Report - Data Analysis

Josef Keilhofer

April 6, 2018

1. Introduction to Probabilistic Reasoning

Exercise 1

Probabilistic reasoning - a first example

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 ?

Due to Laplace's rule of insufficient reason we assign the same probability to both events, getting a girl or getting a boy. We only know, that Janes first child is a girl. However, with no further information, this is independent on the sex of her second child. Therefore the probability, that on the second picture is a girl is 50%.
Now we know that Jane has at least one girl. This leads to three possible outcomes:

1. P(GG): Jane got first a girl and then again a girl.
2. P(BG): Jane got first a boy and then a girl.
3. P(GB): Jane got first a boy and then a girl.

Every outcome has again, applying Laplace's rule of insufficient reason, the same probability: $P(GG) = P(BG) = P(GB) = \frac{1}{3}$. However we have to remind that the events lead to different answers of our question, whether the second child is also a girl. In the case, that Jane has two girls, the condition is fulfilled instead of the other two outcomes.

$$P(Second\ child\ is\ girl) = \frac{P(GG)}{P(GG) + P(BG) + P(GB)} = \frac{\frac{1}{3}}{1} = \frac{1}{3} \tag{1}$$

Exercise 2
Interpreting data

Let us compare the probability we assign to an experimental result where we flip a coin ten times and have the following results:

- S1: THTHHHTHTTH
- S2: TTTTTTTTTT

Give possible definitions for the probability of the data.

First of all we again assume due to Laplace's rule of insufficient reason that the probability for getting tail or head is in both cases 50%.
There are a lot of definitions which lead to different probabilities:
One approach is to ask for the probability to get the given sequence. Since for each flip the probability to get heads or tails is $\frac{1}{2}$, the probabilities of the two events are the same:

$$P(S1|M) = P(S2|M) = 0.5^{10}$$

However, we can also calculate the probability to get the number of T in the sequence:

$$P(S1|M) = \binom{10}{5} 0.5^{10} = 252 \cdot 0.5^{10}$$
$$P(S2|M) = \binom{10}{10} 0.5^{10} = 0.5^{10}$$

Another approach is to ask for the probability to get tails at least a certain number of times:

$$P(S1|M) = \sum_{i=5}^{10} \binom{i}{10} 0.5^{10}$$
$$P(S2|M) = 1 - P(\overline{S1}|M) = 1 - 0.5^{10}$$

Further approaches which we can choose:

- What is the probability to get an even number of T?
- What is the probability to get the subsequence TTT in the sequence?
- What is the probability to get a T on the 8^{th} throw?
- ...

Exercise 3

Truth of measurements and confidence level

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?

First of all we can not assign a certain probability to a certain energy. The probability to get a certain (exact) energy is zero. Therefore we have to define an interval around the measured energy E and assign a probability $1 - \alpha$ to it. This is the probability that the true value lies within the interval. The higher the probability $1 - \alpha$ is chosen, the wider the interval evolves. If the resolution of the detector is 10%, we can define an interval $[E \cdot 0.9; E \cdot 1.1]$. For this interval we can state that the true value lies within it with a certain probability.

Exercise 4
Conditional probability

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?

Let us write the different events in a more easier way. Here M.s. means Mongolian swamp and being healthy means at least not to have Mongolian swamp.

Probability to have M.s.:	$P(M) =$	10^{-4}
Probability to have spots if you have M.s.:	$P(SP\|M) =$	1
Probability to have acute lethargy if you have M.s.:	$P(L\|M) =$	1
Probability to suffer from raging thirst if you have M.s.:	$P(T\|M) =$	0.6
Probability to suffer from violent sneezes if you have M.s.:	$P(VS\|M) =$	0.2
Probability to have spots if you are healthy:	$P(SP\|\overline{M}) =$	0.03
Probability to have acute lethargy if you are healthy:	$P(L\|\overline{M}) =$	0.1
Probability to suffer from raging thirst if you are healthy:	$P(T\|\overline{M}) =$	0.02
Probability to suffer from violent sneezes if you are healthy:	$P(VS\|\overline{M}) =$	0.05

For solving this problem we need Bayes' Theorem. It states, that if $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\sum_{i=1}^{N} A_i = S$, equation 2 holds.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{\sum_{i=1}^{N} P(B_i|A_i) \cdot P(A_i)} \tag{2}$$

For the probability to have M.s., if you go to the doctor with all symptoms we get:

$$\begin{aligned}
&P(M|SP \cap L \cap T \cap VS) = \\
&= \frac{P(SP \cap L \cap T \cap VS|M) \cdot P(M)}{P(SP \cap L \cap T \cap VS|M) \cdot P(M) + P(SP \cap L \cap T \cap VS|\overline{M}) \cdot P(\overline{M})} \\
&= \frac{1 \cdot 1 \cdot 0.6 \cdot 0.2 \cdot 10^{-4}}{1 \cdot 1 \cdot 0.6 \cdot 0.2 \cdot 10^{-4} + 0.03 \cdot 0.1 \cdot 0.02 \cdot 0.05 \cdot (1 - 10^{-4})} \approx 80.0\%
\end{aligned}$$

Let us define to have any three out of four symptoms as:
$P(E) = (SP \cap L \cap T \cap \overline{SN}) \cup (SP \cap L \cap \overline{T} \cap SN) \cup (SP \cap \overline{L} \cap T \cap SN) \cup (\overline{SP} \cap L \cap T \cap SN)$.

Then the probabilities are:

$$
\begin{aligned}
P(E|M) &= (1 \cdot 1 \cdot 0.6 \cdot 0.8 + 1 \cdot 1 \cdot 0.4 \cdot 0.2 + 0 + 0) = 0.56 \\
P(E|\overline{M}) &= (0.03 \cdot 0.1 \cdot 0.02 \cdot 0.95 + 0.03 \cdot 0.1 \cdot 0.98 \cdot 0.05 \\
&\quad + 0.03 \cdot 0.9 \cdot 0.02 \cdot 0.05 + 0.97 \cdot 0.1 \cdot 0.02 \cdot 0.05) = 5.71 \cdot 10^{-4} \\
P(M|E) &= \frac{P(E|M) \cdot P(M)}{P(E|M) \cdot P(M) + P(E|\overline{M}) \cdot P(\overline{M})} \\
&= \frac{0.56 \cdot 10^{-4}}{0.56 \cdot 10^{-4} + 5.71 \cdot 10^{-4} \cdot (1 - 10^{-4})} \approx 8.9\%
\end{aligned}
$$

2. Binomial and Multinomial Distribution

Exercise 8

Probability distribution

For the following function

$$P(x) = xe^{-x} \qquad 0 \leq x \leq \infty \tag{3}$$

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.

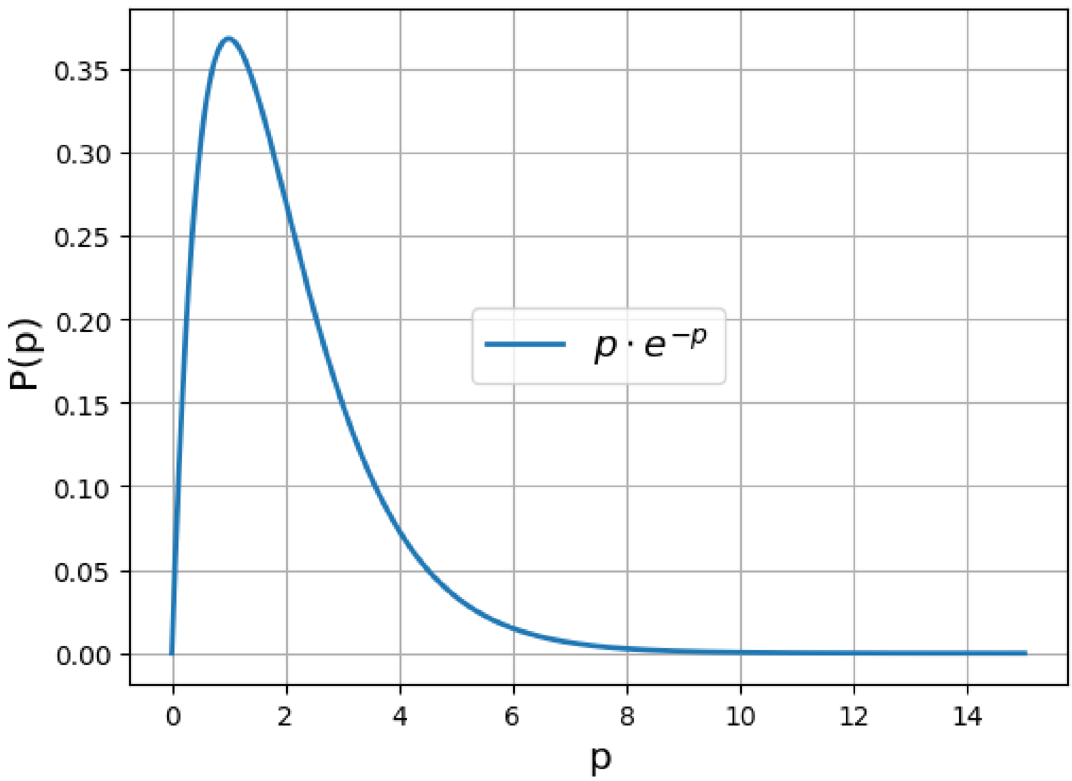


Figure 1: Plot of the given function

The **mean**, also called **expectation value**, is defined as follows:

$$E[x] = \int_{-\infty}^{\infty} xP(x)\mathrm{d}x \tag{4}$$

Inserting equation 3 into 4 and integrating gives:

$$E[x] = [-e^{-x}(x^2 + 2x + 2)]_0^{\infty} = 2$$

The **standard deviation** σ can be calculated with

$$\sigma = \sqrt{V[x]} \tag{5}$$

where V[x] is the **variance**, which is given by

$$V[x] = E[(x - E[x])^2] = E[x^2] - E[x]^2 \tag{6}$$

Therefor $E[x^2]$ needs to be calculated:

$$E[x^2] = \int_0^{\infty} x^3 e^{-x}\mathrm{d}x = \left[-e^{-x}(x^3 + 3x^2 + 6x + 6)\right]_0^{\infty} = 6$$

This leads to the standard deviation:

$$\sigma = \sqrt{6-4} = \sqrt{2}$$

To get the probability content of the given interval $x \in [\mu - \sigma, \mu + \sigma]$, one has to integrate the probability distribution over this interval. In our case this means

$$P(x) = \int_{2-\sqrt{2}}^{2+\sqrt{2}} xe^{-x}\mathrm{d}x = \left[-e^{-x}(x+1)\right]_{2-\sqrt{2}}^{2+\sqrt{2}} \approx 0.74$$

The median X_{med} is defined as the value of x for which the cumulative probability reaches 50%:

$$F(x_{med}) = \int_{-\infty}^{x_{med}} P(x)\mathrm{d}x = 0.5 \qquad \left(= \int_{x_{med}}^{\infty} P(x)\mathrm{d}x \right) \tag{7}$$

For our case this means that following equation has to be solved:

$$\begin{aligned} 0.5 &\overset{!}{=} \left[-e^{-x}(x+1)\right]_0^{x_med} \\ \Leftrightarrow 0.5 &\overset{!}{=} -e^{-x}(x+1) + 1 \\ \Rightarrow x_{med} &\approx 1.68 \end{aligned}$$

The $1 - \alpha$ central interval is defined as the interval $[x_1; x_2]$ with following condition.

$$F(x_1) = \int_{-\infty}^{x_1} P(x)\mathrm{d}x = 1 - F(x_2) = \int_{x_2}^{\infty} P(x)\mathrm{d}x = \frac{\alpha}{2} \tag{8}$$

This means that we have to solve these two equations:

$$\begin{aligned} 0.16 &\overset{!}{=} -e^{-x_1}(x_1 + 1) + 1 \\ 0.16 &\overset{!}{=} e^{-}x_2(x_2 + 1) \\ &\Rightarrow x_1 \approx 0.71 \\ &\Rightarrow x_2 \approx 3.29 \end{aligned}$$

The mode is defined as the value x^* with the highest probability. Therefore the root of the probability distribution's derivative has to be obtained.

$$0 \stackrel{!}{=} \frac{\mathrm{d}}{\mathrm{d}x} P(x) = e^{-x} - x e^{-x}$$
$$\Rightarrow x^* = 1$$

The smallest interval is defined as the interval $[x_1, x_2]$ with following conditions.

$$P(x_1) = P(x_2)$$
$$\int_{x_1}^{x_2} P(x)\mathrm{d}x \stackrel{!}{=} 0.68 \tag{9}$$

In our case this means, that we have to solve:

$$x_1 e^{-x_1} = x_2 e^{-x_2}$$
$$e^{-x_1}(x_1 + 1) - e^{-x_2}(x_2 + 1) = 0.68$$

These equations can not be solved analytically. Therefore we use a numerical algorithm (e.g. multidimensional Newton-Raphson-Algorithm) and come to the following result:

$$\Rightarrow x_1 \approx 0.27$$
$$\Rightarrow x_2 \approx 2.49$$

Exercise 10
Conditional probability

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.

Energy	Trials	Success
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

For this exercise we use the Binomial Distribution, since in each experiment there is a fixed number N of trials and there are two possible outcomes with probability p and $q = 1 - p$. The probability to get r successes with N trials and a success paramter of p is given as:

$$P(r|N,p) = \binom{N}{r} p^r q^{N-r} \tag{10}$$

However we want to obtain the probability distribution for the success parameter p. Therefore we use Bayes' Theorem in integral form where $P_0(p)$ is the prior probability function:

$$P(p|N,r) = \frac{P(r|N,p)P_0(p)}{P(r|N)} = \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)\mathrm{d}p} \tag{11}$$

Since there is no additional information we use a flat prior for $P_0(p)$. Then equation 11 simplifies to:

$$\begin{aligned} P(p|N,r) &= \frac{p^r(1-p)^{N-r}}{\int_0^1 p^r(1-p)^{N-r}\mathrm{d}p} \\ &= ... = \frac{(N+1)!}{r!(N-r)!} p^r (1-p)^{N-r} \end{aligned} \tag{12}$$

We can easily obtain the mode of the function with: $p^* = r/N$. In order to get the smallest interval, we use a numerical approach. Therefore we implement a list with evenly spaced numbers between 0 and 1. For each number the value of equation 11 is determined. Afterwards the list is sorted by the probability $P(p|N,r)$. The list points with highest probability are accumulated as long as the probability content of the integral is smaller than 68%. Automatically the first condition in equation 9 is fulfilled as well as the second one. The results are shown in table 1 and in figure 2

Energy	Trials	Success	mode	p_1	p_2
0.5	100	0	0.0	0.0	0.0112
1.0	100	4	0.04	0.0233	0.0685
1.5	100	20	0.2	0.1625	0.2414
2.0	100	58	0.58	0.5309	0.6281
2.5	100	92	0.92	0.8902	0.9442
3.0	1000	987	0.987	0.9831	0.9903
3.5	1000	995	0.995	0.9924	0.9969
4.0	1000	998	0.998	0.9962	0.9991

Table 1: Intervals for Binomial Probability Distribution

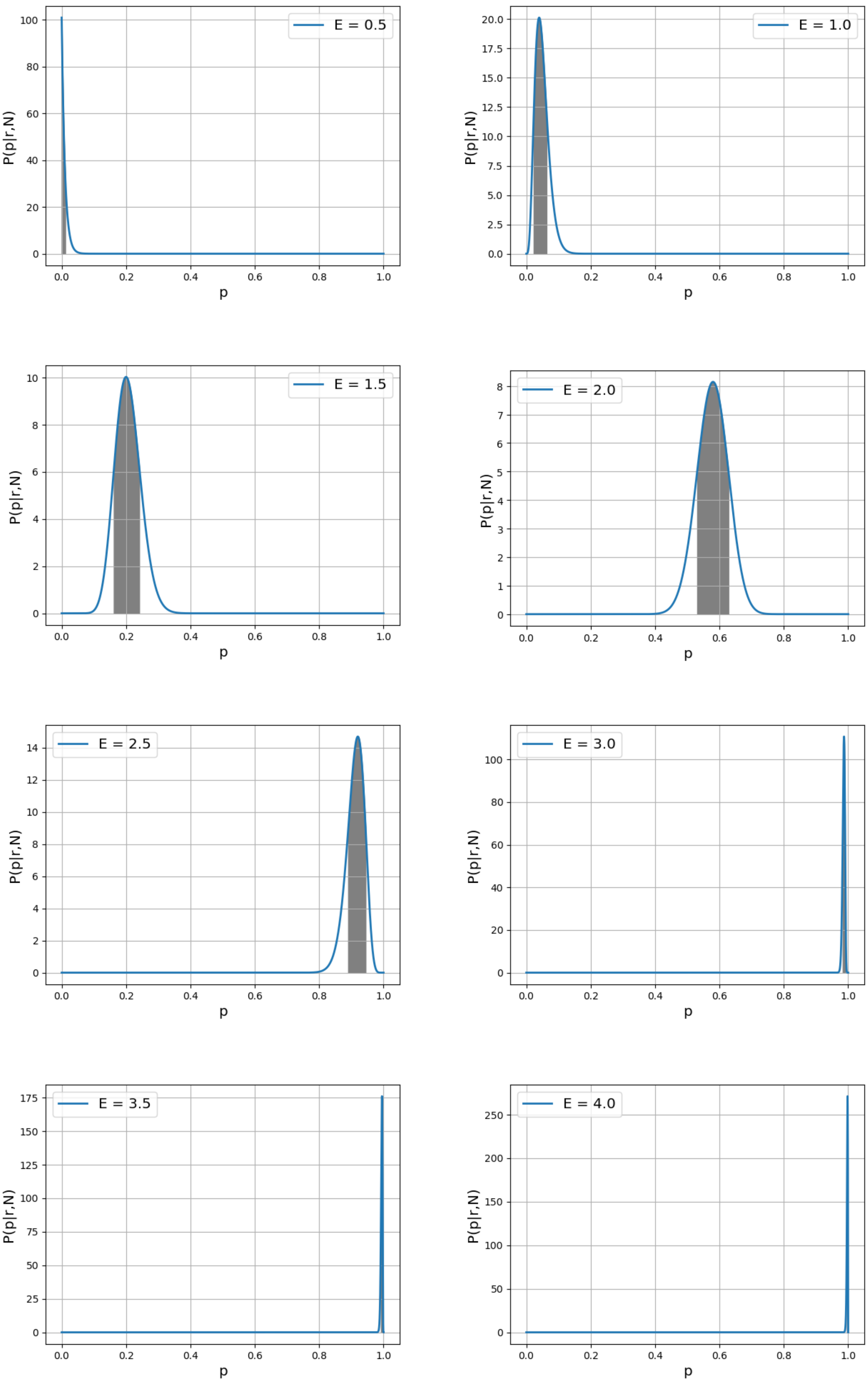


Figure 2: Graphs of Binomial Probability Distribution for each Energy with 68% smallest interval

Exercise 11
Frequentist Analysis

Analyze the data in the table of Exercise 10 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.

We obtain for each experiment with a numerical approach the 90% central interval, which is defined in equation 8. The probability distribution we used is given in equation 12. This leads to the limits of the intervals for each energy (cf. table 2). Plotting the intervals of p for each Energy gives two curves, the lower limit and the upper limit. The mode of each success parameter is marked with a red cross. Since there are more trials for energies above 2.5, the interval is smaller due to the sharper peak in the probability distribution.

Energy	p_{low}	p_{high}
0.5	0.0004	0.0293
1.0	0.0196	0.0884
1.5	0.1437	0.2747
2.0	0.4947	0.6577
2.5	0.8615	0.9528
3.0	0.9793	0.9916
3.5	0.9894	0.9975
4.0	0.9936	0.9993

Table 2: Limits of 90% confidence intervals

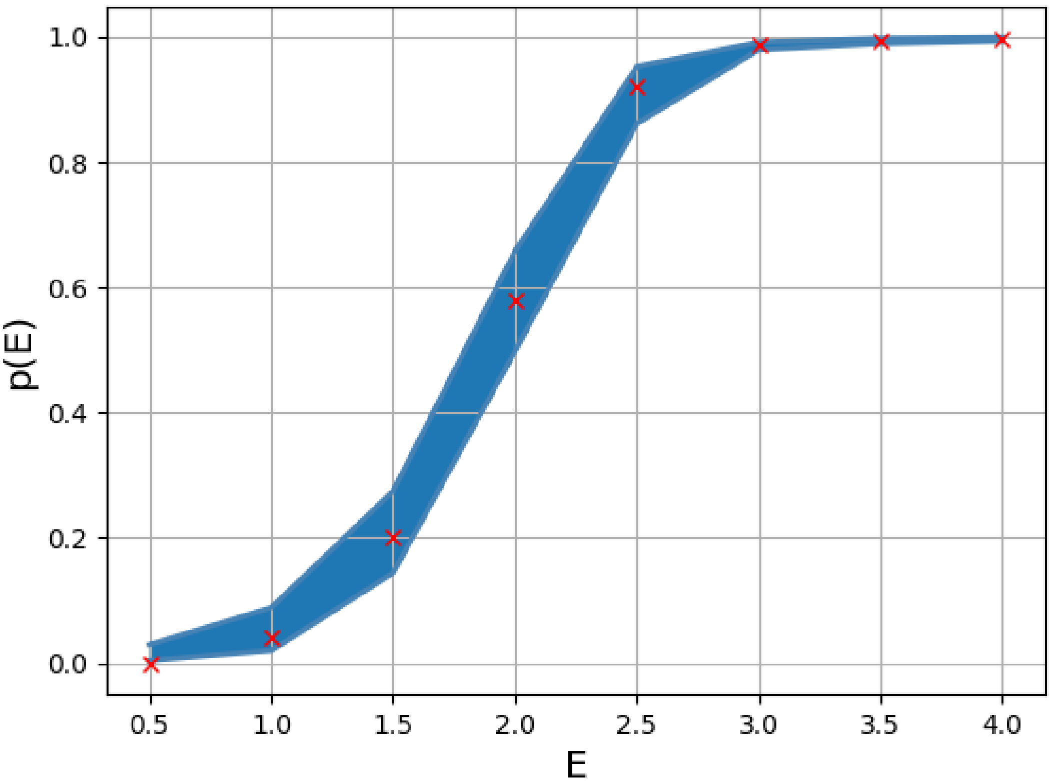


Figure 3: 90% confidence level of data vs. energy

Exercise 13
Using data multiple times

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

$$P_n(p|r,N) = \frac{(nN+1)!}{(nr)!(nN-nr)!} p^{nr}(1-p)^{n(N-r)} \tag{13}$$

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

For $n = 1$ equation 13 is equal to Bayes equation for the Binomial model with a flat prior:

$$P_1(p|r,N) = \frac{(N+1)!}{r!(N-r)!} p^r(1-p)^{N-r} \tag{14}$$

In order to proof equation 13 by conduction, we consider Bayes' Theorem and replace the prior by the posterior of the previous step.

$$\begin{aligned}
P(p|N,r) &= \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)\mathrm{d}p} \\
P_{n+1}(p|N,r) &= \frac{P(r|N,p)P_n(p)}{\int P(r|N,p)P_n(p)\mathrm{d}p} \\
&= \frac{\cancel{\binom{N}{r}} p^r(1-p)^{N-r}p^{nr}(1-p)^{n(N-r)}}{\int \cancel{\binom{N}{r}} p^r(1-p)^{N-r}p^{nr}(1-p)^{n(N-r)}\mathrm{d}p} \\
&= \frac{p^{(n+1)r}(1-p)^{(n+1)(N-r)}}{\beta((n+1)r+1,(n+1)(N+r)+1)} \\
&= \frac{((n+1)r+(n+1)(N-r)+1)!}{((n+1)r)!((n+1)(N-r))!} p^{(n+1)r}(1-p)^{(n+1)(N-r)} \\
&= \frac{((n+1)N+1)!}{((n+1)r)!((n+1)N-(n+1)r)!} p^{(n+1)r}(1-p)^{(n+1)(N-r)} \\
&= P_{n+1}(p|N,r) \qquad \text{q.e.d.}
\end{aligned}$$

To calculate the expectation value, dependent on n we use equation 4 with the definition of the β-function:

$$\int_0^1 x^a(1-x)^b\mathrm{d}x =: \beta(a+1,b+1) = \frac{a!b!}{(a+b+1)!} \tag{15}$$

$$
\begin{aligned}
E_n[p] &= \int p P_n(p|N,r)\mathrm{d}p \\
&= \frac{(nN+1)!}{(nr)!(nN-nr)!}\int_0^1 \underbrace{p^{nr+1}}_{=p\cdot p^{nr}}(1-p)^{n(N-r)} \\
&= \frac{(nN+1)!}{(nr)!(nN-nr)!}\beta(nr+2, n(N-r)+1) \\
&= \frac{(nN+1)!}{(nr)!(nN-nr)!}\frac{(nr+1)!(n(N-r))!}{(nr+1+n(N-r)+1)!} \\
&= \frac{\cancel{(nN+1)!}}{\cancel{(nr)!}\cancel{(n(N-r))!}}\frac{(nr+1)\cancel{(nr)!}\cancel{(n(N-r))!}}{(nN+2)\cancel{(nN+1)!}} \\
&= \frac{nr+1}{nN+2}
\end{aligned}
\tag{16}
$$

applying the limes on 16 leads to:

$$
\lim_{n\to\infty}\frac{nr+1}{nN+2} = \frac{r}{N} \tag{17}
$$

In order to determine the variance we need to investigate $E[p^2]$. Therefore we use the following property of the β-function:

$$
\beta(a+1,b) = \beta(a,b)\cdot\frac{a}{a+b+1} \tag{18}
$$

$$
\begin{aligned}
\lim_{n\to\infty} E_n[p^2] &= \lim_{n\to\infty}\frac{(nN+1)!}{(nr)!(nN-nr)!}\int_0^1 \underbrace{p^{nr+2}}_{=p^2\cdot p^{nr}}(1-p)^{n(N-r)} \\
&= \lim_{n\to\infty}\frac{(nN+1)!}{(nr)!(nN-nr)!}\beta(nr+3, n(N-r)+1) \\
&= \lim_{n\to\infty}\underbrace{\frac{(nN+1)!}{(nr)!(nN-nr)!}\beta(nr+2, n(N-r)+1)}_{=E_n[p]}\cdot\frac{nr+1}{nr+1+n(N-r)+1} \\
&= \lim_{n\to\infty}\left(\frac{nr+1}{nN+2}\right)^2 = \frac{r^2}{N^2}
\end{aligned}
$$

Using equation 6 for the variance, we obtain $\lim_{n\to\infty} V_n[p] = 0$.

3. Poisson Distribution

Exercise 4
Probability distribution

Consider the function:

$$f(x) = \frac{1}{2}e^{-|x|} \quad \text{for} - \infty < x < \infty \tag{19}$$

a) Find the mean and standard deviation fo 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?

Similar to Exercise 8 of Chapter 2, we determine the mean by equation 4.

$$\begin{aligned}
E[x] &= \int_{-\infty}^{\infty} \frac{1}{2} x e^{-|x|} \mathrm{d}x \\
&= \frac{1}{2}\left(\int_{-\infty}^{0} x e^{x} \mathrm{d}x + \int_{0}^{\infty} x e^{-x} \mathrm{d}x \right) \\
&\overset{\text{p.I.}}{=} \frac{1}{2}\left([xe^{x}]_{-\infty}^{0} - \int_{-\infty}^{0} e^{x} \mathrm{d}x + \left[-xe^{-x}\right]_{0}^{\infty} - \int_{0}^{\infty} -e^{-x} \mathrm{d}x \right) \\
&= \frac{1}{2}\left(0 - [e^{x}]_{0}^{\infty} + 0 - \left[e^{-x}\right]_{0}^{\infty} \right) \\
&= \frac{1}{2}\big(0 - 1 + 0 - (-1)\big) = 0
\end{aligned}$$

In order to get the standard deviation and the variance, we need to integrate equation 4 multiplied with x^2:

$$\begin{aligned}
E[x^2] &= \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|} \mathrm{d}x \\
&= \frac{1}{2} \cdot 2 \int_{0}^{\infty} x^2 e^{-x} \mathrm{d}x \\
&= \left[-x^2 e^{-x}\right]_{0}^{\infty} - \int_{0}^{\infty} -2xe^{-x} \mathrm{d}x \\
&= 0 - \left[2e^{-x}\right]_{0}^{\infty} \\
&= -[-2] = 2
\end{aligned}$$

By equations 6 and 5 we obtain the variance and the standard deviation:

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

The FWHM for a symmetric probability distribution with one maximum is defined as the interval between the two points x_1 and x_2, for which the following condition holds:

$$f(x_1) = f(x_2) \stackrel{!}{=} \frac{1}{2} f(x_{max}) \tag{20}$$

Solving the equation for the probability distribution given in equation 19 for this condition leads to:

$$\begin{aligned} f(x) &\stackrel{!}{=} \frac{1}{2} f(0) \\ \Leftrightarrow \frac{1}{2} e^{-|x|} &= \frac{1}{4} \\ \ln\left(\frac{1}{2}\right) &= \begin{cases} -x & \text{for} \quad x > 0 \\ x & \text{for} \quad x < 0 \end{cases} \\ &\Rightarrow x_1 \approx -0.69 \\ &\Rightarrow x_2 \approx 0.69 \end{aligned}$$

The Full Width at Half Maximum is smaller than the standard deviation.
In order to get the probability content of the integral in the interval $[-\sqrt{2}; \sqrt{2}]$ we have to integrate equation 19 over this interval:

$$\begin{aligned} P &= \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{2} e^{-|x|} \mathrm{d}x \\ &= 2 \cdot \frac{1}{2} \int_0^{\sqrt{2}} e^{-x} \mathrm{d}x \\ &= -[e^{-\sqrt{2}} - 1] \\ &\approx 75,7\% \end{aligned}$$

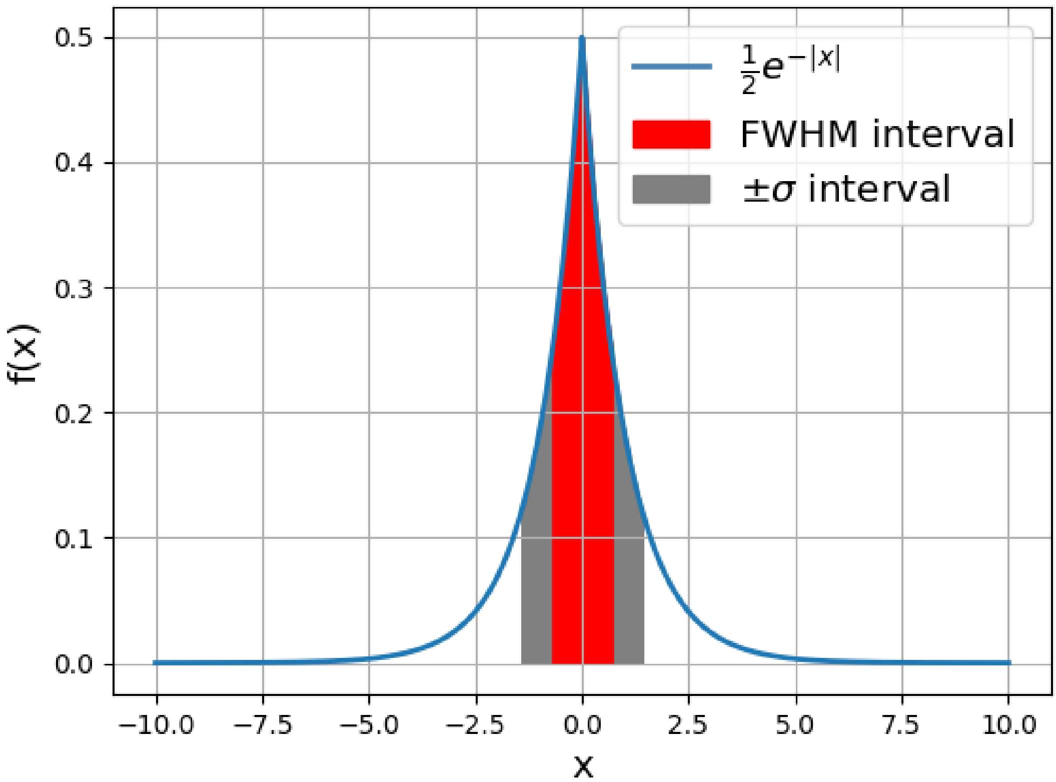


Figure 4: Function with FWHM-interval and $\pm\sigma$-interval

Exercise 7
Probability contents of Poisson Distribution

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?

The Poisson Distribution for a flat prior where n is the number of observed events depends only on one parameter, i.e. ν:

$$P(\nu|n) = \frac{\nu^n e^{-\nu}}{n!} \tag{21}$$

The cumulative probability of ν for a flat prior is given as:

$$F(\nu|n) = 1 - e^{-\nu} \sum_{i=0}^{n} \frac{\nu^i}{i!} \tag{22}$$

In order to obtain the 95% lower limit on the expectation value ν we have to solve the following equation

$$\begin{aligned}
&0.05 \overset{!}{=} F(\nu|n) \\
\Leftrightarrow\ &0.05 = 1 - e^{-\nu} \sum_{i=0}^{9} \frac{\nu^i}{i!} \\
\Rightarrow\ &\nu_{min} \approx 5.425
\end{aligned}$$

This means that with an probability of 95% the value for ν is higher than ν_{min}. However physically this is not meaningful. A more interesting result is the minimum rate R_{min} of events. We can calculate it with the Time in which the 9 events were detected: $R = \frac{\nu_{min}}{T}$.
In the Neyman construction for every possible value of ν the probability distribution (Poisson) is determined. Afterwards for every calculated distribution the 68% smallest interval is obtained. This gives for every value of ν an interval $[n_1, n_2]$ which states that with an 68% probability a measurement will detect a number of events contained in this interval. Plotting the results gives the Neyman Graph shown in figure 5. For 9 detected events we obtain: $6.50 < \nu < 13.30$.

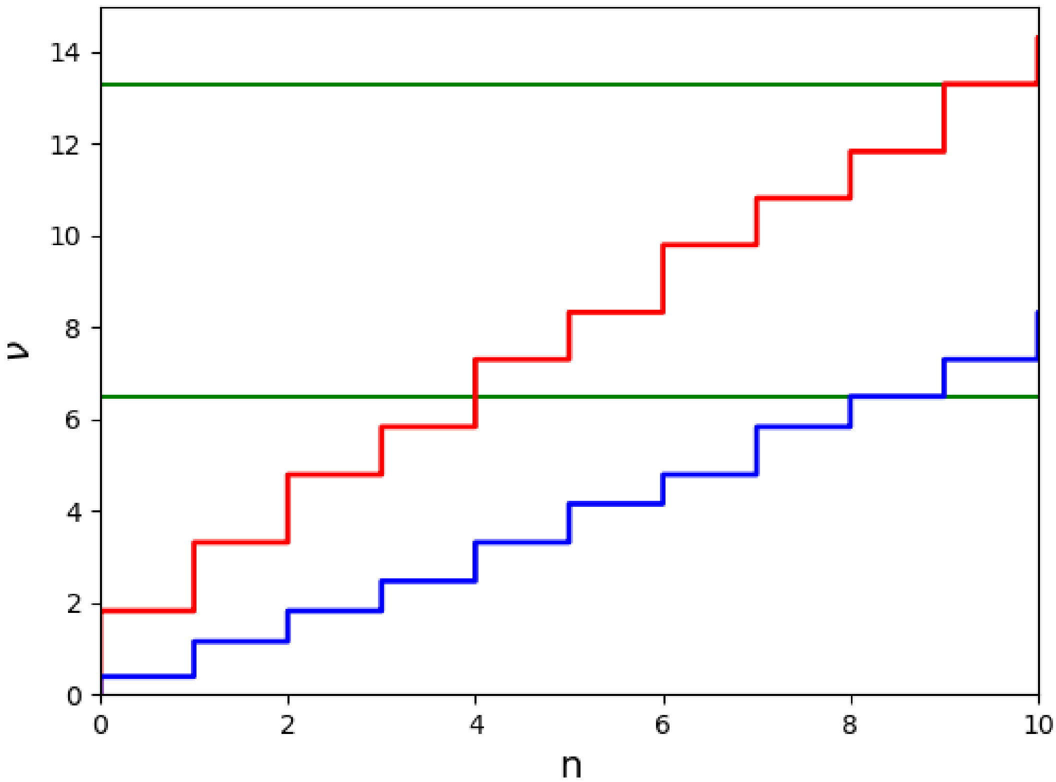


Figure 5: Poisson 68% Confidence Bands

Exercise 8

Superposition of Poisson Processes

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 ν

The combination of two (ore more) Poisson Processes yields another Poisson Process with $\nu_{combined} = \sum_i \nu_i$. In our case with background λ and events ν this leads to the resulting Poisson parameter $\mu = \lambda + \nu$.
The Feldman-Cousins Construction introduces a new ranking parameter r:

$$r = \frac{P(n|\mu = \lambda + \nu)}{P(n|\hat{\mu})} \tag{23}$$

Where $\hat{\mu}$ is the value, which maximizes the outcomes probability. $\hat{\mu} \leq \lambda$ always holds. Similar to the smallest interval approach the rank is used to accumulate the interval. For this exercise we first determine the Neyman Graph for μ afterwards we can simply subtract λ (as long as $\nu \geq 0$) and get the Neyman Graph for ν (cf. figure **??**). For $n = 9$ we obtain: $6.35 \leq \mu \leq 12.81$ and $3.15 \leq \nu \leq 9.61$.
The Neyman Confidence Level Bands (Choosing smallest interval) for μ should be similar to those for exercise 7. However the bands for ν should be shifted downwards by λ (cf. figure 7). As predicted, for $n = 9$ we obtain: $6.50 \leq \mu \leq 13.30$ and $3.30 \leq \nu \leq 10.10$.

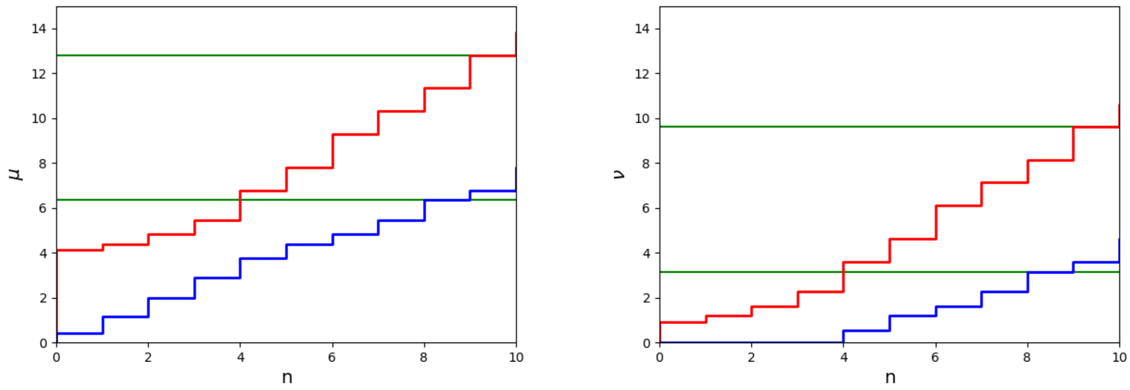


Figure 6: Feldman-Cousins 68% Confidence Level Bands. Left side: μ; right side: ν.

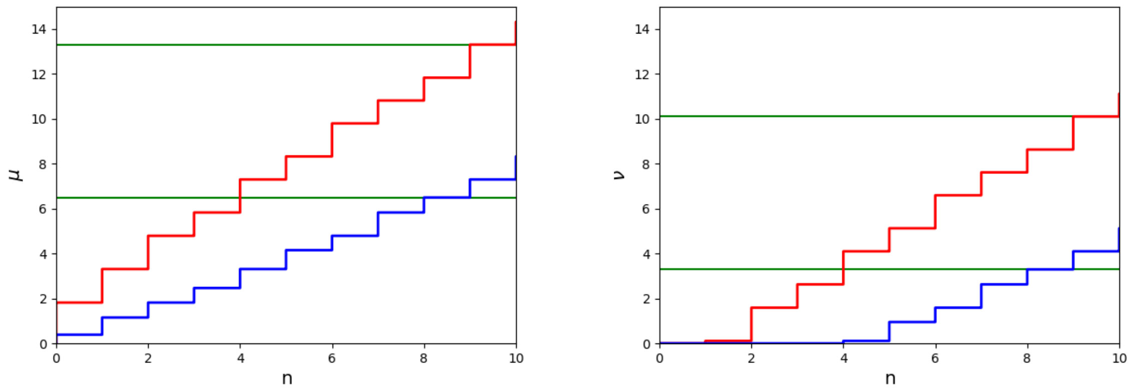


Figure 7: Neyman 68% Confidence Level Bands. Left side: μ; right side: ν.

Exercise 13

Infinitely fine binning

In this problem, we look at the relationship between an unbinned likelihood and a binned Poisson probability. We start with an one dimensional density $f(x|\lambda)$ depending on a parameter λ and defined and normalized in a range $[a, b].n$ events are measured with x values x_i $i = 1, 2, .., n$. The unbinned likelihood is defined as the product of the densities

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i|\lambda)\mathrm{d}x \tag{24}$$

Now we consider that the interval $[a, b]$ is divided into K subintervals (bins). Take for the expectation in bin j

$$\nu_j = \int_{\Delta_j} f(x|\lambda)\Delta \tag{25}$$

where the integral is over the x range in interval j, which is denoted as Δ_j. Define the probability of the data as the product of the Poisson probabilities in each bin. We consider the limit $K \to \infty$ and, if no two measurements have exactly the same value of x, then each bin will have either $n_j = 0$ or $n_j = 1$ event. Show that this leads to

$$\lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} = \prod_{i=1}^{n} f(x_i|\lambda)\Delta \tag{26}$$

where Δ is the size of the interval in x assumed fixed for all j. I.e., the unbinned likelihood is proportional to the limit of the product of Poisson probabilities for an infinitely fine binning.

If we $K \to \infty$, then $\Delta \to 0$ and equation 25 simplifies to

$$\nu_j \approx \underbrace{f(x_j)|\lambda}_{:=f_j}\Delta$$

Since n_j is either 0 or 1, $n_j! = 1$ and we get:

$$\begin{aligned}
\lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} &= \lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-f_j\Delta}(f_j\Delta)^{n_j}}{1} \\
&= \lim_{K\to\infty} \prod_{j=1}^{K} (1 - f_j\Delta)(f_j\Delta)^{n_j} \\
&= \lim_{K\to\infty} \prod_{j=1}^{K} \left((f_j\Delta)^{n_j} - (f_j\Delta)^{n_j+1}\right)
\end{aligned} \tag{27}$$

In the second step we applied $e^{x+1} \approx x + 1$, since $f_j\Delta$ is very small. Since $f_j\Delta$ is very small and n_j is either 0 or 1, we can neglect the second part in the product. For an infinite number of bins n_j can only yield 0 or 1. Now there are two cases:

1. $n_j = 0$: The factor is 1

2. $n_j = 1$: The factor is $f_j \Delta = f(x_j|\lambda)\Delta$

This leads to the result:

$$lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} = \lim_{K\to\infty} \prod_{j=1}^{K} (f(n_j|\lambda)\Delta)^{n_j} = \prod_{i=1}^{n} f(x_i|\lambda)\Delta \qquad \text{q.e.d.} \tag{28}$$

Exercise 16

Thinned Poisson Process

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

$$X = \sum_{n=1}^{N} X_n \tag{29}$$

Show that the probability distribution is given by

$$P(X) = \frac{e^{-\nu p}(\nu p)^X}{X!} \tag{30}$$

The Poisson distribution has the rate $R = \frac{\nu}{T}$. We divide the time T into n intervals of length $\Delta t = \frac{T}{n}$. The probability to get no success $(X = 0)$ in time T is

$$\begin{aligned}
P(\text{no success in } T) &= P(\text{no success in } \Delta t)^n \\
&= \Big((1 - R\Delta t) + R\Delta t(1-p)\Big)^n \\
&= (1 - pR\Delta t)^n \\
\Rightarrow \lim_{n\to\infty}(1 - pR\Delta t)^n &= \lim_{n\to\infty}\left(1 - \frac{pRT}{n}\right)^n \\
&= e^{-pRT} = e^{-p\nu}
\end{aligned}$$

Now we obtain the probability to get one success in the time span T. Imagine that the success occurs at time t in the time interval dt. We can decompose the time interval T into three parts: The time from $0 \to t$, where no success occurred, the time from $t \to t + \mathrm{d}t$, where we have an event, and the time from $t + \mathrm{d}t \to T$, where again no success occurred. The probability for this sequence is:

$$P(\text{one success at time } t) = e^{-pRt}(pR\mathrm{d}t)e^{-pR(T-(t+\mathrm{d}t))} = e^{-pRT}e^{pR\mathrm{d}t}pR\mathrm{d}t \tag{31}$$

We assume the success occurs in infinitesimal time, so that $e^{pR\mathrm{d}t} = 1$, and then integrate over all possible times t within T when the success could have occurred:

$$P(\text{one success in } T) = \int_0^T e^{-pRT}pR\mathrm{d}t = pRTe^{-pRT} \tag{32}$$

We now consider the case for X successes in the time interval T. Imagine the successes occurred at times $t_1, t_2, ..., t_X$. We have probabilities $pR\mathrm{d}t_1, pR\mathrm{d}t_2, ..., pR\mathrm{d}t_X$ for the successes that occurred. We now have to integrate over all possible times, taking care that success n occurred after success $n-1$:

$$
\begin{aligned}
P(X|p,R,T) &= e^{-pRT}\int_0^T pR\mathrm{d}t_1 \int_{t_1}^T pR\mathrm{d}t_2 \,...\int_{t_{X-1}}^T pR\mathrm{d}t_X \\
\overset{\text{see lecture}}{=} ... \quad &= e^{-pRT}\frac{p^X R^X T^X}{X!} \\
\Rightarrow P(X) &= \frac{e^{-\nu p}(\nu p)^X}{X!} \qquad \text{q.e.d.}
\end{aligned}
$$

4. Gauss Distribution

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

$$p(x) = \lambda e^{-\lambda x} \tag{33}$$

Then, try out the CLT for at least 3 different choices of n and λ and discuss the results. To generate random numbers according to the exponential distribution, you can use

$$x = -\frac{ln(U)}{\lambda} \tag{34}$$

Where U is a uniformly distributed random number between $[0, 1)$.

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

$$f(x) = \frac{1}{\pi\gamma}\frac{\gamma^2}{(x - x_0)^2 + \gamma^2} \tag{35}$$

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

$$x = \gamma \tan(\pi U - \pi/2) + x_0 \tag{36}$$

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.

According to the Central Limit Theorem $\mu = E[\bar{x}] = E[x]$:

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} x\lambda e^{-\lambda x} \cdot \begin{cases} 1 & \text{for} \quad x \geq 0 \\ 0 & \text{for} \quad x \leq 0 \end{cases} \mathrm{d}x \\ &= \int_0^{\infty} x\lambda e^{-\lambda x} \mathrm{d}x = \frac{1}{\lambda} \end{aligned} \tag{37}$$

In order to get $\sigma_x^2 = E[x^2] - E[x]^2$ we calculate:

$$\begin{aligned} E[x^2] &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda} \\ \Rightarrow \sigma_x &= \sqrt{\frac{2}{\lambda^2} - \frac{1}{\lambda^2}} = \frac{1}{\lambda} \end{aligned} \tag{38}$$

Further the CLT states:

$$\sigma_z = \frac{\sigma_x}{\sqrt{n}} = \frac{1}{\lambda\sqrt{n}} \tag{39}$$

Let us consider we make several identical experiments with n independent measurements of the same real-valued quantity x, which has a probability density given in equation 33. The Central Limit Theorem stats, that with increasing number of experiments the (binned) probability distribution of the means of the experiment $\bar{x}$ converges to the Gaussian function $\mathcal{G}(x|\mu, \sigma_z)$.

$$\mathcal{G}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(\mu-x)^2}{\sigma^2}} \tag{40}$$

In order to show this, we plot a normalized histogram of randomly generated and with equation 34 evaluated numbers. This should merge quite good the estimated Gaussian function. In figure 8 the results for 10000 experiments with n measurements and parameter λ are shown. The distribution is separated in 50 bins. Although it is not a proof, the graphs confirm the Central Limit Theorem. The convergence is clearly shown.

If we want to calculate the expectation value of 35, we recognize that the integral diverges. The Cauchy Distribution has no well-defined expectation value. With the same argument the Variance can not be determined. A fixed variance and mean are required for the Central Limit Theorem. Therefor we can not apply it to the Cauchy distribution. The Cauchy Distribution does not fall quick enough in the limits $-\infty$ and ∞. There is a residual chance to get random numbers far away from the maximum x_0. In figure 9 on the right, only a small interval of the whole histogram is shown. There are also bars in the left and right regime. However those bars are not extended very much and can not be seen on the scale of the graph. Recommendable is that a much higher bin count was needed since the bars are distributed over a large range and thus every bin would be very wide. However the bin size alternates with every new set of random numbers, since the deviation from the mode is very high and the bin number was fixed.

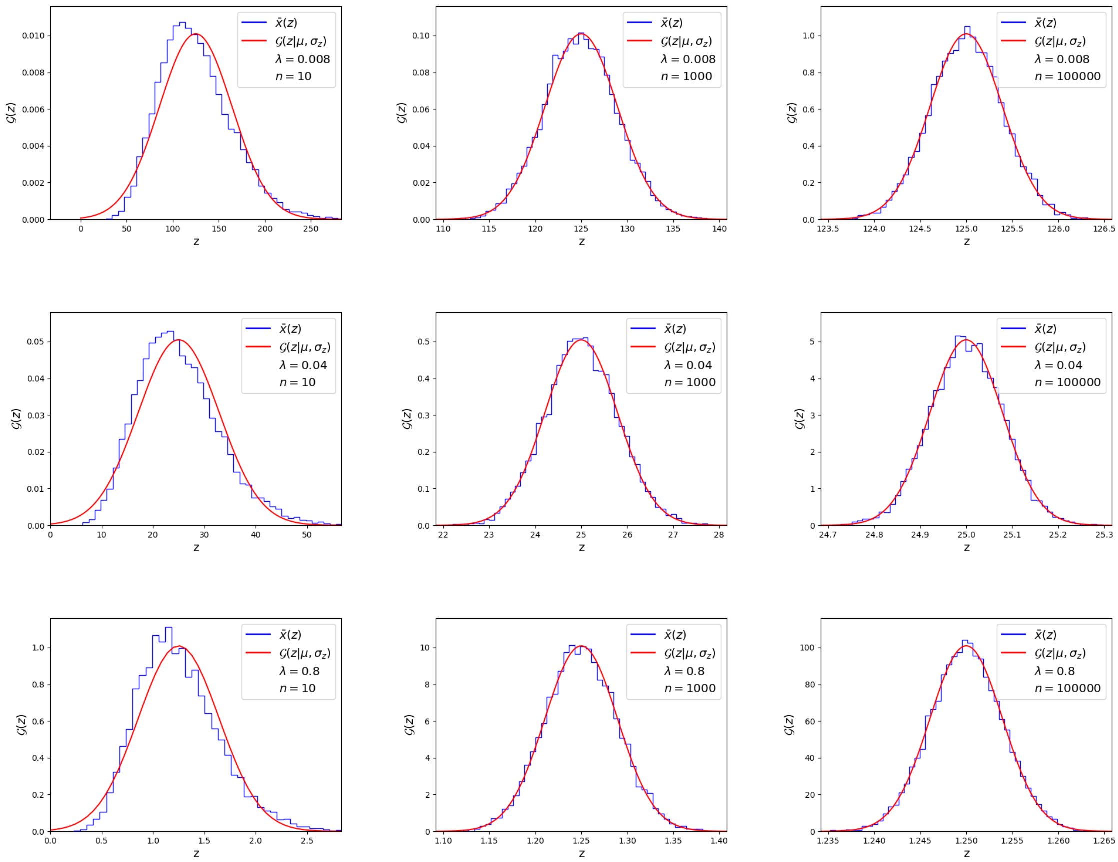


Figure 8: Distributions of $\bar{x}$ for several λ and n compared to the expectation from the Central Limit Theorem

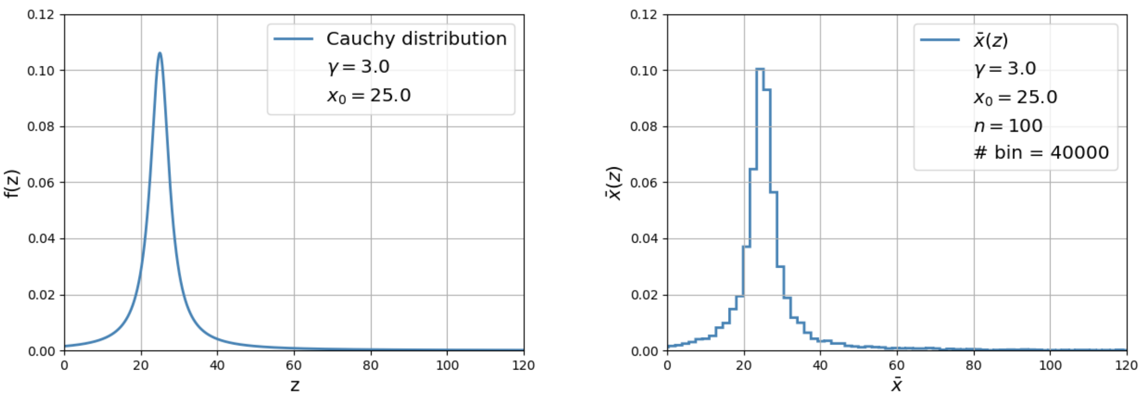


Figure 9: Cauchy Distribution (left) and distribution of mean (right)

Exercise 11
Contours of Bivariate Gauss Function

With a plotting program, draw contours of the bivariate Gauss function for the following parameters:

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

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

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

The 2D Bivariate Gauss Distribution is given as:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left(-\frac{1}{2(1-\rho_{xy}^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho_{xy}xy}{\sigma_x\sigma_y^2}\right)\right) \tag{41}$$

The contours of this function are shown in figure 10.

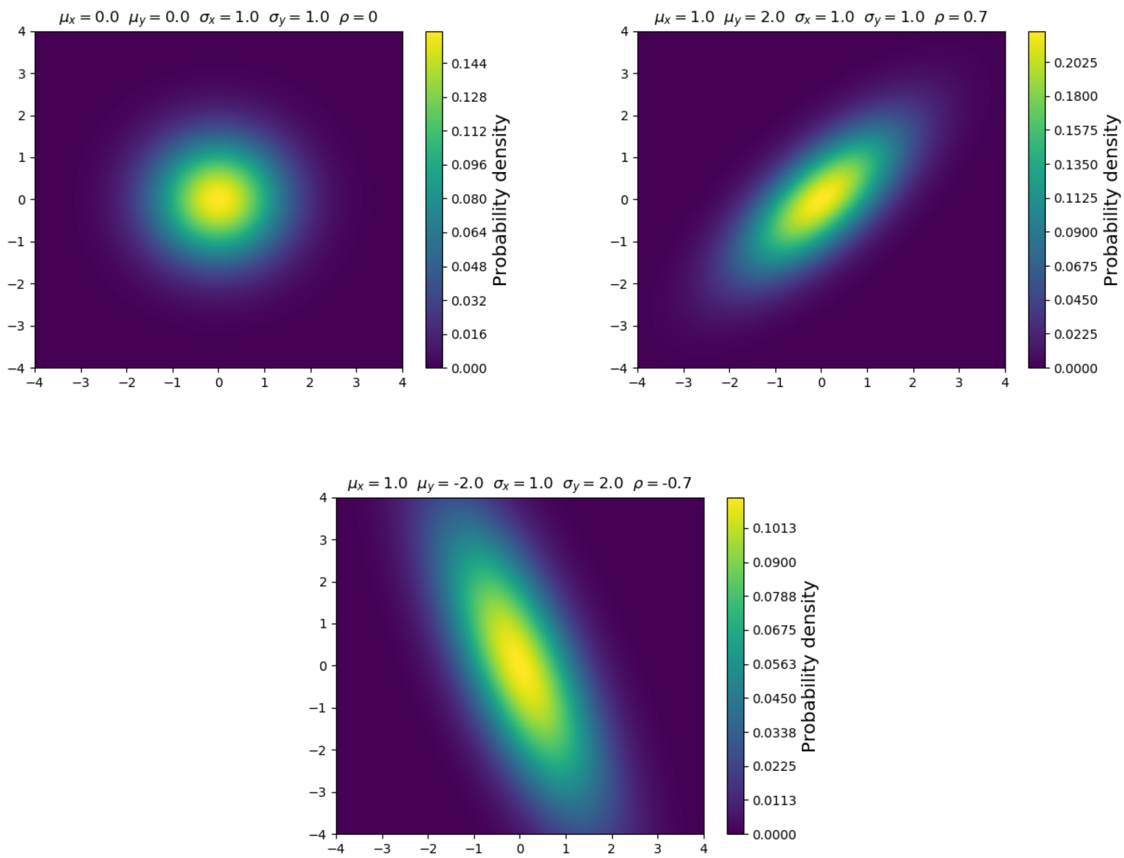


Figure 10: Contours of Bivariate Gauss Function

Exercise 12
Bivariate Gauss Distribution

Bivariate Gauss probability distribution

a) Show that the pdf can be written in the form shown in equation 41.

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

$$\begin{aligned} \mu_z &= \mu_x - \mu_y \\ \sigma_z^2 &= \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \end{aligned}$$

We start with the expression of the joint probability distribution for n correlated Gauss variables

$$P(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})} \tag{42}$$

The covariance matrix is defined as follows

$$\Sigma_{ij} = cov(x_i, x_j) \tag{43}$$

with the covariance:

$$cov(x, y) = E[(x - E[x])(y - E[y])]_{P(x,y)} = cov(y, x) \tag{44}$$

The definition of correlation coefficient ρ_{xy} is

$$\rho_{xy} = \frac{cov(x, y)}{\sigma_x\sigma_y} \tag{45}$$

Since we consider the two-dimensional case, the calculations for the determinant simplifies to

$$|\Sigma| = cov(x, x)cov(y, y) - cov(x, y)cov(y, x) = cov(x, x)cov(y, y) - cov(x, y)^2 \tag{46}$$

With equations 6 and 5 we see

$$\begin{aligned} cov(x, x) &= V[x] \\ \Rightarrow \sigma_x^2 &= cov(x, x) \end{aligned}$$

so that equation 46 can be rewritten using equation 45.

$$\begin{aligned} |\Sigma| &= \sigma_x^2\sigma_y^2 - \sigma_x^2\sigma_y^2\rho_{xy}^2 \\ &= \sigma_x^2\sigma_y^2(1 - \rho_{xy}^2) \end{aligned} \tag{47}$$

Again due to the two-dimensional case, the matrix inverse simplifies

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} cov(y, y) & -cov(y, x) \\ -cov(x, y) & cov(x, x) \end{pmatrix} \tag{48}$$

so that we can calculate the exponent of equation 42

$$\begin{aligned}
\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix} &= \frac{1}{|\Sigma|} \begin{pmatrix} (x-\bar{x})cov(y,y) - (y-\bar{y})cov(y,x) \\ (y-\bar{y})cov(x,x) - (x.\bar{x}cov(x,y) \end{pmatrix}^T \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix} \\
&= \frac{1}{|\Sigma|}\Big((x-\bar{x})^2 cov(y,y) + (y-\bar{y})^2 cov(x,x) - 2(x-\bar{x})(y-\bar{y})cov(x,y)\Big) \\
&= \frac{1}{(1-\rho_{xy}^2)}\left(\frac{(x-\bar{x})^2}{\sigma_x^2} + \frac{(y-\bar{y})^2}{\sigma_y^2} - \frac{2(x-\bar{x})(y-\bar{y})\rho_{xy}}{\sigma_x\sigma_y}\right) \qquad (49)
\end{aligned}$$

Assuming $\bar{x} = \bar{y} = 0$ and setting in equations 47 and 49 into equation 42 with $n = 2$, this leads to the shown form:

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left(-\frac{1}{2(1-\rho_{xy}^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho_{xy}xy}{\sigma_x\sigma_y^2}\right)\right) \quad \text{q.e.d.} \qquad (50)$$

For the second part of the exercise we replace y by $x - z \Rightarrow P(x,y) \rightarrow P(z,x)$. As mentioned in the tutorial the shift due to μ_x and μ_y leads to and prefactor of the Gauss function and causes she shift μ_z in the exponent. Then the Gauss distribution of z is calculated with:

$$
\begin{aligned}
P(z,x) &= \int_{-\infty}^{\infty} P(z,x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \underbrace{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}}_{:=A=\text{const.}} \exp\Bigg(- \underbrace{\frac{1}{2(1-\rho^2)}}_{:=B=\text{const.}} \underbrace{\Bigg(\frac{x^2}{\sigma_x^2} + \frac{(x-z)^2}{\sigma_y^2} - \frac{2\rho(x-z)}{\sigma_x\sigma_y} \Bigg)}_{:=C(x,z)} \Bigg)\mathrm{d}x
\end{aligned}
$$

$$
\begin{aligned}
C &= \frac{x^2}{\sigma_x^2} + \frac{x^2}{\sigma_y^2} + \frac{z^2}{\sigma_y^2} - \frac{2xz}{\sigma_y^2} - \frac{2\rho x^2}{\sigma_x\sigma_y} + \frac{2\rho zx}{\sigma_x\sigma_y} \\
&= x^2 \underbrace{\Bigg(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} - \frac{2\rho}{\sigma_x\sigma_y} \Bigg)}_{:=D=\text{const.}} + 2x \underbrace{\Bigg(\frac{\rho z}{\sigma_x\sigma_y} - \frac{z}{\sigma_y^2} \Bigg)}_{:=E(z)} + \frac{z^2}{\sigma_y^2} \\
&= D\Bigg(x^2 + 2x\frac{E}{D} + \frac{E^2}{D^2} - \frac{E^2}{D^2} + \frac{z^2}{D\sigma_y^2} \Bigg) \\
&= D\Bigg(x + \frac{E}{D} \Bigg)^2 - \frac{E^2}{D} + \frac{z^2}{\sigma_y^2}
\end{aligned}
$$

$$
\begin{aligned}
P(z) &= \int_{-\infty}^{\infty} \underbrace{A \exp\Bigg(- B\Bigg(\frac{z^2}{\sigma_y^2} - \frac{E^2}{D} \Bigg)\Bigg)}_{:=F(z)} \exp\Bigg(- BD\Bigg(x + \frac{E}{D} \Bigg)^2 \Bigg)\mathrm{d}x && x+\tfrac{E}{D}\to x \\
&= AF(z) \int_{-\infty}^{\infty} \exp\Big(- BDx^2 \Big)\mathrm{d}x && \textstyle\int_{-\infty}^{\infty}\exp(-\alpha x^2)\mathrm{d}x = \sqrt{\frac{\pi}{\alpha}} \\
&= AF(z)\sqrt{\frac{\pi}{BD}} \\
&= \frac{1}{2\pi\cancel{\sigma_x\sigma_y\sqrt{1-\rho^2}}} \sqrt{\frac{2\pi\cancel{(1-\rho^2)}}{\cancel{\frac{1}{\sigma_x^2\sigma_y^2}} \underbrace{(\sigma_y^2+\sigma_x^2-2\rho\sigma_x\sigma_y)}_{=\sigma_z^2}}} F(z) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Bigg(- B\Bigg(\frac{z^2}{\sigma_y^2} - \frac{E^2}{D} \Bigg)\Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Bigg(- B\Bigg(\frac{\Big(\frac{\rho z}{\sigma_x\sigma_y} - \frac{z}{\sigma_y} \Big)^2}{\frac{1}{\sigma_x^2\sigma_y^2} \underbrace{\sigma_y^2\sigma_x^2 - 2\rho\sigma_x\sigma_y}_{=\sigma_x}} \Bigg)\Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Bigg(- B\Bigg(\frac{z^2}{\sigma_y^2} - \Bigg(\frac{\frac{\rho^2 z^2}{\sigma_x^2\sigma_y^2} + \frac{z^2}{\sigma_y^4} - \frac{2\rho z^2}{\sigma_x\sigma_y^3}}{\frac{\sigma_z^2}{\sigma_x^2\sigma_y^2}} \Bigg)\Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Big(- B\frac{z^2}{\sigma_y^2\sigma_z^2}\Big(\sigma_z^2 - \rho^2\sigma_y^2 - \sigma_x^2 + 2\rho\sigma_x\sigma_y \Big)\Big) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Bigg(- \frac{z^2}{2\sigma_z^2}\frac{1}{\sigma_y^2(1-\rho^2)}\Big(\cancel{\sigma_x^2} + \sigma_y^2 - \cancel{2\rho\sigma_x\sigma_y} - \rho^2\sigma_y^2 - \cancel{\sigma_x^2} + \cancel{2\rho\sigma_x\sigma_y} \Big)\Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\Bigg(- \frac{z^2}{2\sigma_z^2}\frac{\cancel{\sigma_y^2(1-\rho^2)}}{\cancel{\sigma_y^2(1-\rho^2)}} \Bigg) = \quad \mathcal{G}(z|\mu_z,\sigma_z) \qquad \text{q.e.d.}
\end{aligned}
$$

Exercise 13
Convolution of Gaussians

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-x_0)^2}{2\sigma_x^2}} \tag{51}$$

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

$$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-x)^2}{2\sigma_y^2}} \tag{52}$$

What is the predicted distribution for the observed quantity y?

The convolution of two functions $f(x)$ and $g(x)$ is defined by

$$(f(x) * g(x))(x) = \int_{-\infty}^{\infty} f(\tau)g(x-\tau)\mathrm{d}\tau \tag{53}$$

A useful property is, that the Fourier transformation of the convolution of two functions is the same as the product of their Fourier transformation

$$\mathcal{F}(f(x) * g(x))(k) = (\mathcal{F}f(x))(k) \cdot (\mathcal{F}g(x))(k) \tag{54}$$

with the the Fourier Transformation $(\mathcal{F}f)(k)$, which is, except for a prefactor the same as the characteristic function $\phi(k)$.

$$(\mathcal{F}f)(k) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} f(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\phi(k) \tag{55}$$

The Fourier transformation of a Gauss function is:

$$\begin{aligned}(\mathcal{F}\mathcal{G}(x))(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2\sigma^2 + ik\mu}\end{aligned} \tag{56}$$

This allows us to simplify the calculation for the convolution of the two functions:

$$\begin{aligned}(f(x) * P(y|x))(y,x) &= \mathcal{F}^{-1}((\mathcal{F}f(x))(k) \cdot (\mathcal{F}P(y|x))(k)) \\ &= \left(\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2\sigma_x^2 + ikx_0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2\sigma_y^2 + ikx}\right)\right)(k) \\ &= \left(\mathcal{F}^{-1}\left(\frac{1}{2\pi} e^{-\frac{1}{2}k^2(\sigma_x^2+\sigma_y^2) + ik(x_0+x)}\right)\right)(k) \\ &= \frac{1}{2\pi\sqrt{\sigma_x^2+\sigma_y^2}} e^{-\frac{1}{2}\frac{(y-x-x_0)^2}{2(\sigma_x^2+\sigma_y^2)}} = P(x,y)\end{aligned}$$

Exercise 14
Fitting data

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

θ	30°	45°	90°	120°	150°
Cross section	11	13	17	17	14
Error	1.5	1.0	2.0	2.0	1.5

The units of cross section are $10^{-30}\mathrm{cm}^2/\mathrm{steradian}$. Assume the quoted errors correspond to one Gaussian standard deviation. The assumed model has the form

$$f(\theta) = A + B\cos(\theta) + C\sin(\theta^2) \tag{57}$$

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?

The equation of the posterior is {theta}

$$P(\vec{\lambda}|\{r\},\{\sigma\}) = \frac{\mathcal{G}(\{r\}|\vec{\lambda},\{\sigma\})P_0(\vec{\lambda})}{\int \mathcal{G}(\{r\}|\vec{\lambda},\{\sigma\})P_0(\vec{\lambda})\mathrm{d}\vec{\lambda}} = \frac{\mathcal{G}(\{r\}|f(\{\theta\}|\vec{\lambda}),\{\sigma\})}{\int \mathcal{G}(\{r\}|f(\{\theta\}|\vec{\lambda}),\{\sigma\})\mathrm{d}\vec{\lambda}} \tag{58}$$

with $\vec{\lambda}$ representing the parameters $(a,b,c)^T$, $\{r\}$ representing the measured cross sections, $\{\theta\}$ representing the measurement points and $\{\sigma\}$ representing the errors. In order to get the best values for A,B,C, we are only interested in the maximum of the posterior. Furthermore the variables are independent and the covariance matrix simplifies to the diagonal matrix $\Sigma_{ij} = \sigma_i^2$. This leads to a simplified posterior function with n data points:

$$P'(\vec{\lambda}|\vec{r},\vec{\sigma}) = \frac{1}{(2\pi)^{n/2}\prod_{i=1}^n \sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{(r_i - f(\theta_i|\vec{\lambda}))^2}{\sigma_i^2}\right) \tag{59}$$

In order to obtain $\lambda = (A,B,C)^T$, a 3D-grid was implemented. The position with the highest probability density determines the parameters A,B,C. However the theory function $f(\theta|\vec{\lambda})$ does not fit the data points well. The resulting fit with the according parameters are shown in figure 11.

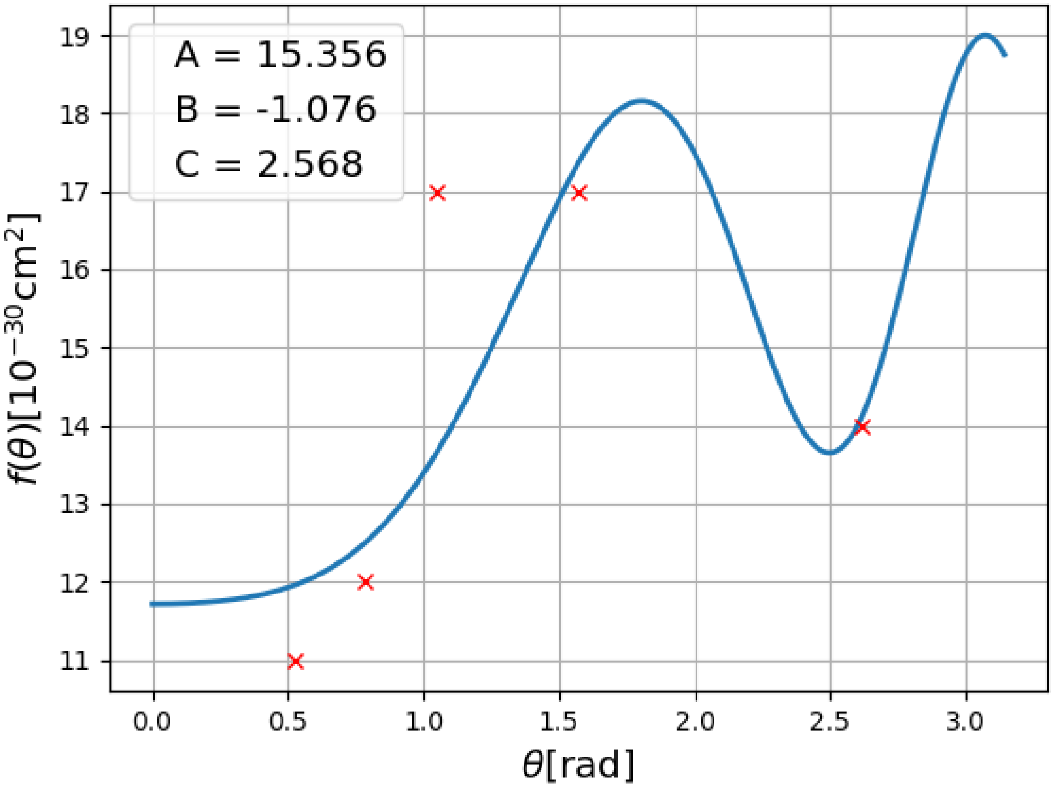


Figure 11: Fit of Cross section with parameters A, B, C

5. Gauss Distribution

Exercise 1
Fitting Sigmoid function

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

Energy (E_i)	Trials(N_i)	Successes (r_i)
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

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).

The posterior probability distribution is

$$P(\lambda|\{r\},\{N\}) = \frac{P(\{r\}|\{N\},\lambda)P_0(\lambda)}{\int P(\{r\}|\{N\},\lambda)P_0(\lambda)\mathrm{d}\lambda} \tag{60}$$

with

$$P(\{r\}|\{N\},\lambda)P_0(\lambda) = \prod_{i=1}^{k} \binom{N_i}{r_i} f(x_i|\lambda)^{r_i}(1-f(x_i|\lambda))^{N_i-r_i} \tag{61}$$

where λ represents the two parameters (A, E_0) of the modifyed Sigmoid function:

$$f(E) = \frac{1}{1+e^{-A(E-E_0)}} \tag{62}$$

The Sigmoid function reaches 0.5, if the argument is zero. For setting up the prior P_0, we assume that for $E_0 = 2.0$ the efficiency is about 0.5. So we choose a Gauss function around $E_0 = 2.0$. As standard deviation we choose 0.3. In order to get the prior for A we consider the derivative of equation 62:

$$\frac{\mathrm{d}f}{\mathrm{d}E} = \frac{Ae^{-A(E-E_0)}}{(1+e^{-A(E-E_0)})^2} = \frac{A}{4} \quad \text{for } E = E_0 \tag{63}$$

Since for an energy change of 0.5 the efficiency changes by around 30%, we conclude for A:

$$0.5\frac{A}{4} \approx 0.3 \Rightarrow A \approx 2.4 \tag{64}$$

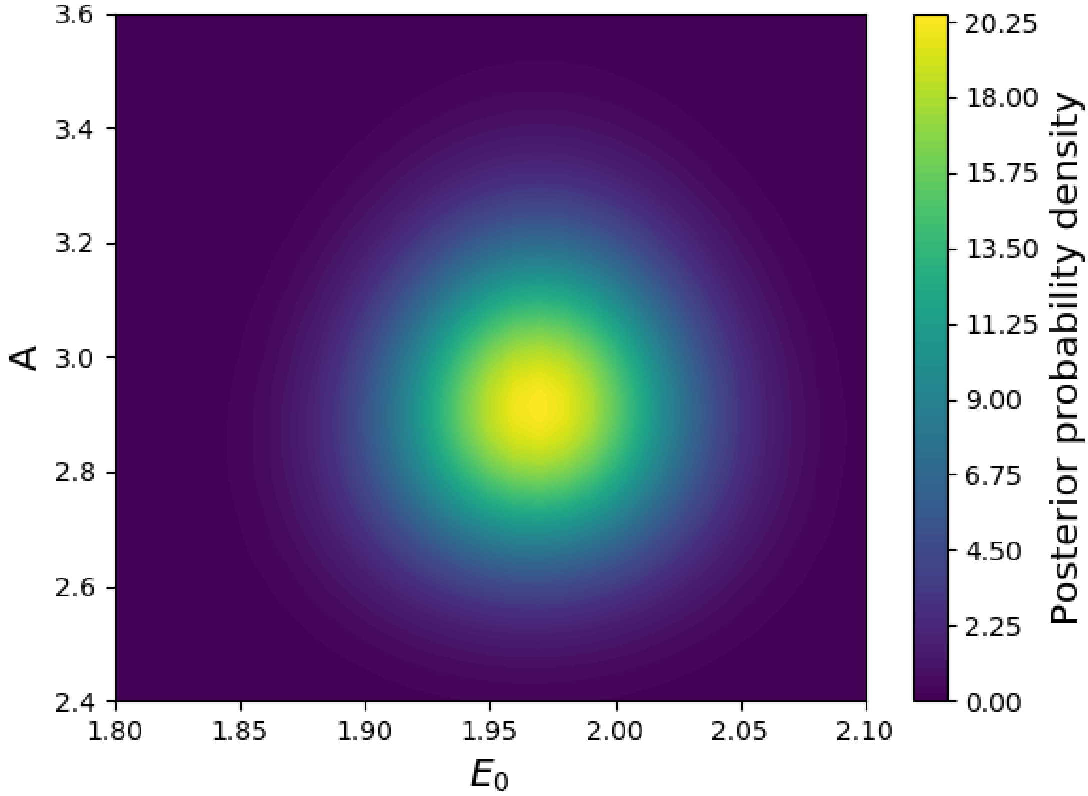


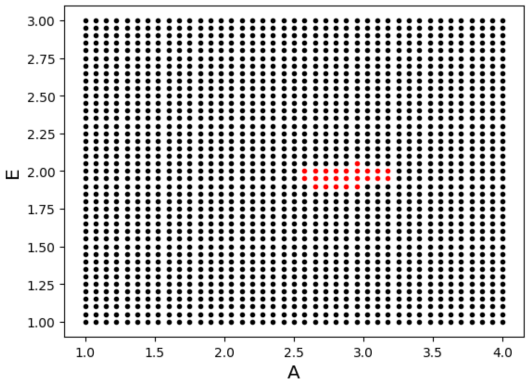
Figure 12: Contour plot of Posterior probability distribution for parameters (A, E_0)

Again we suppose a Gauss distributed prior with $\mu = 2.4$ and $\sigma = 0.5$. The standard deviation again was chosen arbitrarily. However we have a reasonably good information concerning its best value. A contour plot of the posterior probability density is shown in figure 12.

As test statistic we use the one given in the lecture:

$$\xi(\{r_i\}; \lambda) \equiv \prod_{i=k}^{k} \binom{N_i}{r_i} p_i(\lambda)^{r_i} (1 - p_i(\lambda))^{N_i - r_i} \tag{65}$$

Then we set up a 2D grid for possible values of A and E_0. Each grid point represents a fixed tuple of parameters. For each grid point we calculate the success probability p_i via equation 62. Afterwards 1000 experiments are performed in the following way: For each energy an r_i is randomly generated according to the binomial probability distribution. These values are used to calculate ξ with equation 65. All calculated values for ξ are stored an a list and sorted. Then the value of ξ that is succeeded by 68% of measurements is determined. This grid point is in the 68% confidence level, if ξ^{Data} is in the accepted range. The confidence levels are shown in figure 13 and a plot of the fit with the most probable parameters is shown in figure 14.



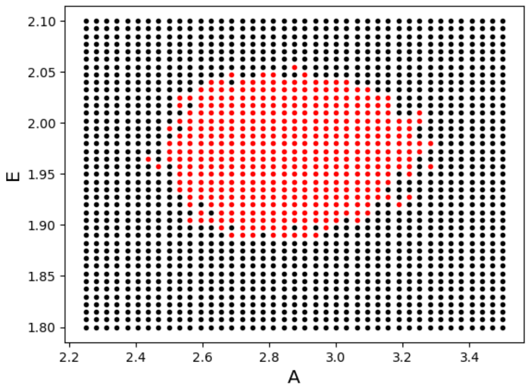


Figure 13: 68% confidence levels. Coarse grid (left) and fine grid (right)

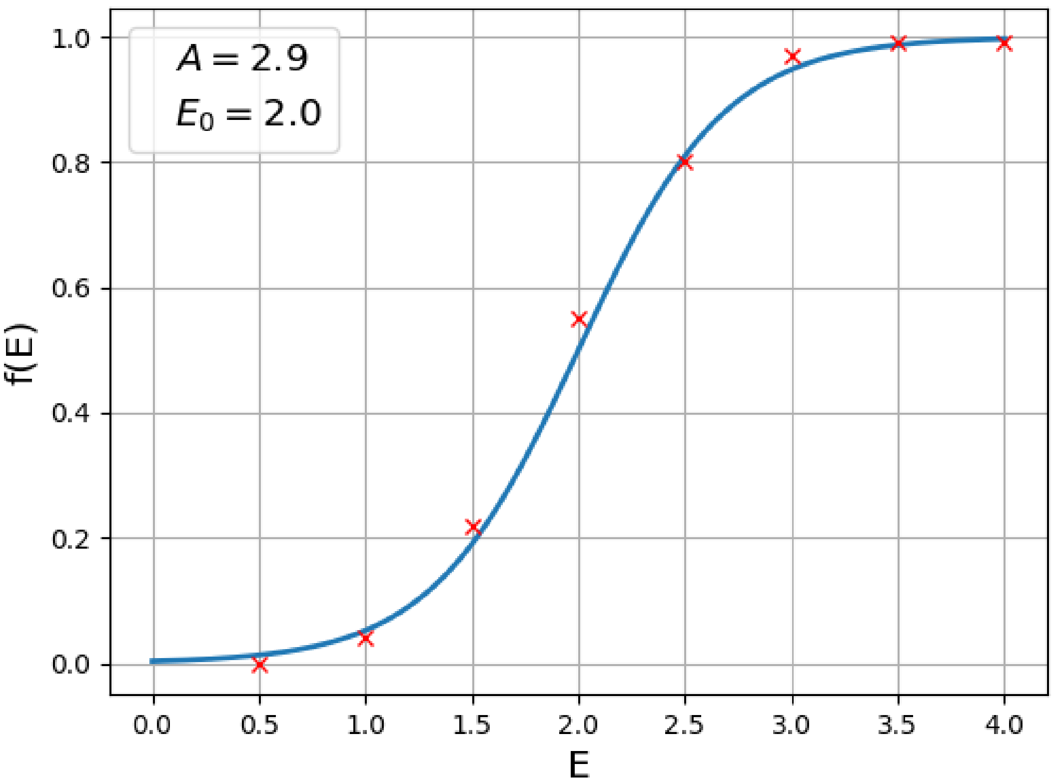


Figure 14: Fit of sigmoid function

Exercise 2
Fitting sinus

Repeat the analysis of the data given in Exercise 1 of chapter 5 with the function:

$$f(E) = \sin(A(E - E_0)) \tag{66}$$

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

b) Find the 68% CL region for (A, E_0)

c) Discuss the result

In a first step we have to guess the priors. Again we set up gaussian functions. For $E = 0$ the function should return 0. Thus we define $E_0 = 0$. Furthermore the function should return a positive number in the interval [0,1] for Energys in $[0, 4]$. This means that we should only examine a quarter period of the sinus function: $A(E-E_0) = A \cdot 4 \overset{!}{=} \frac{\pi}{2} \Rightarrow A = \frac{\pi}{8}$. The function seems to be a poor approximation since it does not change its curvature on the given interval. Therefore we choose higher standard deviations than in Exercise 1: $\sigma_A = 1.0; \sigma_{E_0} = 0.8$. In addition we modify equation 66: For $\sin(A(E - E_0)) < 0$: $f(E) = 0$, because an efficiency smaller than zero is unphysical. The contour plot of the posterior is shown in figure 15. In order to obtain the 68% CL, we implement a list with

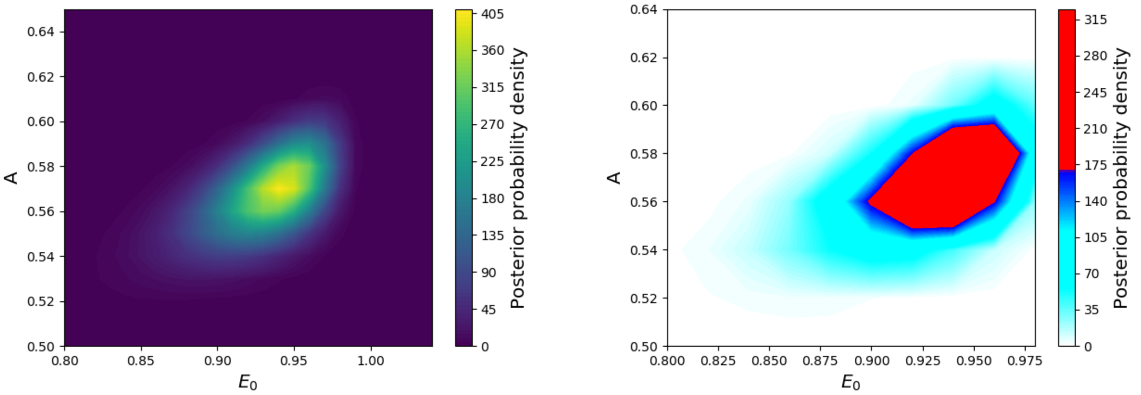


Figure 15: Contour plots of posterior probability distribution for parameters (A, E_0). Red marks the 68% confidence level.

all probabilities for each grid point and sort it. The entries are accumulated as long as the sum does not exceed 0.68. Then the probability of the last added grid point gives the contour whose area tags the 68% confidence level. The Posterior probability distribution is shown in figure 15 and the fit is shown in figure 16.

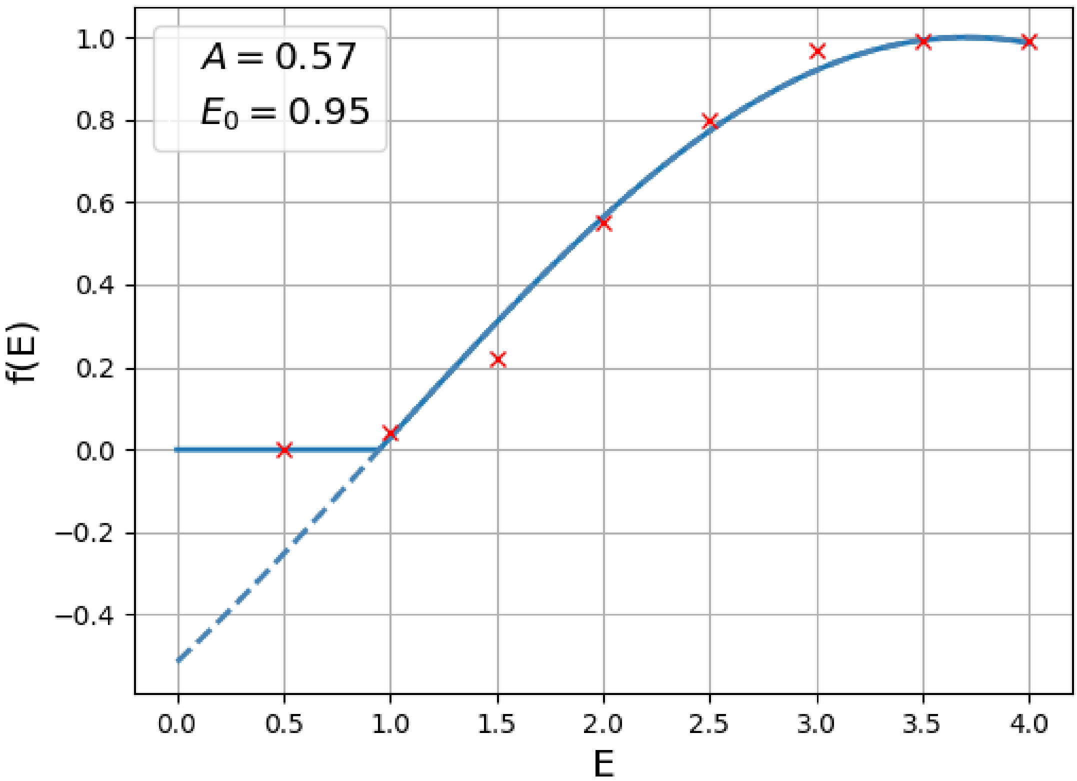


Figure 16: Fit of theory function $f(E)$

Exercise 3

Probability distribution of χ^2

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

First we derive the probability distribution for χ^2. Let us change the variable $\chi^2 = y$. This Change is not monotonic, because every value of χ has two corresponding values of y (one positive and negative). However, because of symmetry, both halves will transform identically and we add a factor 2. According to the change of variable formula

$$\begin{aligned}
P(\chi^2)|\frac{\mathrm{d}\chi^2}{\mathrm{d}y}| &= 2P(y) \qquad \text{with } y \geq f(x|\lambda) \\
\frac{\mathrm{d}\chi^2}{\mathrm{d}y} &= \frac{2(y - f(x|\lambda))}{\sigma^2} \\
\Rightarrow \left|\frac{\mathrm{d}\chi^2}{\mathrm{d}y}\right| &= \frac{2\sqrt{\chi^2}}{\sigma} \\
\Rightarrow P(\chi^2) &= 2\frac{\sigma}{2\sqrt{\chi^2}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\chi^2} \\
&= \frac{1}{\sqrt{2\pi\chi^2}}e^{-\frac{1}{2}\chi^2}
\end{aligned}$$

Now we can calculate the mean with equation 4:

$$\begin{aligned}
E[\chi^2] &= \int_0^\infty \chi^2 P(\chi^2)\mathrm{d}\chi^2 \\
&= \int_0^\infty \frac{\sqrt{\chi^2}}{\sqrt{2\pi}}e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \\
&= \frac{1}{\sqrt{2\pi}}\left[-2\sqrt{\chi^2}e^{-\frac{1}{2}\chi^2}\right]_0^\infty - \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{-2}{2\sqrt{\chi^2}}e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \qquad \text{p.I.} \\
&= [0 - 0] + \int_0^\infty \frac{1}{\sqrt{2\pi\chi^2}}e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \\
&= \int_0^\infty P(\chi^2)\mathrm{d}\chi^2 = 1
\end{aligned}$$

The Variance is determined via equation 6:

$$\begin{aligned}
E[(\chi^2)^2] &= \int_0^\infty (\chi^2)^2 P((\chi^2)^2)\mathrm{d}\chi^2 \\
&= \frac{1}{\sqrt{2\pi}}\left[-2(\chi^2)^{\frac{3}{2}}e^{-\frac{1}{2}\chi^2}\right]_0^\infty - \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{-2\cdot 3}{2}\sqrt{\chi^2}e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \qquad \text{p.I.} \\
&= [0 - 0] + \frac{1}{\sqrt{2\pi}}\left[-6\sqrt{\chi^2}e^{-\frac{1}{2}\chi^2}\right]_0^\infty - \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{-2\cdot 3}{2\sqrt{\chi^2}}e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \qquad \text{p.I.} \\
&= 3\int_0^\infty P(\chi^2)\mathrm{d}\chi^2 = 3 \\
\Rightarrow V[\chi^2] &= 3 - 1 = 2
\end{aligned}$$

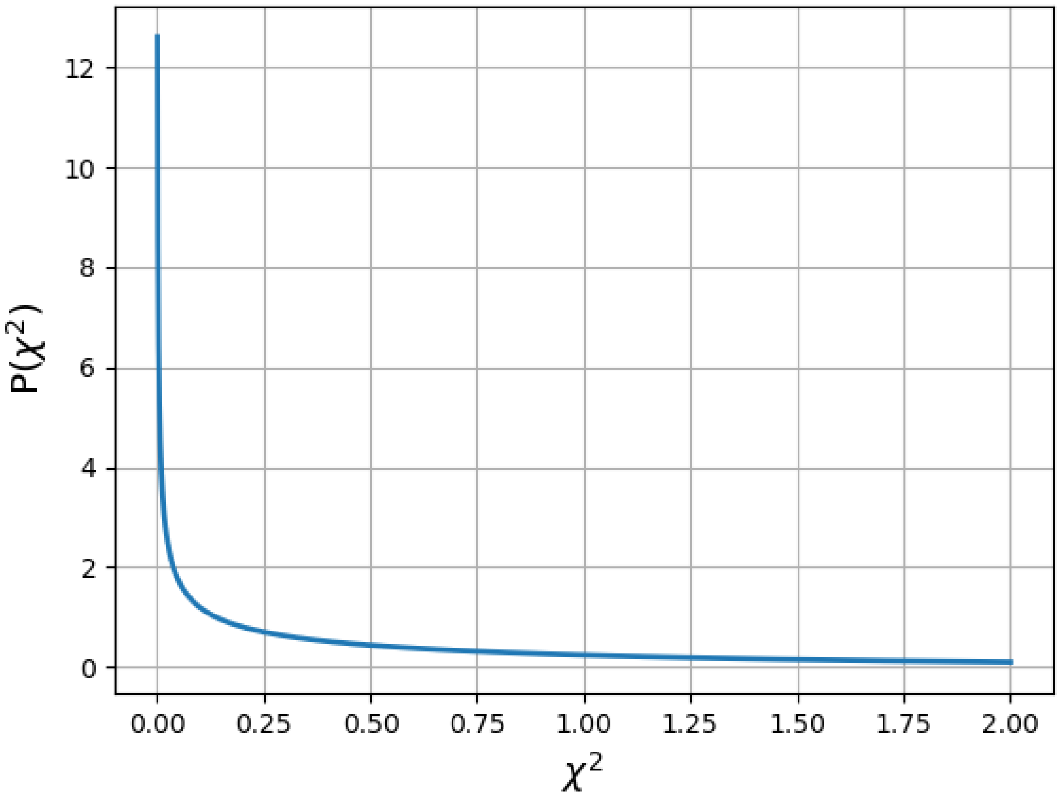


Figure 17: Probability distribution of χ^2

In order to get the mode we obtain the maximum of the pdf:

$$\frac{\mathrm{d}}{\mathrm{d}\chi^2}P(\chi^2) = -\frac{1}{2\sqrt{2\pi\chi^2}}e^{-\frac{1}{2}\chi^2}(1+\frac{1}{x}) < 0 \tag{67}$$

Since the derivative is smaller than zero for all χ^2, we have to search for the global maximum. Figure 17 shows that the mode of the pdf is at zero.

Exercise 8
Maximum Likelihood estimation

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:

a) quadratic, representing background only:

$$f(x|A, B, C) = A + Bx + Cx^2 \tag{68}$$

a) quadratic + Breit-Wigner representing background + signal:

$$f(x|A, B, C, x_0, \Gamma) = A + Bx + Cx^2 \frac{D}{(x - x_0)^2 + \Gamma^2 0} \tag{69}$$

b) Perform a chi-squared minimization fit, and find the best values of the parameters as well as the covariance matrix for the parameters. What is the p-value of the fits?

c) 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?

x	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
y	11.3	19.9	24.9	31.1	37.2	36.0	59.1	77.2	96.0
x	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
y	90.3	72.2	89.9	91.0	102.0	109.7	116.0	126.6	139.8

Assuming that the data can be modeled by a Gauss probability distribution we get for the first model:

$$f(x|A, B, C, D, \mu, \sigma_m) = A + Bx + Cx^2 + \frac{D}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_m^2}\right) \tag{70}$$

In order to obtain the optimum parameters, we perform a chi-squared minimization fit. Therefore the minimum of the following function has to be obtained with the number of data points n and the set of parameters λ. Pay attention that σ is the uncertainty of the data and σ_m is a parameter of the assumed model.

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i|\lambda))^2}{\sigma^2} \tag{71}$$

The minimization is done numerically with an n-dimensional grid. The results are shown in table 3 and figure 18. The covariance matrix is defined by:

parameter	Gauss	Wigner-Breigt
A	2.523	3.0
B	94.918	82.0
C	48.730	61.059
D	0.493	0.122
μ	0.049	0.493
σ_m	4.290	0.056
χ^2	12.478	16.015

Table 3: Results of χ^2-fit according to Gauss model and Wigner-Breigt model

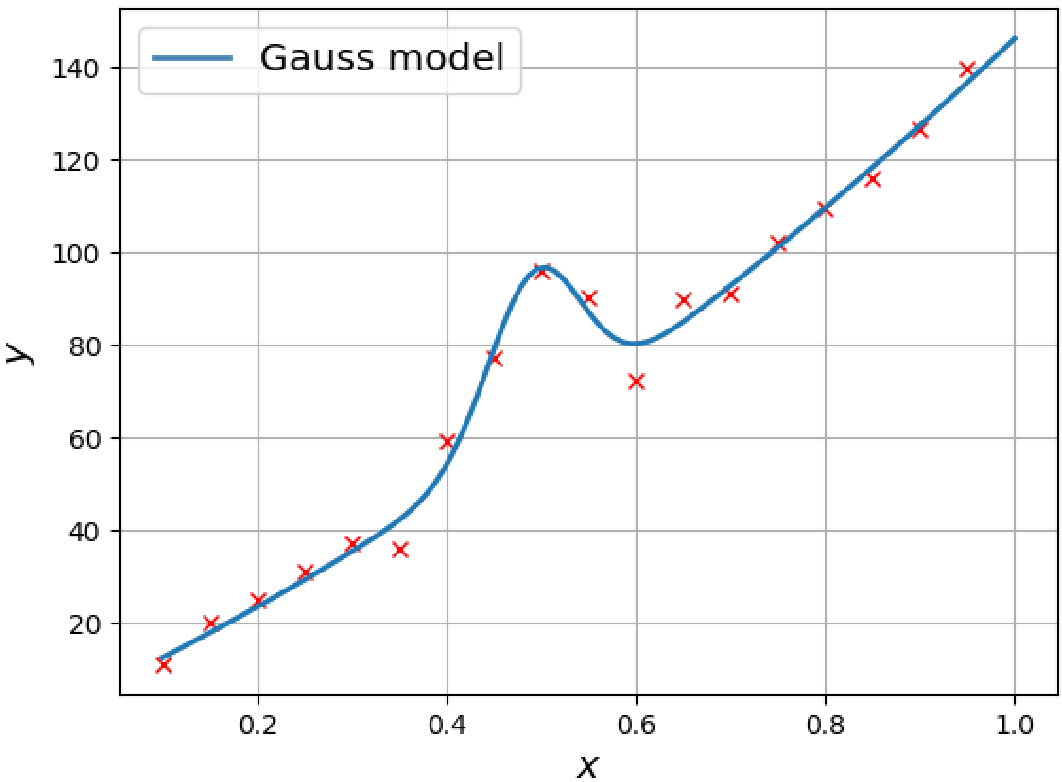


Figure 18: χ^2-fit of Gauss model

$$\Sigma_{ij}^{-1} \approx \frac{1}{2}\frac{\partial^2 \chi^2}{\partial\lambda_i \partial\lambda_j} \tag{72}$$

Since the second derivatives are very long terms, we use the numerical derivative at the optimum parameters λ. Here e_i and e_j is a six-dimensional unit vector in direction i and j respectively and h a very small number.

$$\frac{\partial^2 f(\lambda)}{\partial_i \partial_j} = \frac{f(\lambda + h(e_i + e_j)) + f(\lambda - h(e_i + e_j)) - f(\lambda + h(e_i - e_j)) - f(\lambda + h(e_j - e_i))}{(2h)^2} \tag{73}$$

As inverse matrix we get:

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} 2.245 & 1.182 & 0.776 & 0.0 & 0.0 & 2.501 \\ 1.182 & 0.570 & 10.725 & 0.0 & 1.234 & \\ 0.776 & 0.570 & 0.437 & 10.574 & 1.050 & 0.613 \\ 0.0 & 10.725 & 10.574 & 54110 & 11803 & 3.464 \\ 0.0 & 0.0 & 1.050 & 11803 & 99710 & -642.15 \\ 2.501 & 1.234 & 0.613 & 3.464 & -642.15 & 14.388 \end{pmatrix} \tag{74}$$

There are two entries with a negative value. The second derivative, i.e. the curvature should always be positive. So the matrix and its inverse is incorrect.
The p-value is defined as:

$$p = 1 - F(\xi^{exp}) = \int_{\xi^{exp}}^{\xi^{max}} P(\xi|\lambda)\mathrm{d}\xi \tag{75}$$

with ξ^{exp} the value obtained in the experiment for χ^2 and $\xi^{max} = \infty$, the maximum value for ξ. As Test statistic we use the probability distribution of χ^2:

$$P(\chi^2|\lambda) = \frac{(\chi^2)^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{1}{2}\chi^2} \tag{76}$$

Then the p-value gives

$$\begin{aligned} p &= \int_{\chi^{2,D}}^{\infty} \frac{(\chi^2)^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{1}{2}\chi^2}\mathrm{d}\chi^2 \\ &= 1 - \frac{\gamma\left(\frac{n}{2}, \frac{\chi^{2,D}}{2}\right)}{\Gamma(\frac{n}{2})} \approx 0.822 \end{aligned}$$

where $\gamma(t, s)$ is the lower-incomplete gamma function defined as:

$$\gamma(s, t) = \sum_{i=0}^{\infty} \frac{t^s e^{-t} t^i}{s(s+1)...(s+i)} \tag{77}$$

For the Bayesian Fit we set up the posterior probability assuming flat priors

$$P(\lambda|\{x\}, \{y\}) = \frac{P(\{y\}|\{x\}, \lambda)}{\int P(\{y\}|\{x\}, \lambda)\mathrm{d}\lambda} = \frac{P(\{y\}|f(\{x\}, \lambda))}{\int P(\{y\}|f(\{x\}, \lambda))\mathrm{d}\lambda} \tag{78}$$

For the Likelihood we choose:

$$P(\{y\}|f(\{x\}, \lambda)) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(\sum_{i=1}^{n} \frac{(y_i - f(x_i|\lambda))^2}{\sigma^2}\right) \tag{79}$$

parameter	Gauss	Wigner-Breigt
A	2.928	2.982
B	94.893	81.834
C	48.836	61.410
D	4.138	0.1217
μ/x_0	0.0493	0.0493
σ_m/Γ	0.0433	0.0563

Table 4: Results of Bayesian fit according to Gauss model and Wigner-Breigt model

where n is the number of data points and σ is the uncertainty of the data. Again a six-dimensional grid was implemented an the maximum was determined. Since the calculations afford very strong computing power a coarse grid was chosen. After determining a good approximation for the parameters, each parameter was maximized by fixing the other ones. A further problem was, that the integral of the Posterior probability could not be solved with sufficient accuracy. Since we are not interested in the probability of every single grid point, we normalized the probability of each parameter by its marginalized probability distribution. The distributions with 68%, 90% and 95% CL are shown in figure 19, the fit is shown in figure 20 and the resulting parameters are shown in table 4.

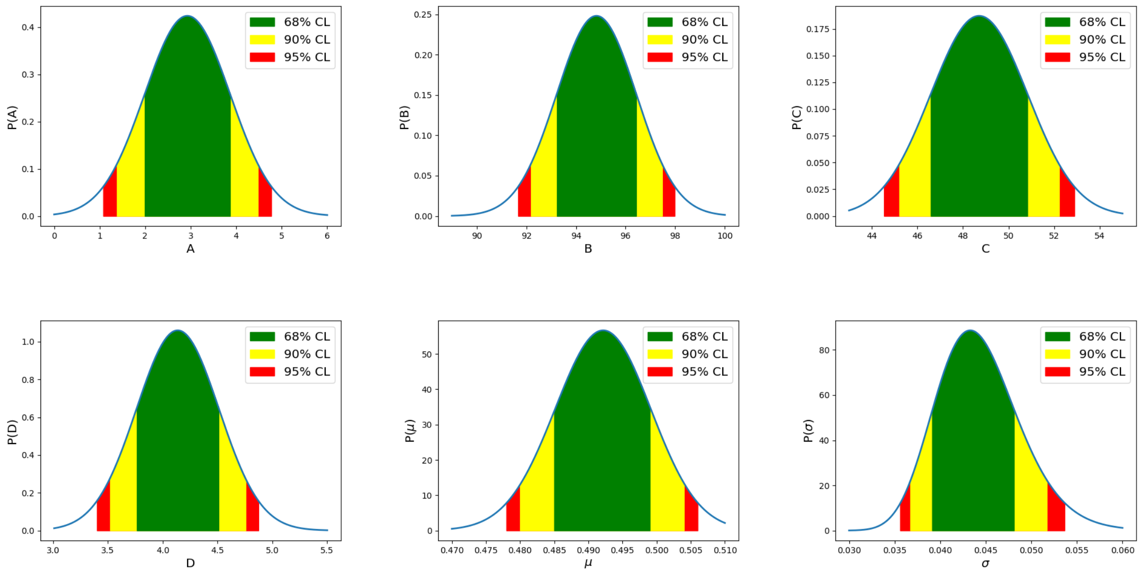


Figure 19: Marginalized probability distributions of the parameters for Gauss model

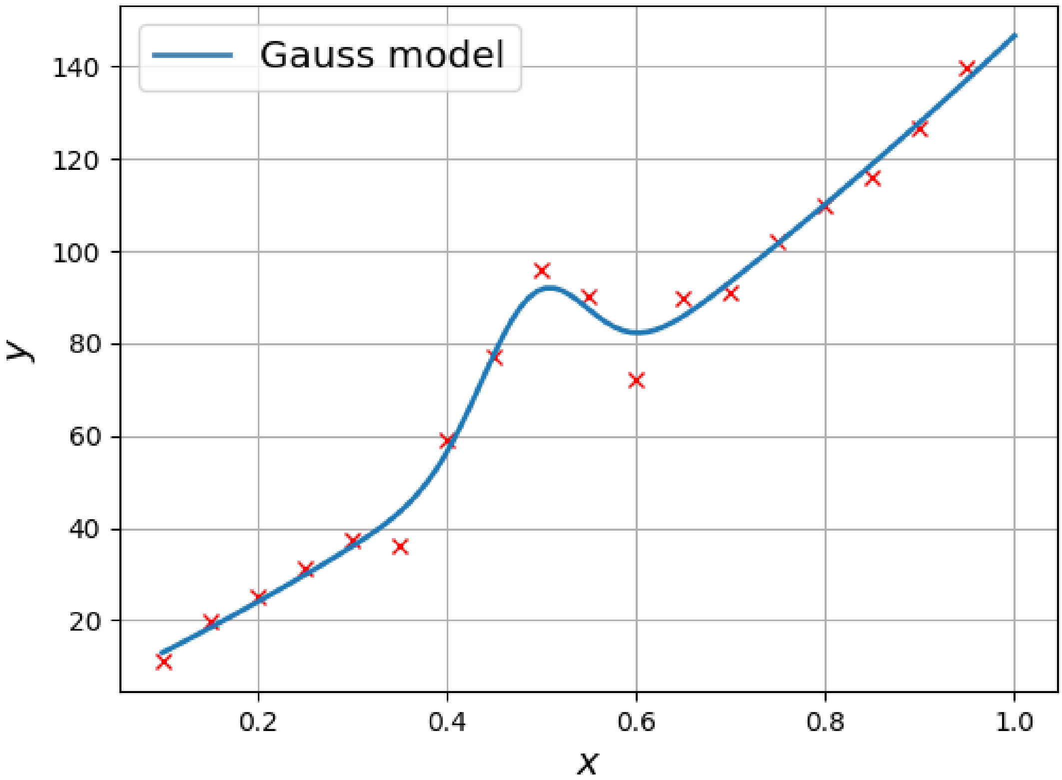


Figure 20: Bayesian fit of Gauss model

Now we consider the second function as model and go through the same procedure. The χ^2 fit results are shown in table 3 and figure 21. The calculated covariance matrix is:

$$\begin{pmatrix} 2.252 & 1.186 & 0.773 & 129.31 & 3.335 & -308.4 \\ 1.186 & 0.764 & 0.567 & 64.113 & 15.943 & -152.1 \\ 0.773 & 0.567 & 0.435 & 33.630 & 14.990 & -75.853 \\ 129.31 & 64.113 & 33.630 & 22544 & 1962.9 & -75594 \\ 3.335 & 15.943 & 14.990 & 1962.9 & 40904 & -11946 \\ -308.4 & -152.1 & -75.853 & -75594 & -11946 & 289781 \end{pmatrix} \tag{80}$$

Again there are negative values in the matrix so that it is incorrect.
The Bayes Factor is defined as:

$$B = \frac{P(y|\lambda_1)}{P(y|\lambda_2)} \tag{81}$$

with the two posterior probabilities at their maximum probability. Since the posteriors could not be normalized due to insufficient computing power a calculation is not possible. However if we compare the marginalized probability distributions, we see that the second model has slightly sharper peaks. Therefore we would conclude that the second model is more probable and the bias factor would be smaller than one. The p-value according to the second χ^2-Fit is: $p \approx 0.592$.

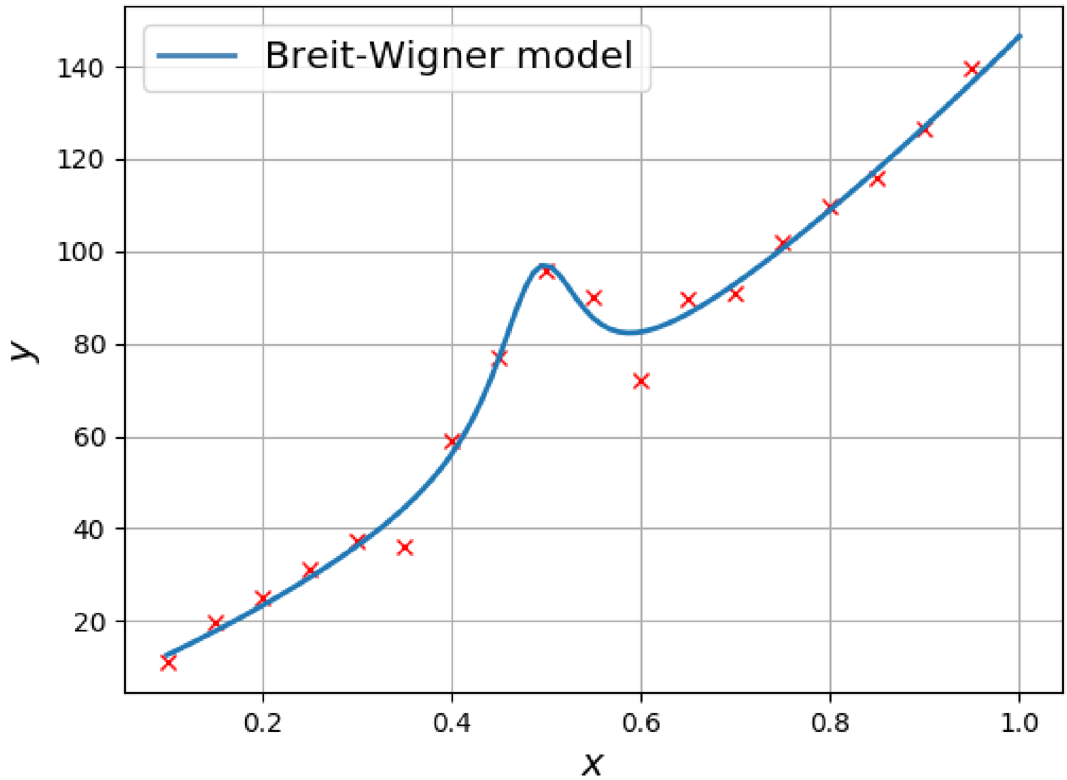


Figure 21: χ^2-Fit of Breit-Wigner model

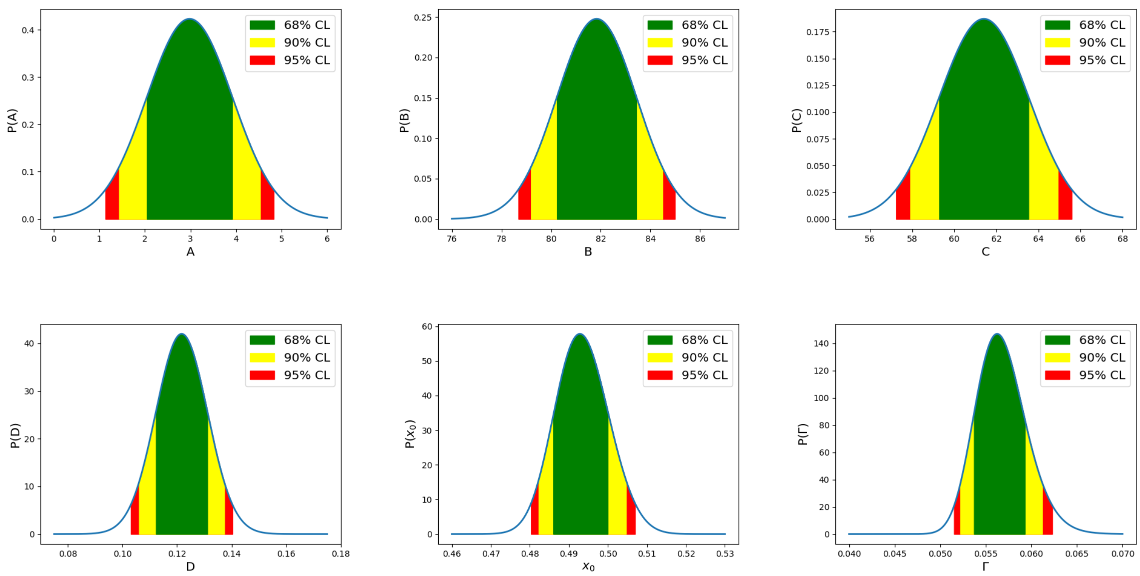


Figure 22: Marginalized probability distributions for the parameters of Breit-Wigner model

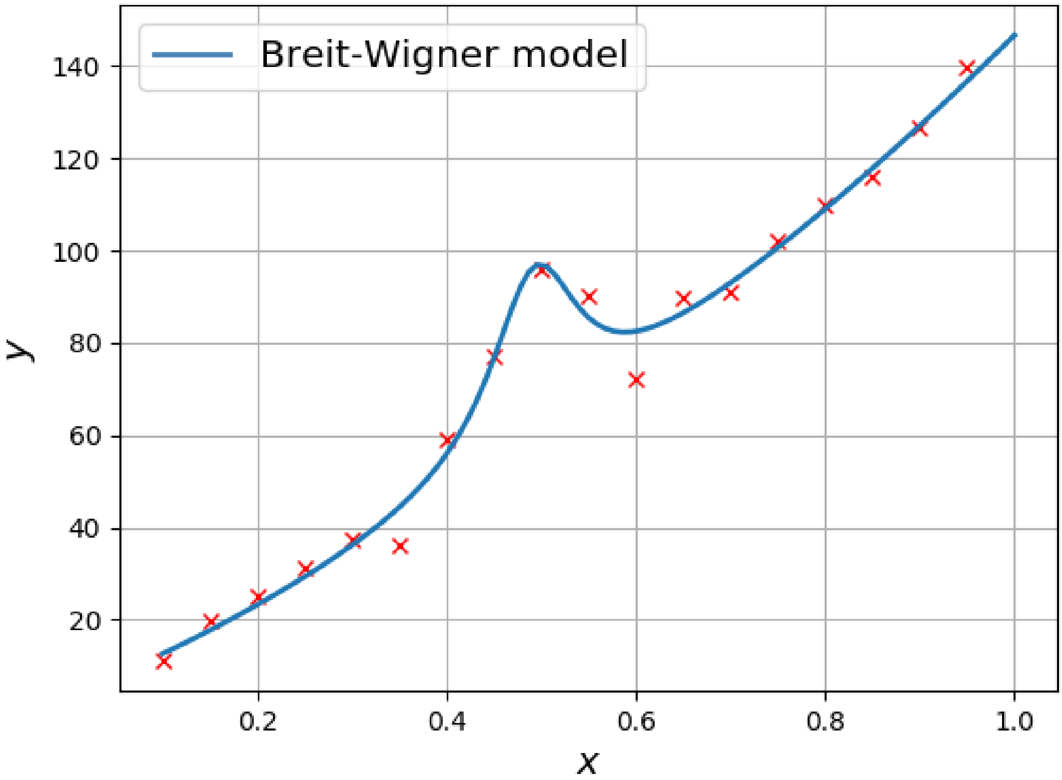


Figure 23: Bayesian fit of Breit-Wigner model

6. Gauss Distribution

Exercise 1
Maximum Likelihood estimation

The family of exponential distributions have pdf $P(x|p) = \lambda e^{-\lambda x} \qquad x \geq 0$

a) Generate $n = 2, 10, 100$ values of x using $x = -\ln U$ where U is a uniformly distributed random number between $(0, 1)$. Find the MLE estimator from your generated data. Repeat this for 1000 experiments and plot the distribution of the maximum likelihood estimator, $\hat{\lambda}$ (note that the true value in this case is $\lambda_0 = 1$.

b) Compare the distributions you found for the MLE to the expectation from the Law of Large Numbers and CLT (see lecture notes) and discuss.

The Likelihood is defined as product of the pdf for every measured x_i:

$$f(\lambda) = \prod_{i=1}^{n} P(x_i|\lambda) \tag{82}$$

Since we are interested in the mode, i.e. the maximum of the Likelihood, and the logarithm is monotonic, we can examine the following function:

$$\begin{aligned}
\mathcal{L}(\lambda) &= \ln\Big(\prod_{i=1}^{n} P(x_i|\lambda)\Big) \\
&= \ln(\lambda^n) + ln(e^{-\lambda \sum_{i=1}^{n} x_i}) \\
&= n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \\
\Rightarrow \frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \overset{!}{=} 0 \\
\Rightarrow \lambda &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}$$

The Central Limit Theorem and Law of Large Numbers tells us, that the distribution of the MLE converges to a Gauss function with mean $\lambda_0 = 1$ and variance $\sigma^2 = \frac{1}{nI(\lambda_0)}$ The

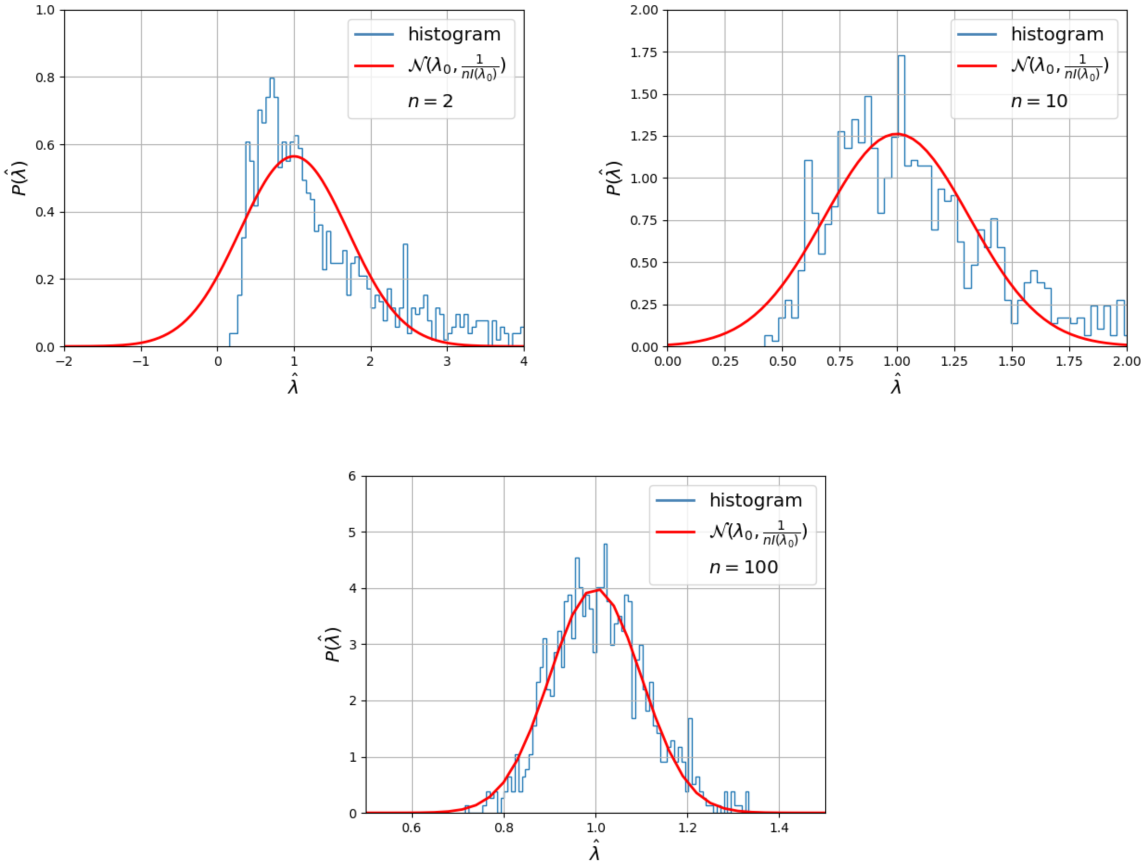


Figure 24: Distribution of $\hat{\lambda}$ and Gauss approximation for different n.

Fischer Information is:

$$
\begin{aligned}
I(\lambda_0) &= -E\left[\frac{\partial^2 \ln P(x|\lambda)}{\partial \lambda^2}\right] \\
&= -E\left[\frac{\partial^2}{\partial \lambda^2}(\ln \lambda - \lambda x)\right] \\
&= -E\left[\frac{\partial}{\partial \lambda}(\frac{1}{\lambda} - x)\right] \\
&= E\left[\frac{1}{\lambda^2}\right] \\
&= \int_0^\infty \frac{1}{\lambda} e^{-\lambda x} \mathrm{d}x \\
&= \left[-e^{-\lambda x}\right]_0^\infty \quad = 1
\end{aligned}
$$

In figure 24 are shown the histograms of the mean values of the experiments with the according Gaussian approximation for $n = 2, 10, 100$. The graphs show that for small n the Central Limit Theorem and Law of Large Numbers is not fulfilled. With increasing n the distribution is converging more and more to the Gaussian function. However with $n = 10$ the approximation is still not clearly visible. With a number of 100 means of λ the CLT and LLN show a good conformity.0

Exercise 2
Fischer Information

The family of Bernoulli distributions have the probability density $P(x|p) = p^x(1-p)^{n-x}$.

a) Calculate the Fischer Information $I(p) = -E\left[\frac{\partial^2 \ln P(x|p)}{\partial p^2}\right]$

b) What is the maximum likelihood estimator for p?

c) What is the expected distribution for $\hat{p} - p_0$?

The mean value theorem states, that for a strictly monotone, continuous and a continuously differentiable function $f(x)$ the derivative at one point $c \in [a, b]$ is:

$$f'(c) = \frac{f(b) - f(a)}{(b-a)} \tag{83}$$

We define:

$$L_n(p) := \frac{1}{n}\sum_{i=1}^{n} \ln P(x_i|p) \tag{84}$$

and:

$$L(p) = E[\ln P(x|p)]_{p_0} \tag{85}$$

We set $f(p) = \frac{\partial L_n}{\partial p} = L'_n(p)$ with $a = \hat{p}$ and $b = p_0$, where p_0 is the mode of equation 85 and $\hat{p}$ the mode of $L_n(p)$. Thus $L'_n(\hat{p}) = 0$ holds. With the mean value theorem we conclude further:

$$\begin{aligned} 0 &= L'_n(\hat{p}) + L''_n(p_1)(\hat{p} - p_0) \\ \Leftrightarrow (\hat{p} - p_0) &= -\frac{L'_n(p_0)}{L''_n(p_1)} \end{aligned} \tag{86}$$

with $p_1 \in [\hat{p}, p_0]$. The law of large numbers tells us for independent and identically distributed values x_i:

$$\lim_{n\to\infty} \bar{x}_n = \lim_{n\to\infty} \frac{x_1 + ... + x_n}{n} \to E[x] \tag{87}$$

This leads to

$$L''_n(p) = \frac{1}{n}\sum_{i=1}^{n} \frac{\ln \partial^2 P(x_i|p)}{\partial p^2} \overset{n\to\infty}{\to} E\left[\frac{\partial^2 \ln P(x|p)}{\partial p^2}\right]_{p_0} \tag{88}$$

The Maximum Likelihood Estimator is the mode of $\ln P(x|p)$. Thus $p = x$ and we find that

$$
\begin{aligned}
I(p) &= -\lim_{n\to\infty} L''_n(p) \\
&= -\frac{\partial}{\partial p}\Big(-\frac{x}{p} - \frac{1-x}{1-p}\Big) \\
&= \frac{x}{p^2}\frac{1-x}{(1-p)^2} \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}
\end{aligned}
$$

In order to get the expected distribution we consider equation 86. We already showed that for large n $p_1 \to p_0$ and therefore $L''_n(p_0) = -I(p)$.

Since p_0 is the mode of $L(p)$, we have $L'(p_0) = 0$ so,

$$
\begin{aligned}
L'n(p_0) &= L'_n(p_0) - L'(p_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \ln P(x_i|p)}{\partial p} - E\Big[\frac{\partial \ln P(x_i|p)}{\partial p}\Big]\right)
\end{aligned}
$$

We can apply the Central Limit Theorem and get a Gauss probability distribution for $L'_n(p_0)$ with mean 0. The variance ratio is

$$
V\left[\frac{L'_n(p_0)}{L''_n(p_0)}\right] = \frac{V\big[\frac{\partial \ln P(x_i|p)}{\partial p}\big]}{nI(p_0)^2} = \frac{1}{nI(p_0)} \tag{89}
$$

This leads to the distribution:

$$
\begin{aligned}
\hat{p} - p_0 &= -\frac{L'_n(p_0)}{L''_n(p_1)} \\
&= P(\hat{p} - p_0) = \mathcal{N}\Big(0, \frac{1}{nI(p_0)}\Big)
\end{aligned}
$$