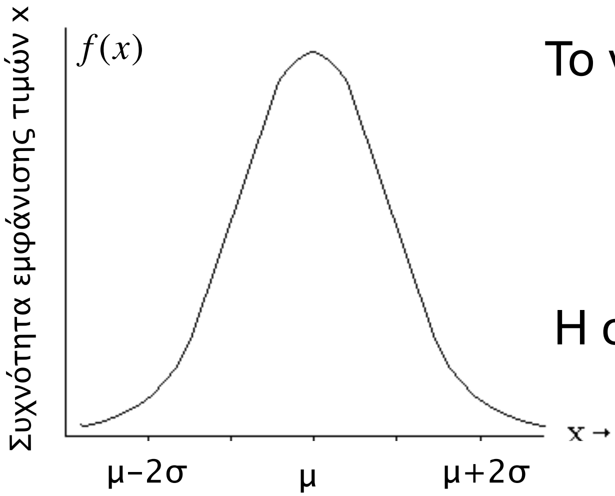
Ανάλογα ο καλύτερος υπολογισμός της τυπικής απόκλισης βρίσκεται ότι είναι:

$$\frac{dP(x_1, x_2, \cdots, x_N)}{d\sigma} = 0 \Rightarrow \sigma = \sqrt{\frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}$$

Gaussian κατανομή

Τυχαία σφάλματα υπάρχουν ακόμα και όταν όλες οι πηγές άλλων σφαλμάτων έχουν εξαλειφθεί. Συχνά, χρειάζεται να επαναλάβουμε μετρήσεις αρκετές φορές ώστε να εκτιμήσουμε την πιθανότητα κάποιος που επαναλαμβάνει μια πανομοιότυπη μέτρηση θα βρει διαφορετικό αποτέλεσμα.

Οι εκτιμήσεις αυτές στηρίζονται σε κάποιο μαθηματικό μοντέλο το οποίο και περιλαμβάνει τις περισσότερες περιπτώσεις. Το μοντέλο αυτό περιγράφεται από την Gaussian ή κανονική (normal) κατανομή.



Το γράφημα του σχήματος δίνεται από τη συνάρτηση:

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

Η συνάρτηση περιγράφεται από δύο παραμέτρους, μ και σ

Η μεταβλητή μ περιγράφει που βρίσκεται η κορυφή της καμπύλης στον $x-$ άξονα.

Η μεταβλητή σ περιγράφει τη διασπορά των τιμών του x, δηλαδή το εύρος της καμπύλης.

Ουσιαστικά η καμπύλη αυτή δείχνει τη συχνότητα με την οποία θα εμφανίζονταν διάφορες τιμές του x αν επαναλαμβάναμε μια μέτρηση πολλές φορές.

Κανονική κατανομή

Η μέση τιμή, μ, δίνει ουσιαστικά την περισσότερη πιθανή τιμή των x και αυτή είναι η μέση τιμή. Παραπλήσιες τιμές έχουν μια μικρότερη αλλά αρκετά μεγάλη πιθανότητα.

Η κατανομή εκτείνεται σε άπειρες τιμές του x (με μικρή ωστόσο πιθανότητα). Αποδεικνύεται ότι το εμβαδό της καμπύλης μεταξύ μ-σ και μ+σ ισούται με το 68.3% του συνολικού εμβαδού της καμπύλης [-∞,+∞].

Το εμβαδό της καμπύλης που περικλύεται μεταξύ μ-2σ και μ+2σ ισούται με το 95.4% του εμβαδού της καμπύλης. Ενώ για ±3σ το ποσοστό είναι 99.7%

Όπως γράψαμε την παραπάνω εξίσωση το ολοκλήρωμά της από -∞,+∞

$$\int_{-\infty}^{+\infty} G(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} dx = 1$$

$$\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

$$\hat{\sigma}^2 = \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left(\hat{\mu} - x_i\right)^2$$

Για N μεγάλο $\hat{\mu} = \mu$, $\hat{\sigma} = \sigma$

Πολλές φορές χρησιμοποιούμε αντί για την ποσότητα σ, το πλήρες εύρος της κατανομής στο μέσο του μέγιστου ύψους της (FWHM) που συμβολίζεται με Γ.

Από τον ορισμό: $G\left(\mu \pm \frac{\Gamma}{2}\right) = \frac{1}{2}G(\mu)$ έχουμε ότι $\Gamma = 2\sigma\sqrt{2\ln 2} \approx 2.35\sigma$