Data Analysis Report

Oskar Moras

April 7, 2018

Contents

1 Chapter 1

1.1 Problem 1

Jane has two children, and pictures of them in her pocket. She shows us one picture; it is a girl. The probability that the second child also is a girl is 50%. The fact that we know that the first child is a girl does not influence the probability of the second child being a girl. The probability of a child being a boy or a girl is 50/50, just because we know that one of her children is a girl, does not influence the probability of the second child being a girl. Therefore, the probability that her second child is a girl is 50%. Variation: Jane has to show us a picture of a girl if she has one. This changes the situation, before we had four choices BB, BG, GB, GG, each with a probability of 1/4. Now Jane shows us a picture of a girl, so we know she can not have two boys, leaving us with three choices BG, GB, GG, each with a probability of 1/3. Since the first child is a girl, the probability that the second child also is a girl is 1/3.

1.2 Problem 2

Now we look at possible definitions for the probability of the data from section 1.2.3 in the lecture script. For example one can calculate the probability to...:

1. Get a certain number of sub-sequences with a certain length, e.g. TTT, within the sequence.
2. Get at least a certain number of T's or H's.
3. Get an even or odd number of T's or H's.
4. Get a T or an H at certain places within the sequence.

1.3 Problem 3

In this Problem we have a detector with a 10% energy resolution and we measure an energy E. To assign probability values to possible true values of the energy is not possible in this case because we do not have enough information. The better the resolution, the closer we can say that we are to the true energy. If we only measure one event, we can not say much about the true energy. Of course, it is more probable to measure the true energy, or a value close to it. Nevertheless, we could just as well have happened to measure an energy very far from the true value. Furthermore, we don't know how much energy the particle actually deposited in the detector, nor how the detector signal is converted to an energy signal that we read out. Basically, we don't have enough information to say anything about the probabilities of possible true values of the energy.

1.4 Problem 4

In this problem we investigate the probability of having a fever (F) under certain conditions, such as spots (SP), acute lethargy (AL), rapid thirst (RT) and violent sneezes (VS). The probabilities given are:

$$P(SP \mid F) = 1; \quad P(AL \mid F) = 1; \quad P(RT \mid F) = 0.6; \quad P(VS \mid F) = 0.2$$
$$P(SP \mid !F) = 0.03; \quad P(AL \mid !F) = 0.10; \quad P(RT \mid !F) = 0.02; \quad P(VS \mid !F) = 0.05$$
$$P(F) = 1/10000; \quad P(!F) = 1 - P(F)$$

We use Bayes theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{\sum_{i=1}^{N} P(B \mid A_i)P(A_i)}$$

We investigate the probability of having the fever if we have all four of the symptoms (S):

$$P(F \mid 4S) = \frac{P(F) \cdot \prod_i^4 P(S_i \mid F)}{P(F) \cdot \prod_i^4 P(S_i \mid F) + P(!F) \cdot \prod_i^4 P(S_i \mid !F)} = 0.8$$

For calculating the probability of having the fever and any three of the symptoms we use the same formula, but the product is only over the three symptoms:

$$P(F \mid SP, AL, RT) = 50.0\%; \quad P(F \mid SP, RT, VS) = 28.5\%;$$
$$P(F \mid AL, RT, VS) = 10.7\%; \quad P(F \mid SP, AL, VS) = 11.8\%$$

2 Chapter 2

2.1 Problem 8

$$P(x) = xe^{-x}, \qquad 0 \leq x \leq \infty$$

a) The integral for the mean (μ) is calculated using a Feynman parameter t[2]:

$$\mu = E[x] = \int_0^\infty x \cdot P(x)dx = \int_0^\infty x^2 \cdot e^{-x} dx = \frac{d^2}{dt^2}[\int_0^\infty e^{-xt} dx] \mid_{t=1} = \frac{d^2}{dt^2}[-\frac{1}{t}e^{-xt}]_0^\infty \mid_{t=1} = \frac{d^2}{dt^2}[\frac{1}{t}] \mid_{t=1} = 2$$

Standard deviation:

$$\sigma = \sqrt{E[x^2] - (E[x])^2}$$

Analogous to the mean:

$$E[x^2] = \int_0^\infty x^2 \cdot P(x)dx = \int_0^\infty x^3 \cdot e^{-x} dx = -\frac{d^3}{dt^3}[\frac{1}{t}] \mid_{t=1} = 6$$

Therefore:

$$\sigma = \sqrt{2}$$

The probability content ($P_{content}$) in the interval ($\mu - \sigma$, $\mu + \sigma$) is given by:

$$P_{content} = \int_{a=\mu-\sigma}^{b=\mu+\sigma} P(x)dx$$

Again, using the Feynman parameter we are left with:

$$P_{content} = e^{-a} \cdot (1 + a) - e^{-b} \cdot (1 + b) = 0.738 \quad \hat{=} \quad 73.8\%$$

b) The median (x_{med}) is calculated via:

$$F(x_{med}) = \int_0^{x_{med}} P(x)dx \stackrel{!}{=} 0.5$$

Which gives the expression:

$$1 - e^{-x} \cdot (x + 1) = 0.5$$

This equation is satisfied by $x_{med} = 1.68$.

For the 68% central interval one must find a maximum lower and minimum upper x-value where the tail integrals of P(x) contain 16% or less each. The lower interval limit is found by integrating P(x) from 0 to x_{min} and raising x_{min} as far as the integral is smaller than or equal to 0.16. The upper interval limit is found by integrating P(x) from x_{max} to infinity and lowering x_{max} as far as the integral value is smaller than or equal to 0.16. The following x_{min} and x_{max} values that make the tail integrals exceed 0.16 are taken as the limits. This procedure gives a 68% central interval of: [0.71, 3.29]. A step size of 0.0001 was chosen for raising and lowering x_{min} and x_{max} respectively.

c) The mode ($x*$) is the x-value that maximizes P(x), which satisfies the following relation:

$$\frac{d}{dx}P(x) \mid_{x=x*} \stackrel{!}{=} 0$$

The mode is therefore $x* = 1$.

The 68% smallest interval is found by starting from the mode, and expanding the interval to either the left or the right by the one x-value with the highest probability, until 68% of the probability is in the interval. The left and right limits are extended according to the probability ranking. As long as the interval contains less than 68% probability, in the next step, the next left or right x-value with highest ranking gets added to the interval. In practice this is done by integrating P(x) over the interval and checking its probability content. If the probability content is too small one of the limits is extended by a step size of 0.0001 and the probability integral is calculated again until the 68% limit is reached. To decide which one of the upper and lower limits gets extended, the $P(x_{low} - 0.0001)$ and $P(x_{high} + 0.0001)$ are compared and the corresponding x-value yielding higher probability value gets included. This procedure gives a 68% smallest interval of: [0.270, 2.490].

2.2 Problem 10

In this problem, the characteristics of a flat prior derived in the lecture script are used. The posterior probability P for the success parameter p is given by:

$$P(p \mid N, r) = (N+1) \cdot P(r \mid N, p)$$

$P(r \mid N, p)$ is the binomial probability distribution of r, where r is the number of successes and N is the number of trials. The mode is given by $p* = r/N$. The values of the mode for the different energies are given in Table 2.2.

To find the 68% probability range, the smallest interval is used. The procedure is described in Problem 8 of Chapter 2. The lower(p1) and upper(p2) limits of the 68% smallest intervals for the different energies are given in Table 2.2.

Energy	Mode(p*)	p1	p2
0.5	0	0	0.0113
1.0	0.04	0.0233	0.0629
1.5	0.2	0.1625	0.2415
2.0	0.58	0.5308	0.6281
2.5	0.92	0.8902	0.9442
3.0	0.987	0.9831	0.9903
3.5	0.995	0.9923	0.9969
4.0	0.998	0.9961	0.9991

Table 1: Lower(p1) and upper(p2) limits of the 68% probability smallest interval for p for the different energies.

These results are plotted in Fig. 1.

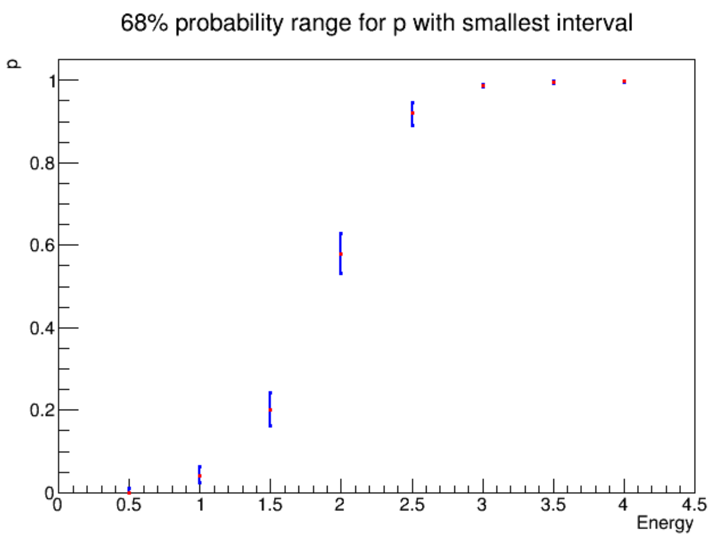

Figure 1: The 68% probability smallest intervals for p for the different energies. The modes are marked with red dots, the intervals with vertical blue lines.

2.3 Problem 11

In this problem the data from problem 10 is analyzed using the frequentist method. To avoid creating a band plot for each energy, and then searching for a probability range corresponding to a certain number of successes, the known number of successes is used directly to search for the probability range. This way only one range has to be calculated corresponding to one success value for each energy. Therefore no complete band plot will be produced in this problem. The interval used for this task is the central interval with 90% probability content. The 90% CL interval for p are those values of p, for which the measured number of successes is contained within the 90% central interval for the success parameter r. The p range is found by calculating the 90% central intervals of r for values of p between 0 and 1, with a step size of 0.0001 and then checking if the measured success parameter lies within those intervals. The lowest possible p, for which the measured success parameter is the smallest member of the central interval, is set as the lower p limit. The highest possible p, for which the measured success parameter is the largest member of the central interval, is set as the upper p limit. This defines the 90% CL interval for p.

The 90% CL intervals for the different energies can be seen in Table 2.3 and are plotted in Fig. 2.

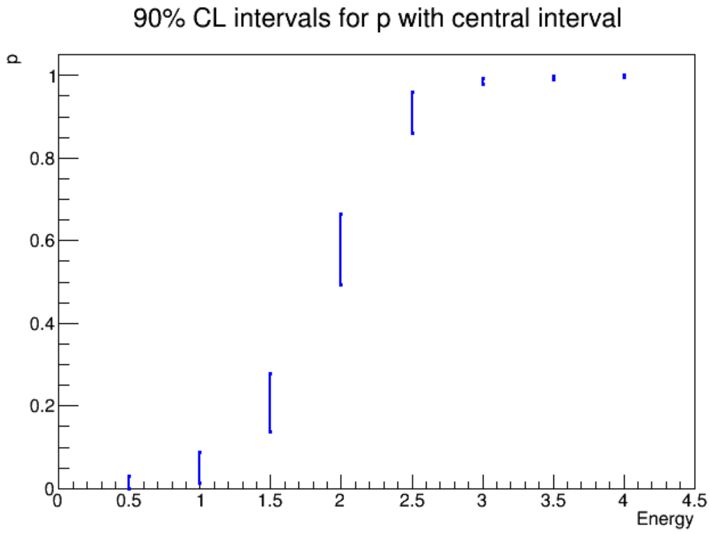

Figure 2: The 90% CL intervals for p for the different energies. The intervals are marked with vertical blue lines.

Energy	Lower limit	Upper limit
0.5	0	0.0295
1.0	0.0138	0.0892
1.5	0.1367	0.2772
2.0	0.4929	0.6635
2.5	0.8604	0.9594
3.0	0.9795	0.9922
3.5	0.9896	0.998
4.0	0.9938	0.9996

Table 2: Lower and upper limits of the 90% CL central intervals for p for the different energies.

2.4 Problem 13

We start with Bayes theorem:

$$P(p \mid N, r) = \frac{P(r \mid N, p)P_0(p)}{\int P(r \mid N, P)P_0(p)dp}$$

Using a flat prior to start with we can write:

$$P(p \mid N, r) = \frac{P(r \mid N, p)}{\int P(r \mid N, p) dp}$$

For the Binomial distribution, the coefficients cancel and we are left with:

$$P(p \mid N, r) = \frac{p^r (1-p)^{N-r}}{\int_0^1 p^r (1-p)^{N-r} dp}$$

To solve the integral in the denominator we use the definition of the standard beta function, which for instance can be found in the lecture script:

$$\beta(r+1, N-r+1) = \int_0^1 p^r (1-p)^{N-r} dp = \frac{r!(N-r)!}{(N+1)!}$$

For the posterior we therefore get:

$$P_1(p \mid N, r) = \frac{(N+1)!}{r!(N-r)! \cdot p^r (1-p)^{N-r}}$$

We now use this posterior probability as the prior in the next iteration and calculate P_2. The prefactors that do not depend on p get canceled and we are left with the expression for the second posterior probability distribution:

$$P_2(p \mid N, r) = \frac{p^{2r} (1-p)^{2(N-r)}}{\int_0^1 p^{2r} (1-p)^{2(N-r)} dp}$$

The denominator therefore corresponds to the beta function:

$$\beta(2r+1, 2(N-r)+1) = \int_0^1 p^{2r} (1-p)^{2(N-r)} dp = \frac{(2r)!(2(N-r))!}{(2N+1)!}$$

The posterior is therefore:

$$P_2(p \mid N, r) = \frac{(2N+1)!}{(2r)!(2N-2r)!} p^{2r} (1-p)^{2(N-r)}$$

For n iterations the beta function is:

$$\beta(nr+1, n(N-r)+1) = \int_0^1 p^{nr} (1-p)^{n(N-r)} dp = \frac{(nr)!(n(N-r))!}{(nN+1)!}$$

And the posterior is:

$$P_n(p \mid N, r) = \frac{(nN+1)!}{(nr)!(nN-nr)!} p^{nr} (1-p)^{n(N-r)} \qquad q.e.d.$$

The expectation value and the variance are defined as:

$$E[p] = \int_0^1 p \cdot P(p \mid N, r) dp$$
$$V[p] = E[p^2] - E[p]^2$$

Using the following definitions of the standard beta function and the gamma function [1]:

$$\int_0^1 p^{\alpha-1} (1-p)^{\beta-1} dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}; \qquad \Gamma(n) = (n-1)!$$

For the expectation value we get:

$$E[p] = \frac{(nN+1)!}{(nr)!(nN-nr)!} \int_0^1 p^{nr+1} (1-p)^{n(N-r)} dp =$$

$$\frac{(nN+1)!}{(nr)!(nN-nr)!} \cdot \frac{\Gamma(nr+2)\Gamma(n(N-r)+1)}{\Gamma(nr+n(N-r)+3)} = \frac{(nN+1)!}{(nr)!(nN-nr)!} \cdot \frac{(nr+1)!(n(N-r))!}{(nr+n(N-r)+2)!} = \frac{nr+1}{nN+2}$$

Taking the limit $n \to \infty$ and using L'Hospital's rule we get:

$$\lim_{n\to\infty} \frac{nr+1}{nN+2} = \frac{r}{N}$$

For the variance, we first calculate $E[p^2]$:

$$E[p^2] = \frac{(nN+1)!}{(nr)!(nN-nr)!} \int_0^1 p^{nr+2}(1-p)^{n(N-r)} dp$$

Using the same steps as for $E[p]$ but with a different argument in one gamma function we are left with:

$$E[p^2] = \frac{(nr+1)(nr+2)}{(nN+2)(nN+3)}$$

Taking the limit $n \to \infty$ and using L'Hospital's rule we get:

$$\lim_{n\to\infty} \frac{n^2r^2+3nr+2}{n^2N^2+5nN+6} = \lim_{n\to\infty} \frac{2nr^2+3r}{2nN^2+5N} = \frac{r^2}{N^2}$$

Therefore the variance is:

$$V[p] = E[p^2] - E[p]^2 = \frac{r^2}{N^2} - \frac{r^2}{N^2} = 0$$

3 Chapter 3

3.1 Problem 4

In this problem we look at some properties of the function $f(x) = \frac{1}{2}e^{-|x|} \quad \text{for} -\infty < x < \infty$.

a) The mean and standard deviation are defined as follows:

$$\mu = E[x]; \qquad \sigma = \sqrt{V[x]}$$

The integrals for $E[x]$and$V[x]$ can be solved using partial integration or Wolframalpha:

$$E[x] = \frac{1}{2}\int_{-\infty}^{\infty} x \cdot e^{-|x|} dx = \frac{1}{2}\int_{-\infty}^{0} x \cdot e^{x} dx + \frac{1}{2}\int_{0}^{\infty} x \cdot e^{-x} dx = 0$$

Or one can simply see that the integrand is anti-symmetrical around 0, and therefore the $E[x]$ integral is 0. For the variance, first $E[x^2]$ is calculated, also using partial integration and split up the same way as for $E[x]$:

$$E[x^2] = \frac{1}{2}\int_{-\infty}^{\infty} x^2 \cdot e^{-|x|} dx = \frac{1}{2}\int_{-\infty}^{0} x^2 \cdot e^{x} dx + \frac{1}{2}\int_{0}^{\infty} x^2 \cdot e^{-x} dx = 2$$

Therefore the standard deviation is:

$$\sigma = \sqrt{V[x]} = \sqrt{2}$$

b) The maximum is found by solving:

$$\frac{d}{dx} f(x) \stackrel{!}{=} 0$$

Using Wolframalpha:

$$0 \stackrel{!}{=} \frac{d}{dx}\frac{1}{2}e^{-|x|} = -\frac{x \cdot e^{-|x|}}{|\, x \,|}$$

Solving the above equation for the cases $x < 0$ and $x > 0$ yields, in both cases, a maximum at $x = 0$ of $f(x = 0) = \frac{1}{2}$.

To find the FWHM we need to solve the equation:

$$\frac{1}{4} = \frac{1}{2}e^{-|x|}$$

From this we get two x-values, $x_1 < 0$ and $x_2 > 0$: $x_1 = -ln(2)$ and $x_2 = ln(2)$. This yields a FWHM of $2ln(2) = 1.39$, compared to the standard deviation $\sigma = \sqrt{2} = 1.41$.

c) The probability content in the $\pm 1\sigma$ interval around the peak at 0 is given by:

$$\int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{2}e^{-|x|}dx = \frac{1}{2}\int_{-\sqrt{2}}^{0} e^{x}dx + \frac{1}{2}\int_{0}^{\sqrt{2}} e^{-x}dx = 1 - e^{-\sqrt{2}} = 0.76$$

To summarize we got the following properties:

$$\mu = 0; \qquad \sigma = \sqrt{2}; \qquad FWHM = 2ln(2); \qquad P_{content}([0-\sigma, 0+\sigma]) = 0.76$$

3.2 Problem 7

In this problem we take a look at the Poisson probability distribution for the case of 9 observed events.

a) The Poisson probability distribution is given by:

$$P(n \mid \nu) = \frac{e^{-\nu\nu^n}}{n!}$$

Assuming a flat prior, the posterior probability distribution for ν is given by(see lecture script for derivation):

$$P(\nu \mid n) = \frac{e^{-\nu\nu^n}}{n!}$$

Here n is fixed and ν is the variable. To find the 95% probability lower limit for ν, one has to integrate $P(\nu \mid n)$ for n=9 from ν_{low} to infinity and set ν_{low} such that the probability content is at least 95%. Basically, one solves the following equation:

$$\int_{\nu_{low}}^{\infty} P(\nu \mid n = 9) \stackrel{!}{=} 0.95$$

In practice this is implemented in code by lowering ν_{low} so far until the integral is bigger or equal to 0.95. ν_{low} is lowered in steps of size 0.0001.

This procedure yields a lower limit on ν of $\nu = 5.4254$, meaning that with 95% probability $\nu > 5.4254$.

b) To find the 68% CL smallest interval for ν, one needs to create 68% smallest intervals of the parameter n for different νs and find those ν values for which the measured number of successes n=9 is contained in that smallest interval for n given a certain ν. This ν interval is then the 68% CL smallest interval for ν.

This is implemented in the code by raising ν in steps of 0.001, creating the 68% smallest interval for n for the probability distribution $P(n \mid \nu_{fixed})$, then checking if n=9 is included in this n-interval. If n=9 is contained in the 68% smallest interval for n, then that ν_{fixed} value is added to the ν interval. In Fig. 3 the CL band plot can be found.

This procedure yields a 68% CL smallest interval for ν of [6.499,13.301].

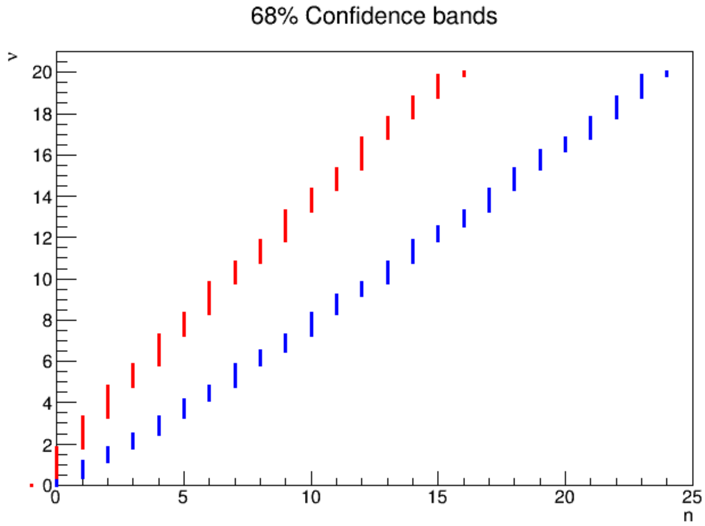

Figure 3: Neyman CL Bands.

3.3 Problem 8

In this problem we take a look at the data from the previous problem and consider the case that we have a known background λ of 3.2 events. For the intervals that we create we need to take this into account, and there are different ways of doing so, here we look at three of them.

a) When creating a Feldman-Cousins confidence level interval, one proceeds in the same way as for the smallest interval, but with a different ranking when deciding which next value of n to add to the interval if the probability content is to small. The ranking parameter r is defined as:

$$r = \frac{P(n \mid \mu = \lambda + \nu)}{P(n \mid \hat{\mu})}$$

where $\hat{\mu}$ is the value that maximizes P for a given n, with the condition $\hat{\mu} \geq \lambda$. This way, empty intervals can be avoided.

Using this ranking, and proceeding as for the smallest interval, we create the ν range for n=9 measured successes. The Feldman-Cousins 68% CL interval for ν that we get using a step size of 0.001 is [6.335,12.790]. The Feldman-Cousins CL bands can be found in Fig. 4.

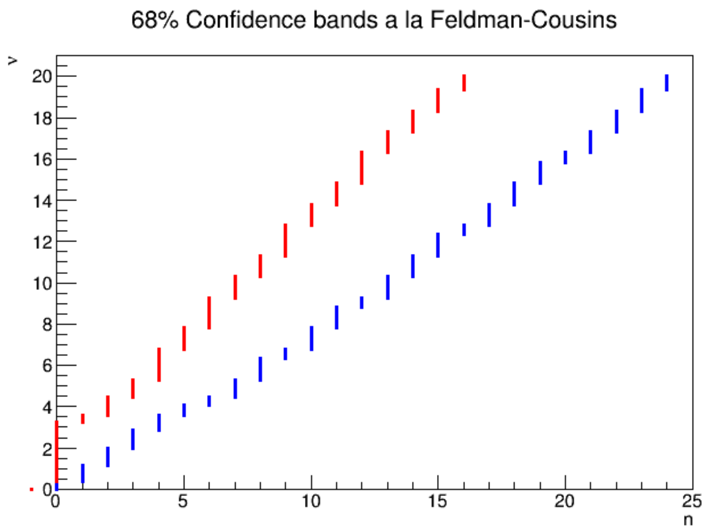

Figure 4: Feldman-Cousins CL bands.

b) When creating a Neyman CL interval, here we use the smallest interval, one proceeds just as without background, but at the end the background is subtracted from the interval limits. If the background is greater than the lower limit, this is set to zero. If the background is greater than the upper limit as well, then both are set to zero. In this case it is possible to get an empty interval. Here we use the interval from the previous problem and subtract the background of 3.2 from the lower and upper limits. The Neyman 68% CL interval that we get for ν is now [3.299,10.101].

c) When doing the Bayesian analysis of the data, and considering a known background one receives a posterior that depends on the background λ. The derivation can be found in the lecture script. The posterior probability for a flat prior is given by:

$$P(\nu \mid n, \lambda) = \frac{e^{-\nu} \cdot (\lambda + \nu)^n}{n! \cdot \sum_{i=0}^{n} \frac{\lambda^i}{i!}}$$

The mode is given by:

$$\nu^* = max\{0, n - \lambda\}$$

Using this posterior probability and the smallest interval definition, starting at the mode of 5.8 with a step size of 0.0001, we get a 68% credible interval for ν of [3.1208,9.1466]. The smallest interval and the posterior probability are plotted in Fig. 5.

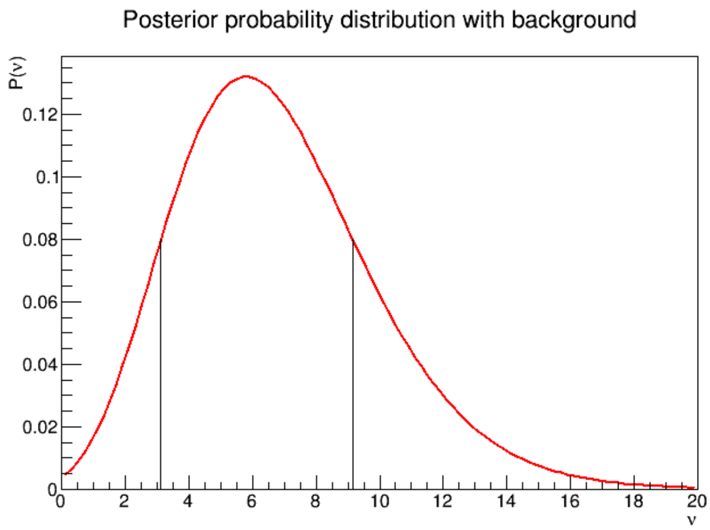

Figure 5: Smallest interval with background.

3.4 Problem 13

In this problem, we investigate the relationship between an unbinned likelihood and a binned Poisson probability. The Poisson distribution is given by:

$$P(n \mid \nu) = \frac{e^{-\nu} \nu^n}{n!}$$

For n_j events per bin we get:

$$P(n \mid \nu_1 ... \nu_K) = \prod_{j=1}^{K} \frac{e^{-\nu_j} \nu_j^{n_j}}{n_j!}; \quad \sum_{j=1}^{K} = n$$

For $K \to \infty$ and assuming each x is not measured more than once we get that each bin only contains either 1 or 0 events, $n_j = 1, 0$.

We start with the relations given in the script. For $K \to \infty$ and $\Delta_j \to \infty$ we have:

$$\nu_j = \int_{\Delta_j} f(x \mid \lambda) dx \to f(x_j \mid \lambda) \cdot \Delta$$

We later let $\Delta \to 0$. Now n is the number of filled bins, we write:

$$\lim_{K\to\infty} \prod_{j=1}^{K} = \lim_{K\to\infty} \prod_{i=1}^{n} \frac{e^{-\nu_i}\nu_i}{1!} \prod_{j=1}^{K-n} \frac{e^{-\nu_j}\nu_j^0}{0!} = \lim_{K\to\infty} \prod_{j=1}^{K} e^{-\nu_j} \prod_{i=1}^{n} \nu_i$$

The exponential term can be written as:

$$\lim_{K\to\infty} \prod_{j=1}^{K} e^{-\nu_j} = \exp(-\sum_{j=1}^{K} f(x_j \mid \lambda) \cdot \Delta$$

We know that f is normalized and let $\Delta \to 0$, then the exponential term goes to 1. Finally we have:

$$\lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} = \prod_{i=1}^{n} \nu_i = \prod_{i=1}^{n} f(x_i \mid \lambda) \cdot \Delta; \quad q.e.d.$$

3.5 Problem 16

In this problem we investigate a thinned Poisson process, and derive the probability distribution for X. Let R be the number of time that $X_n = 1$. Then:

$$X = \sum_{n=1}^{R} 1 + \sum_{n=R+1}^{N} 0 = R$$

We can rewrite $P(X \mid N, p) = P(R \mid N, p)$, which is a binomial distribution. N is given by a Poisson distribution and we must sum over N from 0 to ∞. Let N=X+L, where L is the number of unsuccessful trials. Then:

$$P(X) = \sum_{N=0}^{\infty} P(X \mid N, p) P(N \mid \nu)$$

Where P(X)=0 for N<X. We can therefore shift the sum to start at N=X or let it start at L=0, which is all equivalent. We write:

$$P(X) = \sum_{L=0}^{\infty} P(X \mid X+L) P(X+L \mid \nu) = \sum_{L=0}^{\infty} \frac{(X+L)!}{X!L!} p^X (1-p)^L \frac{e^{-\nu}\nu^{X+L}}{(X+L)!}$$

Simplifying this expression and doing some algebra leaves us with:

$$P(X) = \frac{e^{-\nu p}(\nu p)^X}{X!}; \quad q.e.d.$$

4 Chapter 4

4.1 Problem 8 (a,b)

In this problem we investigate the central limit theorem (CLT) for two different distributions, one where the necessary conditions apply, and one where they don't apply. For the CLT to be applicable, the samples must be random and independent and identically distributed (iid). Furthermore, the moments of the probability distribution need to be finite. Say we perform a number of experiments, and in each experiment perform n measurements of a quantity distributed according to a certain probability distribution, then calculate a mean from these n measurements. If the distribution satisfies the necessary conditions for the CLT, then we expect that the means will be distributed according to a gauss distribution with a certain mean and sigma.

a) In this part of the problem we try out the CLT on the exponential distribution. For this we look at the function $p(x) = \lambda \cdot e^{-\lambda x}$. In this case the CLT applies, because the moments of p(x) are all finite since the exponential function suppresses any power of x, and for large enough n, these higher moments all go to zero. First, we calculate what parameter values we expect for the Gaussian distribution of the means. The parameters of the exponential distribution are given by:

$$\mu = E[x] = \int_0^\infty x \cdot \lambda e^{-\lambda x} = \frac{1}{\lambda}$$

$$E[x^2] = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} = \frac{2}{\lambda^2}$$

$$\sigma_x = \frac{1}{\lambda}$$

From the lecture script we get the probability distribution of the measured means z:

$$P(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \cdot e^{-\frac{1}{2}\frac{(z-\mu)^2}{\sigma_z^2}}$$

Which is a Gaussian, where μ is the expected mean of the distribution. From the script we know that the expected mean is the same as the mean of our original exponential distribution, meaning $\mu = \frac{1}{\lambda}$. And $\sigma_z = \sigma_x/\sqrt{n}$, meaning that the sigma of the means is proportional to the sigma of the original distribution and scales with $1/\sqrt{n}$. Therefore, we expect $\mu = \frac{1}{\lambda}$ and $\sigma_z = \frac{1}{\lambda\sqrt{n}}$ as the parameters of our Gaussian.

We now simulate 100000 experiments for different choices of λ and sample size n. The sample size n is the number of samples that each experiment contains and from which the mean for that experiment is calculated. We choose three values for the parameter $\lambda = 0.01, 1, 100$. For each choice of λ we choose three different sample sizes n = 5,10,100. The random numbers are generated according to the exponential distribution using:

$$x = -\frac{ln(U)}{\lambda}$$

Where U is a uniformly distributed random number between (0,1]. The resulting distributions of the means are shown in the following Figures 6, 7 and 8.

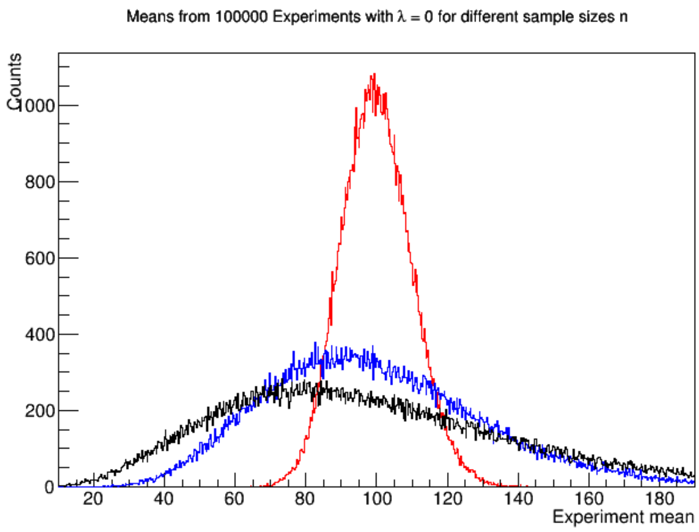

Figure 6: The distribution of the means from 100000 experiments with $\lambda = 0.01$ for the sample sizes n = 5 in black, n = 10 in blue and n = 100 in red.

The distribution of the means is closer to a Gaussian distribution for larger sample sizes. For n=5,10 the distributions are still a bit anti-symmetrical. The parameters for these distributions are of the same order of magnitude, but still differ significantly. For example the mean is shifted by about 20%. For n=100 the distribution is basically a Gaussian with parameter values very close

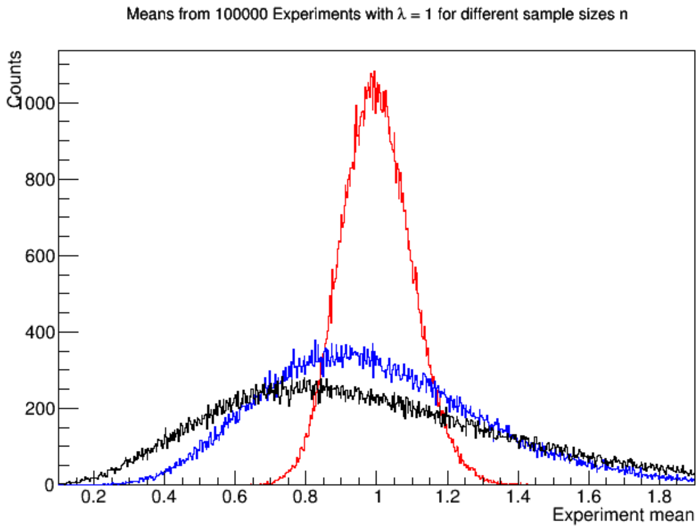

Figure 7: The distribution of the means from 100000 experiments with $\lambda = 1$ for the sample sizes n = 5 in black, n = 10 in blue and n = 100 in red.

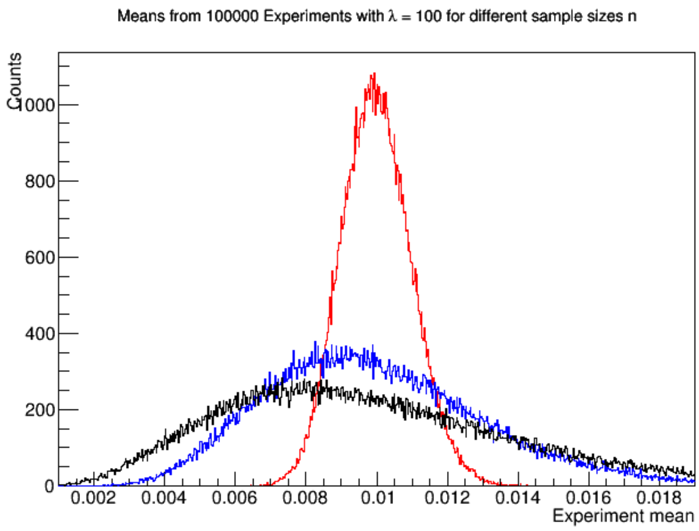

Figure 8: The distribution of the means from 100000 experiments with $\lambda = 100$ for the sample sizes n = 5 in black, n = 10 in blue and n = 100 in red.

to the expected ones. The error on these is smaller than 1%. The expectation for the means is $1/\lambda$. For n=5,10 this is an approximation, for n=100 the value is quite accurate. Also, the sigmas are smaller for larger n's, meaning that the standard deviations are smaller for larger sample sizes, resulting in narrower distributions. The number of experiments only influences the smoothness of the Gaussian, and not its parameters, the more experiments the smoother the Gaussian. Only for large enough n's the distribution has a purely Gaussian shape, since the suppression of the higher moments is greater for large n's. In conclusion, the sample size has to be large enough for the CLT to be applicable.

b) In this part of the problem, we take a look at the Cauchy distribution, which does not satisfy the conditions for the CLT. The Cauchy distribution is described by the function:

$$f(x) = \frac{1}{\pi\gamma} \cdot \frac{\gamma^2}{(x - x_0)^2 + \gamma^2}$$

Where γ and x_0 are parameters. In this problem we use $\gamma = 3$ and $x_0 = 25$. The reason why the CLT can not be applied for the Cauchy distribution is that its higher moments are not finite. From the 2nd moment and higher, the moments are infinite and can not be neglected. The integrand of the nth moment is of the order of x^{n-2} and the integrals are infinite and are not compensated by large n's. Therefore, the CLT does not apply, and the distribution of the means is not a Gaussian. The Cauchy distribution is plotted in Fig. 9.

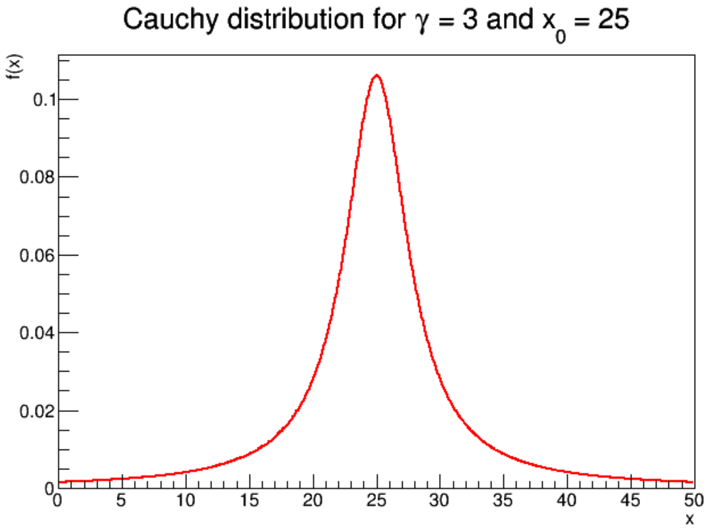

Figure 9: The Cauchy distribution for $\gamma = 3$ and $x_0 = 25$ plotted from 0 to 50.

Now we perform 10000 experiments with a sample size of n=100 in each experiment and plot the distribution of the mean. Random numbers can be generated from the Cauchy distribution via the formula:

$$x = \gamma \cdot \tan(\pi U - \pi/2) + x_0$$

Where U is a uniformly distributed random number between (0,1]. The distribution of the mean is plotted in Fig.

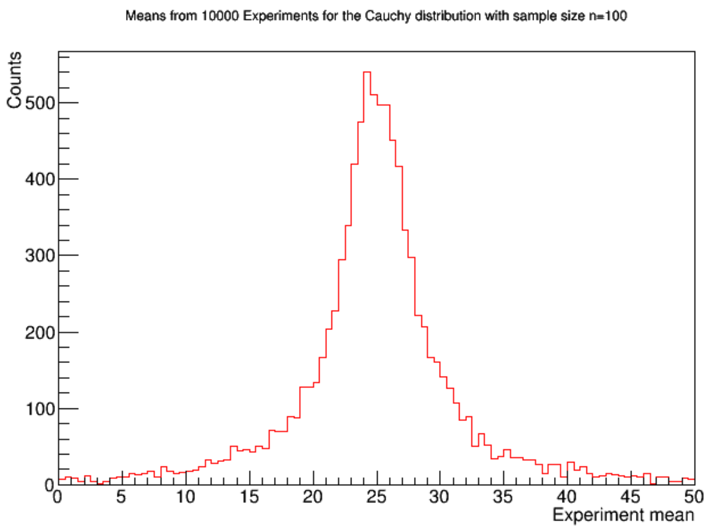

Figure 10: Means from 10000 experiments sampled from the Cauchy distribution for a sample size of n=100.

As we expected, the means are distributed according to the Cauchy distribution. At the top, the Cauchy distribution is generally narrower than the Gaussian. However, the tails of the Gaussian tend towards zero faster than the tails of the Cauchy distribution.

4.2 Problem 11

In this problem we plot the contours of the bivariate Gauss function for different parameters. The bivariate Gauss function is given by:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp[-\frac{1}{2(1-\rho^2)} \cdot (\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y})]$$

With $\mu_{x,y}$ being the means of the variables, $\sigma_{x,y}$ being the sigmas of the variables and ρ being the correlation coefficient, defined by:

$$\rho_{xy} = \frac{cov[x,y]}{\sigma_x\sigma_y}$$

With

$$cov[x,y] = E[xy]_{P(x,y)} - E[x]_{P_x(x)}E[y]_{P_y(y)}$$

Where the index of the expectation value indicates the probability density used when calculating the expectation value. $P_{x,y}$ indicate the marginalized distributions, which are calculated by integrating the distribution depending on all variables over the unwanted variables. For example:

$$P_x(x) = \int P(x,y)dy$$

a) The used parameters are: $\mu_x = 0, \quad \mu_y = 0, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho_{xy} = 0$. The resulting plot of the contours is shown in Fig. 11.

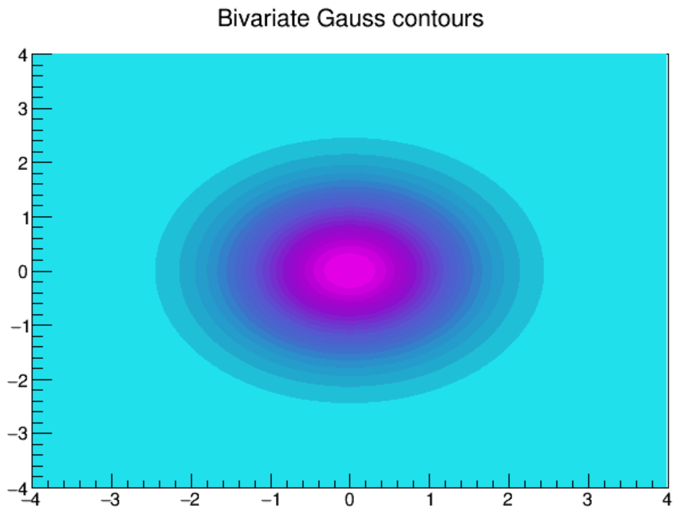

Figure 11: Contour plot of a bivariate Gaussian.

b) The used parameters are: $\mu_x = 1, \quad \mu_y = 2, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho_{xy} = 0.7$. The resulting plot of the contours is shown in Fig. 12.

c) The used parameters are: $\mu_x = 1, \quad \mu_y = -2, \quad \sigma_x = 1, \quad \sigma_y = 2, \quad \rho_{xy} = -0.7$. The resulting plot of the contours is shown in Fig. 13.

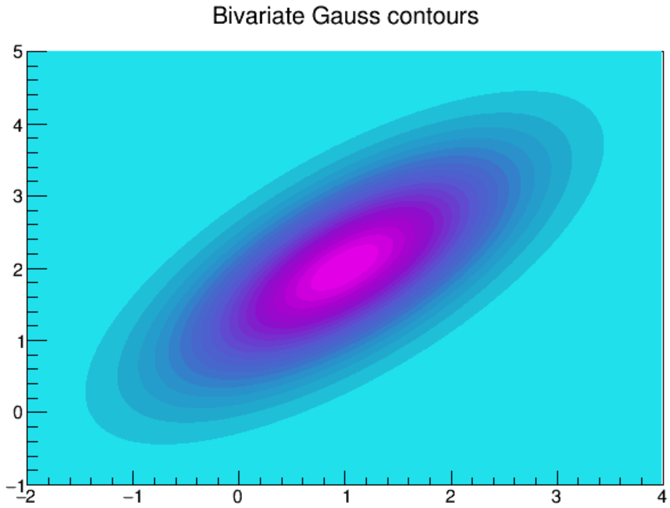

Figure 12: Contour plot of a bivariate Gaussian.

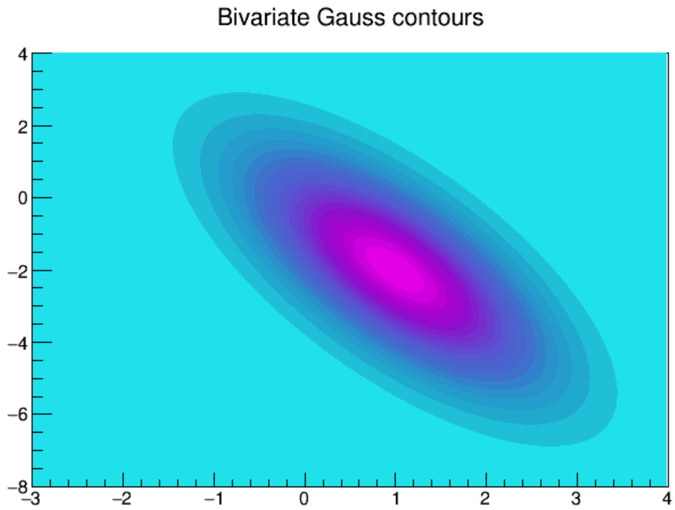

Figure 13: Contour plot of a bivariate Gaussian.

4.3 Problem 12

a) In this part we show that the bivariate Gauss probability distribution can be written as:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp[-\frac{1}{2(1-\rho^2)} \cdot (\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y})]$$

We start from the general formula for the multivariate Gauss distribution given in the lecture script:

$$P(x_1,...,x_n) = \frac{1}{(2\pi)^{n/2} \mid \Sigma \mid^{1/2}} \cdot \exp[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})]$$

Where Σ is the covariance matrix defined by:

$$\Sigma_{ij} = cov[x_i, x_j]$$

And $\mid \Sigma \mid$ is the determinant of the matrix. In the case of two variables we have: $n = 2, \quad x_1 = x, \quad x_2 = y$. Since there are no μ's in the end-formula, we can already set $\mu_x = \mu_y = 0$. We start by calculating the elements of the covariance matrix:

$$\Sigma_{xx} = cov[x,x] = Var[x] = \sigma_x^2, \quad \Sigma_{yy} = \sigma_y^2, \quad \Sigma_{xy} = \Sigma_{yx} = cov[x,y]$$

With $cov[x,y] = \rho\sigma_x\sigma_y$. The determinant is given by:

$$\mid \Sigma \mid = \sigma_x^2\sigma_y^2 - \rho^2\sigma_x^2\sigma_y^2$$

$$\mid \Sigma \mid^{1/2} = \sigma_x\sigma_y\sqrt{(1-\rho^2)}$$

The inverse of the matrix is:

$$\Sigma^{-1} = \frac{1}{\mid \Sigma \mid}\begin{pmatrix} \sigma_y^2 & -cov[x,y] \\ -cov[x,y] & \sigma_x^2 \end{pmatrix}$$

Multiplying the matrix with the vectors:

$$\frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} \sigma_y^2 & -cov[x,y] \\ -cov[x,y] & \sigma_x^2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x^2\sigma_y^2 + y^2\sigma_x^2 - 2xy\cdot cov[x,y]}{\sigma_x^2\sigma_y^2(1-\rho^2)} = \frac{1}{(1-\rho^2)}\cdot(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2xy\rho}{\sigma_x\sigma_y})$$

Now we plug all the parts into the general formula and are left with:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\cdot\exp[-\frac{1}{2(1-\rho^2)}\cdot(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2\rho xy}{\sigma_x\sigma_y})] \quad q.e.d.$$

b) To get the probability distribution for z, we first insert y=x-z into the general formula. Then we have to marginalize in x, meaning we integrate over x, to be left with a distribution in z:

$$P(z) = \int P(x, x-z)dx$$

After the insertion and the integration(many steps of algebra), we are left with:

$$\frac{1}{\sqrt{2\pi}\sigma_z}\exp[\frac{z*^2}{2\sigma_y^2(1-\rho^2)}]exp[\frac{-z*^2(\sigma_x+\rho\sigma_y)^2}{2\sigma_y^2\sigma_z^2(1-\rho^2)}]$$

With z*=z-μ_z. This is a Gaussian distribution for z. If we set $\rho = 0$:

$$P(z) = \frac{1}{\sqrt{2\pi}\sigma_z}\exp[\frac{-(z-\mu_z)^2}{2\sigma_z^2}]; \quad q.e.d.$$

4.4 Problem 13

In this problem we investigate the convolution of two Gaussians. If we have two variables x and y, that follow some true distributions, and we measure the distribution of y given x, we know that this measured distribution is the convolution of the true distributions. In this problem we only know the true distribution for x, and the measured distribution for y given x, which are both Gaussian. Since the convolution of two Gaussians is also a Gaussian, we know that the true distribution for y must be Gaussian. So to find the y-distribution we use that we know that:

$$P(y \mid x) = g(y) \star f(x)$$

To make things easier we make use of the Fourier space and the convolution theorem:

$$\mathcal{F}(P(y \mid x)) = \mathcal{F}(g \star f) = (2\pi^{n/2}\cdot\mathcal{F}(g)\cdot\mathcal{F}(f)$$

Since $P(y \mid x)$ is two-dimensional we use $n = 2$. To find g(y) we first calculate the Fourier transforms of f(x) and $P(y \mid x)$, then solve for $\mathcal{F}(g)$ and perform the inverse Fourier transform to finally get g(y). The (inverse) Fourier transforms can be done by hand or by using Wolframalpha for instance.

$$(\mathcal{F}(f))(k) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{\sigma_x^2 k^2}{2} + ix_0 k]$$

For $P(y \mid x)$ we need to perform a 2-dimensional Fourier transform since it depends on the two variables x and y.

$$(\mathcal{F}(P))(k,l) = \frac{1}{\sqrt{2\pi}} \cdot (\mathcal{F}\{\exp[\frac{-\sigma_y^2 l^2}{2} + ixl]\})(k) = \frac{1}{\sqrt{2\pi}} \cdot \exp(\frac{-\sigma_y^2 l^2}{2}) \cdot \mathcal{F}(e^{ixl}) = \exp(\frac{-\sigma_y^2 l^2}{2}) \cdot \delta(l+k)$$

The resulting equation we get is:

$$\sqrt{2\pi} \cdot \mathcal{F}(g) \cdot \exp(\frac{-\sigma_x^2 k^2}{2} + ix_0 k) = \exp(\frac{-\sigma_y^2 l^2}{2}) \cdot \delta(l+k)$$

Solving for the Fourier transform of g and setting k to -l because of the delta function:

$$\mathcal{F}(g) = \frac{1}{2\pi} \cdot \exp(\frac{l^2}{2} * (\sigma_x^2 - \sigma_y^2) + ix_0 l)$$

Doing the inverse Fourier transform and rearranging we are left with:

$$g(y) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{\sigma_y^2 - \sigma_x^2}} \cdot \exp(-\frac{(y-x_0)^2}{2(\sigma_y^2 - \sigma_x^2)})$$

As expected we receive a Gaussian distribution for the variable y, with the mean x_0 and a sigma of $\sigma = \sqrt{\sigma_y^2 - \sigma_x^2}$.

4.5 Problem 14

a) The likelihood probability distributions for the cross sections are assumed to be Gaussians with the variable $\sigma(\Theta) = A + B\cos(\Theta) + C\cos(\Theta^2)$, the mean σ_i, which is the measured cross section for a certain angle, and the corresponding error e_i. The likelihood for each cross section is therefore:

$$P_{prior,i}(\sigma_i, \sigma(\Theta) \mid A, B, C) = Gauss[\sigma(\Theta, A, B, C), \sigma_i, e_i]$$

Since we assume flat priors, we can cancel these in the posterior probability distribution. The posterior is a product of the likelihoods for the different measurements, that is normalized with respect to the parameters A,B and C. The posterior is therefore:

$$P_{post}(A, B, C \mid \sigma_i, \sigma(\Theta_i)) = \frac{\prod_i Gauss[\sigma(\Theta_i, A, B, C), \sigma_i, e_i]}{\int \prod_i Gauss[\sigma(\Theta, A, B, C), \sigma_i, e_i] dAdBdC}$$

b) To find the parameter values that maximize the posterior distribution, one has to marginalize the posterior and find the maximum of the marginalized distribution for each parameter. This can be done by hand, which is quite cumbersome. Therefore, we use the Bayesian Analysis Toolkit (BAT), which is a software package, that uses Markov Chain Monte Carlo and can be used together with ROOT. BAT performs the marginalization and maximization for us and returns the parameter values at the mode of the posterior. BAT produces plots of the marginalized probability distributions which can be found in the Figures 14, 15 and 16.

At the mode of the posterior we find the parameter values to be: A=15.36, B=-1.08, C=-2.57.

5 Chapter 5

5.1 Problem 1

In this Problem we fit a sigmoid function to data points and try to determine the parameters. The function we are using is:

$$\epsilon(E \mid A, E_0) = \frac{1}{1 + \exp(-A(E - E_0))}$$

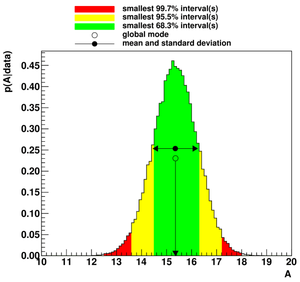

Figure 14: The marginalized probability distribution for the parameter A.

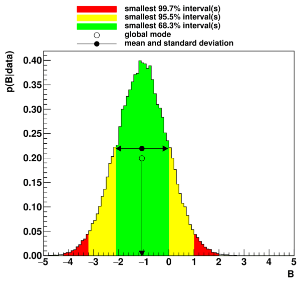

Figure 15: The marginalized probability distribution for the parameter B.

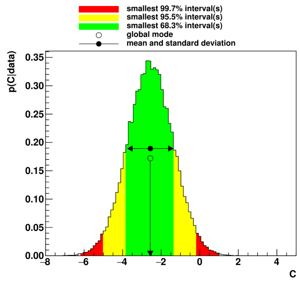

Figure 16: The marginalized probability distribution for the parameter B.

a) Since we are counting the number of successes r amongst N trials, we use the binomial distribution as the likelihood, with the probability parameter p as ϵ. As priors for the parameters A, E_0 we use Gaussians with estimated means. The parameter E_0 is the energy where $p \approx 0.5$. Looking in the table we have p=0.55 for E=2.0, therefore we set the mean of the Gaussian for E_0 to be 2.0. When we increase/decrease the energy at 2.0 by 0.5, the efficiency changes by about 50%. We use the same reasoning as in the lecture script, and since we have similar data, we set the mean of the Gaussian prior for A to 3. The sigmas of the Gaussians are set to 0.5. The total likelihood for all energies is the product of the single binomials for each energy. Therefore, the posterior is given by:

$$P(A, E_0 \mid E) = \frac{\prod_i P(r, N, \epsilon(E_i \mid A, E_0)) \cdot Gauss(A, 3, 0.5) \cdot Gauss(E_0, 2, 0.5)}{\int \prod_i P(r, N, \epsilon(E_i \mid A, E_0)) \cdot Gauss(A, 3, 0.5) \cdot Gauss(E_0, 2, 0.5) dA dE_0}$$

Since this is not analytically solvable, we need to solve it numerically, for this end we use BAT. BAT produces the plots of the marginalized posterior probability distributions and a correlation plot, and returns the parameter values that are at the mode of the posterior. The posteriors can be found in Figures 17 and 18 and the correlation plot in Fig. 19.

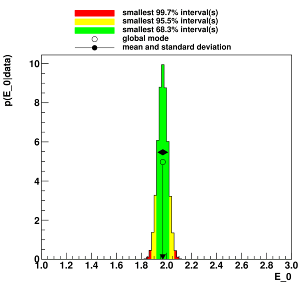

Figure 17: The marginalized probability distribution for the parameter E_0.

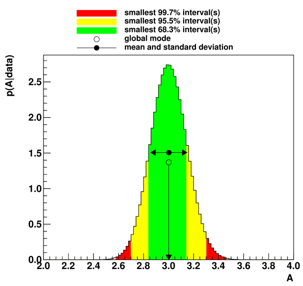

Figure 18: The marginalized probability distribution for the parameter A.

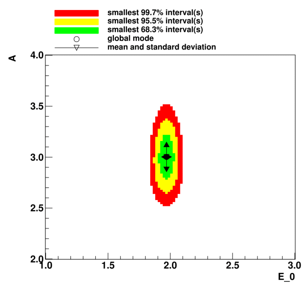

Figure 19: The correlation plot for the parameters A and E_0.

The optimal parameter values are $E_0 = 1.97$ and A=3.0.

b) When doing the frequentist analysis, it is suitable to define a test statistic, that is a scalar function of the parameters. Instead we make confidence levels for the test statistic, and from that in turn deduce confidence levels for the parameters. The detailed proceedings are described in the lecture script. In short, one creates a grid, with different A and E_0 values at each grid point. At each grid point we calculate the success probability according to the sigmoid model. For each energy, indexed k, random success values are generated according to the binomial distribution, with the probability being the sigmoid function. We generate these random numbers to simulate the probability distribution of the test statistic, instead of calculating the value of the test statistic and its distribution for each possible set of results, because there are too many possibilities. We then evaluate the test statistic, which we in this case define as:

$$\xi = \prod_{i=1}^{k} \binom{N_i}{r_i} \cdot \epsilon^{r_i} \cdot (1-\epsilon)^{N_i - r_i}$$

We sort these values in decreasing order, and generate enough data for our ensemble. Here we generate 100 experiments. We note the ξ value for which 68,90,95 % of our experiments are above this value. Finally, we check if the ξ^{data} is in our range, if it is then the values of A,E_0 are in our CL interval.

The result can be seen in Fig. 20.

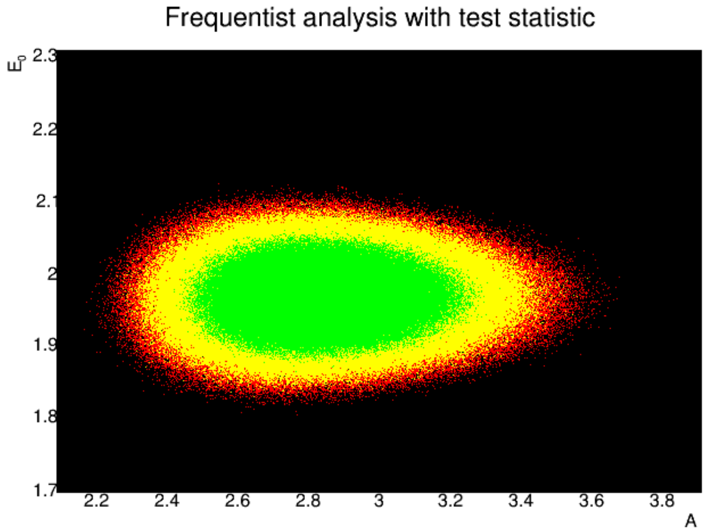

Figure 20: Frequentist analysis. The values of A,E_0 colored green, yellow and red respectively give the 68,90 and 95% CL intervals for the parameters.

As the 68% CL interval for the parameter A we get [2.418,3.371] and for E_0 we get [1.867,2.067].

5.2 Problem 2

In this Problem we try to fit the data from the previous problem with the function:

$$\epsilon = sin(A(E - E_0))$$

a) To find the posterior probability we again use BAT. The sine causes a problem, since it can return negative values, and a probability can not be negative. Therefore we have to exclude these cases in BAT. In the end, after a lot of playing around with the parameters, BAT does not crash, but still only returns values for the parameters as given in the Gaussian priors, which is equivalent to not getting any results. Therefore there is no point showing the plots here.

b) Also in the frequentist analysis, the negative values of the sine cause problems. After excluding all bad cases the following plot is produced in Fig. 21.

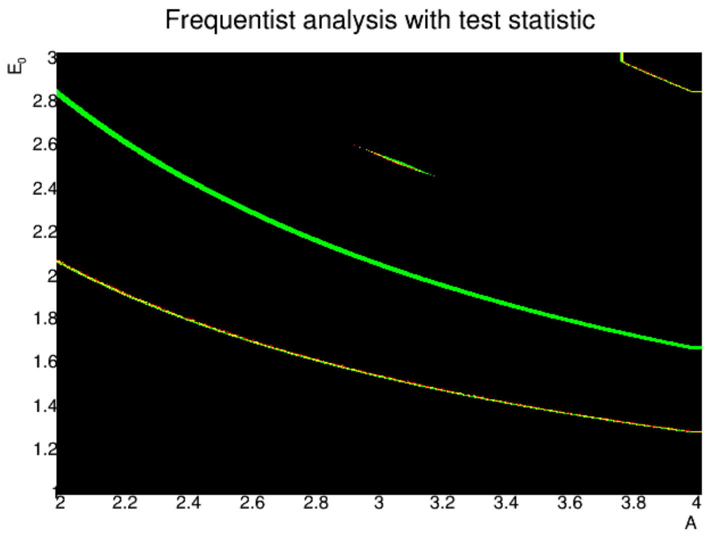

Figure 21: Frequentist analysis. Trying to fit a sine to a sigmoid function.

From this, no CL intervals can be produced for the parameters.

c) The main problem is that the sine can take on negative values, which a probability can not, and should not. This negative success probability causes further problems in the binomial distribution. To work around this problem one either has to exclude the negative values, in which case the whole plot is black, or take the absolute value if the sine is negative, which does not result in a meaningful plot. The conclusion is that the sine is a bad activation function and the sigmoid works very well as one.

5.3 Problem 3

In this problem we investigate the chi-squared distribution for one data point.

$$\chi^2 = \sum_i \frac{(y_i - f(x_i \mid \lambda))^2}{\omega_i^2}$$

If the measurement is Gauss distributed:

$$P(y) = G(y \mid f(x \mid \lambda), \sigma(x \mid \lambda))$$

With $\omega_i = \sigma_i$
For a single measurement and including positive and negative measurement values:

$$P(\chi^2)\frac{d\chi^2}{dy} = 2P(y)$$

We are left with:

$$P(\chi^2) = \frac{1}{\sqrt{2\pi\chi^2}} \exp(\frac{-\chi^2}{2})$$

The mode is $E[\chi^2] = 1$ and the variance is $Var[\chi^2] = 2$. The mode of the distribution is located at $\chi^2 = 0$, then the probability decreases on both sides.

5.4 Problem 8

In this problem we try to fit data with two different models.

a) The first model that we use is the background model:

$$f = A + Bx + Cx^2$$

Frequentist For the frequentist analysis we perform a chi-squared minimization fit using ROOT. The result and fit that we get are shown in Fig. 22.

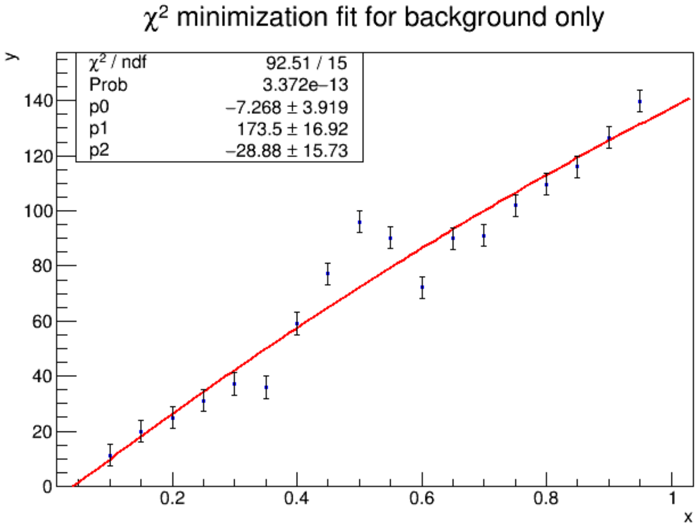

Figure 22: Chi-squared minimization fit for background model.

The parameter values can be seen in the Figure where A=p0,B=p1,C=p2. The p-value of the fit is marked Prob in the stats table. The covariance coefficients are given by ROOT in the covariance matrix. The relevant ones are: $\rho_{AB} = -0.924$; $\rho_{AC} = 0.839$; $\rho_{BC} = -0.977$.

Bayesian For the Bayesian analysis we use BAT. The posteriors are shown in Figures 23, 24 and 25 and the correlation plots are shown in Figures

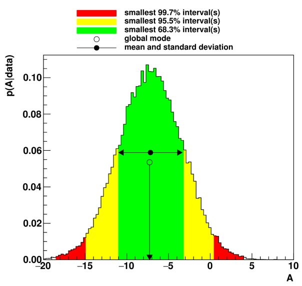

Figure 23: The marginalized probability distribution for the parameter A.

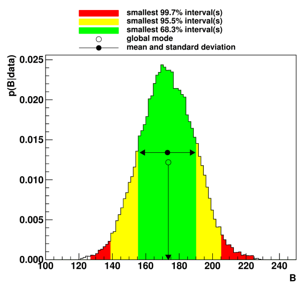

Figure 24: The marginalized probability distribution for the parameter B.

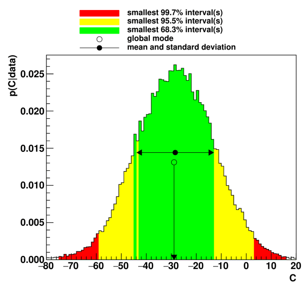

Figure 25: The marginalized probability distribution for the parameter C.

The values of the parameters and uncertainties that BAT returns are given by: $A = -7.27111 \pm 3.86593$; $B = 173.484 \pm 16.7845$; $C = -28.893 \pm 15.485$.

b) The second model used is the background + signal model:

$$f = A + Bx + Cx^2 + \frac{D}{(x - x_0)^2 + \Gamma^2}$$

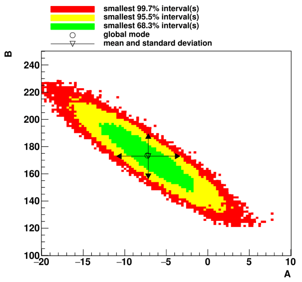

Figure 26: The correlation plot for the parameters A and B.

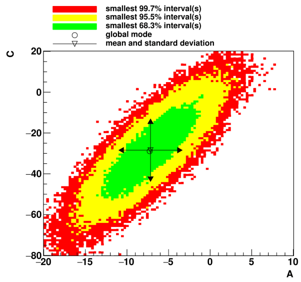

Figure 27: The correlation plot for the parameters A and C.

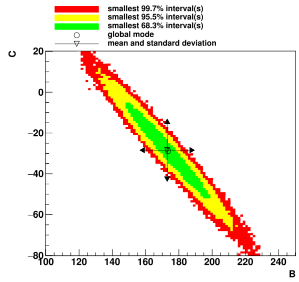

Figure 28: The correlation plot for the parameters B and C.

Frequentist The chi-squared minimization fit does not converge at first, therefore the parameters A,B and C are set to the values from the previous model. The fit converges and returns the following plot, which is to be seen in Fig. 29.

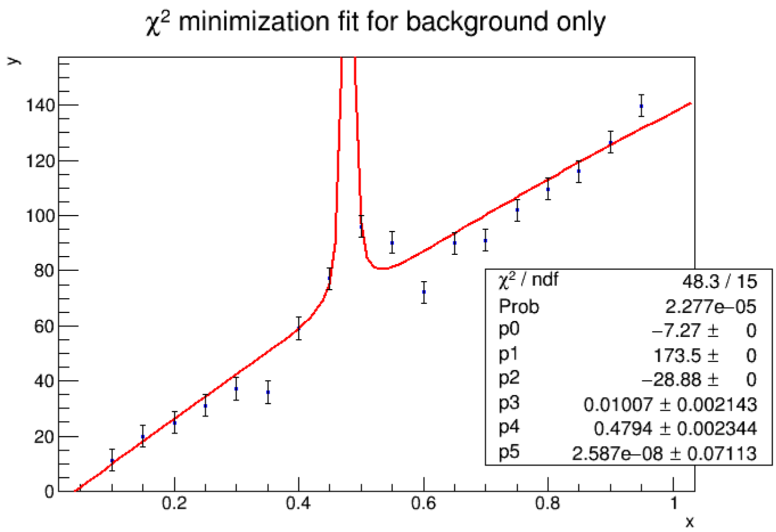

Figure 29: Chi-squared minimization fit for background + signal model.

This fit does not follow the data points at the peak very well. The parameters for this fit can be found in the plot, with p3=D,p4=x_0,p5=Γ. We can see that the peak should be around 0.5, and therefore we can set $x_0 = 0.5$ to ensure this. This produces the fit that can be seen in Fig. 30.

This time the peak is covered by the fit, nevertheless, the lower points around the peak are not quite well fitted in any of the plots. The covariance matrix elements that we get for x_0 not fixed are given by: $\rho_{Dx_0} = -0.707$; $\rho_{D\Gamma} = 0.025$; $\rho_{x_0\Gamma} = 0.012$. For the case where x_0 is fixed the relevant covariance matrix element is: $\rho_{D\Gamma} = 0.949$. The problem here could be that there are not enough points around the peak for the background + signal fit to properly converge.

Bayesian In the Bayesian case we use BAT and get a fit that can be seen in Fig. 31.

Just as in the frequentist fit, this fit does not quite converge and we get a similar plot as for the frequentist case. The fact that the fit is not quite correct can be seen in the error band, which does not follow the best fit. Again, this could be due to not enough data points around the peak. From this fit we get similar parameter values as for the frequentist, and since the fit is quite bad there is no point in reporting the exact values.

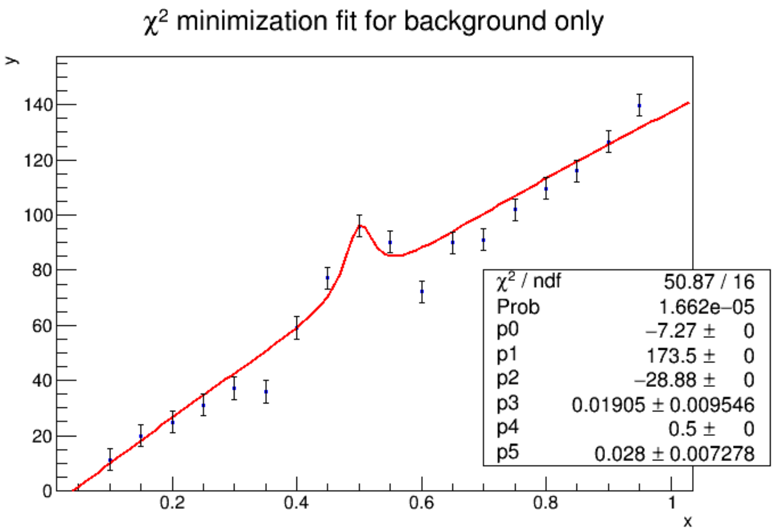

Figure 30: Chi-squared minimization fit for background + signal model with x_0 fixed.

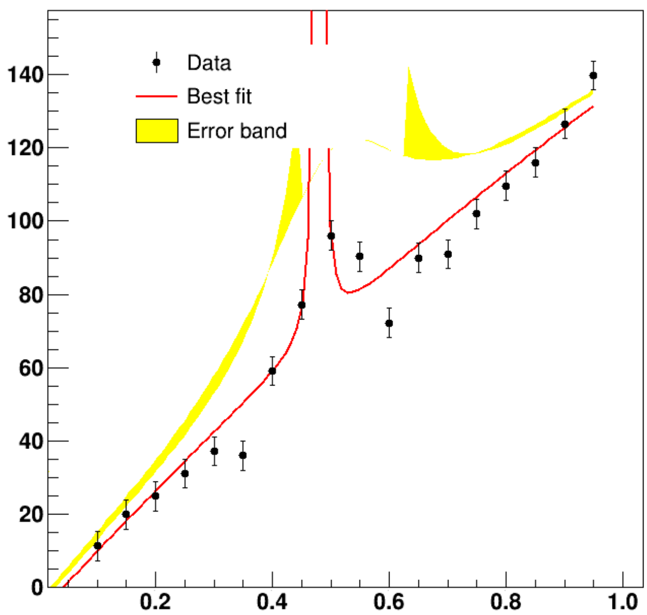

Figure 31: Fit using BAT.

References

[1] Beta distribution. `https://en.wikipedia.org/wiki/Beta_distribution#Characterization`. Accessed: 2018-03-16.

[2] Feynman-parameter. `https://de.wikipedia.org/wiki/Feynman-Parameter`. Accessed: 2018-02-24.