Report: Data Analysis, WS17/18

Korbinian Urban, Matr.: 03648366

April 5, 2018

Contents

0.1 Introduction . . . 2

1 Introduction to Probabilistic Reasoning **3**
1.1 Summary . . . 3
1.2 Exercises . . . 3
1.2.1 Exercise 1: Jane with two children . . . 3
1.2.2 Exercise 2: More definitions of the coin probability . . . 4
1.2.3 Exercise 3: Resolution and probabilities . . . 4
1.2.4 Exercise 4: Mongolian swamp fever . . . 4

2 Binomial and Multinomial Distribution **6**
2.1 Summary . . . 6
2.2 Exercises . . . 6
2.2.1 Exercise 8: Mean and standard deviation . . . 6
2.2.2 Exercise 10 . . . 7
2.2.3 Exercise 11 . . . 8
2.2.4 Exercise 13: Reuse of data . . . 9

3 Poisson Distribution **11**
3.1 Summary . . . 11
3.2 Exercises . . . 11
3.2.1 Exercise 4: Mean and standard deviation . . . 11
3.2.2 Exercise 7: Poisson with Bayes and Neyman . . . 12
3.2.3 Exercise 8: Background and Feldman . . . 13
3.2.4 Exercise 13 . . . 15
3.2.5 Exercise 16 . . . 16

4 Gaussian Probability Distribution Function **17**
4.1 Summary . . . 17
4.2 Exercises . . . 17
4.2.1 Exercise 8: Central Limit Theorem . . . 17
4.2.2 Exercise 11: Plotting bivariate Gauss . . . 19
4.2.3 Exercise 12: Bivariate Gauss probability distribution . . . 19
4.2.4 Exercise 13: Convolution of Gauss . . . 21
4.2.5 Exercise 14: Probability of three parameters . . . 22

5 Model Fitting and Model Selection **25**
5.1 Summary . . . 25
5.2 Exercises . . . 25
5.2.1 Exercise 1: Fit of the Sigmoid Function . . . 25
5.2.2 Exercise 2: Fit of another Function . . . 28

5.2.3 Exercise 3: χ^2 distribution for one data point 29
5.2.4 Exercise 8: Fit with many parameters 30

0.1 Introduction

This report is summarizing the content of the lecture "Data Analysis" in wintersemester 2017/18 by Prof. Alan Cadwell. To every chapter of the lecture, there are executed some exercises about the content of the chapter. The first chapter introduces probabilistic reasoning and gives some important approaches like Bayes theorem. Chapters 2 to 4 are about the most important probability distributions, which are binomial, Poisson and Gauss. To analyze data following these distributions Bayes approach and the frequentists approach are used. Chapter 5 is about how to get not the distribution parameters, but model parameters out of measured data. At the end there is a short addition about the maximum likelihood estimator.

Chapter 1

Introduction to Probabilistic Reasoning

1.1 Summary

In this section, the term probability was introduced. Probability is no absolute value, but one can define the conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \tag{1.1}$$

This is known as **Bayes Theorem**. There is also the law of total probability:

$$P(B) = \sum_{i=1}^{N} P(B|A_i)P(A_i) \qquad \text{where} \qquad A_i \cap A_j = \emptyset \quad i \neq j, \quad \sum_{i=0}^{N} A_i = S \tag{1.2}$$

This can be written as the **Bayes-Lapace Theorem**:

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)} \tag{1.3}$$

1.2 Exercises

1.2.1 Exercise 1: Jane with two children

You meet Jane on the street. She tells you she has two children, and has pictures of them in her pocket. She pulls out one picture, and shows it to you. It is a girl. What is the probability that the second child is also a girl?

Variation: Jane takes out both pictures, looks at them, and is required to show you a picture of a girl if she has one. What is now the probability that the second child is also a girl ?

Solution Notation: $P(BG)$ is the probability to get a boy on the first picture and a girl on the second. Assumption: $P(GG) = P(BG) = P(GB) = P(BB) = \frac{1}{2}$. Using (1.1) one gets:

$$P(GG|G?) = \frac{P(G?|GG)P(GG)}{P(G?)} = \frac{1 \cdot 0.25}{0.5} = \frac{1}{2}$$

The probability to get a girl on the second picture if you know that the first picture is a girl is 50%.

Variation:

$$P(GG|\bar{B}\bar{B})) = \frac{P(\bar{B}\bar{B}|GG)P(GG)}{P(\bar{B}\bar{B})} = \frac{1 \cdot 0.25}{0.75} = \frac{1}{3}$$

Now the probability is lower compared to the first case.

1.2.2 Exercise 2: More definitions of the coin probability

Go back to section 1.2.3 and come up with more possible definitions for the probability of the data.

Solution We could for example also define the probability of the following events:

- Get a certain number of switches from T to H
- Get no sequences longer than three throws with the same side
- ...

1.2.3 Exercise 3: Resolution and probabilities

Your particle detector measures energies with a resolution of 10 %. You measure an energy, call it E. What probabilities would you assign to possible true values of the energy ? What can your conclusion depend on?

Solution With the given information you can not say a lot about the true value of E. An unknown property is the distribution of the measured values. It is not clear what the 10% means, e.g. the FWHM of a Gauss distribution.

1.2.4 Exercise 4: Mongolian swamp fever

Mongolian swamp fever is such a rare disease that a doctor only expects to meet it once every 10000 patients. It always produces spots and acute lethargy in a patient; usually (I.e., 60 % of cases) they suffer from a raging thirst, and occasionally (20 % of cases) from violent sneezes. These symptoms can arise from other causes: specifically, of patients that do not have the disease: 3 % have spots, 10 % are lethargic, 2 % are thirsty and 5 % complain of sneezing. These four probabilities are independent. What is your probability of having Mogolian swamp fever if you go to the doctor with all or with any three out of four of these symptoms ?

Solution Notation: s: spots, l: lethargy, t: thirst, v: violent sneezes, F: Swamp fever. Given probabilities:

$$\begin{aligned}
P(F) =& \ 0.0001 \\
P(\bar{F}) =& \ 0.9999 \\
P(s|F) =& \ 1 \\
P(l|F) =& \ 1 \\
P(t|F) =& \ 0.6 \\
P(v|F) =& \ 0.2 \\
P(s|\bar{F}) =& \ 0.03 \\
P(l|\bar{F}) =& \ 0.1 \\
P(t|\bar{F}) =& \ 0.02 \\
P(v|\bar{F}) =& \ 0.05
\end{aligned}$$

The probabilities of the symptoms are independent, which means:

$$P(sltv|X) = P(s|X)P(l|X)P(t|X)P(v|X)$$

Using the Bayes-Laplace Theorem (1.3) one gets:

$$P(F|sltv) = \frac{P(sltv|F)P(F)}{P(sltv|\bar{F})P(\bar{F}) + P(sltv|F)P(F)} = \frac{1 \cdot 1 \cdot 0.6 \cdot 0.2 \cdot 0.0001}{0.03 \cdot 0.1 \cdot 0.02 \cdot 0.05 + 1 \cdot 1 \cdot 0.6 \cdot 0.2 \cdot 0.0001} = 0.8$$

The probability of having swamp fever if you have all four symptoms is 80%.

$$P(F|\bar{s}ltv) = \frac{P(\bar{s}ltv|F)P(F)}{P(\bar{s}ltv|\bar{F})P(\bar{F}) + P(\bar{s}ltv|F)P(F)} = 0 \quad \text{as} \quad P(\bar{s}|F) = 0$$

The probability of having swamp fever is you have lethargy, thirst and violent sneezes but no spots is zero.

$$P(F|s\bar{l}tv) = \frac{P(s\bar{l}tv|F)P(F)}{P(s\bar{l}tv|\bar{F})P(\bar{F}) + P(s\bar{l}tv|F)P(F)} = 0 \quad \text{as} \quad P(\bar{l}|F) = 0$$

The probability of having swamp fever is you have spots, thirst and violent sneezes but no lethargy is zero.

$$P(F|sl\bar{t}v) = \frac{P(sl\bar{t}v|F)P(F)}{P(sl\bar{t}v|\bar{F})P(\bar{F}) + P(sl\bar{t}v|F)P(F)} = \frac{1 \cdot 1 \cdot 0.4 \cdot 0.2 \cdot 0.0001}{0.03 \cdot 0.1 \cdot 0.98 \cdot 0.05 + 1 \cdot 1 \cdot 0.4 \cdot 0.2 \cdot 0.0001} = 0.052$$

The probability of having swamp fever is you have spots, lethargy and violent sneezes but no thirst is 5.2%.

$$P(F|slt\bar{v}) = \frac{P(slt\bar{v}|F)P(F)}{P(slt\bar{v}|\bar{F})P(\bar{F}) + P(slt\bar{v}|F)P(F)} = \frac{1 \cdot 1 \cdot 0.6 \cdot 0.8 \cdot 0.0001}{0.03 \cdot 0.1 \cdot 0.02 \cdot 0.95 + 1 \cdot 1 \cdot 0.6 \cdot 0.8 \cdot 0.0001} = 0.46$$

The probability of having swamp fever is you have spots, lethargy and thirst but no violent sneezes is 46%.

Chapter 2

Binomial and Multinomial Distribution

2.1 Summary

The **binomial distribution** is given by:

$$P(r|N,p) = \binom{N}{r} p^r (1-p)^{N-r} \qquad \binom{N}{r} = \frac{N!}{r!(n-r)!} \tag{2.1}$$

Where N is the fixed number of trials, p the fixed success probability and r the number of successes.
Some properties:
Expectation value:

$$E[r] = \sum_{r=0}^{N} rP(r|N,p) = Np \tag{2.2}$$

Variance:

$$V[r] = E[(r-E[r])^2] = E[r^2] - (E[r])^2 = Np(1-p) \tag{2.3}$$

There are different intervals in probability distributions:
Central interval: All r except the ones who cover $\frac{\alpha}{2}$ from each end of the distribution.
Smallest interval: Start at mode, according to rank (sorted by P) of a member.
In this chapter, the Neyman Confidence Level Construction was introduced.

2.2 Exercises

2.2.1 Exercise 8: Mean and standard deviation

For the following function

$$P(x) = xe^{-x} \qquad 0 \leq x < \infty \tag{2.4}$$

(a) Find the mean and standard deviation. What is the probability content in the interval (mean-standard deviation, mean+standard deviation).

(b) Find the median and 68 % central interval

(c) Find the mode and 68 % smallest interval

Solution

(a) The mean can be found by the following equation:

$$x_{mean} = E[x] = \int_0^\infty x \cdot xe^{-x}dx = [(x^2 - 2x - 2)e^{-x}]_0^\infty = 2 \tag{2.5}$$

The standard derivation σ can by calculated by:

$$\sigma = \sqrt{E[x^2] - E[x]^2} = \sqrt{\int_0^\infty x^2 \cdot xe^{-x}dx - x_{mean}^2} = \sqrt{[-x^3 - 3x^2 - 6x - 6]_0^\infty - 4} = \sqrt{2} \tag{2.6}$$

The probability content in the interval ($x_{mean} - \sigma$,$x_{mean} + \sigma$) is 74%:

$$P(x_{mean} \pm \sigma) = \int_{2-\sqrt{2}}^{2+\sqrt{2}} xe^{-x}dx = 0.74 \tag{2.7}$$

See figure 2.1.

(b) The median can be calculated by:

$$\int_0^{x_{median}} xe^{-x}dx = 1 - (x_{median} + 1)e^{x_{median}} = 0.5 \tag{2.8}$$

This was solved numerically and the result is $x_{median} = 1.68$. The 68% central interval is calculated by:

$$\int_0^{x_1} xe^{-x}dx = (-x_1 - 1)e^{-x_1} + 1 = \frac{\alpha}{2} = 0.16 \tag{2.9}$$

$$\int_{x_2}^\infty xe^{-x}dx = (x_2 + 1)e^{-x_2} = \frac{\alpha}{2} = 0.16 \tag{2.10}$$

x_1 and x_2 were calculated numerically to $x_1 = 0.71$ and $x_2 = 3.29$. See figure 2.1.

(c) The mode is the maximum of $P(x)$: $x_{mode} = 1$ The 68% smallest interval was found numerically, by starting at the mode and adding $\pm dx$ to x_1 or x_2, in direction of the largest value of P until the is 0.68. The found limits are $x_1 = 0.27$ and $x_2 = 2.49$. See figure 2.1.

2.2.2 Exercise 10

Consider the data in the table: Starting with a flat prior for each energy, find an estimate for the efficiency (success parameter p) as well as an uncertainty. For the estimate of the parameter, take the mode of the posterior probability for p and use the smallest interval to find the 68% probability range. Make a plot of the result.

Solution To get a probability distribution of p, one has to use Bayes-Law 1.3. In our case it looks like:

$$P(p|N,r) = \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)dp} \tag{2.11}$$

And with a flat prior $P_0(p) = 1$ and the binomial distribution for $P(r|N,p)$ one gets:

$$P(p|N,r) = \frac{p^r(1-p)^{N-r}}{\int_0^1 p^r(1-p)^{N-r}dp} = \frac{(N+1)!}{r!(N-r)!}p^r(1-p)^{n-r} \tag{2.12}$$

This gives a probability distribution P(p) for each of the measured energies. To plot some p-E dependence, the mode and the 68 % smallest interval should be calculated for each energy. The mode is where $\frac{d}{dp}P(p|N,r)|_{p=p_{mode}} = 0$. One gets:

$$p_{mode} = \frac{r}{N} \tag{2.13}$$

The smallest interval was calculated numerically. The results are plotted in figure 2.2.

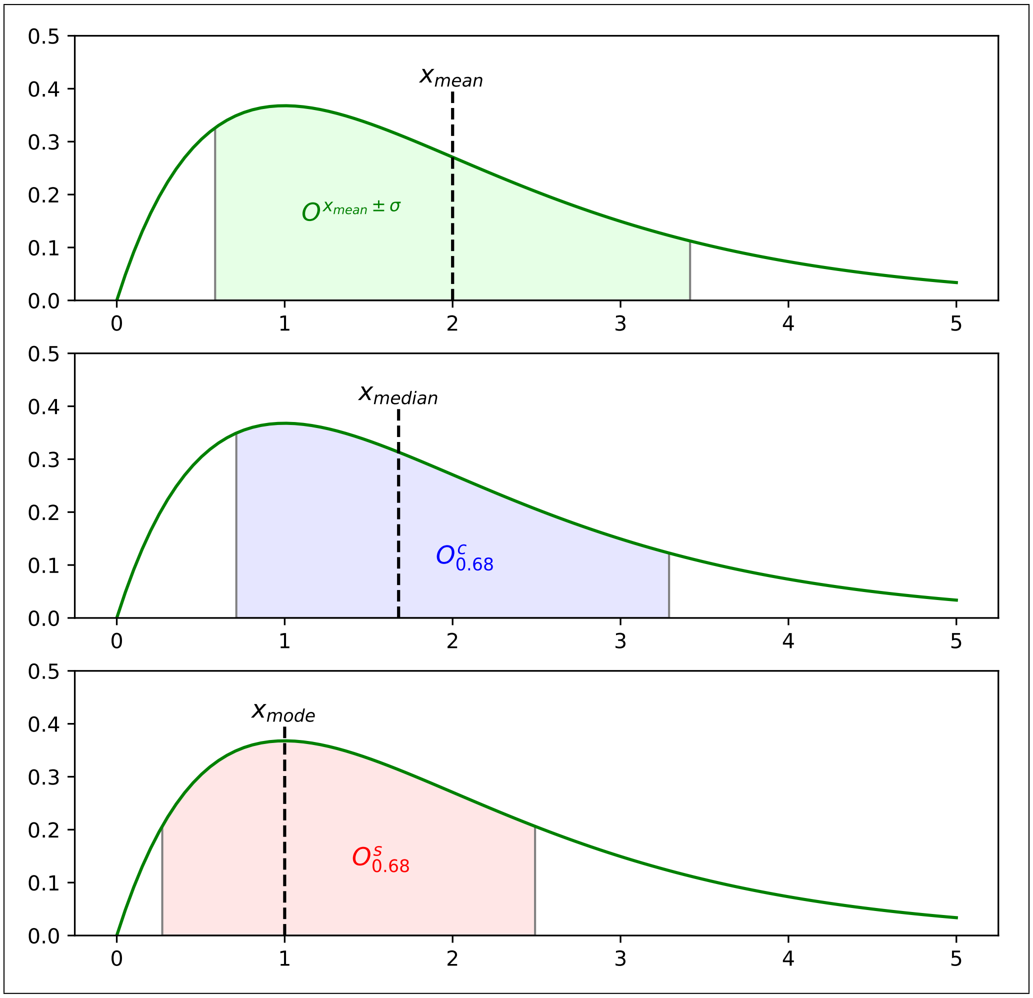


Figure 2.1: Distribition $P(x) = xe^{-x}$ with three different intervals

E	N	r
0.5	100	0
1.0	100	4
1.5	100	20
2.0	100	58
2.5	100	92
3.0	1000	987
3.5	1000	995
4.0	1000	998

Table 2.1: Measurement results for exercise 10, 11 and 13

2.2.3 Exercise 11

Analyze the data in the table from a frequentist perspective by finding the 90 % confidence level interval for p as a function of energy. Use the Central Interval to find the 90 % CL interval for p.

Solution This exercise was mainly solved numerically with python. First, the limits $r_{min}(p, N)$ and $r_{max}(p, N)$ of the 90% central interval of the binomial distribution were calculated. N was chosen to be 100 or 1000, as contained in the measured data table. Figure 2.3 shows the calculated ranges for $N = 100$. To discretize the p axis, a step of 10^{-3} was used.

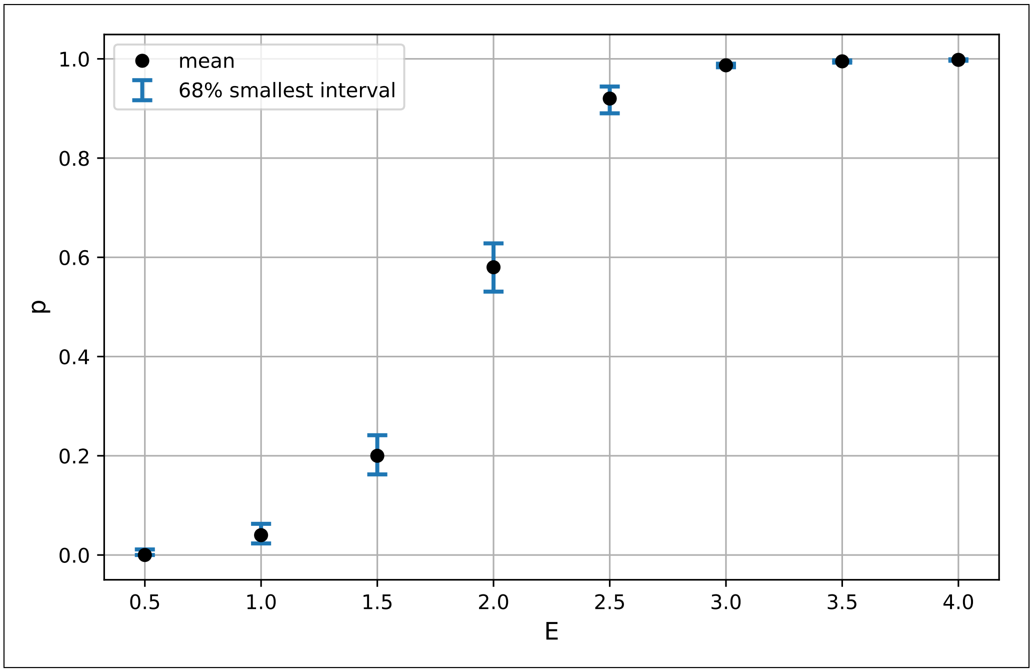


Figure 2.2: The mean and 68% smallest interval for the posterior probability distribution for the parameter p

E	p_{mode}	p_{min}	p_{max}
0.5	0.0	0.01122	0.0
1.0	0.02331	0.06286	0.04
1.5	0.16246	0.24144	0.2
2.0	0.53087	0.62809	0.58
2.5	0.89024	0.94424	0.92
3.0	0.98308	0.99026	0.987
3.5	0.9924	0.99693	0.995
4.0	0.99616	0.99913	0.998

Table 2.2: Exercise 10: Calculated values of the 68% smallest interval

In the last step this relation was evaluated for the measured r. This leads to a range of p, where r is in the 90% central interval, the co called 90% confidence level for p. The result for every energy is shown in figure 2.4.

2.2.4 Exercise 13: Reuse of data

Let us see what happens if we reuse the same data multiple times. We have N trials and measure r successes. Show that if you reuse the data n times, starting at first with a flat prior and then using the posterior from one use of the data as the prior for the next use, you get

$$P_n(p|N,r) = \frac{(nN+1)!}{nr!(nN-nr)!} p^{nr}(1-p)^{n(N-r)} \tag{2.14}$$

What are the expectation value and variance for p in the limit $n \to \infty$?

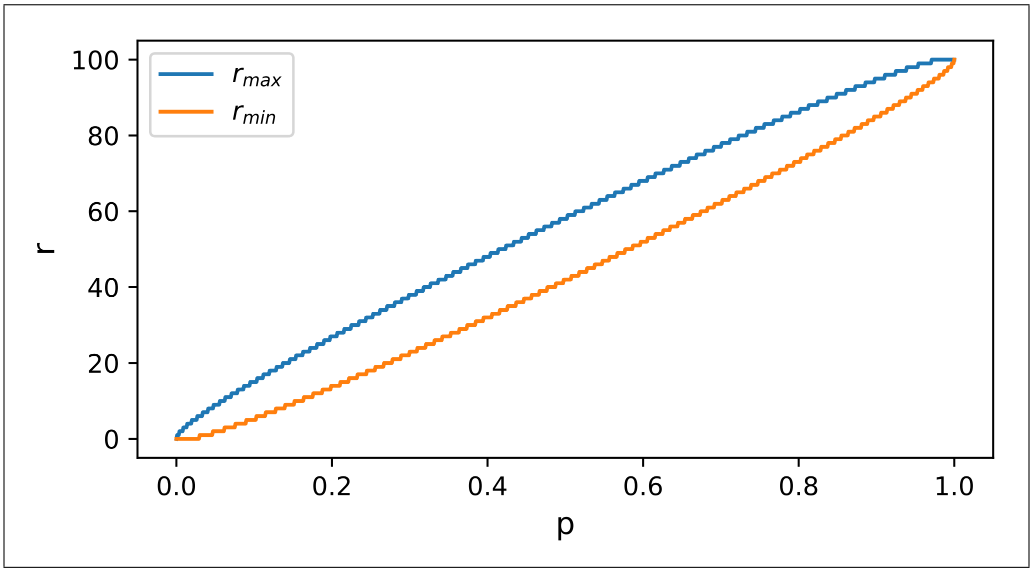


Figure 2.3: 90% central interval bands for N=100

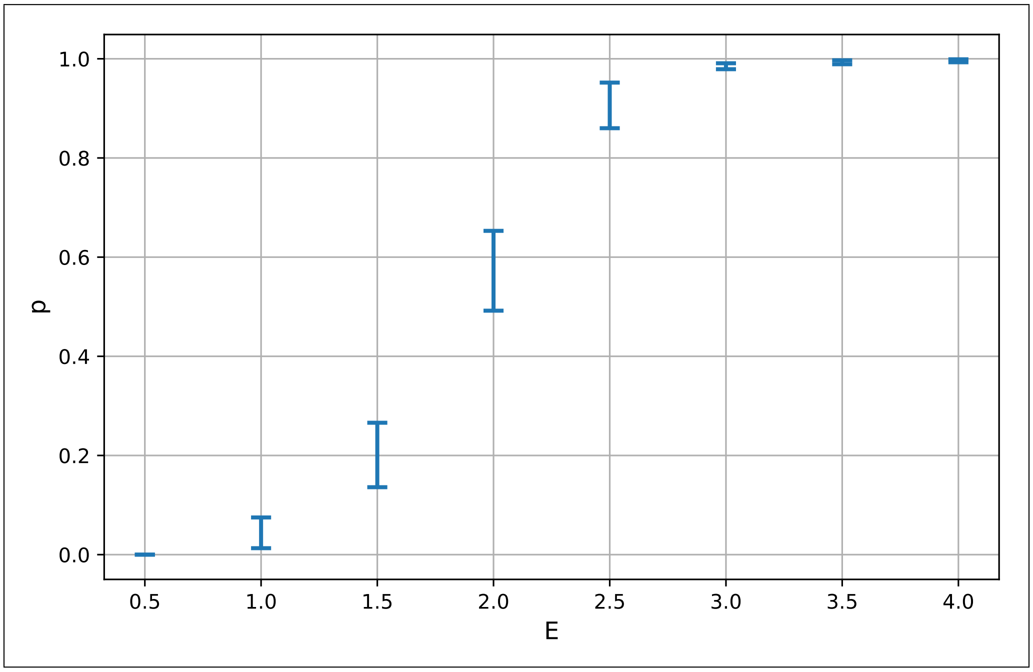


Figure 2.4: For every observed energie, the range of p, for which the result would fit in the 90% centrel interval is shown.

Chapter 3

Poisson Distribution

3.1 Summary

The **Poisson distribution** is used, if the number of trials is very large and unknown, and there occur only a few 'success' events. It can also be seen as the distribution of decays per time interval of a radioactive isotope. Within this interval there is a infinite amount of dt of which each has the same fixed probability to contain a decay. But the number of decays is finite.

The poisson distribution can be derived from the binomial distribution 2.1 if you do the limit for $N \to \infty$ and keep $\nu = Np$ finite. With a bit of cleverly limits one gets:

$$P(n|\nu) = \frac{e^{-\nu}\nu^n}{n!} \tag{3.1}$$

This distribution has the following properties:

Expectation value:

$$E[n] = \nu \tag{3.2}$$

Variance

$$V[n] = \nu \tag{3.3}$$

Mode

$$n^* = \lfloor \nu \rfloor \qquad \nu \notin Z \tag{3.4}$$

There is a double mode for $n \in Z$

3.2 Exercises

3.2.1 Exercise 4: Mean and standard deviation

Consider the function $f(x) = \frac{1}{2}e^{-|x|}$ for $-\infty < x < \infty$.

(a) Find the mean and standard deviation of x.

(b) Compare the standard deviation with the FWHM (Full Width at Half Maximum).

(c) What probability is contained in the ± 1 standard deviation interval around the peak ?

Solution The function is shown in figure 3.1.

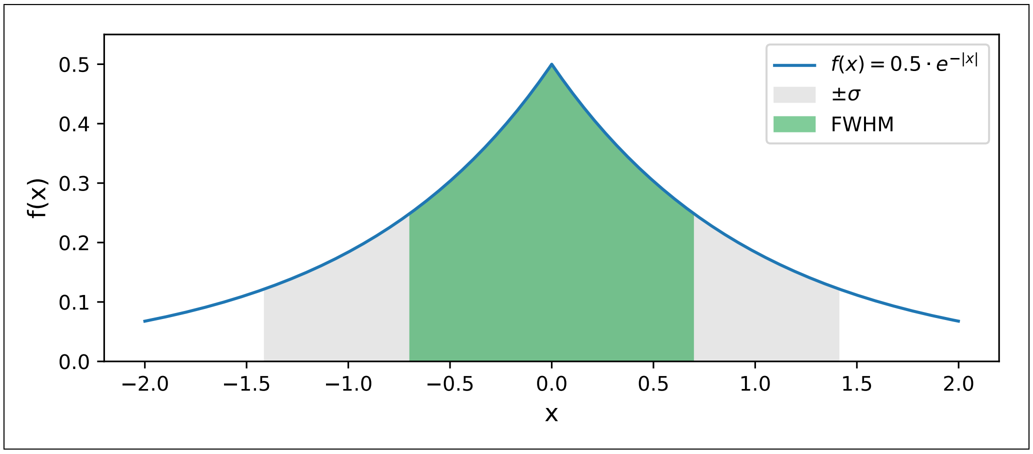


Figure 3.1: Probability distribution for exsercise 8

a) The mean can be found by:

$$E[x] = \int_{-\infty}^{\infty} x \cdot \frac{1}{2} e^{-|x|} dx = 0 \tag{3.5}$$

This can also be seen from the symmetry. For the standard deviation one gets

$$E[x^2] = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{2} e^{-|x|} dx = [-x^2 - 2x - 2]_0^{\infty} = 2 \tag{3.6}$$

$$\sigma = \sqrt{E[x^2] - E[x]^2} = \sqrt{2} \tag{3.7}$$

b) For the FWHM one hast to solve:

$$f(x_{1,2}) = \frac{1}{2} e^{-|x|} = \frac{1}{2} f(x_{mode}) = \frac{1}{4} \tag{3.8}$$

$$\Rightarrow \qquad x_{1,2} = \pm \ln 2 \qquad \Rightarrow \qquad FWHM = x_2 - x_1 = 2 \ln 2 \approx 1.39 \tag{3.9}$$

The interval $\pm\sigma$ is approximately twice as big as the FWHM. For a Gauss distribution the interval $\pm\sigma$ is only 1.18 times as big as the FWHM.

c)

$$I = \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{2} e^{-|x|} dx = \int_0^{\sqrt{2}} e^{-x} dx = 1 - e^{-\sqrt{2}} = 0.76 \tag{3.10}$$

The $\pm\sigma$ interval contains 76% of the distribution

3.2.2 Exercise 7: Poisson with Bayes and Neyman

9 events are observed in an experiment modeled with a Poisson probability distribution.

(a) What is the 95% probability lower limit on the Poisson expectation value ν ? Take a flat prior for your calculations.

(b) What is the 68% confidence level interval for ν using the Neyman construction and the smallest interval definition?

Solution

a) If you use a flat prior the posterior probability distribution for a poisson distribution is:

$$P(\nu|n) = \frac{e^{-\nu}\nu^n}{n!} \tag{3.11}$$

The content of the posterior distribution above the lower limit ν_l must contain 95%. Therefor ν_l can be calculated by:

$$\int_0^{\nu_l} \frac{e^{-\nu}\nu^n}{n!} d\nu = 0.05 \tag{3.12}$$

This was solved for $n = 9$ numerically to $\nu_l = 5.43$.

b) To get the 68% confidence level from the Neyman construction, one has to first get the 68% smallest interval for all possible values of ν. That is done by starting at the mode of equation 3.1 and finding numerically the smallest interval of n for each fixed ν. The step size for ν was chosen to be 0.02. The result is shown in figure 3.2.

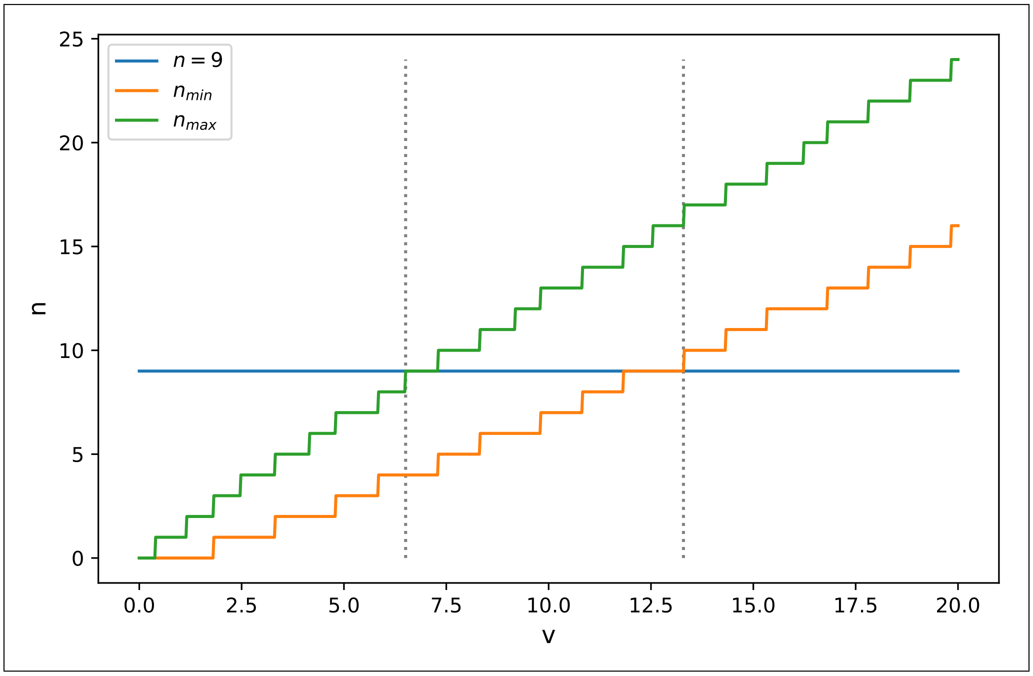


Figure 3.2: 68% smallest interval bands for exercise 7

To get a range for ν, one has now to extract all values of ν, for which $n = 9$ is contained in the 68% smallest interval. This is for $\nu \in [6.5; 13.3]$, as it can be seen from the dashed lines in figure 3.2.

3.2.3 Exercise 8: Background and Feldman

Repeat the previous exercise, assuming you had a known background of 3.2 events.

(a) Find the Feldman-Cousins 68% Confidence Level interval

(b) Find the Neyman 68% Confidence Level interval

(c) Find the 68% Credible interval for interval for ν

Solution

a) The procedure here is the same as in the previous exercise, except that you now use the Feldman Cousins interval and a known background expectation value of $\lambda = 3.2$. To determine the 68% interval, the different n are sorted not according to their probability, but according to the factor

$$r = \frac{P(n|\mu)}{P(n|\hat{\mu})} \tag{3.13}$$

where $\hat{\mu}$ is the value of possible μ which maximizes $P(n|\hat{\mu})$. μ is the expectation value of events, which is related to the expectation value of a not-background event ν by $\mu = \nu + \lambda$, this means $\hat{\mu} \geq 3.2$. In our case $P(n|\mu)$ is the Poisson distribution, so $\hat{\mu} = n$ for ≥ 4 and $\hat{\mu} = 4$ for ≤ 4. The 68 % intervals for each possible ν is shown in figure 3.3. The figure also shows the limits for the 68% smallest interval. The main difference between these two different methods of interval definition is that the Feldman Cousins method will give you a interval for every measured value n, while the smallest interval method can lead to an empty interval for ν. In our case this happens for $n = 0$.

For the exercise we should assume a measured $n = 9$ and therefore get

$$\nu \in [3.14; 9.58] \tag{3.14}$$

and for **b)** using the smallest interval:

$$\nu \in [3.30; 10.09] \tag{3.15}$$

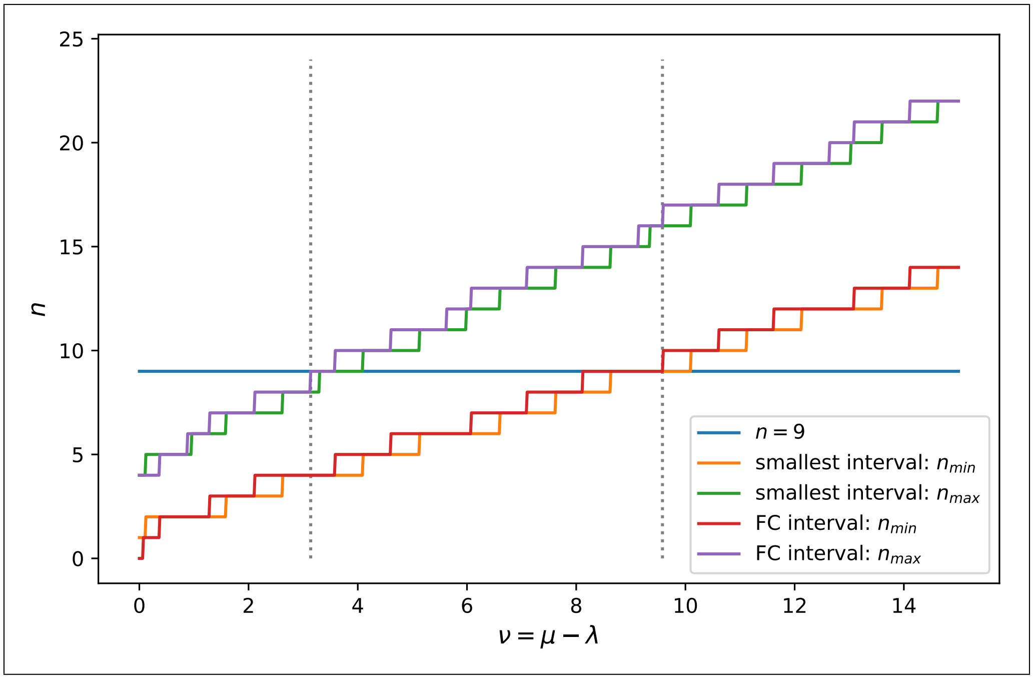


Figure 3.3: 68% smallest interval bands for exercise 8. The dashed lines show the interval for ν using the Fieldman Cousins intervals.

c) The posterior of a Poisson distribution $P(n|\nu, \lambda)$ with background λ and a flat prior is given by

$$P(\nu|n, \lambda) = \frac{e^{-(\nu+\lambda)}(\lambda + \nu)^n}{\int_0^{\nu_{max}} e^{-(\nu+\lambda)}(\lambda + \nu)^n d\nu} = \frac{e^{-\nu}(\lambda + \nu)^n}{n! \sum_{i=0}^{n} \frac{\lambda^i}{i!}} \tag{3.16}$$

In this equation we had to introduce ν_{max}, which is a limit of the flat prior priority that is necessary because the prior has to be normalized. But we can look at the limit $\nu_{max} \to \infty$ and the result makes perfect sense.

In our case we can plug in the given $\lambda = 3.2$ and the measured $n = 9$. The resulting porterior probability distribution is shown in figure 3.4. The mode is at $\nu_{mode} = n - \lambda = 5.8$, as it can be expected. To get a interval for ν, the 68% smallest interval was calculated. The result is $\nu \in [3.12; 9.15]$, also shown in figure 3.4. This result is similar to the ones reached before using the Neyman Construction.

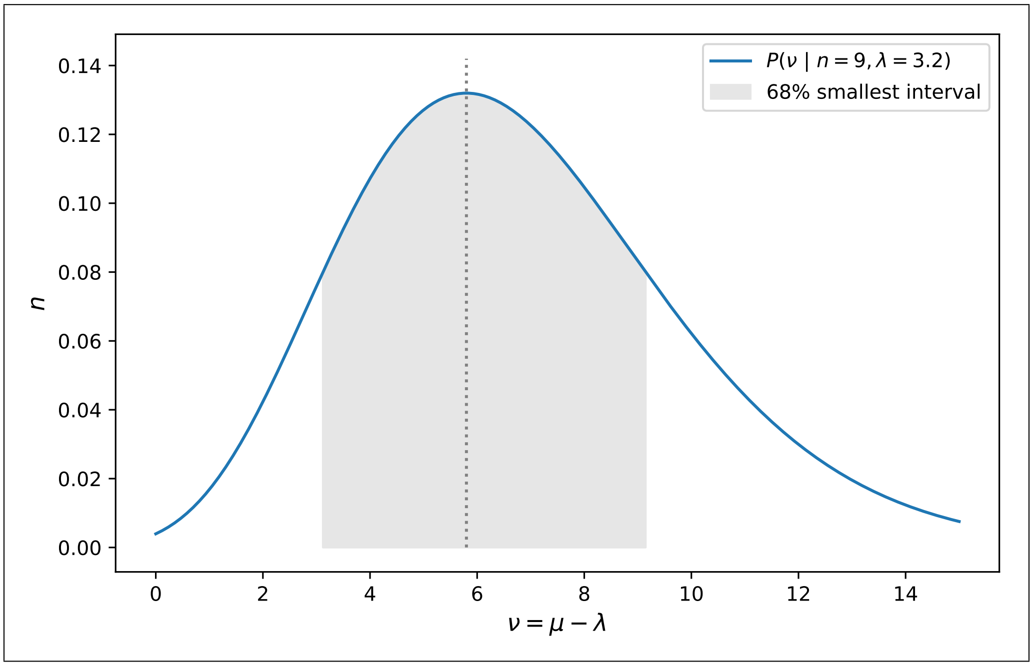


Figure 3.4: The posterian probability distribution for excercise 8c

3.2.4 Exercise 13

[...]

Solution In this exercise we measure n events with values x_i which are distributed according to $f(x|\lambda)$. If we bin the distribution into many intervals with the size Δ, the probability to get a value inside one of the intervals is $\nu_j = \int_\Delta f(x|\lambda)dx$. The probability of getting a certain amount of events in certain bins, or one can say the probability of our dataset x_i, is the the product of the poisson distributions of the bins:

$$P(x_i) = \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} \tag{3.17}$$

n_j is the number of events in bin j. For the limit $K \to \infty$ each n_j will be either 1 or 0, so $n_j! = 1$:

$$P(x_i) = \prod_{j=1}^{K} e^{-\nu_j} \prod_{j=1, n_j=1}^{K} \nu_j \tag{3.18}$$

For $K \to \infty$ the probability ν_j becomes the infinitesimal element $f(x_i|\lambda)\Delta$ and the exponential can be pulled out of the first product:

$$P(x_i) = e^{-\sum_{j=1}^{K}\nu_j} \prod_{i=1}^{n} f(x_i|\lambda)\Delta = \frac{1}{e}\prod_{i=1}^{n} f(x_i|\lambda)\Delta \tag{3.19}$$

The latter product is the unbinned likelihood, so we have shown that that the product of the Poisson probabilities for a infinitely small binning becomes the unbinned likelihood.

3.2.5 Exercise 16

We consider a thinned Poisson process. Here we have a random number of occurrences, N, distributed according to a Poisson distribution with mean ν. Each of the N occurrences, X_n, can take on values of 1, with probability p, or 0, with probability $(1-p)$. We want to derive the probability distribution for

$$X = \sum_{n=1}^{N} X_n \tag{3.20}$$

Show that the probability distribution is given by

$$P(X) = \frac{e^{-\nu p}(\nu p)^X}{X!} \tag{3.21}$$

Solution For a known N, X is distributed by the binomial distribution

$$P(X|N,p) = \frac{N!}{X!(N-X)!}p^r(1-p)^{n-r} \tag{3.22}$$

. The probability to get a certain N and a certain X can be calculated if you combine this with the Poisson distribution of N:

$$P(N,X|\nu,p) = \frac{e^{-\nu}\nu^N}{N!}\frac{N!}{X!(N-X)!}p^X(1-p)^{n-X} \tag{3.23}$$

To get a distribution of X one now only has to sum over all possible N ($N >= r$):

$$P(X|\nu,p) = \sum_{N=X}^{\infty}\frac{e^{-\nu}\nu^N}{N!}\frac{N!}{X!(N-X)!}p^X(1-p)^{n-X} = \frac{e^{-\nu}(p\nu)^X}{X!}\sum_{N=X}^{\infty}\frac{\nu^{N-X}}{(N-X)!}(1-p)^{n-X} \tag{3.24}$$

Now we can shift the index of the sum to $M = N - X$:

$$P(X|\nu,p) = \frac{e^{-\nu}(p\nu)^X}{X!}\sum_{M=0}^{\infty}\frac{(\nu-p\nu)^M}{M!} = \frac{e^{-\nu}(p\nu)^X}{X!}e^{(\nu-p\nu)} = \frac{e^{-\nu p}(\nu p)^X}{X!} \tag{3.25}$$

Chapter 4

Gaussian Probability Distribution Function

4.1 Summary

The **Gaussian probability distribution function** has big importance, as appears in the central limit theorem and is a limit if binomial and Poisson distributions. The Gaussian probability distribution function is written as:

$$G(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} \tag{4.1}$$

μ is the expectation value and σ is the variance of the distribution.

The **central limit theorem** states that the limiting distribution for the sum (or average) of repeated samplings of the same distribution for any distribution with finite moments is a Gauss distribution.

4.2 Exercises

4.2.1 Exercise 8: Central Limit Theorem

In this problem, you try out the Central Limit Theorem for a case where the conditions under which it was derived apply, and a case under which the conditions do not apply.

a) In this exercise, try out the CLT on the exponential distribution. First, derive what parameters of a Gauss distribution you would expect from the mean of n samples taken from the exponential distribution with

$$p(x) = \lambda e^{-\lambda x} \tag{4.2}$$

Then, try out the CLT for at least 3 different choices of n and λ and discuss the results. [...]

b) Now try out the CLT for the Cauchy distribution:

$$f(x) = \frac{1}{\pi\gamma} \frac{\gamma^2}{(x - x_0)^2 - \gamma^2} \tag{4.3}$$

Argue why the CLT is not expected to hold for the Cauchy distribution. You can generate random numbers from the Cauchy distribution by setting

$$x = \gamma \tan(U - \frac{\pi}{2}) + x_0 : \tag{4.4}$$

Try $x_0 = 25$ and $\gamma = 3$ and plot the distribution for x. Now take $n = 100$ samples and plot the distribution of the mean. Discuss the results.

Solution

a) The CLT say that the distribution of the averages is a Gauss distribution with parameters $\mu = \bar{x}$ and $\sigma = \frac{\sigma_x}{\sqrt{n}}$, where n is the number of values used to build the average. For the given exponential distribution one gets (with partial integration):

$$\mu = \bar{x} = E[x] = \int_0^\infty \lambda x e^{-\lambda x} dx = \left[-\left(x + \frac{1}{\lambda} \right) e^{-\lambda x} \right]_0^\infty = \frac{1}{\lambda} \tag{4.5}$$

and

$$E[x^2] = \int_0^\infty \lambda x^2 e^{-\lambda x} dx = \left[\left(-x^2 - \frac{2x}{\lambda} - \frac{2}{\lambda^2} \right) e^{-\lambda x} \right]_0^\infty = \frac{2}{\lambda^2} \tag{4.6}$$

$$\sigma = \frac{\sigma_x}{\sqrt{n}} = \sqrt{\frac{E[x^2] - E[x]^2}{n}} = \frac{1}{\lambda \sqrt{n}} \tag{4.7}$$

To see the central limit theorem, the distribution of the averages was plotted for 9 different configurations for n and λ in figure 4.1. For each plot, 10000 averages were used. The plotted Gaussians have the parameters calculated according the equations above.

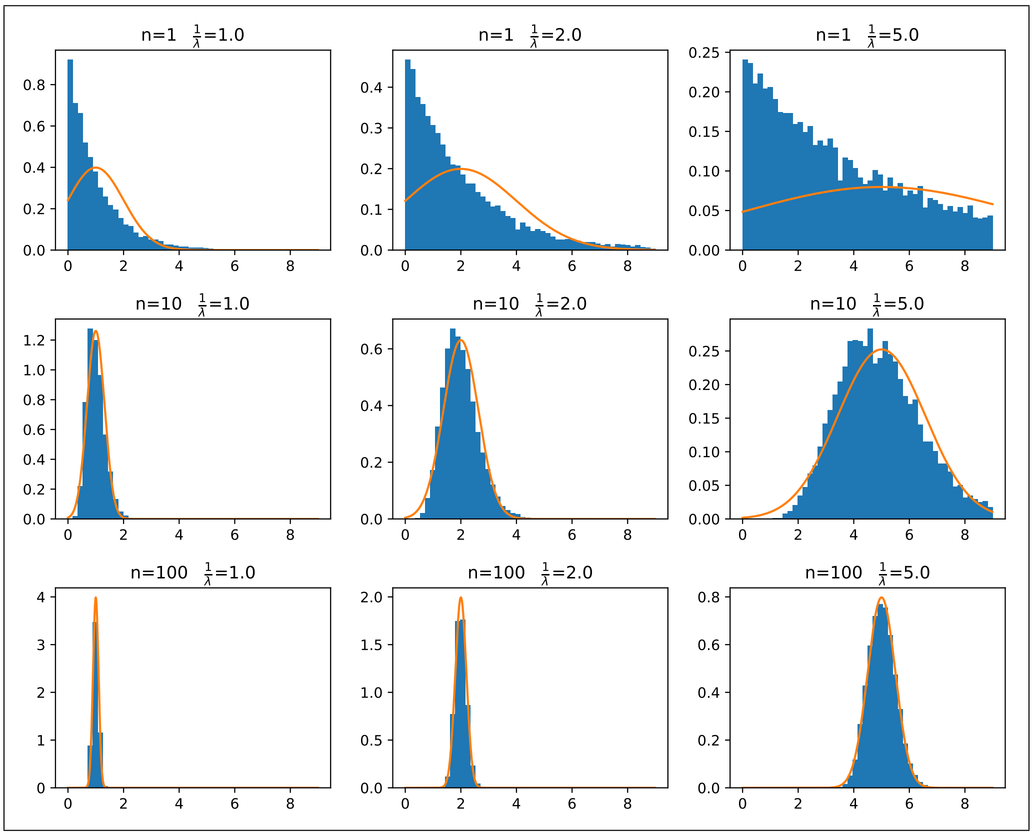


Figure 4.1: The distribution of the averages of n values which are distributed with the exponential distribution with parameter λ.

For $n = 1$ there is actually no average, so the plotted distribution is the exponential distribution. It is obvious that the CLT is not valid there. In the middle row of figure 4.1 with $n = 10$, the distribution of the averages is already quite Gaussian, but still one can see significant differences. In the bottom row with $N = 100$, the distribution fits very good with the Gaussian, as the CLT predicts.

b) For the Cauchy distribution the CLT is not valid, as it can be seen in figure 4.2. The blue distribution at the bottom is the distribution of 100000 averages of 100 Cauchy distributed values each. According to the CLT this should be a Gauss distribution, but one can see easily that it is not, but again the Cauchy distribution. This is due to the non negligible moments of the Cauchy distribution. This also leads to the fact that one cannot calculate the variance of the Cauchy distribution.

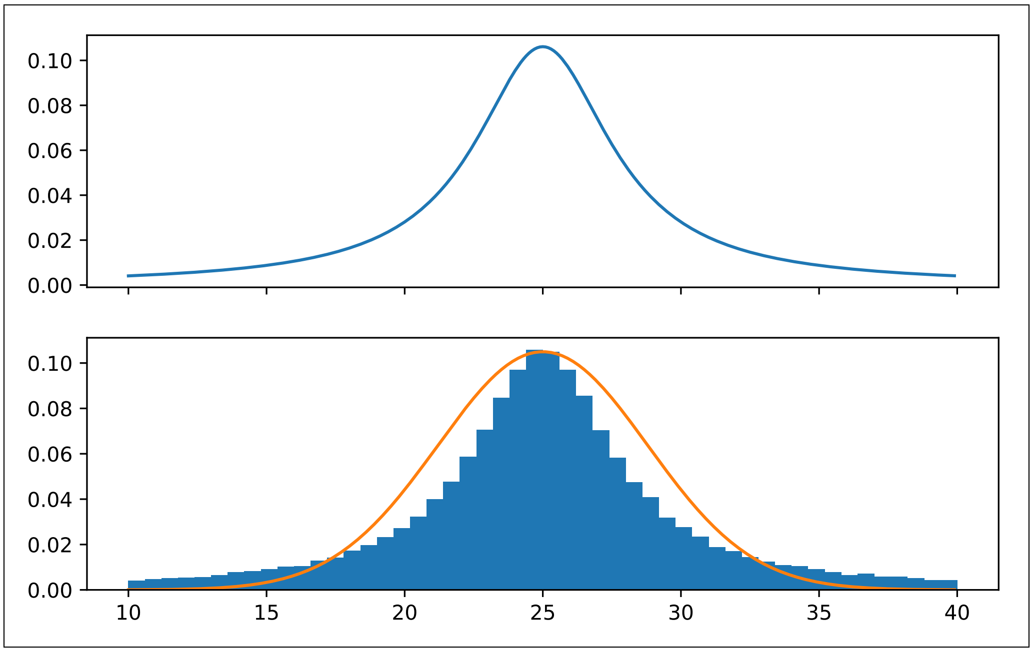


Figure 4.2: The upper plot shows the Cauchy distribution for $x_0 = 25$ and $\gamma = 3$. The lower plot showns the distribution of averages of 100 Cauchy distributed values. The orange plot at the bottom is a Gauss distribution.

4.2.2 Exercise 11: Plotting bivariate Gauss

With a plotting program, draw contours of the bivariate Gauss function (see next exercise for the definition of the function) for the following parameters:

(a) $\mu_x = 0 \quad \mu_y = 0 \quad \sigma_x = 1 \quad \sigma_y = 1 \quad \rho_{xy} = 0$

(b) $\mu_x = 1 \quad \mu_y = 2 \quad \sigma_x = 1 \quad \sigma_y = 1 \quad \rho_{xy} = 0.7$

(c) $\mu_x = 1 \quad \mu_y = -2 \quad \sigma_x = 1 \quad \sigma_y = 2 \quad \rho_{xy} = -0.7$

Solution The bivariate Gauss function, as given in the script, is already implemented in the python matplotlib package. The resulting plots are shown in figure 4.3.

4.2.3 Exercise 12: Bivariate Gauss probability distribution

a) Show that the pdf can be written in the form

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)\right) \tag{4.8}$$

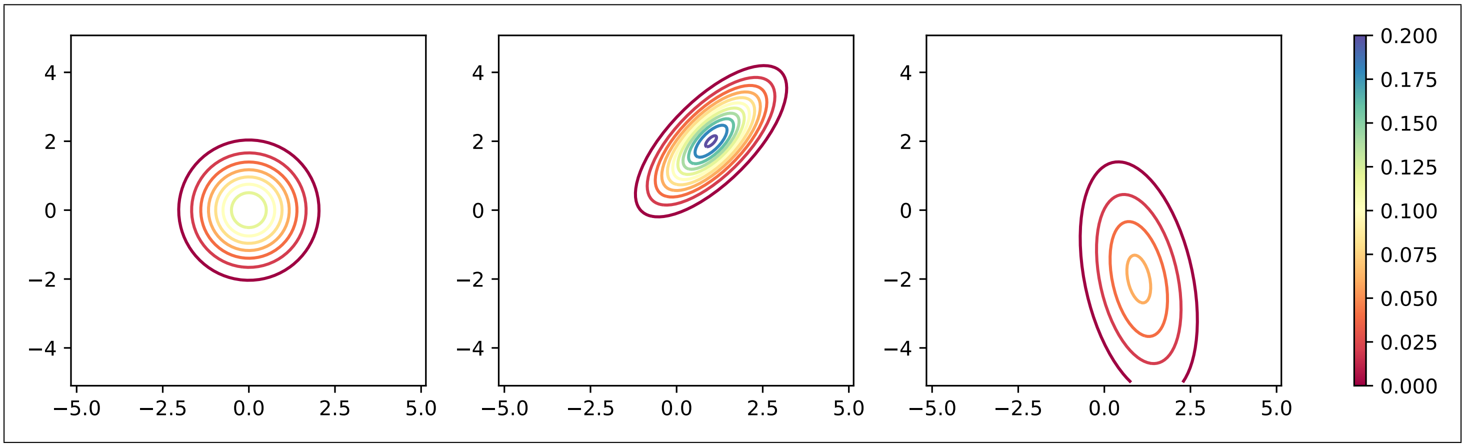


Figure 4.3: The contourplots of the three different bivariate Gauss distribution from exercise 11.

b) Show that for $z = x - y$ and x, y following the bivariate distribution the resulting distribution for z is a Gaussian probability distribution with

$$\mu_z = \mu_x - \mu_y \tag{4.9}$$
$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \tag{4.10}$$

Solution

a) In the script the probability of joint correlated Gauss distributions is given. With $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ this is

$$P(x,y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\vec{x}^T\Sigma\vec{x}\right) \tag{4.11}$$

where we assumed $\vec{\mu} = 0$ which is just a shift. Σ is the covariance matrix

$$\Sigma = \begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \tag{4.12}$$

$$\det\Sigma = \sigma_x^2\sigma_y^2(1-\rho^2) \quad \text{and} \quad \Sigma^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} \tag{4.13}$$

If one inserts this into equation 4.11 one gets:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)\right) \tag{4.14}$$

b) For the probability of z one has integrate the bivariate pdf in one dimension with $y = x - z$:

$$P(z) = \int P(x, x-z)dx \tag{4.15}$$

With the equation of bivariate Gauss one gets:

$$P(z) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)\right)dx \tag{4.16}$$

For a better overview we define:

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{4.17}$$

$$B = \frac{1}{2(1-\rho^2)}\left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} - \frac{2\rho}{\sigma_x\sigma_y}\right) \tag{4.18}$$

$$C = \frac{1}{2(1-\rho^2)}\left(\frac{-2z}{\sigma_y^2} + \frac{2\rho z}{\sigma_x\sigma_y}\right) \tag{4.19}$$

$$D = \frac{1}{2(1-\rho^2)}\left(\frac{z^2}{\sigma_y^2}\right) \tag{4.20}$$

This leads to:

$$P(z) = A\int \exp\left(-[Bx^2 + Cx + D]\right)dx = Ae^{-D}\int \exp\left(-Bx\left(x + \frac{C}{B}\right)\right)dx \tag{4.21}$$

As the integration limit are ∞ we can now make a substitution $x = x' - \frac{C}{2B}$ which leads to:

$$P(z) = Ae^{-D}\int e^{-\left(Bx^2 - \frac{C^2}{4B}\right)}dx = Ae^{\frac{C}{4B}-D}\int e^{-Bx^2}dx \tag{4.22}$$

This integral can be solved to $\sqrt{\frac{\pi}{B}}$, and with resubstituting A, B, C and D one gets (with a lot of auxiliary calculation):

$$P(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y}}e^{-\frac{z^2}{2(\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y)}} \tag{4.23}$$

This is a Gauss distribution for z with

$$\mu_z = \mu_x - \mu_y \tag{4.24}$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \tag{4.25}$$

4.2.4 Exercise 13: Convolution of Gauss

Convolution of Gaussians: Suppose you have a true distribution which follows a Gaussian distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x}e^{\frac{-(x-x_0)^2}{2\sigma_x^2}} \tag{4.26}$$

and the measured quantity, y follows a Gaussian distribution around the value x.

$$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y}e^{\frac{-(y-x)^2}{2\sigma_y^2}} \tag{4.27}$$

What is the predicted distribution for the observed quantity y ?

Solution To get the distribution for y one has to integrate over all x:

$$P(y) = \int_{-\infty}^{\infty} P(y|x)f(x)dx = \int_{-\infty}^{\infty} \mathcal{G}(y-x|0,\sigma_y)\mathcal{G}(x|x_0,\sigma_x)dx \tag{4.28}$$

Change to the characteristic function:

$$\Phi_y(k) = \int_{-\infty}^{\infty} e^{iky}\int_{-\infty}^{\infty} \mathcal{G}(y-x|0,\sigma_y)\mathcal{G}(x|x_0,\sigma_x)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iky}\mathcal{G}(y-x|0,\sigma_y)\mathcal{G}(x|x_0,\sigma_x)dxdy \tag{4.29}$$

Now we substitute with $z = y - x$. This can be done easily as the integrals go to infinity.

$$\Phi_y(k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ik(x+z)}\mathcal{G}(z|0,\sigma_y)\mathcal{G}(x|x_0,\sigma_x)dxdz = \int_{-\infty}^{\infty} e^{ikz}\mathcal{G}(z|0,\sigma_y)dz \int_{-\infty}^{\infty} e^{ikx}\mathcal{G}(x|x_0,\sigma_x)dx \quad (4.30)$$

With the definition of the characteristic functions for Gaussians of the script the two integrals can be written as

$$\Phi_z(k) = \int_{-\infty}^{\infty} e^{ikz}\mathcal{G}(z|0,\sigma_y)dz = e^{-\frac{k^2\sigma_y^2}{2}} \quad (4.31)$$

$$\Phi_x(k) = \int_{-\infty}^{\infty} e^{iky}\mathcal{G}(x|x_0,\sigma_x)dx = e^{ikx_0-\frac{k^2\sigma_x^2}{2}} \quad (4.32)$$

the gives:

$$\Phi_y(k) = \Phi_z(k)\Phi_x(k) = e^{ikx_0-\frac{k^2(\sigma_y^2+\sigma_x^2)}{2}} \quad (4.33)$$

This is of the same form as the characteristic function of a Gaussian again. This means that the resulting probability distribution for y is a Gaussian with $\mu = x_0$ and $\sigma^2 = \sigma_x^2 + \sigma_y^2$

4.2.5 Exercise 14: Probability of three parameters

Measurements of a cross section for nuclear reactions yields the following data.

θ	$Crossection$	$Error$
30°	11	1.5
45°	13	1.0
90°	17	2.0
120°	17	2.0
150°	14	1.5

The units of cross section are 10^{-30} cm2/steradian. Assume the quoted errors correspond to one Gaussian standard deviation. The assumed model has the form

$$\sigma = A + B\cos(\theta) + C\cos(\theta)^2 \quad (4.34)$$

a) Set up the equation for the posterior probability density assuming flat priors for the parameters A, B, C.

b) What are the values of A, B, C at the mode of the posterior pdf?

solution The bayesian analysis says that the probability density of the model parameters A, B, C can be calculated by:

$$P(A,B,C|r_i) = \frac{P(r_i|A,B,C)P_0(A,B,C)}{\int P(r_i|A,B,C)P_0(A,B,C)dAdBdC} \quad (4.35)$$

For a flat prior P_0 cancels out. Here r_i is a single datapoint. The probability of $P(r_i|A,B,C)$ is given by the Gauss distribution

$$P(r_i|A,B,C) = \frac{1}{\sqrt{2\pi}\delta}e^{\frac{(\theta-\sigma(\vec{\lambda}))^2}{2\delta^2}} \quad (4.36)$$

where δ is the error of the data. If you don't care about the normalization of the data, as we are only interested in the mode of the distribution, the pdf for A, B, C including all data points is

$$P(A,B,C|\{r_i\}) \propto \prod_i P(r_i|A,B,C) \quad (4.37)$$

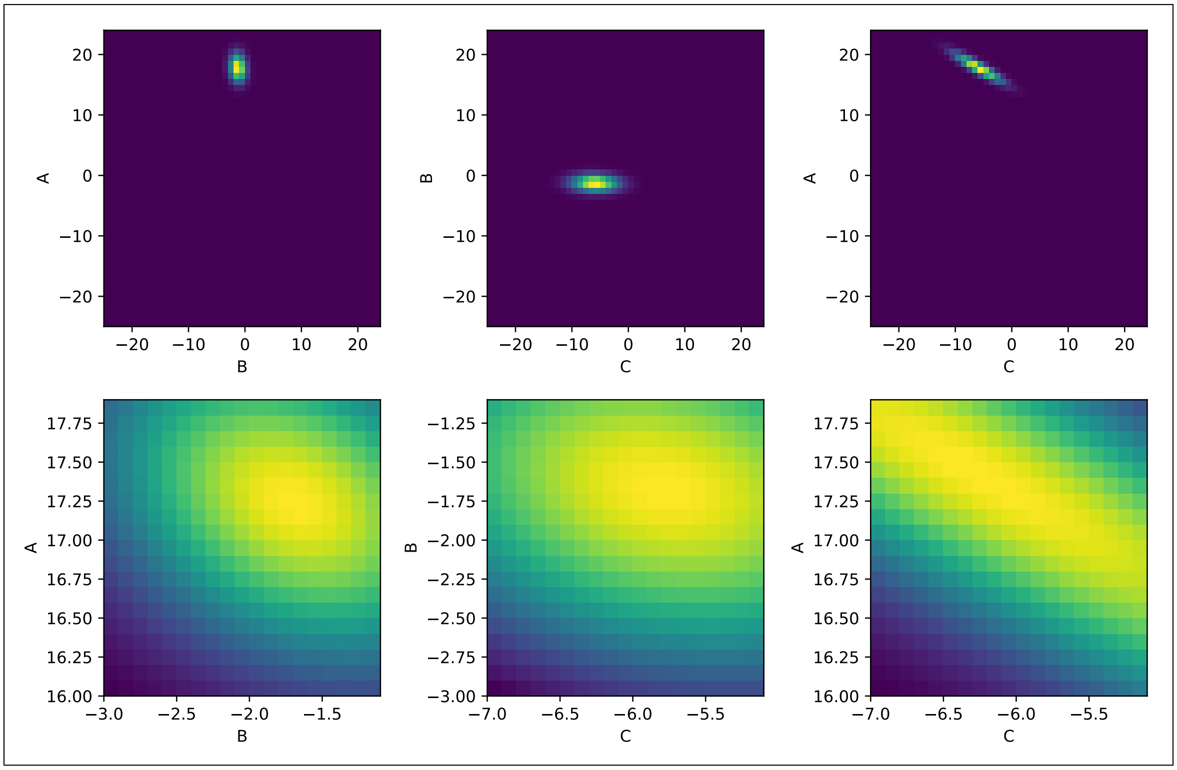


Figure 4.4: The probability distribution of the parameters A, B, C as marginals of the 3D space. The second three diagrams show a zoomed in area at the mode.

This was calculated numerically. The result is shown in figure 4.4. The mode (value with the highest probability density) was calculated to be at $A = 17.3$, $B = -1.7$ and $C = -6.3$.

The model function that corresponds to these values is shown in figure 4.5. One can see that the function fits quite well to the measured data.

Note: For this exercise I used the model function $\sigma(\theta) = A + B\cos(\theta) + C\cos(\theta)^2$ and not $\sigma(\theta) = A + B\cos(\theta) + C\cos(\theta^2)$ as in the original exercise, because in my opinion this makes more sense.

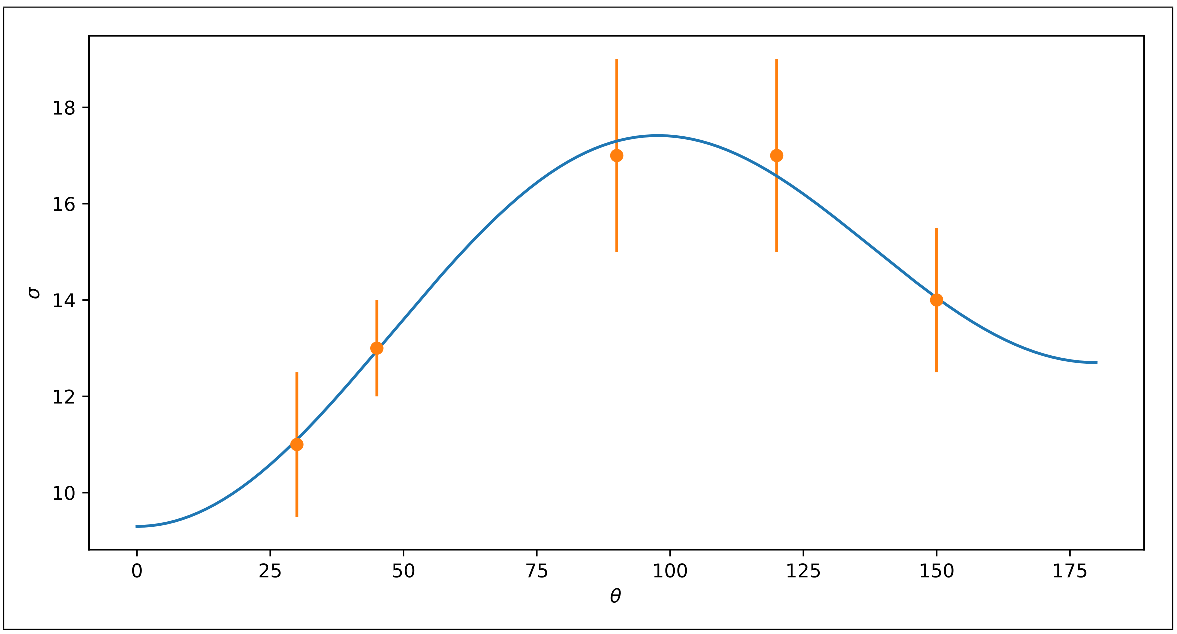


Figure 4.5: The model function for $A = 17.3$, $B = -1.7$ and $C = -6.3$ which is the mode of the pdf. The orange points show the measured data.

Chapter 5

Model Fitting and Model Selection

5.1 Summary

This chapter is about how to find a probability distribution for parameters of a **model**. With the frequentists approach parameter confidence level regions are calculated. One very popular method is a χ^2 **fit**, which is a frequentists approach with a certain test statistic. The last part part of the chapter concerns the **maximum likelihood estimator**.

5.2 Exercises

5.2.1 Exercise 1: Fit of the Sigmoid Function

Follow the steps in the script to fit a Sigmoid function to the following data:

E	N	r
0.5	100	0
1.0	100	4
1.5	100	22
2.0	100	55
2.5	100	80
3.0	100	97
3.5	100	99
4.0	100	99

Table 5.1: Measurement results for exercise 1 and 2

(a) Find the posterior probability distribution for the parameters (A, E_0)

(b) Define a suitable test statistic and find the frequentist 68% Confidence Level region for (A, E_0)

Solution

a) To get the posterior probability distribution Bayes Theorem is used:

$$P(A, E_o|\{r_i\}) = \frac{P(\{r_i\}|A, E_0)P_0(A, E_0)}{\int P(\{r_i\}|A, E_0)P_0(A, E_0)dAdE_0} \tag{5.1}$$

The probability $P(\{r_i\}|A, E_0)$ is the probability to get exactly the result $\{r_i\}$ we observed if we assume our system behaves like the model $p(E_i|A, E_0)$ with the parameters A, E_0. In our case this is a product of binomial probabilities:

$$P(\{r_i\}|A, E_0) = \prod_i \binom{N_i}{r_i} \cdot p(E_i|A, E_0)^{r_i} \cdot (1 - p(E_i|A, E_0))^{N_i - r_i} \tag{5.2}$$

Our model is a sigmoid function:

$$p(E|A, E_0) = \frac{1}{1 + e^{-A(E-E_0)}} \tag{5.3}$$

As priors I choose flat priors in the region of interest. For E_0, which is the center of the function, this is 1.5 to 2.5. For A, which corresponds to the slope of the function, a region from 2.0 to 4.0 turns out make sense. To get the PDF the product 5.2 is calculated for every point on a fine grid in this region. Finally the result is normalized. The result can be seen in figure 5.1.

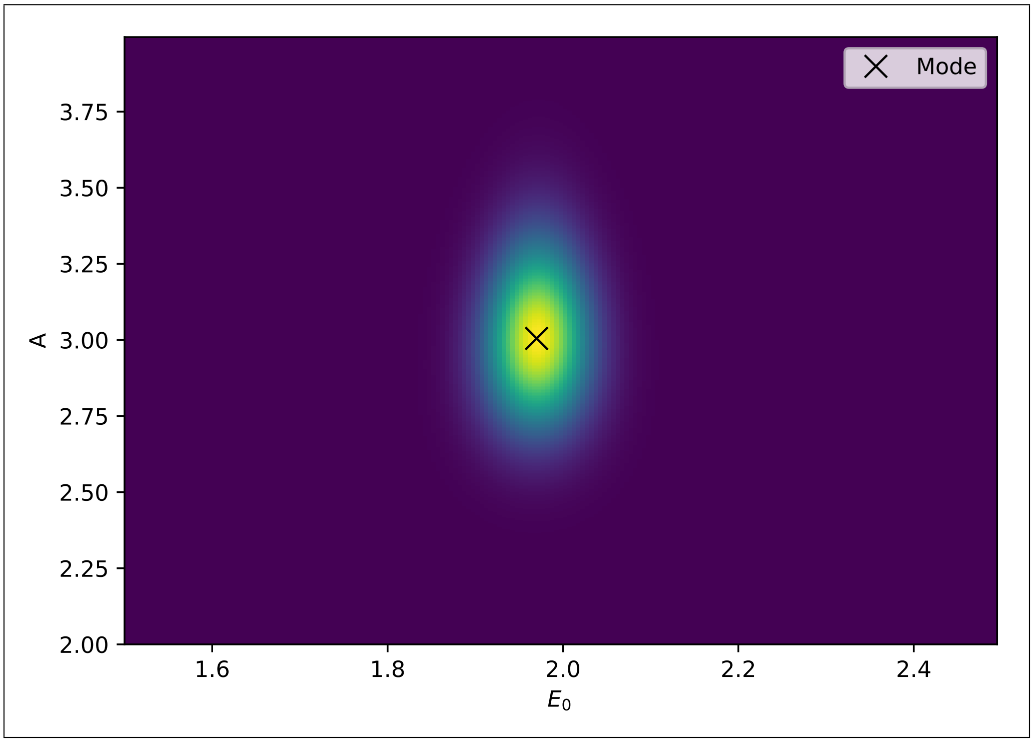


Figure 5.1: The posterior probability density of the model parameters E_0 and A for exercise 1a)

The mode of the posterior probability density is at $A = 3.00$ and $E_0 = 1.97$. To check the result the model function for the mode is plotted in figure 5.2. The sigmoid function fits very good to the measured data.

b) Now the same data is fitted using the frequentists approach: Therefor one hast to calculate for every point in parameter space weather the measured data is accepted for these parameters or not. To do that a test statistic for the data is defined, and it is then tested if the test statistic of the measured data is within a certain probability range of the test statistic distribution for these parameters. As a test statistic in this exercise the following is used, which is similar to the likelihood:

$$\xi(r_i, A, E_0) = \sum_{i=1}^{n} \ln(\binom{N_i}{r_i} \cdot p(E_i|A, E_0)^{r_i} \cdot (1 - p(E_i|A, E_0))^{N_i - r_i} \tag{5.4}$$

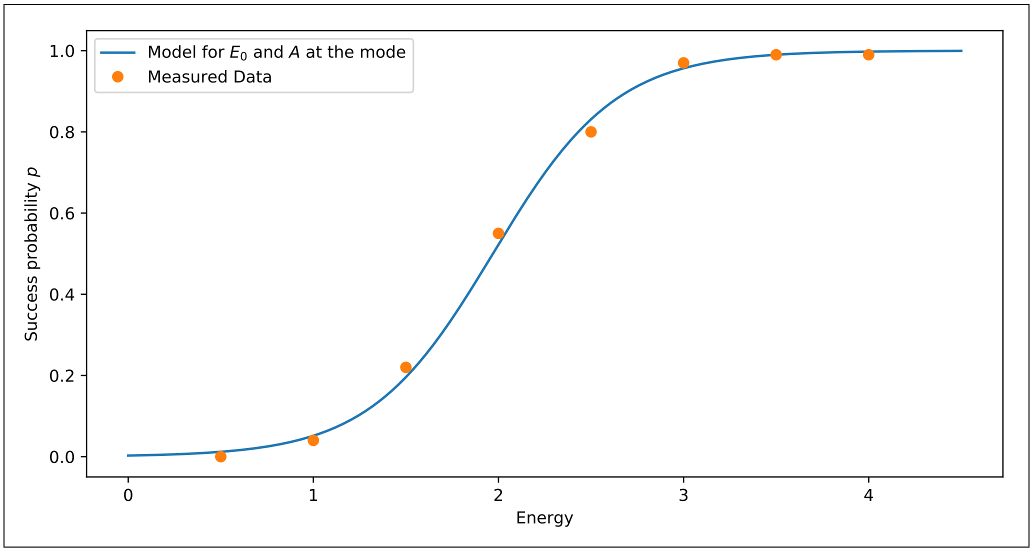
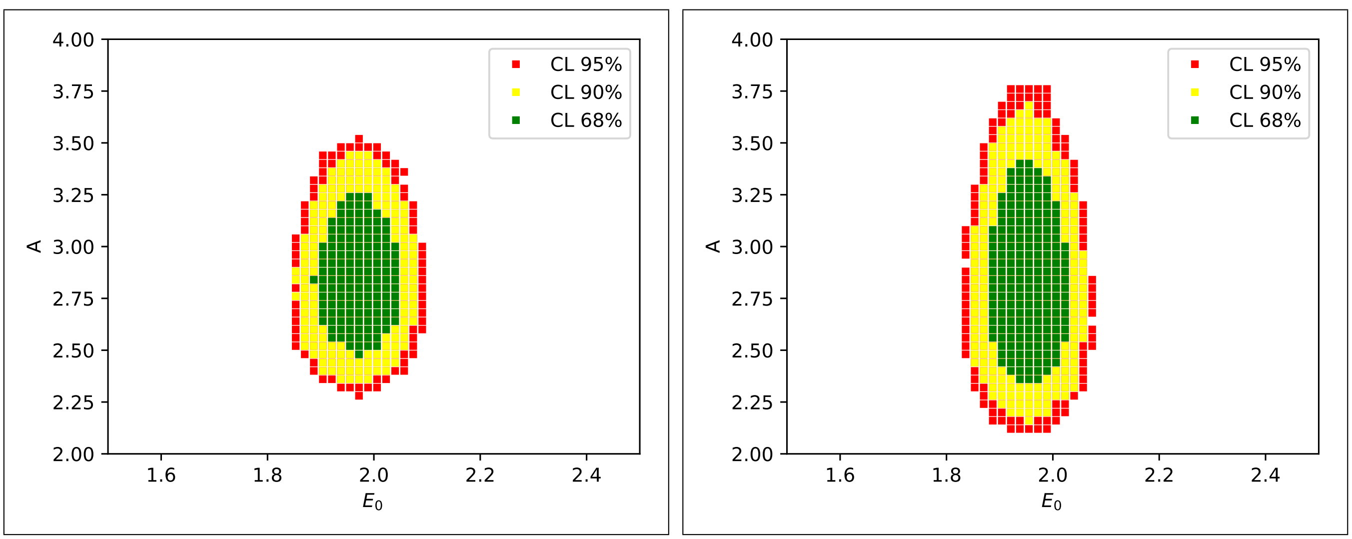


Figure 5.2: Model Function $p(E|A = 3.00, E_0 = 1.97)$ compared to the measured data.

Here $p(E_i|A, E_0)$ is the model used (in our case the sigmoid function). The distribution of this test statistic is calculated by taking 1000 random generated data sets according to the model. The CL reagions are found by looking for parameters where the test statistic of the data is bigger than $1 - \alpha\%$ of the simulated test statistics of this parameter. The resulting regions for α of 68%, 90% and 95% is shown in figure 5.3a.



(a) likelihood test statistic

(b) test statistic according to equation 5.5

Figure 5.3: The calculated CL regions for the sigmoid function model for different test statistics for exercise 1b).

The result fits the the result of the bayesian analyis in exercise a).
Additionally another test statistic was used, which was the squared sum of differences to the model data:

$$\xi(r_i, A, E_0) = \sum_{i=1}^{n} (r_i - p(E_i|A, E_0))^2 \tag{5.5}$$

The result is shown in figure 5.3b. The shape of the regions is a bit different, but the position and size fit quite well.

5.2.2 Exercise 2: Fit of another Function

Repeat the analysis of the data in the previous problem with the function

$$\epsilon(E) = \sin(A(E - E0)) \tag{5.6}$$

(a) Find the posterior probability distribution for the parameters $(A, E0)$

(b) Find the 68% CL region for $(A, E0)$

(c) Discuss the results

Solution Now the used model has a difficult property: It can give negative values. As the model describes the success probability of a binomial distribution, a negative value does not make sense. For the following I assumed the model to be zero if equation 5.6 leads to a negative value.

a) The calculation was similar to the exercise 1a). As priors it was used a flat prior from 1 to 15 for E_0 and -1 to 1 for A. The resulting PDF is shown in figure 5.4. The PDF is not normalized. The "streams" in the PDF correspond to the same resulting function, as the sinus is periodic. The mode on every stream has the same value, and the resulting function looks the same for all streams.

The mode in the first stream is at $A = -0.57$ and $E_0 = 6.46$. The corresponding model function is shown in figure 5.5.

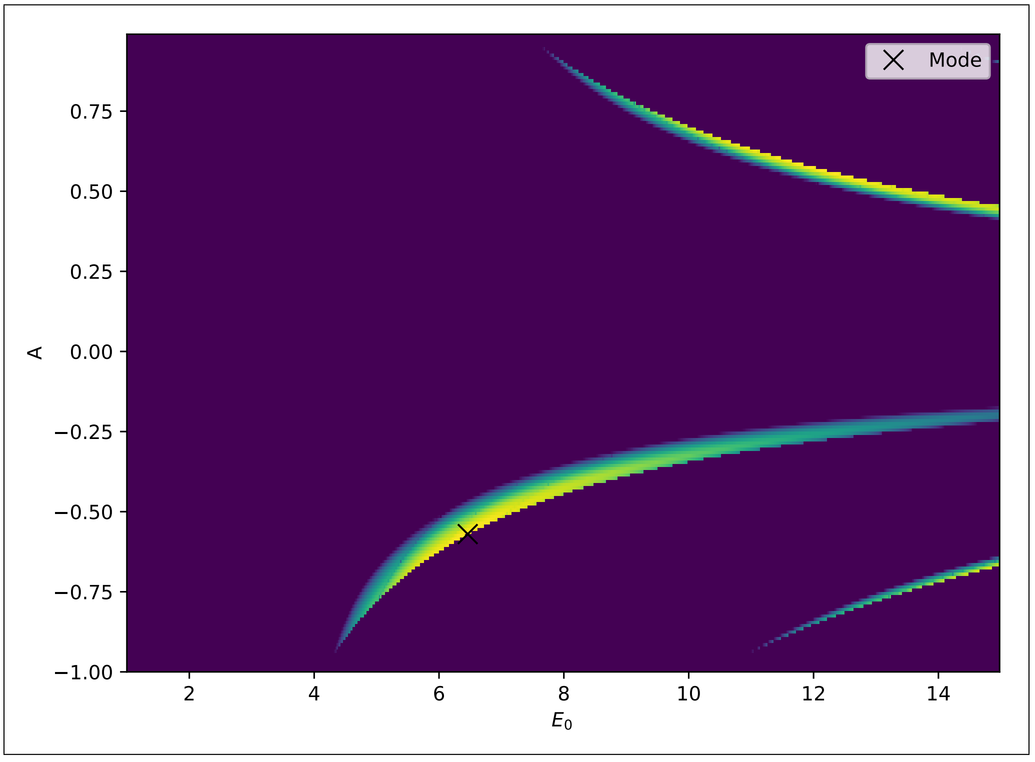


Figure 5.4: The calculated PDF for the model 5.4. The color scale is logarithmic.

The result is what one would expect: The fit is not as good as in a), and the shape is arbitrary, as the cutoff of negative values is only motivated by making the calculation possible. One could also ignore these values, or find another method to handle negative values.

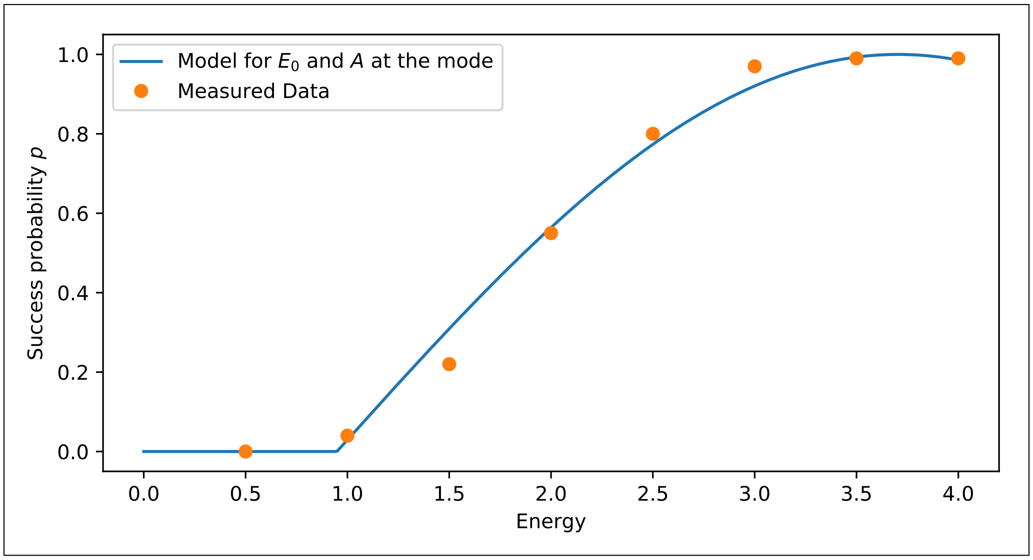


Figure 5.5: Resulting model with the parameters of the mode of the distribution in figure 5.4. Nagative model values were not allowed.

b) Now the model 5.6 is fitted to the data using the frequentists approach as in exercise 1b). As a test statistic, the total likelihood was used following equation 5.4. To get the test statistic distribution, 1000 random data sets were used. The resulting CL regions are shown in figure 5.6. The plotted region in parameter space is a zoomed in area of the plot in figure 5.4. The grey squares indicate that there occurred at least one negative value in 5.6 which was assumed to be zero.

There are no parameters for which fulfill the 68% CL condition. This means that for all possible parameters, the test statistic of the data is outside of the 68% upper region of the test statistic distribution. This is an indicator, the it is not very probable that the model could lead to the measured data.

c) The second model does seem to be reasonable for the measured data. But the two approaches handle the "bad" model differently:

Even for a "bad" model bayes method leads to a posterior probability function and it can not be directly seen in the result makes sense or not. So the PDF describes the probability of the parameters assuming the model is correct.

The frequentists approach can give you more information weather the model is reasonable. So if it is not probable that the model could lead to the measured data, the CL interval is just empty.

Another problem of the second model is that it can lead to negative values, which physically do not make sense. One had to find a way to handle this. In this case negative values were replaced by zero.

5.2.3 Exercise 3: χ^2 distribution for one data point

Derive the mean, variance and mode for the χ^2 distribution for one data point.

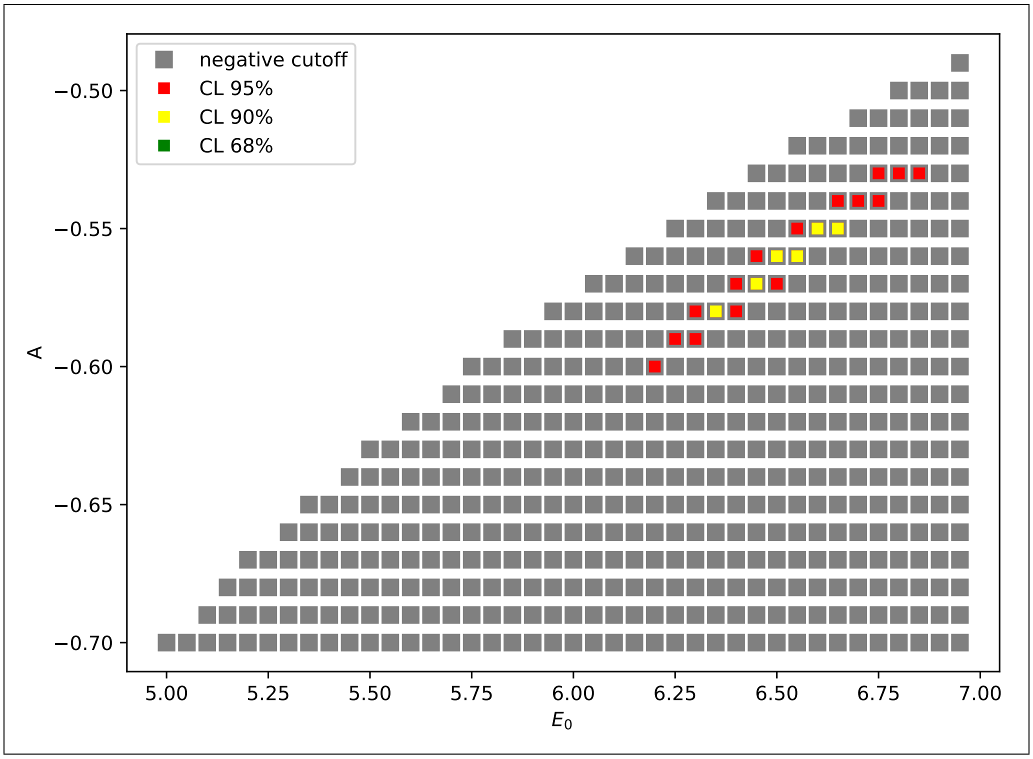


Figure 5.6: CL regions for the model 5.4. A 68% CL did not occur.

Solution

$$P(\chi^2) = \frac{1}{\sqrt{2\pi\chi^2}} e^{-\frac{\chi^2}{2}} \qquad \chi^2 = x \geq 0 \qquad x = 2y^2 dx \quad dx = 4y dy$$

$$E[x] = \int_0^\infty x \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} = \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{2} y^2 4 e^{-y^2} dy$$

$$= \frac{2}{\sqrt{\pi}} \int_0^\infty y \cdot 2y e^{-y^2 dy} \overset{\text{P.I.}}{=} -\frac{2}{\sqrt{\pi}} [y e^{-y^2}]_0^\infty + \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-y^2} dy$$

$$= \frac{2}{\sqrt{\pi}} [\frac{\sqrt{\pi}}{2} erf(y)]_0^\infty = erf(\infty) - erf(0) = 1 - 0 = 1$$

$$E[x^2] = \int_0^\infty x^2 \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} = \frac{4}{\sqrt{\pi}} \int_0^\infty y^3 2 e^{-y^2} dy$$

$$\overset{\text{P.I.}}{=} \frac{6}{\sqrt{\pi}} \int_0^\infty y 2 e^{-y^2} dy = \frac{6}{\sqrt{\pi}} [\frac{\pi}{2} erf(y)]_0^\infty = 3$$

$$Var[x^2] = E[x^2] - E[x]^2 = 3 - 1 = 2$$

$P(\chi^2)$ is monotonically decreasing:

$$\frac{P(x+h)}{P(x)} = \frac{\sqrt{x}}{\sqrt{x+h}} e^{-\frac{1}{2}h} < 1 \qquad \forall x, h > 0 \qquad \to x_{mode} = 0$$

5.2.4 Exercise 8: Fit with many parameters

Analyze the following data set assuming that the data can be modeled using a Gauss probability distribution where all data have the same uncertainty given by $\sigma = 4$. Try the two models:

(I) quadratic, representing background only:

$$f(x|A,B,C) = A + Bx + C^2$$

(II) quadratic + Breit-Wigner representing background+signal:

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

Perform a Bayesian fit assuming flat priors for the parameters. Find the best values of the parameters as well as uncertainties based on the marginalized probability distributions. What is the Bayes Factor for the two models?

Solution For the Bayesian fit one has to calculate the probability for all points in parameter space according to bayes theorem. The procedure in this exercise is the same as in exercise 14 of chapter 4 (4.2.5). With flat priors in a certain region this reduces to calculating the likelihood, which is easy to calculate with the known distribution of the parameters. The normalization can be done afterwords.

Model I The following parameter regions were chosen:

Parameter	Range	Number of points
A	-25 to -15	50
B	100 to 250	50
C	-100 to 50	50

This leads to 125000 points in the parameter space. The calculated PDF is marginalized and shown in figure 5.7. The best values for the parameters is assumed to be the mode of the PDF. The uncertainties are taken from the 68% smallest confidence level interval of the marginalized probability distributions.

Parameter	Best values	68% regions	approximated uncertainties
A	-7	$(-11; -4)$	± 4
B	173	$(158; 189)$	± 15
C	-30	$(-45; -14)$	± 15

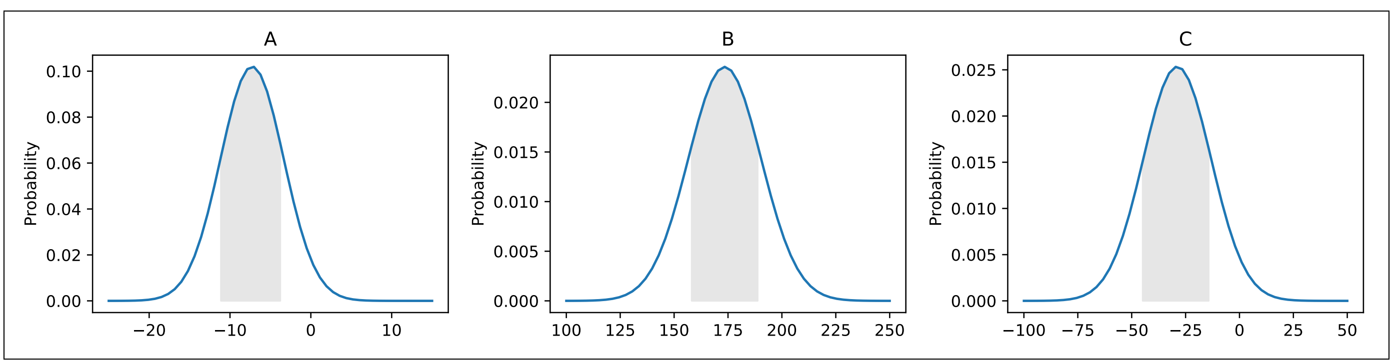


Figure 5.7: Marginalized probability distributions for the model I. The shaded areas show the 68% smallest intervals.

The resulting model function of the mode and the measured data is shown in figure 5.8

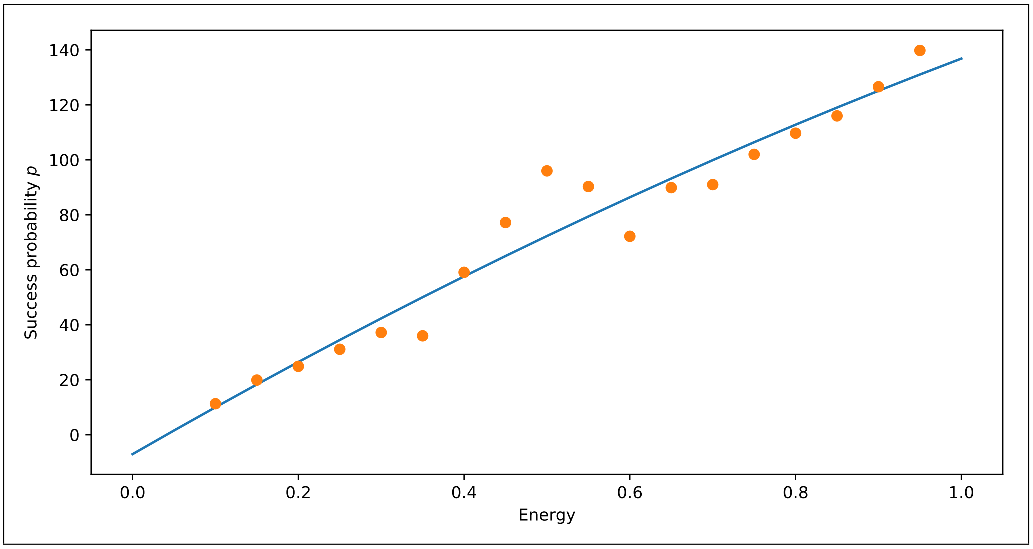


Figure 5.8: Measured data and the best fitted model I for exercise 8

Parameter	Range	Number of points
A	-15 to 25	20
B	-50 to 150	20
C	0 to 200	20
D	0 to 0.5	20
x_0	0.41 to 0.53	20
Γ	0.02 to 0.14	20

Model II The following parameter regions were chosen:

This leads to 64×10^6 points in the parameter space. The calculated PDF is marginalized and shown in figure 5.9. The best values for the parameters is assumed to be the mode of the PDF. The uncertainties are taken from the 68% smallest confidence level interval of the marginalized probability distributions.

Parameter	Best values	68% regions	approximated uncertainties
A	8	$(6; 12)$	± 3
B	55	$(24; 76)$	± 30
C	84	$(74; 116)$	$+30; -10$
D	0.16	$(0.11; 0.24)$	$+0.08; -0.05$
x_0	0.492	$(0.486; 0.498)$	± 0.006
Γ	0.064	$(0.058; 0.083)$	$+20; -6$

The resulting model function of the mode and the measured data is shown in figure 5.10

Bayes factor The bayes factors of the two models can be calculated by integrating over the whole parameter space. For the two models one reaches:

$$F_I = 9 \times 10^{-19} \tag{5.7}$$

$$F_{II} = 5 \times 10^{-2} \tag{5.8}$$

The comparison of the two factors indicates which model fits better to the data, in our case $\frac{F_{II}}{F_I} = 1.8 \times 10^{17}$ which means that the more complicated model II is preferred. In the plot 5.8 one can see that the peak in the data, which is not good explainable by background leads to the low Bayes factor. inf model II (figure 5.8) this peak is well explained by the model.

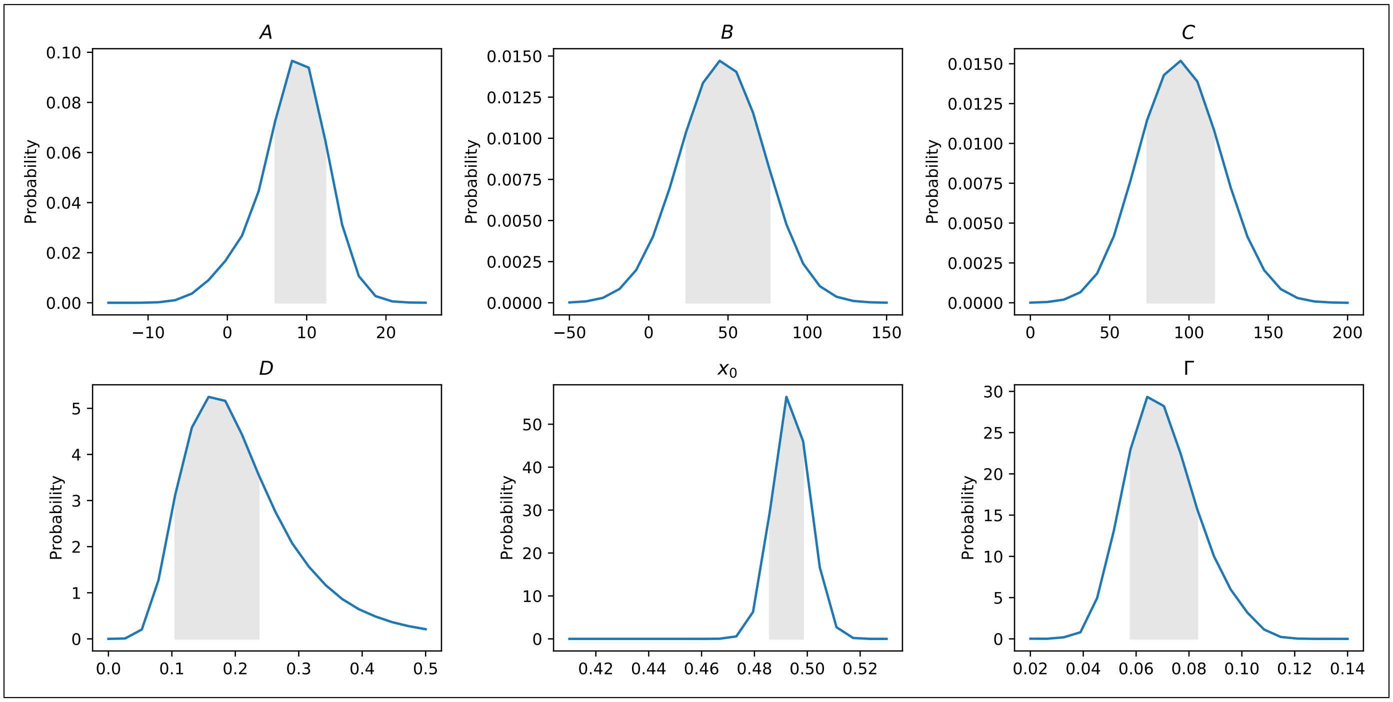


Figure 5.9: Marginalized probability distributions for the model II. The shaded areas show the 68% smallest intervals.

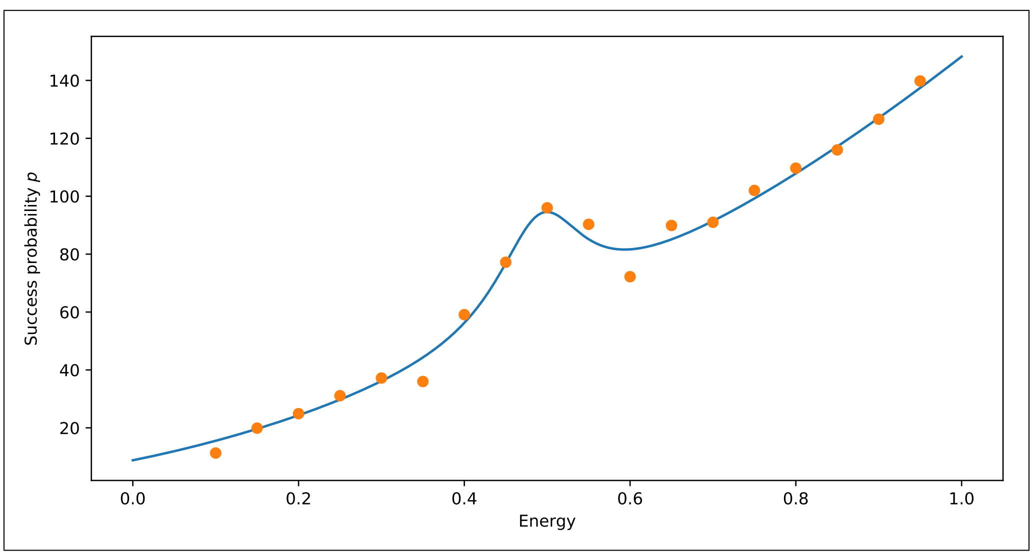


Figure 5.10: Measured data and the best fitted model II for exercise 8