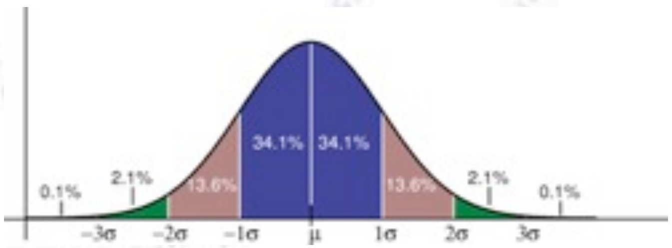
Applied Statistics

ProblemSet - Solution Example






Troels C. Petersen (NBI)



"Statistics is merely a quantization of common sense"

Problem 1.1

1.1 Let x be distributed according to the PDF $f(x) = x\exp(-x)$ in the interval $[0, \infty]$. What is the mean, mode, median, and RMS of this distribution? Possibly verify with simulation!

Generally usefull stuff:

$$f(x) = x \cdot \exp(-x) \quad (1)$$
$$f'(x) = (x-1) \cdot \exp(-x) \quad (2)$$
$$f''(x) = (x-2) \cdot \exp(-x) \quad (3)$$
$$F(x) = -\exp(-x) \cdot (x+1) \quad (4)$$

Mean m

$$m = \int_0^\infty f_m(x)\,dx, \text{ with } f_m(x) = x \cdot f(x) \quad (5)$$
$$F_m(x) = -\exp(-x) * (x^2 + 2x + 2) \quad (6)$$
$$m = \lim_{a\to\infty} F_m(a) - F_m(0) \quad (7)$$
$$= \lim_{a\to\infty} 2 - \exp(-x) * (x^2 + 2x + 2) = 2 \quad (8)$$

Mode e

Three conditions have to be fulfilled for any value e to be a maximum:

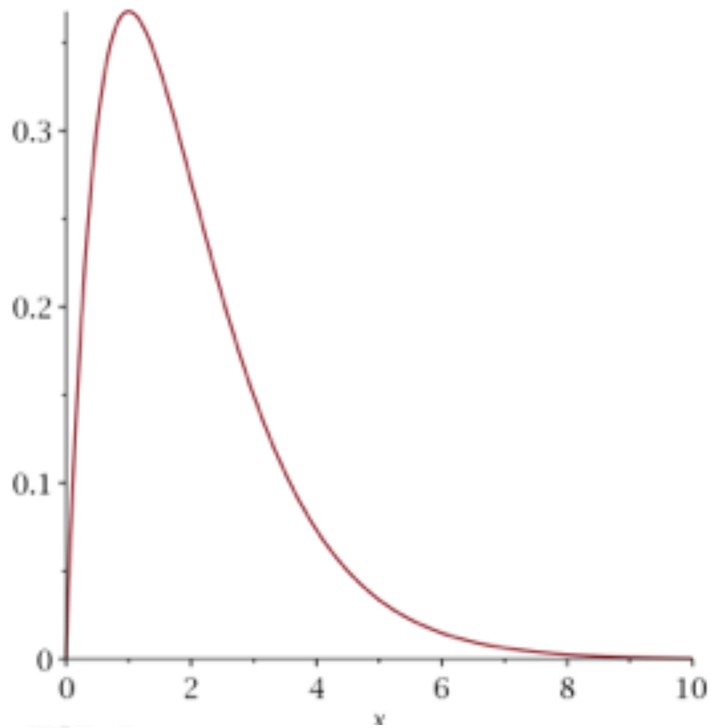
$$f'(e) = (e-1) \cdot \exp(-e) = 0 \quad (9)$$
$$f''(e) = (e-2) \cdot \exp(-e) < 0 \quad (10)$$
$$e > 0 \quad (11)$$
$$\Rightarrow e \in \{1\} \quad (12)$$

Since there is only one maximum, it is at the same time the mode.

Median c

$$0.5 = \int_0^c f(x)\,dx = F(c) - F(0) \quad (13)$$
$$\Rightarrow 0.5 = 1 - (c+1) \cdot \exp(-c) \quad (14)$$
$$c > 0 \quad (15)$$
$$c \approx 1.67835 \quad (16)$$

The right hand side of 14 is a transcendental function of c. See numerical code for solution.



$$r^2 = \int_0^\infty f_r(x)\,dx, \text{ with } f_r(x) = (x-m)^2 \cdot f(x)$$
$$F_r(x) = -\exp(-x) * (x^2 - x^2 + 2x + 2)$$
$$r^2 = \lim_{a\to\infty} F_r(a) - F_r(0) = F_r(0) = 2$$
$$\Rightarrow r = \sqrt{2}$$

Problem 1.2

1.2 Orcs are attacking Minas Tirith with catapults. Gandalf is (by using magic) capable of destroying 98% of the stones shot at the city.

- If the orcs shoot 100 rocks, what is the chance that the city will be unharmed?
- How many stones do the Orcs need to shoot at Minas Tirith in order to be 95% confident of at least one stone hitting the city?

Assuming a binomial process with probability of success (hitting the city) $p = 0.02$ in each of the $n = 100$ trials. The probability $P_{100,0.02,0}$ of zero successes $k = 0$ is:

$$P_{100,0.02,0} = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \approx 0.13262 \quad (21)$$

The number of trials necessary to have no sucesses with a probability of less than $1 - 0.95$ is bounded by the following inequality:

$$1 - 0.95 < P_{n,0.02,0} = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \quad (22)$$

$$\Rightarrow 0.05 = (1-p)^n = 0.98^n \quad (23)$$

$$\Rightarrow n > 148.284 \quad (24)$$

At least $n = 149$ trials are necessary to have at least one hit with a confidence of 0.95.

Problem 1.3

1.3 In 2011 220 persons were killed in traffic and 4039 were injured. In 2012 the numbers were 167 and 3611, respectively. What is the percentage drop in number of deaths? And injuries? How significant are each of these drops?

The drop in number of deaths d and injuries i are:

$$1 - \frac{d_{2012}}{d_{2011}} = 1 - \frac{167}{220} \approx 0.241 = 24.1\% \tag{25}$$

$$1 - \frac{i_{2012}}{i_{2011}} = 1 - \frac{3611}{4039} \approx 0.106 = 10.6\% \tag{26}$$

The variance on two incident numbers n_1, n_2 is (assuming a Poisson process) the same as the numbers themselves. The variance on the difference is the sum of the individual differences. It follows that the error on the difference is $\sigma_n = \sqrt{|n_1 + n2|}$. The ratio of differences and errors for d and i are:

$$\frac{|d_1 - d_2|}{\sigma_d} \approx 2.7 \tag{27}$$

$$\frac{|i_1 - i_2|}{\sigma_i} \approx 4.9 \tag{28}$$

Problem 2.1

2.1 In a repeated experiment the velocity of a ball v is measured seven times.

Velocity (m/s)	94.1	86.3	93.9	89.8	101.2	97.5	118.3

- What is the average velocity and its uncertainty?
- If the mass of the ball is $m = 0.27 \pm 0.03$ kg, what is the kinetic energy $E_{\text{kin}} = \frac{1}{2}mv^2$ of the ball and its uncertainty?
- If, for sc ... ne velocity and the mass of ... n be?
- Do you f ... your answer.

The mean of the velocity is

$$< v >= \frac{1}{N}\sum_i^N v_i = 97.3m/s$$

The RMS is

$$\sqrt{< v^2 > - < v >^2} = 9.7m/s$$

The error on the mean is

$$\sigma_v = \frac{RMS_v}{\sqrt{7}} = 3.7m/s$$

So the velocity is $97 \pm 4m/s$

The kinetic energy becomes

$$E_{kin} = \frac{1}{2}0.27kg \cdot (97.3m/s)^2 = 1.3kJ$$

The uncertainty is found through error propagation

$$\sigma_E = \sqrt{\left(\frac{\partial E}{\partial m}\right)^2 \sigma_m^2 + \left(\frac{\partial E}{\partial v}\right)^2 \sigma_v^2} = 0.2kJ$$

Problem 2.1

2.1 In a repeated experiment the velocity of a ball v is measured seven times.

Velocity (m/s)	94.1	86.3	93.9	89.8	101.2	97.5	118.3

- What is the average velocity and its uncertainty?
- If the mass of the ball is $m = 0.27 \pm 0.03$ kg, what is the kinetic energy $E_{\text{kin}} = \frac{1}{2}mv^2$ of the ball and its uncertainty?
- If, for some reason, there were a (linear) correlation between the velocity and the mass of the ball of $\rho_{vm} = -0.6$, what would the above answer then be?
- Do you find any of the measurements to be suspecious? Quantify your answer.

With correlation $\rho_{vm} = -0.6$ we get

$$\sigma_E = \sqrt{\left(\frac{\partial E}{\partial m}\right)^2 \sigma_m^2 + \left(\frac{\partial E}{\partial v}\right)^2 \sigma_v^2 + 2\frac{\partial E}{\partial m}\frac{\partial E}{\partial v}\rho_{vm}\sigma_m\sigma_v} = 0.11kJ$$

Thinking along the lines of the following is correct, but...

$$z = \frac{118.3\,\text{m/s} - \mu_v}{\sigma_v} = \frac{118.3\,\text{m/s} - 97.3\,\text{m/s}}{10.5\,\text{m/s}} = 2$$

Problem 2.2

2.2 If $\theta = 0.54 \pm 0.02$, what is the uncertainty on $\cos\theta$, $\sin\theta$, and $\tan\theta$? What if $\theta = 1.54 \pm 0.02$?

Note that for $\theta = 1.54 \pm 0.02$ error propagation using the derivative is not suitable for the tan, because the derivative changes to quickly.

$$\text{For } \theta = 0.54 \pm 0.02 : \tag{32}$$

$$\sin(\theta) = 0.51 \pm 0.02 \tag{33}$$

$$\cos(\theta) = 0.857 \pm 0.010 \tag{34}$$

$$\tan(\theta) = 0.60 \pm 0.03 \tag{35}$$

$$\text{For } \theta = 1.54 \pm 0.02 : \tag{36}$$

$$\sin(\theta) = 0.9995 \pm 0.00006 \tag{37}$$

$$\cos(\theta) = 0.03 \pm 0.02 \tag{38}$$

$$19 < \tan(\theta) < 93 \tag{39}$$

Problem 2.3

2.3 Snell's Law states that $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Find n_2 and its error from the following measurements:

$$\theta_1 = (22.03 \pm 0.2)^\circ \qquad \theta_2 = (14.45 \pm 0.2)^\circ \qquad n_1 = 1.0000$$

First we change the measured values from degrees to radians:

$$\theta_1 = (22.03 \pm 0.2)\frac{2\pi}{360} = 0.3845 \pm 0.0035.$$
$$\theta_2 = (14.45 \pm 0.2)\frac{2\pi}{360} = 0.2522 \pm 0.0035.$$

Rewriting Snell's law yields the value for n_2:

$$n_2 = n_1 \frac{\sin\theta_1}{\sin\theta_2} = 1.50.$$

The error on n_2 is:

$$\sigma_{n_2} = \sqrt{\left[\frac{\partial n_2}{\partial \theta_1}\sigma_{\theta_1}\right]^2 + \left[\frac{\partial n_2}{\partial \theta_2}\sigma_{\theta_1}\right]^2}$$
$$= \sqrt{\left[n_1 \frac{\cos\theta_1}{\sin\theta_2}\sigma_{\theta_1}\right]^2 + \left[-n_1 \frac{\sin\theta_1 \cos\theta_2}{\sin^2\theta_2}\sigma_{\theta_1}\right]^2}$$
$$= 0.02.$$

So the final result for the refractive index is:

$$n_2 = 1.50 \pm 0.02$$

Problem 2.4

2.4 The initial activity N_0 and lifetime τ of a radioactive source is known with a relative uncertainty of 1%. When estimating the activity $N = N_0 e^{-t/\tau}$ the uncertainty will initially be dominated by the uncertainty in N_0 and later by the uncertainty in τ. For what value of t/τ will the to uncertainties contribute equally to the uncertainty on N?

$$\Delta N_{N_0} = \Delta N_0 \exp(-t/\tau)$$

$$\Delta N_\tau = \Delta\tau \cdot N_0 \cdot t \exp(-t/\tau)/\tau^2$$

$$\Delta N_{N_0} = \Delta N_\tau$$

$$\Rightarrow \Delta N_0 \exp(-t/\tau) = \Delta\tau \cdot N_0 \cdot t \exp(-t/\tau)/\tau^2$$

$$\Rightarrow t/\tau = \frac{\tau \Delta N_0}{N_0 \Delta\tau} = 1$$

Problem 3.1

3.1 Let $f(x) = ax^2$ be proportional to a PDF for $x \in [-1, 2]$.

- In order for this PDF to be normalized, what value should a have?
- What is the mean and width of $f(x)$?
- By which method would you generate random numbers according to this PDF?

mbers according to $f(x)$. Let t be
ierate 1000 values of t.
ording to the mean and width of t
ime as the one above?

$$1 = \int_{-1}^{2} a\,x^2\,dx = 3a \quad (46)$$

$$\Rightarrow a = \frac{1}{3} \quad (47)$$

$$\text{mean: } m = \int_{-1}^{2} a\,x^3\,dx = \frac{15a}{4} = \frac{5}{4} \quad (48)$$

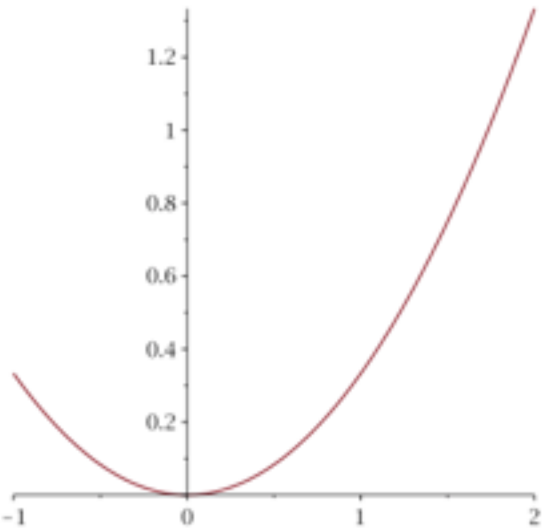
$$\text{square of width: } w^2 = \int_{-1}^{2} (x-m)^2 a\,x^2\,dx = m^2 - \frac{5}{2}m + \frac{11}{5} = \frac{51}{80} \quad (49)$$

$$(50)$$

To generate random numbers according to this distribution, one can calculate the cumulative distribution $c(x) = \frac{1}{9}(x^3+1)$ and invert it:

$$x(c) = \begin{cases} (9c-1)^{\frac{1}{3}} & \text{if } x > \frac{1}{9} \\ -(1-9c)^{\frac{1}{3}} & \text{otherwise} \end{cases} \quad (51)$$

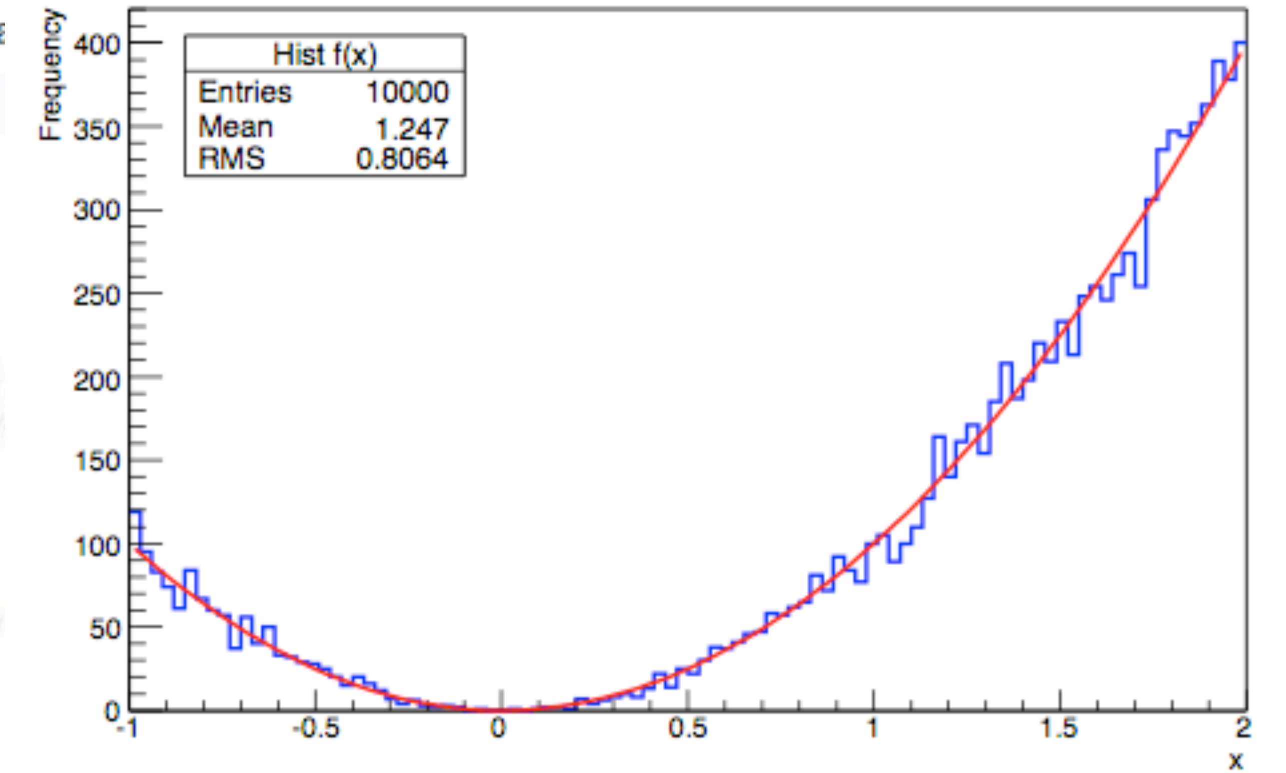
A uniformly distributed random variable c on the interval $[0,1]$ can now be transformed into the desired result using $x(c)$.



Problem 3.1

3.1 Let $f(x) = ax^2$ be proportional to a PDF for $x \in [-1, 2]$.

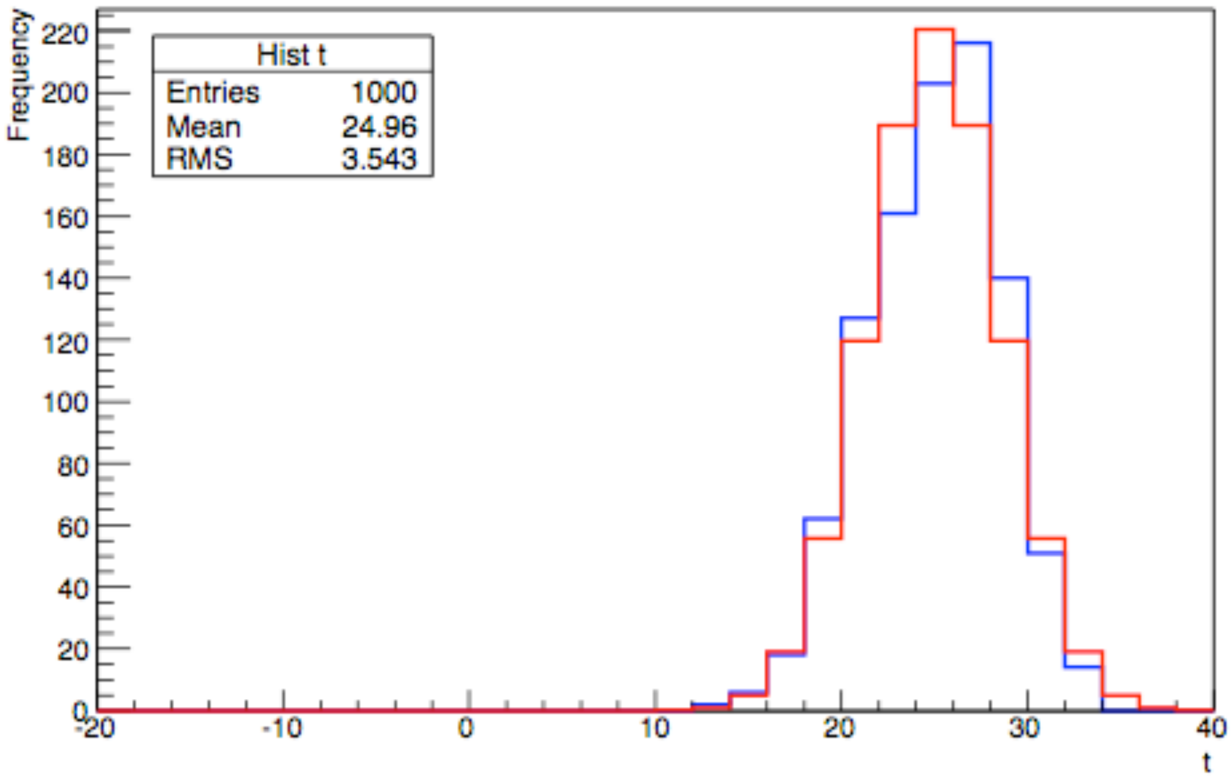
- In order for this PDF to be normalized, what value should a have?
- What is the mean and width of $f(x)$?
- By which method would you generate random numbers according to this PDF?
- Produce an algorithm, which generates random numbers according to $f(x)$. Let t be a sum of twenty random values from $f(x)$, and generate 1000 values of t.
- Ge ... nd width of t (ca ... ?



Problem 3.1

3.1 Let $f(x) = ax^2$ be proportional to a PDF for $x \in [-1, 2]$.

- In order for this PDF to be normalized, what value should a have?
-
- 'DF?
- . Let t be
- width of t



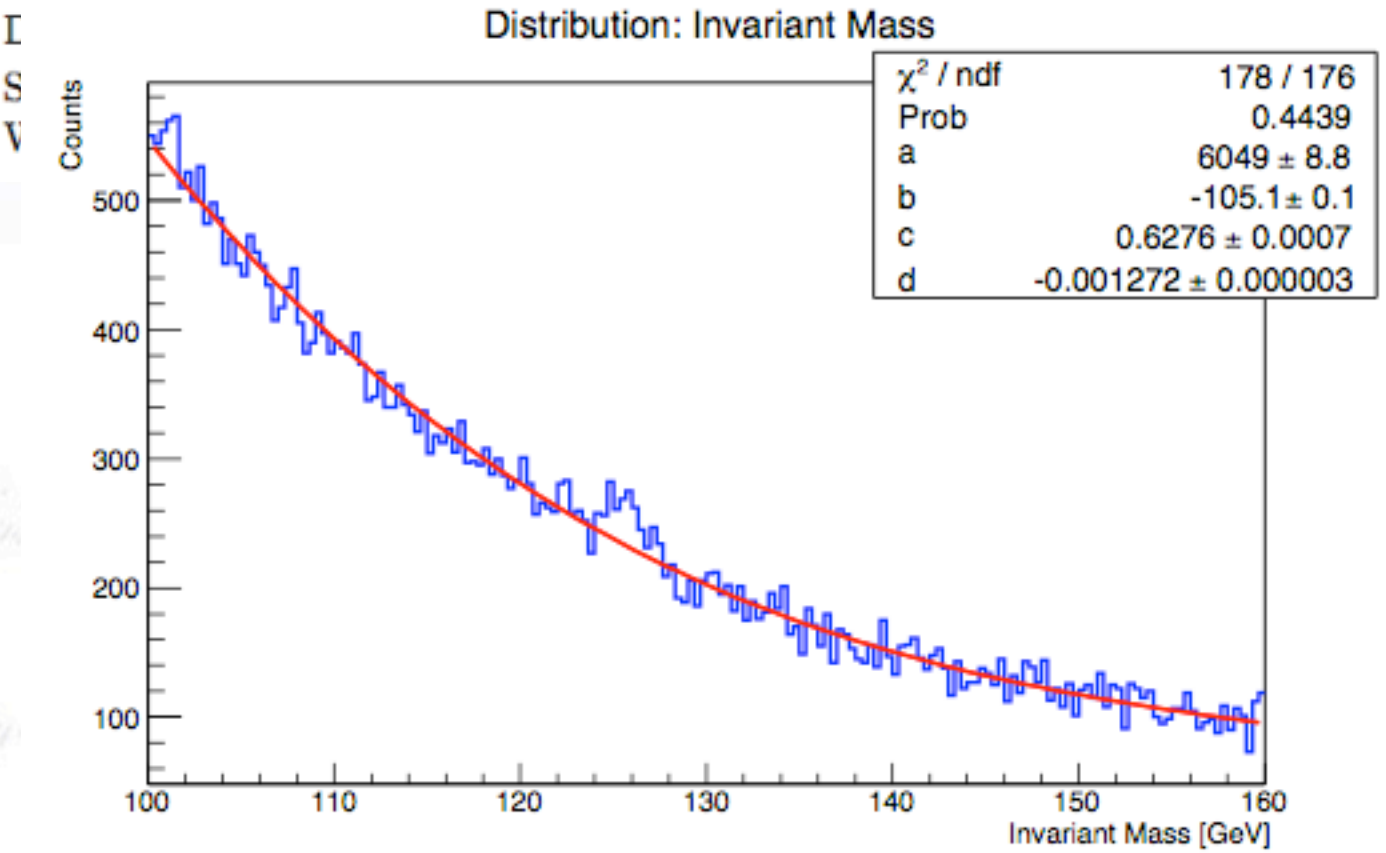
Calculating the χ^2 as:

$$\chi^2 = \sum_{i=17}^{28} \frac{(O_i - E_i)^2}{E_i} = 22.3.$$

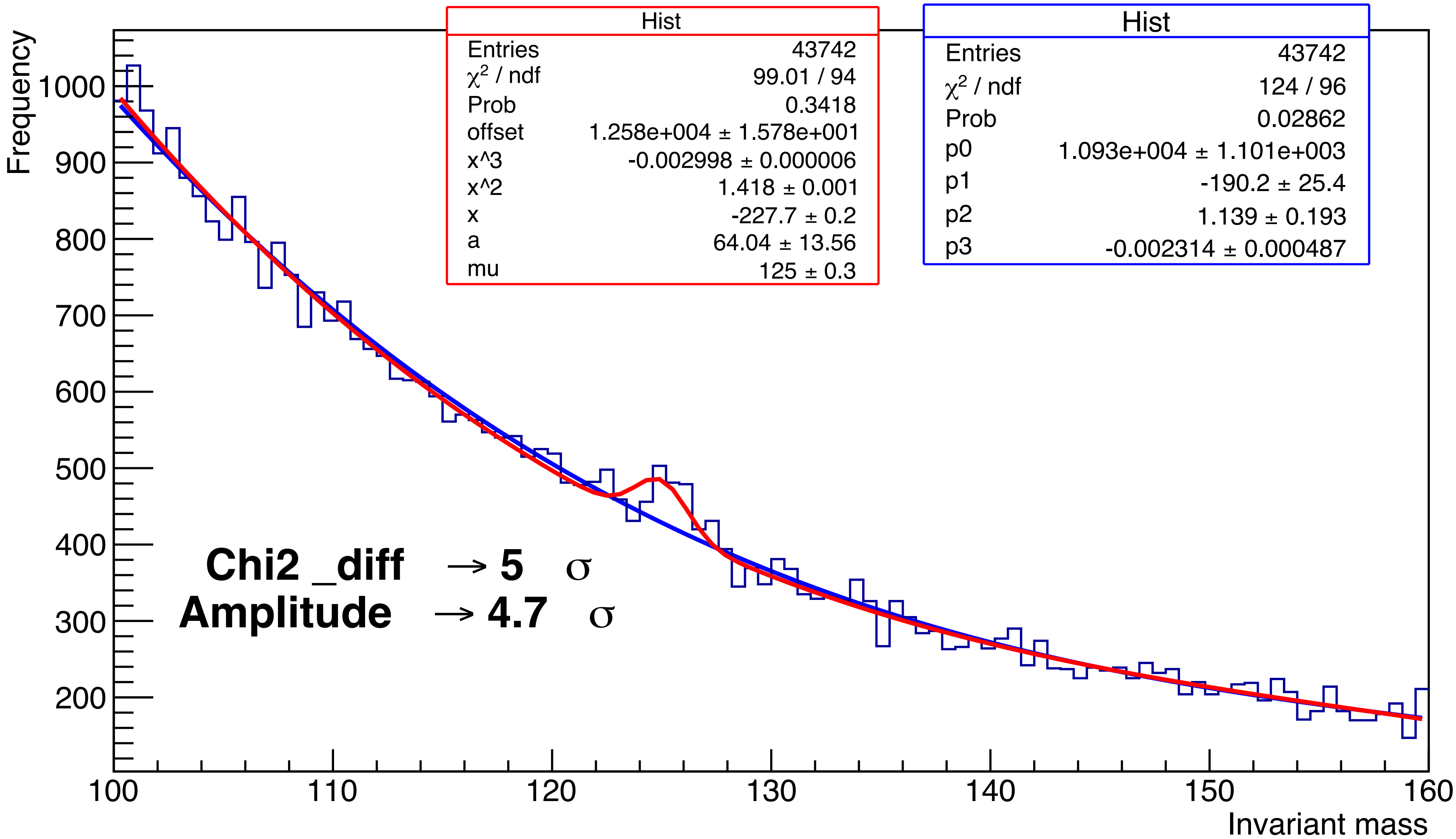
Problem 4.1

4.1 The file [www.nbi.dk/~petersen/data_HiggsGG.txt] contains 43742 measurements of the invariant mass between two photons in the range 100-160 GeV at the ATLAS experiment. (This problem is inspired by the Higgs search and discovery in 2012, which triggered the 2013 Nobel prize in physics).

- Read the file, plot the measurements with a reasonable binning and fit the distribution with a suitable smooth function. How well do you manage to describe this distribution?
- I ... .
- S ... trum.
  V



Problem 4.1



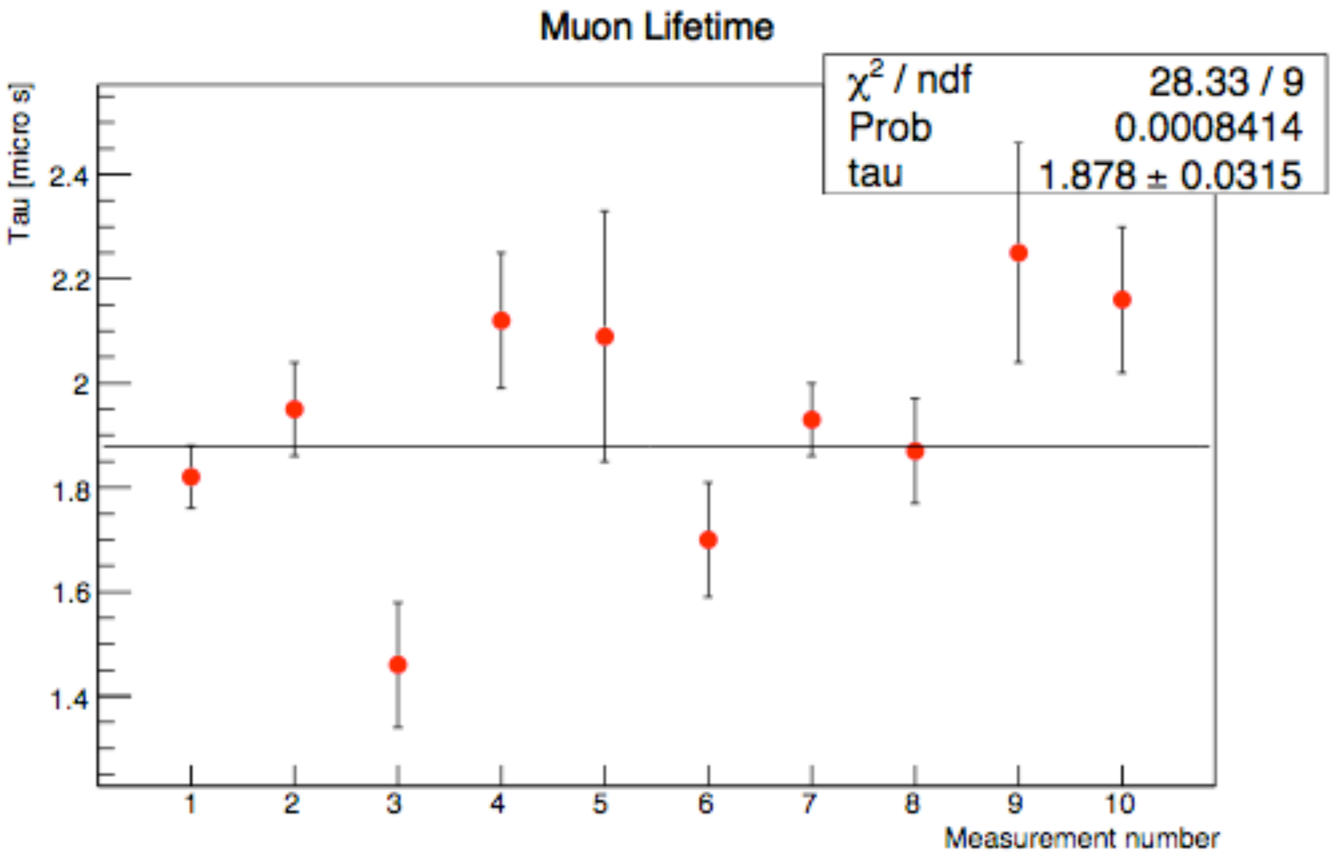
Hist
Entries 43742
χ² / ndf 99.01 / 94
Prob 0.3418
offset 1.258e+004 ± 1.578e+001
x^3 -0.002998 ± 0.000006
x^2 1.418 ± 0.001
x -227.7 ± 0.2
a 64.04 ± 13.56
mu 125 ± 0.3
Hist
Entries 43742
χ² / ndf 124 / 96
Prob 0.02862
p0 1.093e+004 ± 1.101e+003
p1 -190.2 ± 25.4
p2 1.139 ± 0.193
p3 -0.002314 ± 0.000487
Frequency
1000
900
800
700
600
500
400
300
200
100
110
120
130
140
150
160
Invariant mass
Chi2 _diff → 5 σ
Amplitude → 4.7 σ

Problem 4.2

4.2 In the past years several groups of students have been measuring the lifetime of the muon in the basement at NBI. Their results and estimated uncertainties are listed below:

Group	1	2	3	4	5	6	7	8	9	10
Result (μs)	1.82	1.95	1.46	2.12	2.09	1.70	1.93	1.87	2.25	2.16
Uncertainty (μs)	0.06	0.09	0.12	0.13	0.24	0.11	0.07	0.10	0.21	0.14

- (... the χ^2 and
  t
- I
- I ... of τ_μ?



Problem 5.1

5.1 A magnet drops a ball through six timing gates providing t at various distances d to measure the acceleration due to gravity g. The distances are accurately know, while t has an uncertainty of 0.01 s.

Distance (m)	0.200	0.500	1.000	1.500	2.000	2.500
Time (s)	0.150	0.265	0.383	0.503	0.582	0.652

- Assume the magnet drops the ball at $t = 0$, and fit the data to obtain g. Comment on the fit quality.