

Corso integrato di informatica, statistica e analisi dei dati sperimentali Esercitazione 5

Esercizio 1) I files di dati si trovano nella directory **Moto**.

```

accelerato.dat
#t [s]      y [cm]
#err.
# 0.3  0.5
# -----
0.9      0.4
1.7      1.2
2.7      1.8
3.7      2.0
4.5      3.0
    
```

Sul file **accelerato.dat** sono registrate le posizioni di un oggetto che si muove di moto rettilineo lungo l'asse y in funzione del tempo. Al tempo t=0 il corpo passa per y=0. Effettuare un fit dei dati assumendo un moto rettilineo uniformemente accelerato per calcolare l'accelerazione del corpo. Sapendo che il corpo ha massa M = 50 g, calcolare la forza agente.

L'equazione generale del moto rettilineo uniformemente accelerato è:

$$Y^{th}(t) = y_0 + v_0 (t-t_0) + 0.5 a (t-t_0)^2$$

avendo posto t₀=0 si ha:

$$Y^{th}(t) = y_0 + v_0 t + 0.5 a t^2$$

Utilizziamo Gnuplot per la grafica dei dati e per i raffinamento dei parametri. Il file **accel.plt** contiene una serie di script per i grafici e i fit che possono essere usati come esempio.

1) Si definiscono i titoli degli assi e si graficano i dati riportando l'errore sulle posizioni (y) misurate

```

set title "moto accelerato"
set xlabel "t [s]"
set ylabel "y [cm]"

pl "accelerato.dat" u 1:2:(0.5) w yerr
    
```

2) si definisce la funzione da usare per il fit, si definiscono valori ragionevoli per i parametri v₀ e a₀ e assumiamo, per ora, e y₀ = 0

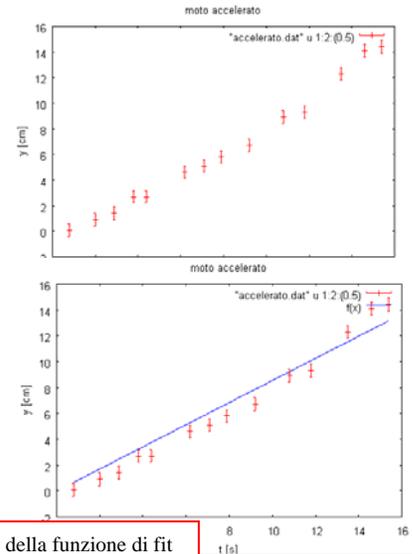
```

f(x) = vo * x + 0.5 * ao * x**2 + yo
vo = 0.1; ao = 0.0; yo=0.0
    
```

Per gli errori assumiamo trascurabile l'errore sulle misure di t. L'errore sulla posizione è costante per tutti i punti: σ_y = 0.5. Facciamo prima un fit assumendo nulla l'accelerazione:

```

fit f(x) "accelerato.dat" u 1:2:(0.5) via vo
pl "accelerato.dat" u 1:2:(0.5) w yerr,f(x) w l 3
    
```



```

After 3 iterations the fit converged.
final sum of squares of residuals : 50.9454
rel. change during last iteration : -1.18269e-007

degrees of freedom (FIT_NDF) : 13
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 1.97962
variance of residuals (reduced chisquare) = WSSR/ndf : 3.91888

Final set of parameters      Asymptotic Standard Error
=====
vo = 0.853654                +/- 0.02892 (3.388%)
    
```

Se l'errore è definito in modo corretto questo è il χ²

Variazione della funzione di fit all'ultima iterazione

Gradi di libertà

Se l'errore è definito in modo corretto questo è il χ² ridotto

Parametro v₀

Errore su v₀

Salviamo i risultati del fit. Gnuplot salva alcuni parametri utilizzati per il fit in opportune variabili:

FIT_NDF = gradi di libertà (n_lib)

FIT_STDFIT = (WSSR/n_lib)^{1/2}

Dove WSSR è la somma dei quadrati degli scarti

Posso usare queste variabili per il calcolo del χ^2 associato al fit:

chi_1 = FIT_STDFIT2 * FIT_NDF**

nl_1 = FIT_STDFIT2 * FIT_NDF**

chi_r1 = FIT_STDFIT2**

queste variabili mi possono essere utili per valutare il fit.

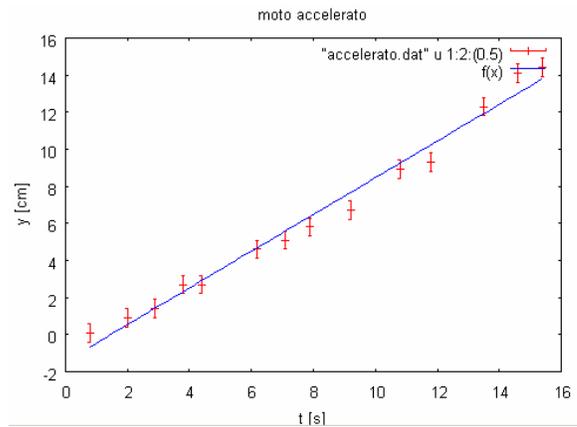
Ricalcoliamo il fit utilizzando due variabili indipendenti: y_0 e v_0 e salviamo i risultati della statistica. Attenzione: nello scegliere un valore diverso da zero per inizializzare un parametro da raffinare, altrimenti le routines di fit possono andare in errore.

```
After 4 iterations the fit converged.
final sum of squares of residuals : 20.3411
rel. change during last iteration : -2.32081e-008

degrees of freedom (FIT_NDF) : 12
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 1.30196
variance of residuals (reduced chisquare) = WSSR/ndf : 1.6951

Final set of parameters Asymptotic Standard Error
-----
vo = 0.991199 +/- 0.03755 (3.788%)
yo = -1.45917 +/- 0.3434 (23.53%)

correlation matrix of the fit parameters:
vo yo
yo 1.000
-0.862 1.000
```



```
yo = 0.1
fit f(x) "accelerato.dat" u 1:2:(0.5) via vo,yo
pl "accelerato.dat" u 1:2:(0.5) w yerr,f(x) w l 3

chi_2 = FIT_STDFIT**2 * FIT_NDF
nl_2 = FIT_STDFIT**2 * FIT_NDF
chi_r2 = FIT_STDFIT**2
```

Il χ^2 diminuisce e il fit sembra migliorare, possiamo controllare se questo miglioramento sia significativo utilizzando la statistica della funzione

$$F(\nu_1, \nu_2) = \frac{\tilde{\chi}_1^2 - \tilde{\chi}_2^2}{\tilde{\chi}_2^2} \quad F:$$

con $\nu_1 = 1$ e $\nu_2 = 12$. Otteniamo $F(1,12) = 1.31$ che confrontata con la statistica di F (uso Excel):

$$\text{DISTRIB.F}(1.31,1,12) = 0.27$$

Se si pone un limite di confidenza al 5%, risulta che il miglioramento ottenuto aggiungendo y_0 come parametro libero non migliora il fit in modo significativo, cioè la probabilità di sbagliare affermando che il fit migliora è ~27%.

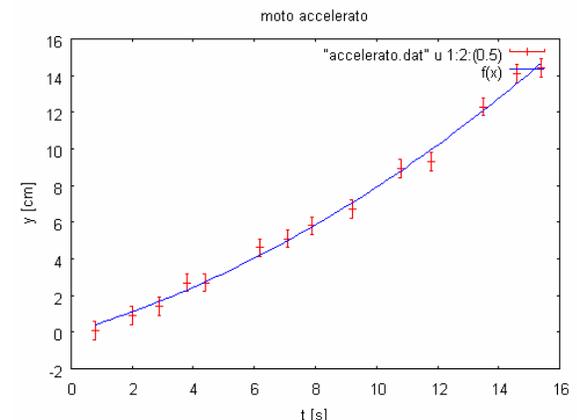
Proviamo a vedere se utilizzare v_0 e a_0 migliora il fit.

```
yo = 0.0
ao=0.1

fit f(x) "accelerato.dat" u 1:2:(0.5) via vo,ao
pl "accelerato.dat" u 1:2:(0.5) w yerr,f(x) w l 3

chi_3 = FIT_STDFIT**2 * FIT_NDF
nl_3 = FIT_STDFIT**2 * FIT_NDF
chi_r3 = FIT_STDFIT**2

# definizione della funzione F:
F_1_12 = (chi_r1 - chi_r3)/chi_r3
pr F_1_12
```



```

final sum of squares of residuals : 6.1745
rel. change during last iteration : -1.72702e-006

degrees of freedom (FIT_NDF) : 12
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.717315
variance of residuals (reduced chisquare) = WSSR/ndf : 0.514541

Final set of parameters      Asymptotic Standard Error
=====
vo = 0.48779                +/- 0.0406      (8.323%)
ao = 0.0604866             +/- 0.006484   (10.72%)

correlation matrix of the fit parameters:

      vo      ao
vo    1.000
ao   -0.966  1.000

```

calcolando F come sopra si ottiene:

$$F = 6.62$$

Che confrontato con la statistica per una variabile F(1,12) da:

$$\text{DISTRIB.F}(6.62,1,12) = 0.024$$

Che mi dice come la probabilità di osservare per caso un tale miglioramento è ~2.4 %,

quindi fissando un limite di confidenza al 95% (rischio di errore 5%) utilizzando a_0 e v_0 migliora il fit.

I risultati del fit sono:

$$v_0 = 0.490 \pm 0.04 \text{ [cm s}^{-1}\text{]} \quad a_0 = 0.060 \pm 0.006 \text{ [cm s}^{-2}\text{]}$$

che sono in buon accordo con i parametri che avevo usato per generare la funzione: $v_0^{\text{th}} = 0.5 \text{ [cm s}^{-1}\text{]}$ e $a_0^{\text{th}} = 0.06 \text{ [cm s}^{-2}\text{]}$.

Il fit mi dice, comunque, che i parametri a_0 e v_0 sono fortemente correlati: correlazione ~ 0.966. Quindi gli errori standard ottenuti dal fit possono essere sottostimati. Un modo corretto per stimare l'intervallo di confidenza sui parametri quando questi risultano fortemente correlati consiste nel definire il valore limite del χ^2 (LIM) che definisce un limite di confidenza $1-\alpha$ sul fit, quindi si cercano i valori limite di ogni parametro per i quali il risulta $\chi^2 < \chi^2$ (LIM). Alcuni programmi lo fanno in modo automatico, qui possiamo cercare di trovare gli errori per tentativi.

Nel nostro caso il valore limite di χ^2 per una confidenza del 67% (errore standard, $\alpha = 0.33$) è: ~13.6

scegliamo di fissare a_0 a diversi valori e raffinare v_0 , es:

ao=0.036; fit f(x) "accelerato.dat" u 1:2:(0.5) via vo ottengo $\chi^2 = 13.51$
 ao=0.085; fit f(x) "accelerato.dat" u 1:2:(0.5) via vo ottengo $\chi^2 = 13.53$

quindi l' intervallo di confidenza per il parametro a_0 al 67% per l'accelerazione è:

$$a_0 = [0.036 ; 0.085] \text{ [cm s}^{-2}\text{]}$$

come si vede l'errore vero è molto più grande dell'errore stimato dal fit (circa di un fattore 10)

Per v_0 :

vo=0.34 ; fit f(x) "accelerato.dat" u 1:2:(0.5) via ao ottengo $\chi^2 = 13$
 vo=0.64; fit f(x) "accelerato.dat" u 1:2:(0.5) via ao ottengo $\chi^2 = 13.4$

quindi l' intervallo di confidenza per il parametro a_0 al 67% per l'accelerazione è:

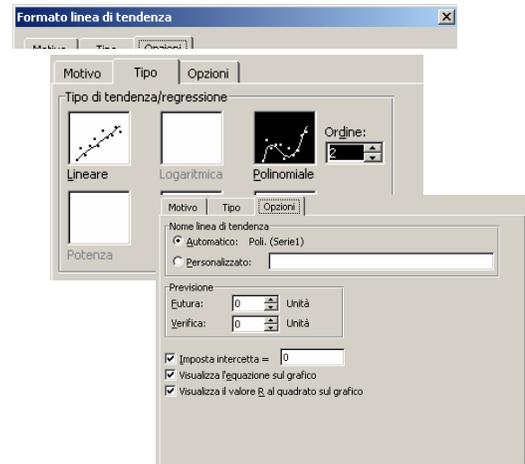
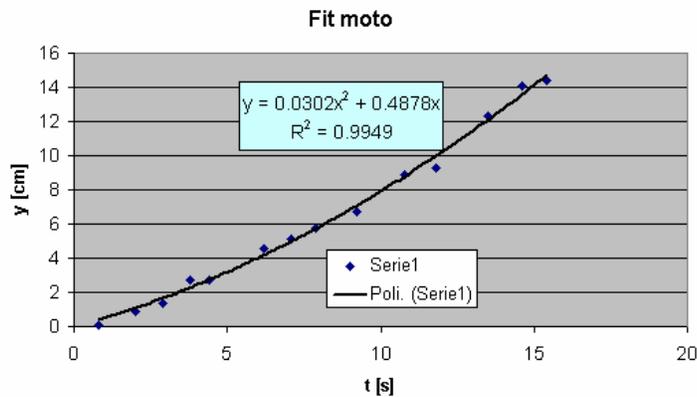
$$v_0 = [0.34 ; 0.64] \text{ [cm s}^{-1}\text{]}$$

anche in questo caso l'errore vero è molto più grande dell'errore stimato dal fi (di circa un fattore 10).

Provate a calcolare gli intervalli di confidenza al 95%.

Quando i parametri sono molti il calcolo preciso degli errori può risultare molto complesso. Di solito si preferisce riportare come risultato l'errore standard determinato dal fit e la matrice di correlazione.

Si può ottenere un risultato simile utilizzando le opzioni di Excel:



Ottenuta utilizzando un polinomio di II grado: e imponendo il passaggio per l'origine. In questo caso non abbiamo però alcuna informazione sugli errori associati ai parametri o sulla correlazione tra essi.

Esercizio2) Sul file **pendolo.dat** sono registrate in funzione del tempo le posizioni di un oscillatore armonico. Determinare ampiezza (A), fase iniziale (f_0), pulsazione (ω) e, eventualmente, smorzamento (τ). L'equazione dell'oscillatore armonico con effetto dell'attrito è:

$$y(t) = A \sin(\omega t + f_0) e^{-\frac{t}{\tau}}$$

Il file **pend.plt** contiene alcuni script per l'analisi utilizzando Gnuplot.

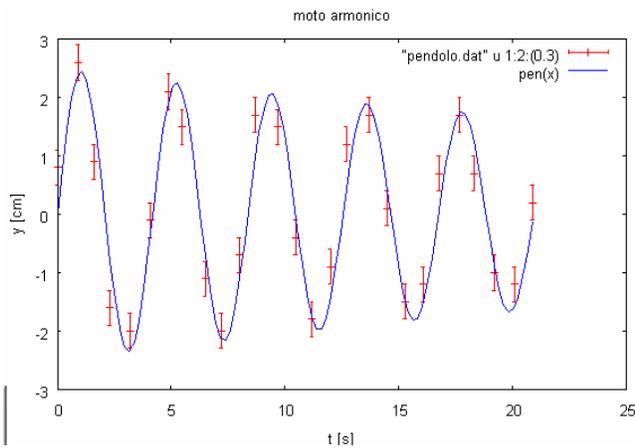
```
#accel.plt
set title "moto armonico"
set xlabel "t [s]"
set ylabel "y [cm]"

pen(x) = A0 * sin(om * x + fo) * exp(-x/tau)
A0 = 5.
om = 1.
fo = 0.
tau = 100.

pl "pendolo.dat" u 1:2:(0.3) w yerr,pen(x) w l 3
```

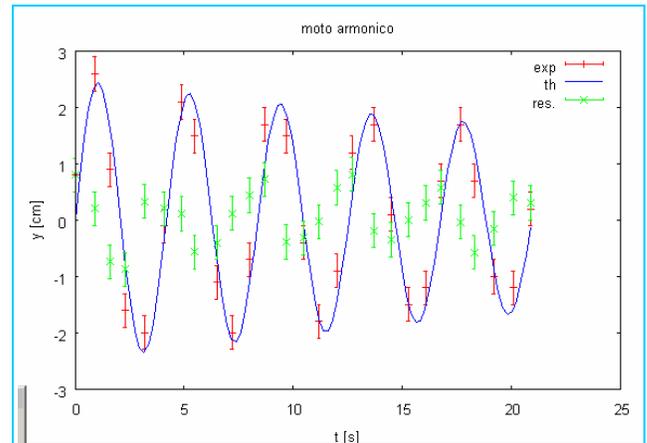
Come prima cosa definiamo i titoli per il grafico e per gli assi.

Quindi definiamo la funzione da utilizzare per il fit: pen(x) dando un seti di valori iniziali ai parametri. Per scegliere i valori iniziali dei parametri Si graficano insieme dati sperimentali e funzione teorica in modo da scegliere un seti di valori iniziali dei parametri che dia un certo accordo tra dato sperimentale e teoria.



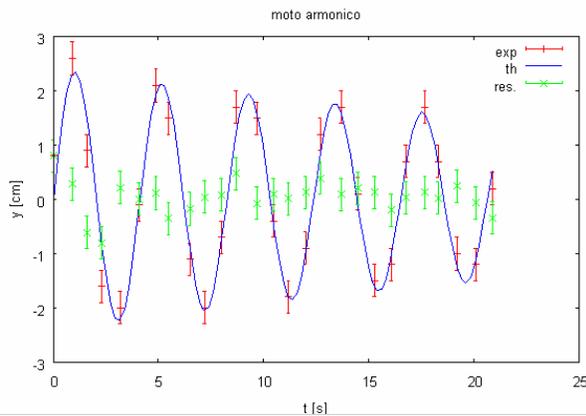
Passiamo quindi al fit trascurando, per ora, f_0

```
fit pen(x) "pendolo.dat" u 1:2:(0.3) via om,Ao, tau
replot
```



```
pl "pendolo.dat" u 1:2:(0.3) t'exp' w yerr,\
pen(x) t'th' w l 3,\
"pendolo.dat" u 1:($2-pen($1)):(0.3) t'res.' w yerr 2
```

Grafica insieme dati sperimentali (rosso), curva teorica (bleu) e residui (verde)



```
After 6 iterations the fit converged.
final sum of squares of residuals : 29.7765
rel. change during last iteration : -6.28787e-007

degrees of freedom (FIT_NDF) : 24
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 1.11386
variance of residuals (reduced chisquare) = WSSR/ndf : 1.24069
```

```
Final set of parameters      Asymptotic Standard Error
=====
om          = 1.521          +/- 0.004267    (0.2805%)
Ao          = 2.41491        +/- 0.2066     (8.557%)
tau         = 43.3583        +/- 15.46      (35.65%)
```

correlation matrix of the fit parameters:

```
om      om      Ao      tau
om      1.000
Ao      -0.015  1.000
tau     0.031  -0.827  1.000
```

Il χ^2 ridotto dell'ordine dell'unità mi dice che il fit è buono. Voglio vedere se è possibile, dai dati, ottenere il valore della fase iniziale f_0 . Per questo rifaccio il fit utilizzando anche il parametro f_0 :

```
fo = 1.
fit pen(x) "pendolo.dat" u 1:2:(0.3) via om,Ao, tau,fo
replot
```

attenzione: per raffinare f_0 iniziarlo ad un valore diverso da 0 (es. $f_0 = 1.$) altrimenti il programma va in errore)

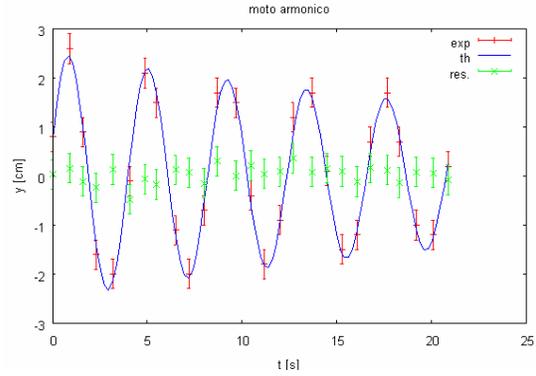
```

degrees of freedom (FIT_NDF) : 23
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.633201
variance of residuals (reduced chisquare) = WSSR/ndf : 0.400944

Final set of parameters
-----
Asymptotic Standard Error
-----
om = 1.49729 +/- 0.004125 (0.2755%)
Ao = 2.50527 +/- 0.1182 (4.717%)
tau = 38.9889 +/- 7.005 (17.97%)
fo = 0.312606 +/- 0.04357 (13.94%)

correlation matrix of the fit parameters:
om Ao tau fo
om 1.000
Ao 0.018 1.000
tau 0.016 -0.822 1.000
fo -0.806 -0.038 0.014 1.000

```



Il χ^2 migliora, ma questo miglioramento è significativo?

La F calcolata è $F(1,23) = 48.3$, a questa è associata una probabilità molto piccola ($4.E-7$) e quindi posso affermare che il miglioramento del fit non è significativo.

```

gnuplot> F = (1.24-0.4)/(0.4)*23.
gnuplot> pr F
48.3

```

Utilizzo i risultati del primo fit per preparare la tabella riassuntiva:

ω [rad / s]	A [cm]	τ [s]
1.521 ± 0.004	2.41 ± 0.21	43 ± 15

E la tabella di correlazione tra i parametri

	om	Ao	tau
om	1		
Ao	-0.015	1	
tau	0.031	-0.83	1

Esercizio 3) Fit di distribuzioni discrete. I files con dati ed esempi sono nella directory **Discrete**

Per valutare la sensibilità di una specie di piante ad un virus si espongono al contagio M gruppi ($M=90$) di $n = 20$ individui ciascuno. Dopo un tempo certo tempo si contano il numero di contagi k per ogni gruppo e si riportano (file: **Binomiale.dat**) le frequenze N_k che rappresentano il numero di volte in cui sono osservati k contagi per gruppo. Se la probabilità di contagio è p posso supporre (modellino semplice) che il numero di contagi per gruppo segua una distribuzione binomiale $f(n; p)$:

$$f(k, n) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Vogliamo effettuare un fit dei dati per ottenere p . Preparo uno script per effettuare il fit usando Gnuplot (in questo caso non sono in grado di effettuare il fit usando Excel:

```
# Bino.plt
set title 'Binomial plot'
set xlabel 'k'
set ylabel 'N_k'
pl 'binomiale.dat' u 1:2
```

Definisce i titoli e mostra il plot dei dati

Mostra il plot dei dati come istogramma con errori

```
Ntot = 90.
pl 'binomiale.dat' u 1:2 with boxes, '' u 1:2:(sqrt($2*(1-$2/Ntot))) t'' w yerr
```

```
n = 20 # numero di elementi per campione
Ntot = 90 # numero totale di campioni
po = 0.2 # valore iniziale per la po

f(x) = Ntot * n! / (int(x)! * (20 - int(x))!) * po**int(x) * (1-po)**(20-int(x))

pl 'binomiale.dat' u 1:2 t'' w boxes, '' u 1:(f($1)) w boxes 3
```

Calcola la distribuzione teorica: Le frequenze attese sono $N_k^{th} = M f(k,n)$

Calcolo del fit

```
# ===== raffinamento senza errori =====
fit f(x) 'binomiale.dat' u 1:2 via po

pl 'binomiale.dat' u 1:2:(sqrt($2*(1-$2/M))) t'exp' w yerr, \
'' u 1:(f($1)) t'th' w l 3, \
'' u 1:($2-f($1)):(sqrt($2*(1-$2/M))) t'res' w yerr 2

pause -1 "fit"
# ===== raffinamento con errori =====

fit [:][0.1:]f(x) 'binomiale.dat' u 1:2:(sqrt($2*(1-$2/M))) via po
replot
```

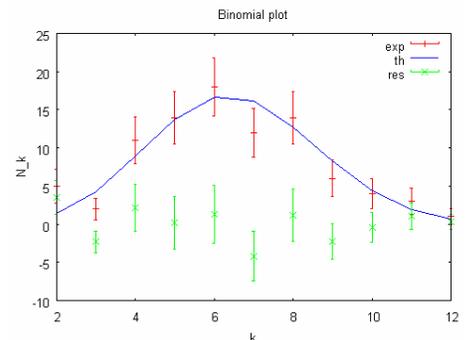
```
final sum of squares of residuals ~: 9.17418
rel. change during last iteration : -1.27931e-007

degrees of freedom (FIT_NDF) : 10
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.957819
variance of residuals (reduced chisquare) = WSSR/ndf : 0.917418

Final set of parameters Asymptotic Standard Error
=====
po = 0.326681 +/- 0.01011 (3.095%)
```

Risultati del fit

Quindi: $p_0 = 0.33 \pm 0.01$

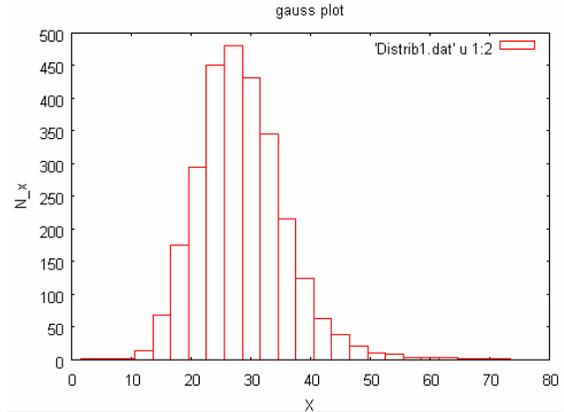


Esercizio 4) Distribuzioni continue. I dati si trovano nella cartella **gauss**.

I dati riportati nel file **Distrib1.dat** mostrano la distribuzione di un insieme di risultati X in forma di istogramma di frequenze assolute: la colonna 1 mostra le classi di valori, in colonna 2 sono riportate le frequenze assolute (n_x) delle osservazioni per ogni classe: 1 osservazione tra 3 e 6, 14 osservazioni tra 9 e 12, etc...

Grafichiamo i dati utilizzando Gnuplot:

```
set title 'gauss plot'
set xlabel 'X'
set ylabel 'N_x'
pl 'Distrib1.dat' u 1:2 with boxes
```



Per calcolare l'errore associato ad ogni osservazione osserviamo che, in mancanza di altre informazioni, possiamo assumere che il numero di osservazioni in nella i-esima classe dell'istogramma segua una distribuzione binomiale con probabilità p_i , valore atteso Np_i e varianza $\sigma_i^2 = Np_i(1-p_i)$. La migliore stima di cui dispongo per il valore delle probabilità p_i è $p_i \sim f_i$, quindi l'errore associato ad ogni frequenza assoluta è:

$$\sigma_{n_i} = \sqrt{Np_i(1-p_i)} \sim \sqrt{N \frac{n_i}{N} \left(1 - \frac{n_i}{N}\right)} = \sqrt{n_i \left(1 - \frac{n_i}{N}\right)}$$

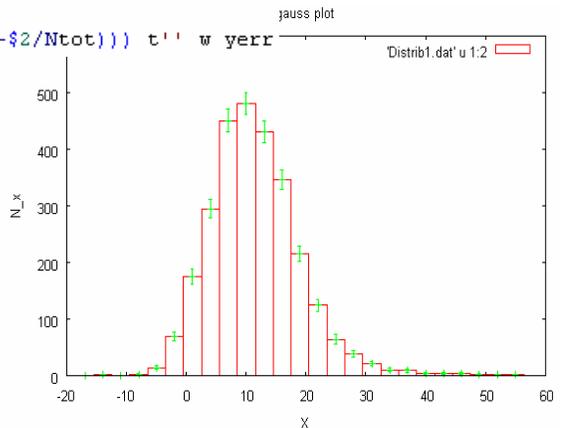
```
# ===== grafico on errori =====
Ntot = 2759.
```

pl 'Distrib1.dat' u 1:2 with boxes, '' u 1:2:(sqrt(\$2*(1-\$2/Ntot))) t '' w yerr

Il comando

u 1 : 2 : (sqrt(\$2 * (1-\$2/Ntot))) w yerr

Permette di calcolare l'errore come radice quadrata dei dati in colonna 2 (\$2) per 1 - i dati in colonna 2 divisi per Ntot.



Calcoliamo il valore medio, varianza e dev. st della distribuzione utilizzando un foglio Excel: ricordiamo

che il valor medio di x è la somma pesata per il numero di osservazioni e quindi può essere calcolata usando le frequenze assolute n_i o le frequenze relative f_i ($f_i = n_i/N$)

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} = \frac{\sum_{i=1}^k x_i n_i}{N} = \sum_{i=1}^k x_i f_i$$

Dove le sommatorie corrono su tutti i valori x delle diverse classi. La varianza sulla popolazione osservata è:

$$\sigma^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{\sum_{i=1}^k n_i} = \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

Vogliamo verificare se la distribuzione è compatibile con una distribuzione Gaussiana.
 In Gnuplot definiamo la distribuzione Gaussiana:

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

la f(x) rappresenta la probabilità di osservare un valore di x compreso tra x e x+dx (densità di probabilità): $f(x) = P(x, x+dx)$.

La probabilità di osservare un valore compreso tra x_{i-1} e x_i è:

$$P(x_i < x \leq x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x) dx$$

In modo rozzo ma efficace posso calcolare l'integrale:

$$P(x_i < x \leq x_{i+1}) \sim \Delta x_i \frac{(f(x_{i+1}) + f(x_i))}{2}$$

con $\Delta x = x_{i+1} - x_i$. Data una popolazione di N individui, il numero di osservazioni (frequenze assolute teoriche) che possiamo attenderci all'interno dell'intervallo $[x_{i-1}, x_i]$ è:

$$n_i^{th} = N \Delta x_i \frac{(f(x_{i+1}) + f(x_i))}{2}$$

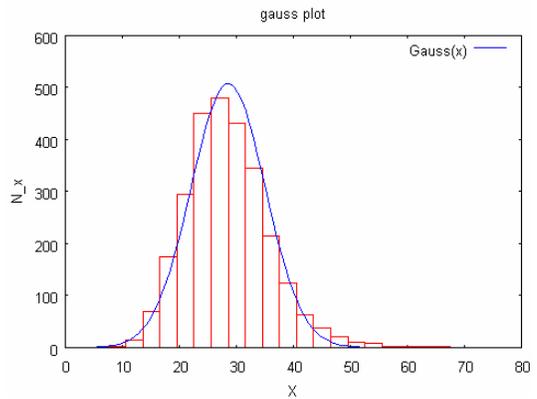
Questi sono i valori teorici che devo confrontare con i miei dati, nel nostro caso $N = N_{tot} = 2759$ e $\Delta x = 3$ per tutti gli intervalli, quindi definisco la funzione Gauss(x):

```
dx = 3.
f(x) = 1. /sqrt(2 * pi * s2) * exp( -((x-xo)**2/s2) )
Gauss(x) = dx * A * (f(x+dx) + f(x))/2.

A = 2760.
s2 = 40.
xo = 10.
```

E posso graficare dato sperimentale e risultati teorici:

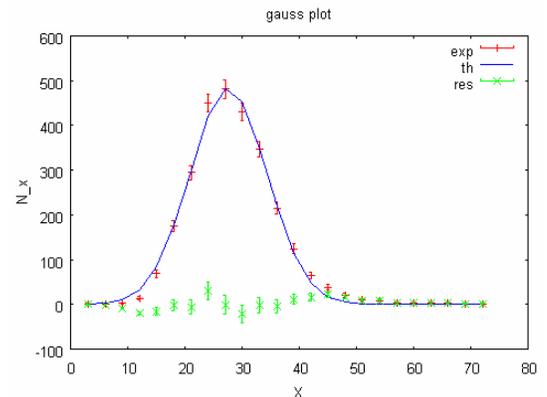
```
pl 'Distrib1.dat' u 1:2 t'' w boxes, \
    Gauss(x) w l 3
```



A questo punto sono pronto per fare il fit.

```
fit Gauss(x) 'Distrib1.dat' u 1:2 via A,xo,s2
pl 'Distrib1.dat' u 1:2:(sqrt($2*(1-$2/Ntot))) t'' w yerr, \
    '' u 1:(Gauss($1)) w l 3, \
    '' u 1:($2-Gauss($1)):(sqrt($2*(1-$2/Ntot))) t'' w yerr 2
```

Il comando permette di graficare dati sperimentali, curva teorica e residui.



After 5 iterations the fit converged.

final sum of squares of residuals : 3422.91
rel. change during last iteration : -5.0975e-006

Somma dei quadrati (se l'errore è ben definito è il χ^2)
Variazione relativa ($d\chi/\chi$) all'ultima iterazione, è una stima della
distanza dal minimo

degrees of freedom (FIT_NDF) : 21
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 12.767
variance of residuals (reduced chisquare) = WSSR/ndf : 162.996

N. gradi di libertà $N - 1 - k$

χ^2 ridotto se l'errore è ben definito

Final set of parameters Asymptotic Standard Error
=====

A	= 2713.17	+/- 44.07	(1.624%)
xo	= 29.0369	+/- 0.126	(0.434%)
s2	= 42.8618	+/- 1.696	(3.958%)

parametri ed errori

correlation matrix of the fit parameters:

matrice di correlazione

	A	xo	s2
A	1.000		
xo	-0.000	1.000	
s2	0.578	-0.002	1.000

Dobbiamo ancora tener conto dell'errore se vogliamo avere un'informazione statistica, per questo utilizziamo:

```
fit Gauss(x) 'Distrib1.dat' u 1:2: (sqrt($2*(1-$2/Ntot))) via A,xo,s2  
replot
```

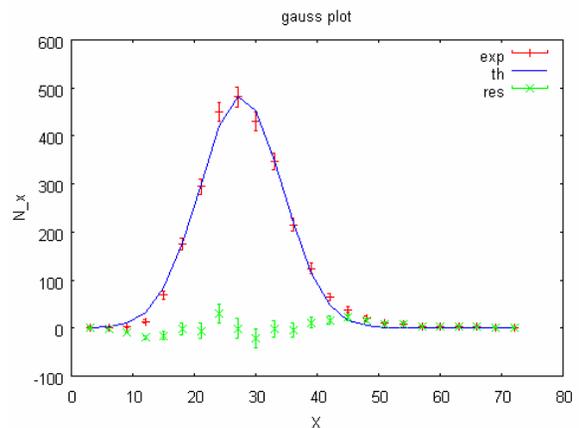
Otteniamo:

final sum of squares of residuals : 96.3016
rel. change during last iteration : -3.05768e-006
degrees of freedom (FIT_NDF) : 21
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 2.14145
variance of residuals (reduced chisquare) = WSSR/ndf : 4.58579

Final set of parameters	Asymptotic Standard Error
=====	=====
A	= 2666.59 +/- 103.4 (3.877%)
xo	= 29.5601 +/- 0.2673 (0.9043%)
s2	= 40.1861 +/- 2.388 (5.943%)

correlation matrix of the fit parameters:

	A	xo	s2
A	1.000		
xo	0.016	1.000	
s2	0.048	0.215	1.000



da notare che, tenendo conto degli errori cambiano

i valori dei parametri, gli errori associati e la matrice di correlazione. Soprattutto si ottiene un χ^2 più ragionevole.

Torniamo sul foglio Excel e valutiamo quale è la probabilità associata al valore del χ^2 osservato:

DISRIB.CHI(96.3, 21) = 1.e-11, molto piccolo, quindi i dati non sono compatibili con una funzione Gaussiana, anche con la migliore scelta dei parametri.

La funzione logNormale rappresenta la funzione di distribuzione di probabilità per una variabile aleatoria il cui logaritmo segue una distribuzione normale:

$$f_{LN}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\log_e x - \bar{x})^2}{2\sigma^2}}$$

Ora posso definire come prima la f_{LN} in Gnuplot e riprovare a fare il fit (il file **LogNorm.plt** contiene uno script Gnuplot per il fit)

```

dx = 3.
L(x) = 1. / (x * sqrt(2 * pi * s2)) / exp(((log(x) - xo)**2 / s2))
LogN(x) = dx * A * (L(x+dx) + L(x)) / 2.

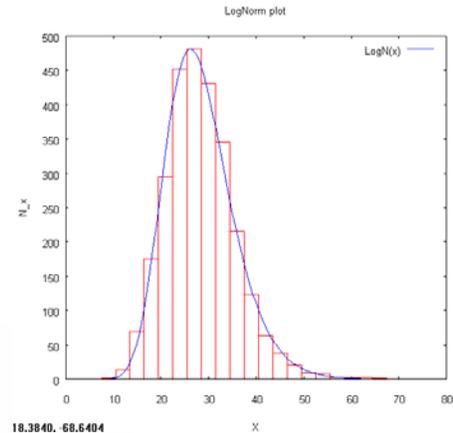
A = 2760.
s2 = .05
xo = 3.

# ===== raffinamento dell'ampiezza =====
fit LogN(x) 'Distrib1.dat' u 1:2 via A,xo,s2

pl 'Distrib1.dat' u 1:2 t '' w boxes, LogN(x) w l 3

pause -1 "fit"
# ===== raffinamento dell'ampiezza =====
Ntot = 2760. # Ntot serve per il calcolo degli errori
fit LogN(x) 'Distrib1.dat' u 1:2:({sqrt($2*(1-$2/Ntot))}) via A,xo,s2
replot

```



```

After 7 iterations the fit converged.
final sum of squares of residuals : 28.1868
rel. change during last iteration : -3.11478e-007

degrees of freedom (FIT_NDF) : 21
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 1.15855
variance of residuals (reduced chisquare) = WSSR/ndf : 1.34223

Final set of parameters Asymptotic Standard Error
=====
A = 2735.54 +/- 56.76 (2.075%)
xo = 3.37151 +/- 0.005165 (0.1532%)
s2 = 0.0546809 +/- 0.001882 (3.442%)

correlation matrix of the fit parameters:

A xo s2
A 1.000
xo 0.008 1.000
s2 0.051 -0.095 1.000

```

Torniamo sul foglio Excel e valutiamo quale è la probabilità associata al valore del χ^2 osservato: **DISRIB.CHI(28.2, 21) = 13%** quindi non posso rifiutare il modello e posso dire che i risultati sono consistenti con una distribuzione LogNormale caratterizzata da

$$A = 2736(57); X_0 = 3.3715(52); \sigma^2 = 0.0547(19)$$