REPORT: DATA ANALYSIS
Lecture by Prof. Allen C. Caldwell

Written By

Johannes Summer
03619814

Technische Universität München

07.04.2018

Contents

Introduction

This report was written for the lecture series "Data Analysis", which was taught by Prof. Caldwell in the winter term 2017/2018 at the TU Munich. In this series of lectures, the concept of probability was introduced, the basic statistical distributions developed and a detailed discussion of data analysis was given. This report focuses on the given exercises of the course and provides solution, which were either solved mathematically by the script or numerically by python. In the exercises the theory of the lectures is also approached and briefly summarized.

1 Introduction to Probabilistic Reasoning

1.1 Exercise 1

In the first exercise we show that it is not only information, but also the way **how** we got it, which matters for the result in the end.

Problem
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.
(a) What is the probability that the second child is also a girl?
(b) 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
First we use **Laplace's rule of insufficient reason**:
"If we have no further information, we have no reason to prefer one elementary event over the other."
Therefore any possible option like "children composition" ((bb),(bg),(gb),(gg)) or the picture of a girl/boy are as likely as their alternatives. With the additional rule that all the branches of one probability path must add up to 1, we get:

$$P(bb) = P(gb) = P(bg) = P(gg) = \frac{1}{4}$$

$$(\text{Random picture is boy}|bg) = P(\text{Random picture is girl}|bg) = \frac{1}{2}$$

Next we introduce **Bayes Theorem**, which says that we can get the conditional probability $P(A|B)$, if we know all other probabilities.

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

(a) Now we can solve the first problem, with $A = gg$, B = Random picture is girl: (rpg):

$$P(gg|rpg) = \frac{P(rpg|gg)P(gg)}{P(rpg|gg)P(gg) + P(rpg|bg)P(bg) + P(rpg|gb)P(gb) + P(rpg|bb)P(bb)}$$

$$P(gg|rpg) = \frac{1 \cdot \frac{1}{4}}{1 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4}} = \frac{1}{2}$$

(b) Here the way of getting the information changes to B = Jane shows you the picture of a girl, if she has at least one girl (pgg). So the new probabilities are $P(pgg|bg) = P(pgg|gb) = 1$ and we get:

$$P(gg|pgg) = \frac{1 \cdot \frac{1}{4}}{1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4}} = \frac{1}{3}$$

So not only the information, but also the way how we got the information influences the probability in the end!

1.2 Exercise 2

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

Solution
In 1.2.3 we flipped a coin ten times and looked at two different outcomes, *S1: THTHHTHTTH* and *S2: TTTTTTTTTT*.
The probabilities of getting exactly one of those sequences are the same, but we can also look at them as representations of certain outcome definitions, which have summed up different probabilities.
These outcomes could for instance be defined as:

- at least n-times T
- exactly n-times T
- a certain number of switches (TH, HT)
- a certain number of T or H in a row (TT, HHH)

1.3 Exercise 3

Problem
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
The given information is not sufficient for a conclusion, since we do not have any knowledge about the probability distribution of the detector. It could for instance be either Gaussian or Poisson or Breit-Wigner distributed. Even if we would know the kind of distribution, we would still not know $P(E_{exp})$. Elsewise, with a Gauss distribution we could

use the given resolution of 10 % to set the FWHM equal to $0.1E_{exp}$ and derive the standard derivation σ out of it. In the Gaussian distribution the probability of the true energy E_{true} lying within a certain range around E_{exp} is defined by σ: $\pm 1\sigma$ means at least 68 %, $\pm 2\sigma$ means at least 95 %, ...
But only knowing the detector resolution is not enough to assign probability values.

1.4 Exercise 4

Problem
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 Mongolian swamp fever if you go to the doctor with all or with any three out of four of these symptoms ? (From R.Barlow)

Solution
We first introduce the definitions of Conditional Probability, Independent Probability, Marginal Likelihood that are needed for **Bayes Theorem** and are used here:

$$\text{Conditional Probability:} \quad P(A|B) = \frac{P(B \cap A)P(A)}{P(B)}$$

$$\text{Independent Probability:} \quad P(A \cap B) = P(A) \cdot P(B)$$

$$\text{Marginalization:} \quad P(A) = \sum_{i=1}^{N} P(B|A_i)P(A_i)$$

$$\text{Bayes Theorem:} \quad P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)}$$

Next we set the probability values for the fever and it symptoms. The other values can be calculated by $P(A|B) + P(\bar{A}|B) = 1$.

$$\begin{array}{lll}
F: Fever & P(F) = 10^{-4} & \\
A: Spots & P(A|\bar{F}) = 3 \cdot 10^{-2} & P(A|F) = 1 \\
B: Lethargy & P(B|\bar{F}) = 1 \cdot 10^{-1} & P(B|F) = 1 \\
C: Thirsty & P(C|\bar{F}) = 2 \cdot 10^{-2} & P(C|F) = 6 \cdot 10^{-1} \\
D: Sneezing & P(D|\bar{F}) = 5 \cdot 10^{-2} & P(A|F) = 2 \cdot 10^{-1}
\end{array}$$

Using the formulas and inserting the values:

$$\begin{aligned}
P(F|A \cap B \cap C \cap D) &= \frac{P(A \cap B \cap C \cap D|F)P(F)}{P(A \cap B \cap C \cap D|F)P(F) + P(A \cap B \cap C \cap D|\bar{F})P(\bar{F})} \\
&= \frac{P(A|F)P(B|F)P(C|F)P(D|F)}{P(A|F)P(B|F)P(C|F)P(D|F) + P(A|\bar{F})P(B|\bar{F})P(C|\bar{F})P(D|\bar{F})} = 80\%
\end{aligned}$$

$$P(F|\bar{A}\cap B\cap C\cap D)=0 \qquad P(F|A\cap\bar{B}\cap C\cap D)=0$$
$$P(F|A\cap B\cap\bar{C}\cap D)=5\% \quad P(F|A\cap B\cap C\cap\bar{D})=46\%$$

2 Binomial and Multinomial Distribution

2.1 Exercise 8

Problem
For the following function

$$p(x)=xe^{-x} \qquad 0\le x\le\infty$$

(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
The terms are briefly defined and calculated using python. Afterwards the plot with the terms is shown in figure 1.

(a) Mean and Standard Deviation
Both, the mean μ and the standard derivation σ can be determined with the written formulas below and by using the partial integration (twice for $E(x)$ and three times for $E(x^2)$). Afterwards the probability of the interval $I_\sigma=[\mu-\sigma,\mu-\sigma]$ can be determined using again partial integration and python iteration.

$$\mu=E(x)=\int_{D(x)}x\cdot p(x)dx=\int_0^\infty x^2e^{-x}dx=...\overset{\text{P.I.}}{=}-(x^{-2}-2x-2)\exp^{-x}|_0^{-\infty}=2$$
$$E(x^2)=\int_{D(x)}x^2\cdot p(x)dx=\int_0^\infty x^3e^{-x}dx=...\overset{\text{P.I.}}{=}-(x^{-3}+3x^2+6x+6)\exp^{-x}|_0^{-\infty}=6$$
$$\sigma=\sqrt{E(x^2)-E(x)^2}=\sqrt{6-2^2}=\sqrt{2}$$
$$P(x\in[\mu-\sigma,\mu+\sigma])=\int_{\mu-\sigma}^{\mu+\sigma}xe^{-x}dx=....\overset{\text{P.I.}}{=}(-x-1)e^{-x}|_{\mu-\sigma}^{\mu+\sigma}\overset{\text{python}}{\approx}0.73753473$$

(b) Median and Central Interval
The median is the value of x for which the cumulative probability reaches 50 %. The actual value of the median was derived by python.
The interval $[x_1;x_2]$ is called the Central Probability Interval and contains probability $1-\alpha$. We would then say that the parameter x is in this interval with $1-\alpha=68\%$ probability. At the central Interval the probabilities of the outside areas left and right from the Interval are equal. So we find the central interval $I_{central}=[x_1,x_2]$ by determining x_1 and

x_2 with $P(x_1) = \frac{\alpha}{2} = 16\%$ and $P(x_2) = 1 - \frac{\alpha}{2} = 84\%$ via python.

$$P(x_{median}) = \int_0^{x_{median}} p(x)dx = \int_0^{x_{median}} xe^{-x}dx = \int_{x_{median}}^{\infty} xe^{-x}dx = 0.5$$

$$x_{median} \overset{\text{python}}{\approx} 1.6784 \qquad x_1 \overset{\text{python}}{\approx} 0.7121 \qquad x_2 \overset{\text{python}}{\approx} 3.2886$$

(c) Mode and Shortest Interval

The mode is the value of x that maximizes $p(x)$. It therefore can be found by $\frac{d}{dx}p(x) \overset{!}{=} 0 \wedge \frac{d^2}{dx^2}p(x) < 0$. The x-values with the highest connected $p(x)$-values are taken to form the smallest interval $I_{smallest} = [x_{1s}, x_{2s}]$ for a given P(x) term. The values were derived by python by multiple selecting between adding the right or left x-term to x_{mode} until the wanted probability threshold was reached.

$$\frac{d}{dx}p(x) = (1-x)e^{-x} = 0 \qquad \rightarrow x = 1 \qquad \wedge \qquad \frac{d^2}{dx^2}p(x) = (x-2)e^{-x} \qquad \rightarrow p''(1) < 0$$

$$x_{mode} = 1 \qquad x_{1s} \overset{\text{python}}{\approx} 0.270600 \qquad x_{1s} \overset{\text{python}}{\approx} 2.490000$$

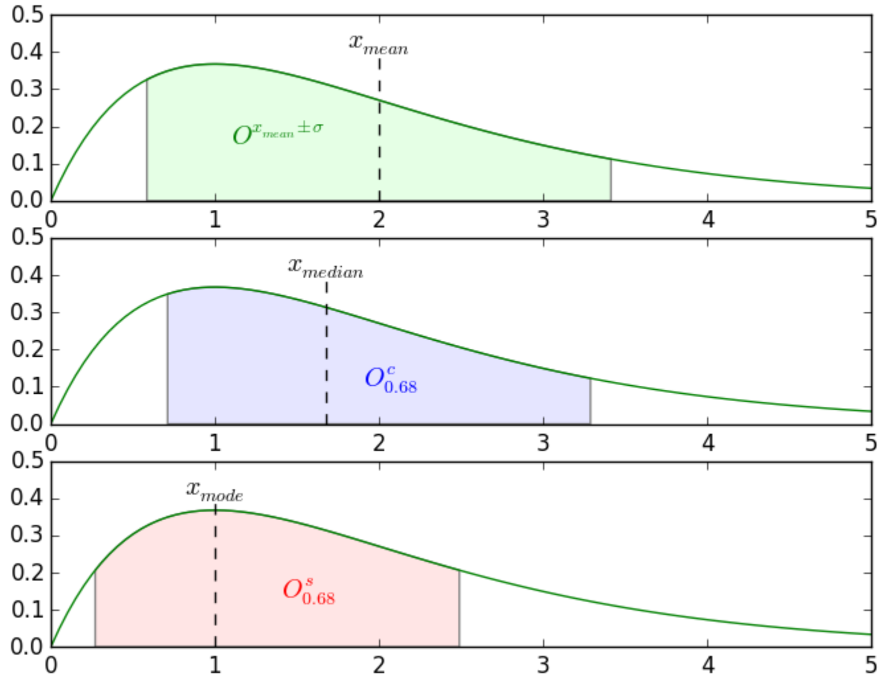

Figure 1: $1-\alpha = 0.68$-Intervals: I_σ with $x_{mean} = \mu$, $I_{central}$ with x_{median} and $I_{smallest}$ with x_{mode}

2.2 Exercise 10

Problem

Consider the data in the table 1. 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 (E_i)	Trials (N_i)	Successes (r_i)
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 1: Energy Dependency Energy of p

Solution

Bayes Theorem with Flat Prior:

$$P_n(p|r,N) = \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)dp} \overset{\text{Flat Prior}}{=} \frac{(N+1)!}{(r)!(N-r)!} p^r (1-p)^{(N-r)}$$

Another logical pair of numbers that can be used to summarize the probability density for p are the mode and the shortest interval containing a fixed probability. The mode is the value of p that maximizes $P(p|N,r)$ and in general depends on the prior distribution chosen. For a flat prior, it is $p^* = \frac{r}{N}$. The shortest interval for a unimodal distribution is defined as $[p_1;p_2]$ with the conditions

$$1-\alpha = \int_{p_1}^{p_2} P(p|r,N)dp \qquad P(p_1|r,N) = P(p_2|r,N) \qquad p_1 < p^* < p_2$$

The intervals $I_{smallest}$ around x_{mode} were calculated numerically by python and are plotted in figure 2:

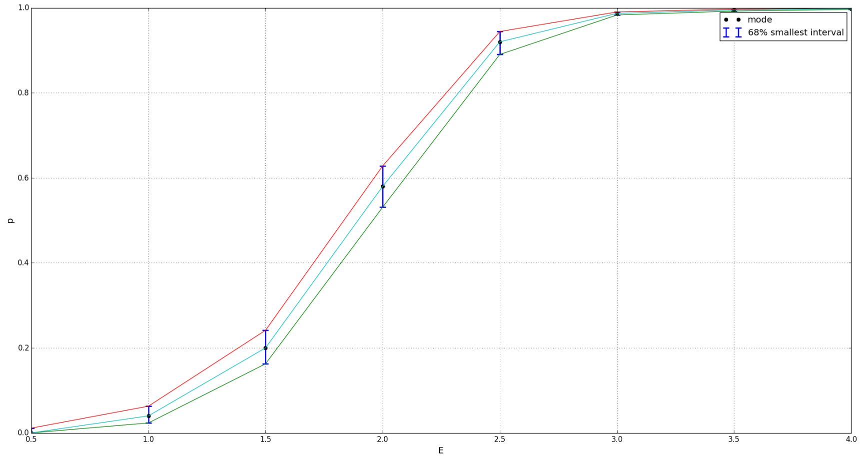

Figure 2: 68 % Smallest interval bands for binomial probability distributions for given N_i and r_i values according to different energy levels E_i taken from table 1.

2.3 Exercise 11

Problem

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

The median is the value of p for which the cumulative probability reaches 50 %. We can define a central probability interval as the region around the median, containing a fixed probability $1-\alpha$ and having equal probabilities at the low and high tails outside.

$$F(p_{med}) = \int_0^{p_{med}} P(p|N,r)dp = 1 - F(p_{med}) = \int_{p_2}^{1} P(p|N,r)dp = \frac{1}{2}$$
$$F(p_1) = \int_0^{p_1} P(p|N,r)dp = 1 - F(p_2) = \int_{p_2}^{1} P(p|N,r)dp = \frac{\alpha}{2}$$

The Confidence Interval specifies a range of p for which E is within the specified probability interval for outcomes E_i. We imagine that there is a "true value" for p. In figure 3 we plotted the data from table 1 by using a central interval and taking $1-\alpha = 0.90$ as confidence level. The values were numerically calculated by python.

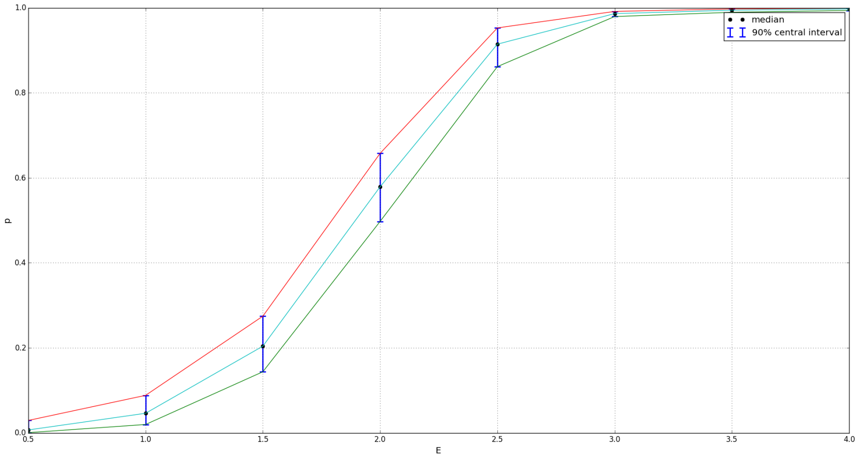

Figure 3: 90 % Confidence Level bands plotted against energy by using the Central Interval for binomial probability distributions for given N_i and r_i values taken from different energy levels E_i from table 1. For a given E_i, the blue line gives the value of p_{min}, the smallest element in the set, while the red line gives p_{max}, the largest element in the set.

2.4 Exercise 13

Problem

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

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

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

Solution

$$P_n(p|r,N) = \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)dp} \overset{\text{Flat Prior}}{=} \frac{(N+1)!}{(r)!(N-r)!} p^r(1-p)^{(N-r)}$$

$$\begin{aligned} P_n(p|r,N) &= \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)dp} \overset{\text{Posterior}}{=} \frac{P(r|N,p)P(r|N,p)}{\int P(r|N,p)P(r|N,p)dp} \\ &= \frac{p^{2r}(1-p^{(2N-2r)}}{\int p^{2r}(1-p)^{2N-2r}dp} = \frac{(2N+1)!}{(2r)!(2N-2r)!} p^{2r}(1-p)^{(2N-2r)} \end{aligned}$$

Doing mathematical induction or by redefining $r = 2r$ and $N = 2N$ n times, we receive the result above.

The expectation value and variance are defined as:

$$E[p] = \frac{r+1}{N+2} \rightarrow E[p]_n = \frac{nr+1}{nN+2} \qquad \lim_{n\to\infty} E_n[p] = \frac{r}{N}$$
$$V[p] = E[p^2] - E[p]^2 = \frac{E[p](1-E[p])}{N+3} \rightarrow V_n[p] = \frac{E_n[p](1-E_n[p])}{nN+3}$$
$$\lim_{n\to\infty} V_n[p] = \frac{\frac{r}{N}(1-\frac{r}{N})}{nN+3} \sim \frac{1}{n} = 0$$

3 Poisson Distribution

3.1 Exercise 4

Problem

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

(a) Mean and standard deviation of x

$$E(x) = \int_{D(x)} xf(x)dx = \int_{-\infty}^{+\infty} x\frac{1}{2}^{-|x|} dx = \frac{1}{2}(\int_{-\infty}^{0} xe^{x}dx + \int_{0}^{+\infty} xe^{-x}dx)$$
$$= \frac{1}{2}(\int_{0}^{+\infty} -xe^{-x}dx + \int_{0}^{+\infty} xe^{-x}dx) = \frac{1}{2}\int_{0}^{+\infty}(x-x)e^{-x}dx = 0$$
$$E(x^2) = \int_{D(x)} x^2 f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{2}x^2 e^{-|x|}dx = \frac{1}{2}(\int_{-\infty}^{0} x^2 e^{x}dx + \int_{0}^{+\infty} x^2 e^{-x}dx)$$
$$= \frac{1}{2}(\int_{0}^{+\infty} x^2 e^{-x}dx + \int_{0}^{+\infty} xe^{-x}dx) = \frac{1}{2}\int_{0}^{+\infty} 2x^2 e^{-x}dx = \int_{0}^{+\infty} 2x^2 e^{-x}dx...$$
$$\stackrel{\text{P.I.}}{=} -(x^{-2} - 2x - 2)\exp^{-x}|_{0}^{-\infty} = 2$$
$$\sigma = \sqrt{E(x^2) - E(x)^2} = \sqrt{2-0} = \sqrt{2}$$

(b) Full Width at Half Maximum (FWHM)

$$\text{FWHM:} \quad f(x_1) = f(x_2) = \frac{1}{2}f(x_{max}) \qquad f(x_{max}) = f(0) = \frac{1}{2}$$
$$f(x_1) = f(x_2) = \frac{1}{2}e^{-x} = \frac{1}{4} \rightarrow x_1 = x_2 = ln2 \qquad FWHM = 2ln2$$
$$\text{Here: } FWHM = \frac{2ln2}{\sqrt{2}}\sigma \qquad \text{Gauss distribution: } FWHM = 2\sqrt{2ln2}\sigma$$

(c) Probability in the $\pm$ 1 standard deviation interval around the peak

$$P(x \in [-\sigma, +\sigma]) = \int_{-\sigma}^{+\sigma} \frac{1}{2} e^{-|x|} dx = \int_{0}^{+\sigma} e^{-x} dx = -e^{-x}|_0^{\sqrt{2}} = 1 - e^{-\sqrt{2}} \approx 0.75688$$

3.2 Exercise 7

Problem

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) The Poisson distribution and the lower limit probability are first shown and then plotted for $n = 9$ and $1 - \alpha = 90\%$ in figure 4 via python:

$$\text{Poisson distribution:} \qquad P(n|\nu) = \frac{e^{-\nu}\nu^n}{n!} \qquad P(\nu|n) = \frac{e^{-\nu}\nu^n}{n!}$$

$$\text{Lower limit:} \qquad F(\nu|n) = \int_0^{\nu_{limit}} \frac{e^{-\nu}\nu^n}{n!} d\nu \stackrel{!}{=} \frac{\alpha}{2} = 5\% \rightarrow \nu_{limit} = 5.44$$

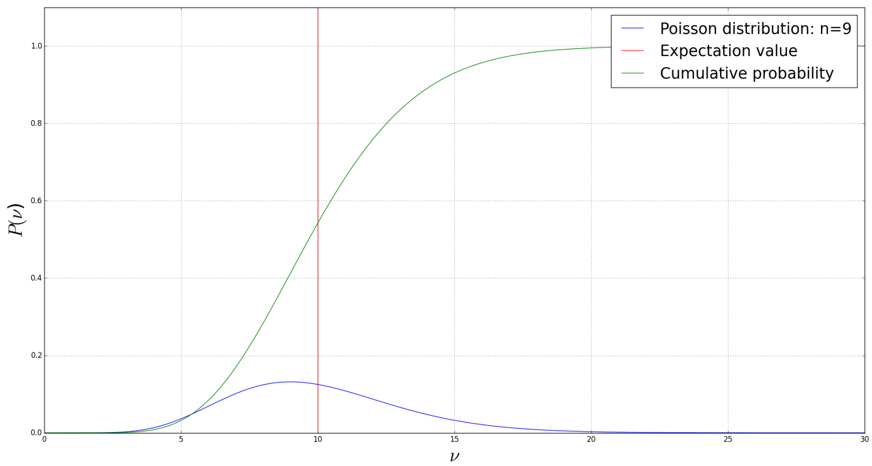

Figure 4: Poisson distribution $P(\nu|n)$ and cumulative $F(\nu|n)$ for $n = 9$. The red line shows the expectation value $E(x)$. The lower limit of $F(\nu|n) = 5\%$ is reached for $\nu = 5.44$

(b) The confidence level is $1-\alpha = 68\%$. The Neyman level for smallest interval is defined as

$$F(x_2) - F(x_1) = 1-\alpha \qquad P(x_1) = P(x_2)$$

For the smallest levels for each n the probabilities were ranked and the highest ones summed up one after another until the cumulative probabilities reached the 68 % value. The values were numerically calculated and plotted via python and can be seen in figure 5. The minimum and maximum value for $1-\alpha = 68\%$ and $n = 9$ are $\nu_{min} = 6.5$, $\nu_{max} = 13.3$.

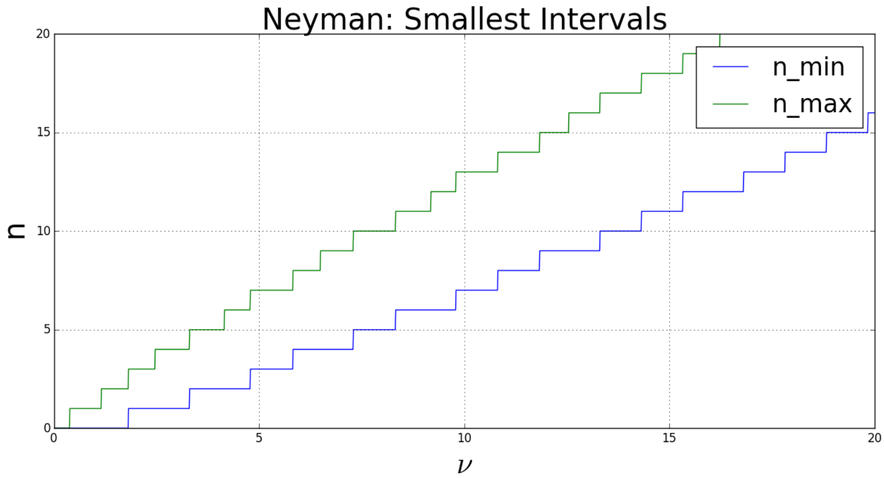

Figure 5: Poisson distribution $P(\nu|n)$ and cumulative $F(\nu|n)$ for $n = 9$. The red line shows the expectation value $E(x)$. The lower limit of $F(\nu|n) \geq 5\%$ is reached for $\nu = 5.44$

3.3 Exercise 8

Problem

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

Solution

(a) The Feldman-Cousins Interval is based on a different ranking of possible data outcomes. Instead of ranking results according to the probability of n given μ , the results are ranked according to the probability ratio r,

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

where $\hat{\mu}$ is the optimal value for μ

(b) In the beginning the Neyman 68% CL bands (with the smallest interval definition) corresponding to each $\mu = \nu + \lambda$ are calculated. They are the same as the confidence bands from the previous exercise (Fig. 5). Due to $\nu = \mu - \lambda$ and $\lambda = 3.2$ the confidence bands corresponding to ν are the bands from the Fig. 5 that are shifted by 3.2 downwards. Note that ν can never be negative because the event rate is always positive, therefore all values of ν that would be negative are set to zero. The new CL bands are presented in figure 6.

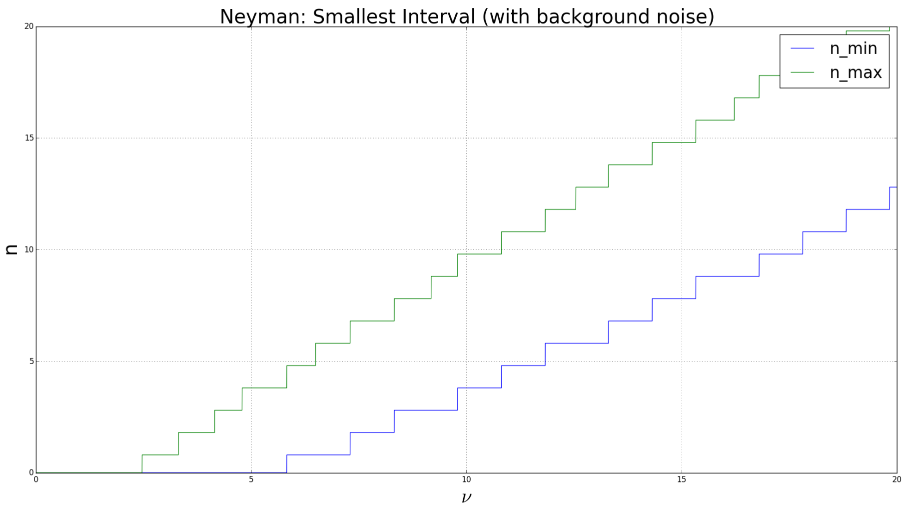

Figure 6: Neyman Confidence Interval bands according to smallest intervals, with Poisson distribution and known background of $\lambda = 3.2$ events

In case (b) there are significantly more empty intervals than before, which come from the Neyman procedure. If one does not want to have any empty intervals, one should use the Feldman-Cousins construction from (a).

(c) In case of a known and fixed background λ the posterior can be written as

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

If we us use the smallest interval definition, the smallest 68% Credible interval for ν can be found by applying the script described in Ex. 2.8 on the posterior $P(\nu|3.2, 9)$. The Credible interval is shown of Fig. 7.

$$\mathcal{O}^S_{0.68} = [3.116, 9.146]$$

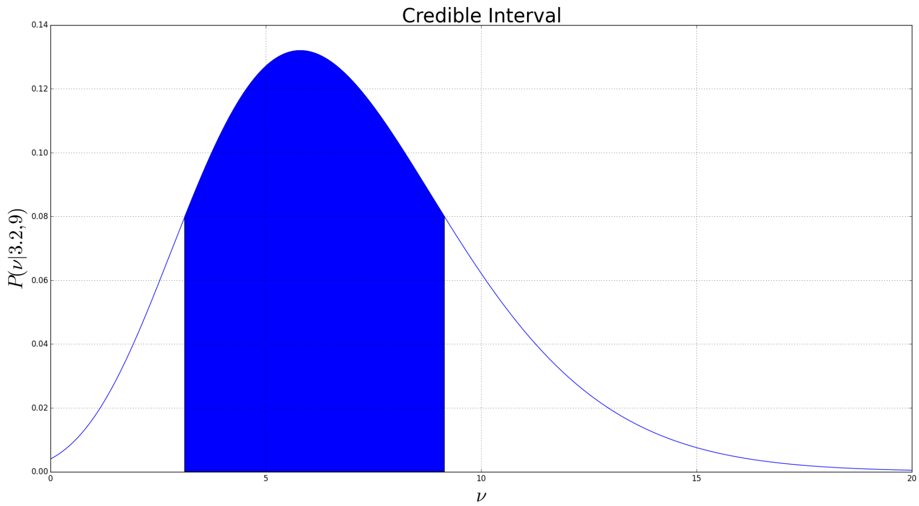

Figure 7: Bayesian posterior as the function of ν with the 68% Credible interval (filled)

3.4 Exercise 13

Problem

In this problem, we look at the relationship between an unbinned likelihood and a binned Poisson probability. We start with a 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 $x_i = 1,,n$. The unbinned likelihood is defined as the product of the densities:

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

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

$$\nu_j = \int_{\Delta_j} f(x|\lambda)dx$$

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

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

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.

Solution
The interval $[a,b]$ is divided into K subintervals. The probability of unbinned Likelihood is defined as the product of the Poisson probabilities for each bin. Use notation: $f_j := f(x_i|\lambda)$ and the following:

(i) $n_j! = 1$ due to $n_j \in \{0,1\}$.

(ii) $\lim\limits_{K\to\infty} v_j = \lim\limits_{K\to\infty} \int_{\Delta_j} f(x|\lambda)dx \approx f(x_i|\lambda)\Delta$ due to Δ being infinitesimally small when $K \to \infty$.

(iii) n intervals with $n_j = 1$ and $K-n$ intervals with $n_j = 0$

(iv) $e^x \approx 1 + x$ for $|x| \ll 1$

(v) Constraints: $\sum_{j=1}^{K} n_j = n$ $\quad \prod_{z=1}^{Z} x_z^0 = 1$ $\quad \sum_{j=1}^{\infty} f(x_i|\lambda)\Delta = 1$

$$L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{j=1}^{K} \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!}$$

$$\lim_{K\to\infty} \prod_{j=1}^{K} \frac{e^{-\nu_j} v_j^{n_j}}{n_j!} \overset{\text{(i)}}{=} \lim_{K\to\infty} \left(\prod_{j=1}^{n} e^{-\nu_j}\nu_j^{n_j} \cdot \prod_{j=n+1}^{K} e^{-\nu_j}\nu_j^{n_j}\right)$$

$$\overset{\text{(ii)(iii)(v)}}{=} \lim_{K\to\infty} \underbrace{\underbrace{e^{-\sum_{j=1}^{K} f(x_i|\lambda)\Delta}}_{e^{-1}}}_{\text{constant from norm (v)}} \cdot \prod_{j=1}^{n} f(x_i|\lambda)\Delta \sim \prod_{j=1}^{n} f(x_i|\lambda)\Delta \qquad q.e.d.$$

We have shown that the unbinned likelihood is proportional to the limit of the product of Poisson probabilities for an infinitely fine binning.

3.5 Exercise 16

Problem
We consider a thinned Poisson process. Here we have a random number of occurrences, N, distributed according to a Poisson distribution with mean.
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

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

Show that the probability distribution is given by

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

Solution:
We can divide X for a given N in two sums, one that takes $r - times$ the value 1 with

probability p and one that takes $(N-r)-times$ the value 0 with probability $(1-p)$.

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

Therefore the probability distribution for X with given N can be rewritten by:

$$P(X|N,p) = P(r|N,p)$$

Now $P(X|N,p)$ is the Binomial distribution for a given N. N is Poisson-distributed. For the distribution $P(X)$ we sum over all $N \in \mathbb{R}_0^+$. We will ignore the "p" in our notation, since it is always given. We set the unsuccessful trials as $L = N - X$. We calculate $P(X)$ by marginalization.

$$P(X) = \sum_{N=0}^{\infty} P(X|N)P(N|\nu) \overset{P(X)=0 \text{ for } N<X}{=} \sum_{N=X}^{\infty} P(X|X+L)P(N|\nu)$$

Changing N with L: $\qquad P(X) = \sum_{L=0}^{\infty} P(X|X+L)P(N|\nu)$

Inserting the distributions, we get:

$$\begin{aligned}
P(X) &= \sum_{L=0}^{\infty} \frac{(X+L)!}{X!L!} p^X (1-p)^L \cdot \frac{e^{-\nu}\nu^{X+L}}{(X+L)!} \\
&= \sum_{L=0}^{\infty} \frac{(\nu p)^X)(\nu(1-p)^L e^{-\nu}}{X!L!} \\
&= \frac{(\nu p)^X}{X!} \sum_{L=0}^{\infty} \frac{e^{-\nu}(\nu(1-p))^L}{L!} \\
&= \frac{(\nu p)^X}{X!} e^{-\nu + p\nu - p\nu} \sum_{L=0}^{\infty} \frac{(\nu(1-p))^L}{L!} \\
&= \frac{(\nu p)^X}{X!} e^{-\nu(1-p) - p\nu} \sum_{L=0}^{\infty} \frac{(\nu(1-p))^L}{L!} \\
&= e^{-p\nu} \frac{(\nu p)^X}{X!} e^{-\nu(1-p)} e^{\nu(1-p)} \\
&= \frac{e^{-p\nu}(p\nu)^X}{X!} \qquad q.e.d.
\end{aligned}$$

4 Gaussian Probability Distribution Function

4.1 Exercise 8

Problem

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

$$p(x) = \lambda \exp^{-\lambda x}$$

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

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

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

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

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

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

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

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

The central limit theorem (CLT) says that, mostly, the sum of independent random variables tends toward a normal distribution even if the original variables themselves are not normally distributed.

(a) Firstly, we derive the theoretical mean and standard deviation for the exponential distribution for the exponential distribution and the resulting normal distribution:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E[\bar{x}] = \mu_x = \int_0^\infty xp(x)dx = \int_0^\infty x\lambda e^{-\lambda x} = \lambda\int_0^\infty -\partial_\lambda e^{-\lambda x}dx = \lambda(-\partial_\lambda)\int_0^\infty -\partial_\lambda e^{-\lambda x}dx$$

$$= \lambda(-\partial_\lambda)[-\frac{1}{\lambda e^{-\lambda x}}]_0^\infty = \lambda\frac{1}{\lambda}(e^0 - e^{-\infty}) = \frac{1}{\lambda}$$

$$E[\bar{x}^2] = \int_0^\infty x^2 p(x)dx = \int_0^\infty x^2\lambda e^{-\lambda x} = \lambda\partial_\lambda^2\int_0^\infty e^{-\lambda x}dx = \lambda(-\frac{2}{\lambda^3})(e^\infty - e^0) = \frac{2}{\lambda^2}$$

$$\sigma_x = \sqrt{E[\bar{x}^2] - E[\bar{x}]^2} = \frac{1}{\lambda}$$

$$\mu_z = \mu_x = \frac{1}{\lambda} \qquad \sigma_z = \frac{\sigma_x}{\sqrt{n}} = \frac{1}{\lambda\sqrt{n}}$$

Then, we try the CLT by creating random numbers according to the exponential distribution and plotting their averaged value $\bar{x}$ for least 3 different choices of n and λ. In figure 8 the CLT can be seen: The distribution approaches a normal distribution.
Increasing λ decreases the mean μ_z.
Additionally to increasing λ, increasing n (the number of values averaged) decreases σ_z.
The results confirm the theoretical values of the CLT.

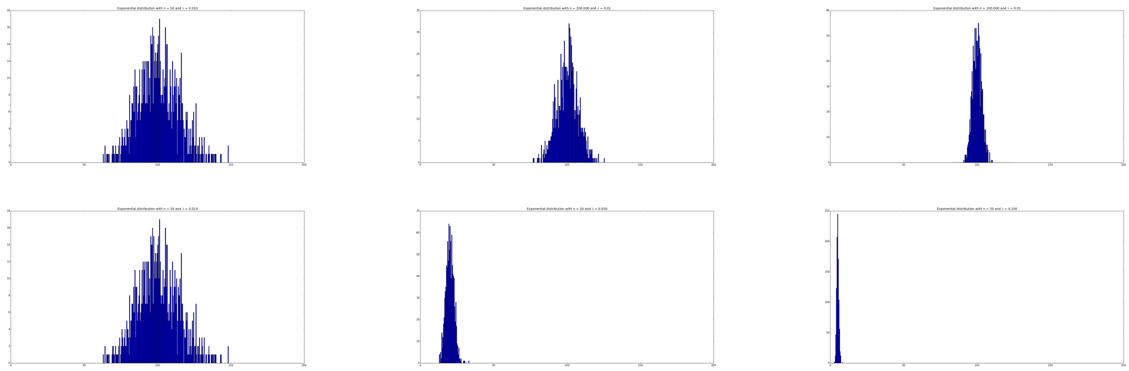

Figure 8: In the first row the dependency from n can be seen. The higher n for the averaged $\bar{x}$ the thinner σ_z becomes.
In the second row the dependency from λ can be seen. If we increase λ μ_z becomes smaller and the variance also decreases.

(b) Central Limit Theorem requires that the data from the distribution has a finite mean and variance. Cauchy distribution is not normalizable and therefore has neither a finite E(x) nor Var(x). Therefore we cannot use the CLT, so $f(x)$ does not converge against Gauss distribution. As can be seen in figure 9 the Cauchy distribution has a shorter, narrower peak than the normal distribution, but has broader tails (= has a higher probability for outliers compared to Gaussian distribution).

$$\int_{-\infty}^{\infty} \frac{1}{(x-x_0)^2+\gamma^2} dx \qquad \text{integral does not converge}$$

$f(x)$ is not normalizable $\rightarrow$ E(x) and Var(x) cannot be calculated

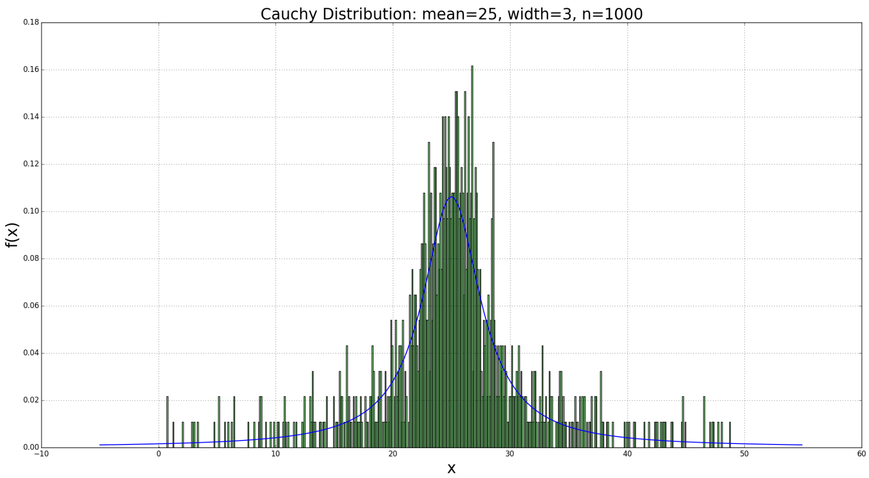

Figure 9: Cauchy distribution for $x_0 = 25$ and $\gamma = 3$. The Cauchy distribution has a shorter, narrower peak and broader tails compared to Gaussian distribution.

4.2 Exercise 11

Problem

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) $\mu_x = 0, \mu_y = 0, \sigma_x = 1, \sigma_y = 1, \rho_{xy} = 0$
(b) $\mu_x = 1, \mu_y = 2, \sigma_x = 1, \sigma_y = 1, \rho_{xy} = 0.7$
(c) $\mu_x = 1, \mu_y = -2, \sigma_x = 1, \sigma_y = 2, \rho_{xy} = -0.7$

Solution

The plotted contours can be seen in figure 10. The effect of the correlation coefficient ρ_{xy} can be seen in the second plot. The effect of the diagonal correlations matrix elements not normalized can be seen in the third plot, where the off diagonal correlation matrix element σxy is twice as large as ρ_{xy} because of σ_y.

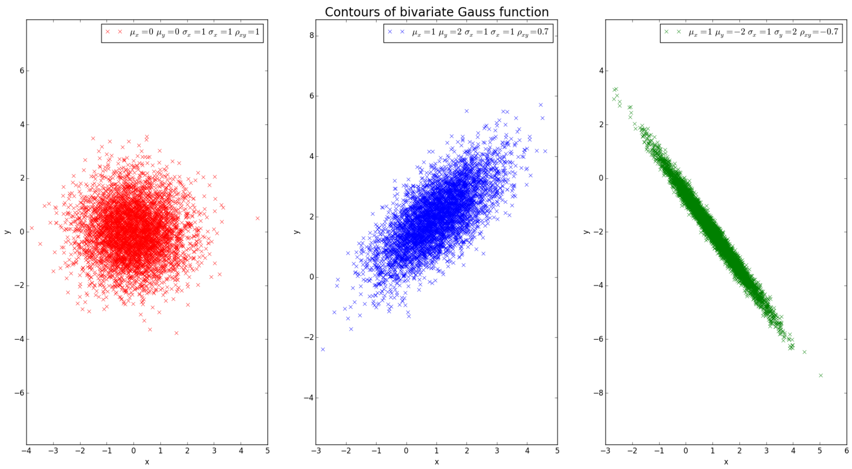

Figure 10: The contours of bivariate Gauss functions are plotted according to μ_i, σ_i and ρ_{xy}.

4.3 Exercise 12

Problem

Bivariate Gauss probability distribution

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

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

(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

$$\mu_z = \mu_x - \mu_y$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

Solution

(a)

$$P(x,y) = \frac{1}{2\pi|\Sigma|^{0.5}} \exp(-\frac{1}{2} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}^t \Sigma^{-1} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}$$

μ_x, μ_y-shift, move distribution to origin:

$$P(x,y) = \frac{1}{2\pi|\Sigma|^{0.5}} \exp(-\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^t \Sigma^{-1} \begin{pmatrix} x \\ y \end{pmatrix})$$

First we look at the pre factor:

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

$$|\Sigma| = \sigma_x^2\sigma_y^2 - cov(x,y)^2) = \sigma_x^2\sigma_y^2(1 - \frac{cov(x,y)^2}{\sigma_x^2\sigma_y^2} = \sigma_x^2\sigma_y^2(1-\rho^2)$$

Now we look at the exponent:

$$(x,y)\frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{pmatrix} \sigma_y^2 & -cov(x,y) \\ -cov(x,y) & \sigma_x^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}(\sigma_y^2 x^2 - 2xycov(x,y) + \sigma_x^2 y^2)$$

Putting everything together we get:

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

If we would "reshift" our function, we would get:

$x \to x - \mu_x \qquad y \to y - \mu_y$

(b)

$z = x - y \rightarrow y = x - z$

$$P(z) = \int_{-\infty}^{+\infty} P(x, x-z)dx = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)}\frac{x^2}{\sigma_x^2} + \frac{(x-z)^2}{\sigma_y^2} - \frac{2\rho x(x-z)}{\sigma_x\sigma_y})dx$$

Substituting with new function:

$F = \beta' x^2 - \gamma' x + \delta'$

Defining terms:

$$\beta' = \frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} - \frac{2\rho}{\sigma_x\sigma_y} \qquad \gamma' = 2z(\frac{1}{\sigma_y^2} - \frac{\rho}{\sigma_x\sigma_y}) \qquad \delta' = \frac{z^2}{\sigma_y^2}$$

$$\beta = \frac{1}{2(1-\rho^2)}\beta' \qquad \gamma = \frac{1}{2(1-\rho^2)}\gamma' \qquad \delta = \frac{1}{2(1-\rho^2)}\delta'$$

$$P(z) = A\int_{-\infty}^{+\infty} \exp(-\beta x^2 + \gamma x - \delta)dx = A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta x^2 + \gamma x)dx$$

$$= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta x(x - \frac{\gamma x}{\beta})dx \overset{x \to x + \frac{\gamma}{2\beta}}{=} A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta(x - \frac{\gamma}{2\beta})(x + \frac{\gamma}{2\beta}))dx$$

$$= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta(x^2 - \frac{\gamma^2}{4\beta^2}))dx = A\exp(-\delta + \frac{\gamma^2}{4\beta^2})\int_{-\infty}^{+\infty} \exp(-\beta x^2)dx$$

Gauss integral: $\int_{-\infty}^{+\infty} \exp(-\beta x^2)dx = \sqrt{\frac{\pi}{\beta}}$

$$P(z) = A\sqrt{\frac{\pi}{\beta}}\exp(-\delta + \frac{\gamma^2}{4\beta^2})$$

with: $A\sqrt{\frac{\pi}{\beta}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y}} \qquad -\delta + \frac{\gamma^2}{4\beta^2} = -\frac{z^2}{2\underbrace{(\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y)}_{\sigma_z^2}}$

$\rightarrow \sigma_z^2 = (\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y)$

$\rightarrow \mu_z = \mu_x - \mu_y$

4.4 Exercise 13

Problem

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

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

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

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

What is the predicted distribution for the observed quantity y?

Solution

The problem can be defined as the convolution of two Gaussian PDFs

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_f} \exp - \frac{(x-x_f)^2}{2\sigma_f^2} \text{and} \qquad g(x) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp - \frac{(x-x_g)^2}{2\sigma_g^2}$$

The convolution of two functions $f(t)$ and $g(t)$ is defined as:

$$\int_0^x f(x-\tau)g(\tau)d\tau = f * g$$

Next we use the convolution theorem:

$$F^{-1}[F(f(x))F(g(x))] = f(x) * g(x)$$

with F the Fourier transform and F^{-1} the inverse Fourier transform:

$$F(f(x)) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi ikx}dx$$

$$F^{-1}(f(k)) = \int_{-\infty}^{+\infty} f(k)e^{2\pi ikx}dk$$

Using the transformation $x' = x - x_f$ the Fourier transform of $f(x)$ is given by

$$F(f(x)) = \frac{1}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^{+\infty} e^{\frac{(x')^2}{2\sigma_f^2}} e^{-2\pi ik(x'-x_f)}dx' = \frac{e^{-2\pi ikx_f}}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^{+\infty} e^{\frac{(x')^2}{2\sigma_f^2}} e^{-2\pi ikx'}dx'$$

Using Eulers formula $e^{-i\theta} = \cos\theta - i\sin\theta$ we can split the term in e' to

$$F(f(x)) = \frac{e^{-2\pi ikx_f}}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^{+\infty} e^{\frac{(x')^2}{2\sigma_f^2}} (\cos(2\pi kx') - i\sin(2\pi kx')dx'$$

The term in $\sin(x')$ is odd and so its integral over all space will be zero, leaving

$$F(f(x)) = \frac{e^{-2\pi ikx_f}}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^{+\infty} e^{\frac{(x')^2}{2\sigma_f^2}} \cos(2\pi kx')dx'$$

$$\text{From } \int_0^{\infty} e^{-at^2}\cos(2xt)dt = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-}\frac{x^2}{a}$$

$$\rightarrow F(f(x)) = e^{-2\pi ikx_f}e^{-2\pi^2\sigma_f^2k^2}$$

$$F(f(x))F(g(x)) = e^{-2\pi ik(x_f+x_g)}e^{-2\pi^2(\sigma_f^2+\sigma_g)k^2}$$

Comparing the last two Fourier transforms we can see that:

$$x_{f*g} = x_f + x_g \qquad \sigma_{f*g} = \sqrt{\sigma_f^2 + \sigma_g^2}$$

$$P_{f*g}(x) = F^{-}1[F(f(x))F(g(x))] = \frac{1}{\sqrt{2(\sigma_f^2+\sigma_g^2)}} \exp(-\frac{(x-(x_f+x_g))^2}{2(\sigma_x^2+\sigma_g^2)})$$

We have shown that the convolution of two Gauss PDF is again a Gauss distribution. Therefore the distribution of y will be Gaussian again.

4.5 Exercise 14

Problem

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 $10^{-30}\frac{cm^2}{steradian}$. Assume the quoted errors correspond to one Gaussian standard deviation. The assumed model has the form

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

(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

(a) The posterior probability for the wanted parameter(s) $a(s)$ with observed data d_i (here crossections σ_i) at the points x_i (here angles θ_i) can be derived by Bayes Theorem.

$$P(a|d_i) = \frac{P(d_i|a)P_0(a)}{\int P(d_i|a)P_0(a)da} \qquad P(d_i|a) = P(d_i|f(x_i|a))$$

For N data points we get:

$$P(a|\{d\}) = \frac{P(\{d\}|a)P_0(a)}{\int P(\{d\}|a)P_0(a)da} = \prod_{i=1}^{N} \frac{P(d_i|a)P_0(a)}{\int P(d_i|a)P_0(a)da}$$

Assuming flat priors, our posterior probability becomes:

$$P(a|d) = \prod_{i=1}^{N} \frac{P(d_i|a)P_0(a)}{\int P(d_i|a)P_0(a)da} \overset{\text{Flat Prior}}{=} \prod_{i=1}^{N} \frac{P(d_i|a)}{\int P(d_i|a)da}$$
$$= \prod_{i=1}^{N} \frac{P(d_i|f(x_i|a)}{\int P(d_i|f(x_i|a)da} = \prod_{i=1}^{N} \frac{\frac{1}{\sqrt{2\pi}\sigma_i}\exp(-\frac{(d_i-\sigma(\theta_i))^2}{2\sigma_i^2})}{\int P(d_i|f(x_i|a)da}$$

(b) The derived posterior probability can now be used to determine the wanted values A,B,C of the function $\sigma(\theta)$. For the determination we maximize the posterior by numerically calculating the best values for A,B,C so that $d_i \to \sigma(\theta_i)$. The calculated values for the variables are $A = 15.356$, $B = -1.076$, $C = -2.568$. The assumed model with the mode values is fitted to the observed data points in figure 11 .

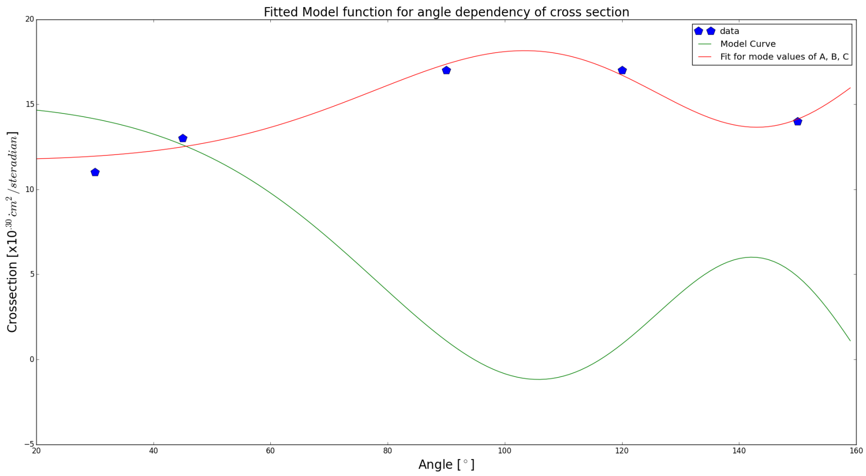

Figure 11: Data points of table [exercise 4.14], $\sigma(\theta) = A + B\cos(\theta) + C\cos(\theta^2)$ was fitted twice, once for a "random guess" and once for the mode values of A, B, C.

5 Model Fitting and Model selection

5.1 Exercise 1

Problem

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).

Solution

(a) The posterior probability density function for the parameters(s) λ is given by:

$$P(\lambda|\{r\},\{N\}) = \frac{P(\{r\}|\{N\},\lambda)P_0(\lambda)}{\int P(\{r\}|\{N\},\lambda)P_0(\lambda)d\lambda} = \frac{P(\{r\}|\{N\},\lambda)P_0(A)P_0(E_0)}{\int P(\{r\}|\{N\},\lambda)P_0(A)P_0(E_0)d\lambda}$$

$$P(\{r\}|\{N\},\lambda) = \prod_{i=1}^{k} \frac{N_i!}{(N_i - r_i)!r_i!}\epsilon(\lambda)_i^r(1-\epsilon(\lambda))^{N_i - r_i}$$

We specify λ and the locations where we measure, $\{x\}$, then we have the prediction for the success probabilities $\{f(x|\lambda)\}$. To perform a Bayesian Analysis and extract information on the parameters in our function, we need to specify the priors. The registered data (see figure) can be fitted by a modified Sigmoid function with 2 parameters:

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

The parameters A and E_0 can be first guessed from the data:

$$P_0(E_0) = G(E_0|\mu = 2.0, \sigma = 0.3) \qquad P_0(A) = G(A|\mu = 2., \sigma = 0.5)$$

and then numerically optimized to get the best values. for the priors we get two Gauss distribution, which were plotted in figure for the mode values $E_0 =$ and $A =$.

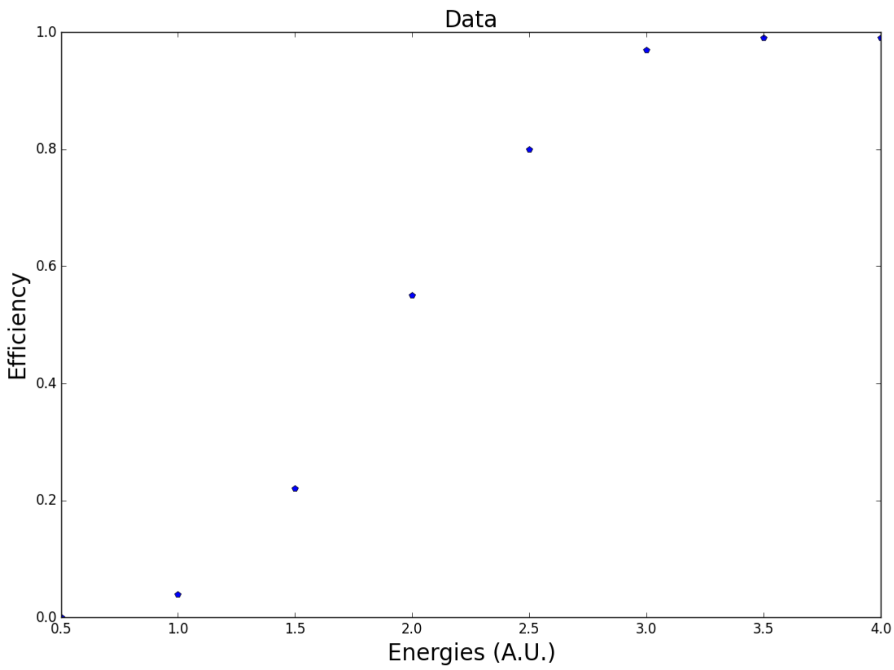

Figure 12: Data points from table [exercise 5.1] plotted against energy.

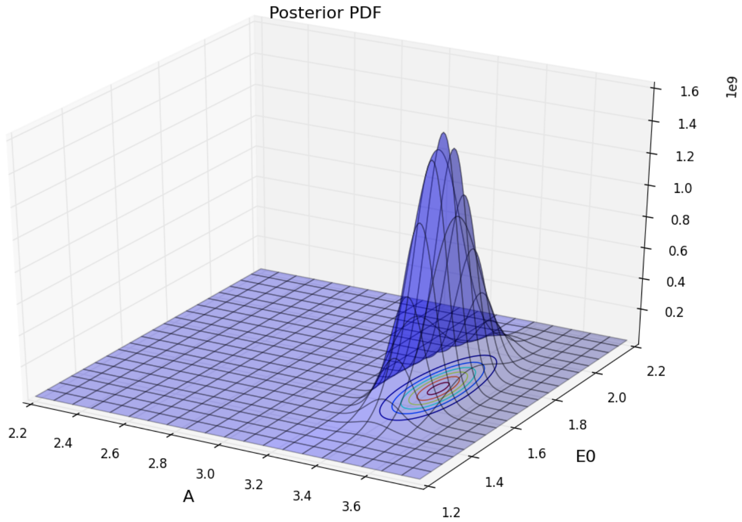

Figure 13: Approximated Gauss distributions for flat priors plotted against E_0 and A

(b) For the test statistic we take the product of all the probabilities of our data points

$$F(\{r_i\}|\lambda) = \prod_i binom(r_i, N_i, P_i\lambda))$$

5.2 Exercise 2

Problem

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

$$\epsilon(E) = \sin(A(E - E_0)) \tag{1}$$

(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 results

Solution

(a) Posterior probability

$$\begin{aligned} P(\lambda|\{r_i\},\{N_j\}) &= \frac{P(\{r_i\}|\{N_i\},\lambda)\cdot P_0(\lambda)}{\int P(\{r_i\}|\{N_i\},\lambda)\cdot P_0(\lambda)d\lambda} \\ &= \frac{P(\{r_i\}|\{N_i\},\lambda)\cdot P_0(A)\cdot P_0(E_0)}{\int P(\{r_i\}|\{N_i\},\lambda)\cdot P_0(A)\cdot P_0(E_0)d\lambda} \end{aligned}$$

$$P(\{r_i\}|\{N_j\},\lambda) = \prod_{i=1}^{k} \binom{N_i}{r_i} \epsilon(\lambda)^{r_i}(1-\epsilon(\lambda))^{N_i - r_i}$$

$$\lambda = [x_i, A, E_0] \qquad \epsilon(\lambda) = \sin(A(E - E_0))$$

The (non normalized) normal distributions of the priors can be seen in figure 14. The combined distribution is plotted for a rough estimation in figure 15

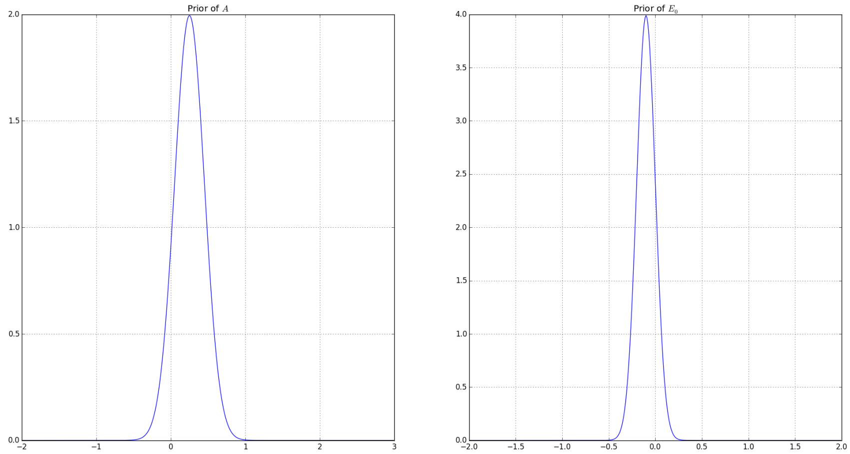

Figure 14: Combined distribution of A and E_0 for rough grid

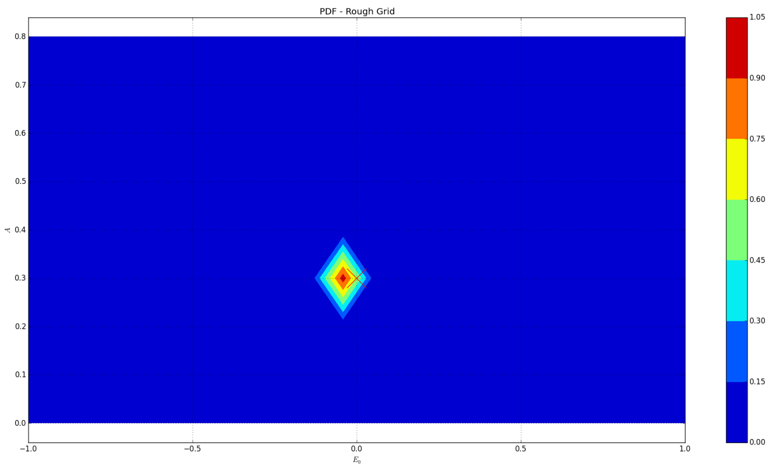

Figure 15: (Not normalized) Gauss distribution of the priors $P(A)$ and $P(E_0)$

(b) The 68% region according to E_0 and A is plotted in figure 16 in addition to the 90 and 95 interval and the mode, which is marked as red cross. (Unfortunately, the interval got a little bit thinner than it should have be)

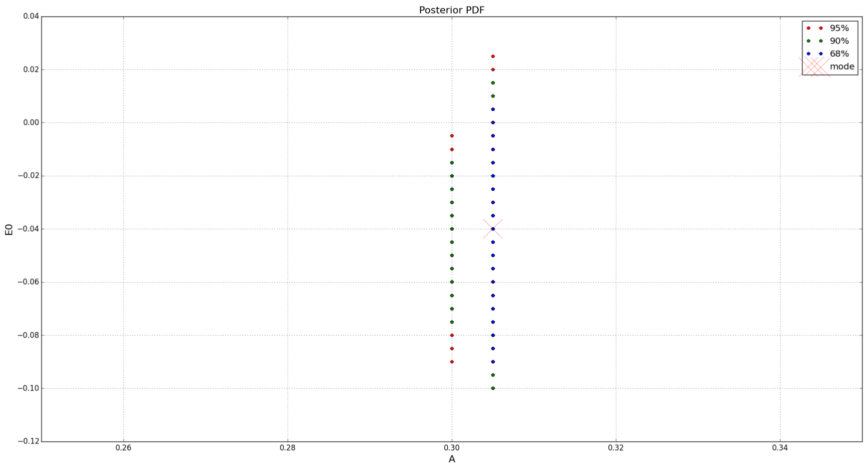

Figure 16: (68% CL interval depending from E_0 and A)

(c) Discussion of the results:
The smallest Interval for E_0 contains positive values, but we got a negative expectation value.
Even the fine grid, does not describe the data very well (see figure 17), the sine model does not seem to work properly. Furthermore the data was modeled only for small energy, therefore the fit cannot show us what happens for larger energies. Consequently it is a bad model function.

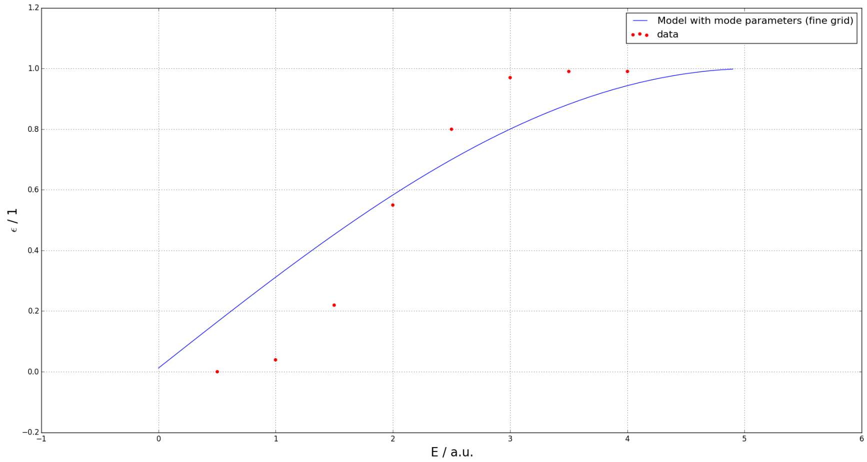

Figure 17: Fine grid for the data points with $A = 0.305$ and $E_0 = -0.04$ still does not fit the data good enough.

5.3 Exercise 3

Problem

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

Solution

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

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

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

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

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

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

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

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

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

5.4 Exercise 8

Problem

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:

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

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

$$f(x|A,B,C,x_0,\Gamma) = A + Bx + Cx^2 + \frac{D}{(x-x_0)^2 + \Gamma^2} \tag{3}$$

(a) 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.

(b) 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

Solution

(b) The values of the parameters were calculated by assuming flat priors and can be seen in figure 18. The resulting Bayes Fit with the most likely parameters for the first formula can be seen in figure 19 . The calculation for the quadratic + Breit-Wigner Fit took too much computational power for my computer, but the python code can be seen in the appendix.There I only wanted to investigate 20 values for the parameters, for the 68% smallest interval.

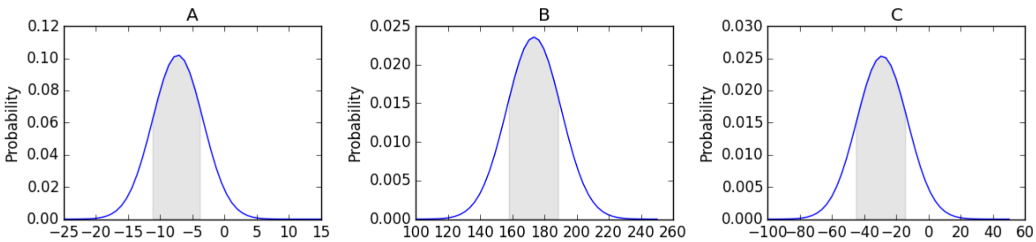

Figure 18: Value probability for the quadratic representing background

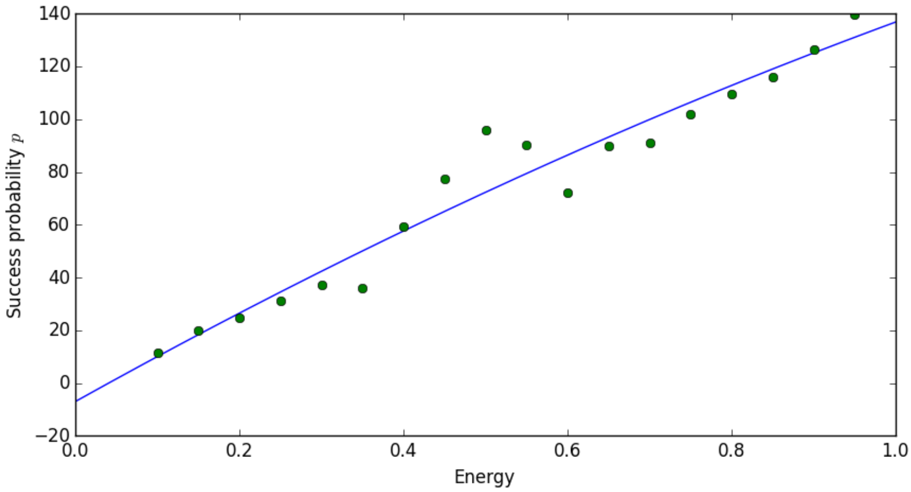

Figure 19: Fit with only quadratic background

6 Maximum Likelihood Estimator

6.1 Exercise 1

Problem

The family of Bernoulli distributions have the probability density

$$P(x|p) = p^x(1-p)^{1-x}$$

(a) Calculate the Fischer information $I(p) = -E[\frac{\partial^2 lnP(x|p)}{\partial p^2}]$.

(b) What is the maximum likelihood estimator for p*?*

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

Solution

(a) Fischer information:
The Fischer information is for the probability distribution of the data for one data point.

It tells us how quickly a distribution is changing near the true value of the parameter.

$$P(x|p) = p^x(1-p)^{1-x} \qquad x \in \{0,1\} \qquad p \in [0,1]$$
$$\partial_p P(x|p) = \partial_p ln(p^x(1-p)^{1-x}) = \frac{x}{p} - \frac{1-x}{1-p}$$
$$\partial_p^2 P(x|p) = \partial_p(\frac{x}{p} - \frac{1-x}{1-p}) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$
$$I(p) = -E[-\frac{x}{p^2} - \frac{1-x}{(1-p)^2}] = \frac{1}{p^2}E[x] + \frac{1}{(1-p^2)}(E[1]-E[x]) = \frac{1}{p} + \frac{1-p}{(1-p)^2} = \frac{1}{(1-p)^2}$$

(b) Maximum likelihood estimator:
We need to find the $\hat{p}$ that maximizes the likelihood $L(p)$. We will use a "trick" that often makes the differentiation a bit easier. The value $\hat{p}$ that maximizes the natural logarithm of the likelihood function $ln(L(p))$ is also the value of p that maximizes the likelihood function $L(p)$. So, the "trick" is to take the derivative of $ln(L(p))$ (with respect to p) rather than taking the derivative of $L(p)$.

$$L(p) = p^x(1-p)^{1-x} \qquad ln(L(p)) = xln(p) + (1-x)ln(1-p)$$
$$\partial_p L(p)|_{\hat{p}} = 0 = \partial_p ln(L(p))|_{\hat{p}}$$
$$\partial_p ln(L(p)) = \frac{x}{p} - \frac{1-x}{1-p} = 0 \qquad x(1-p) - p + px = 0$$

For 1 experiment: $\hat{p} = x$ In general, for n experiments : $\hat{p} = \frac{\sum_{i=1}^n x_i}{n}$

(c) Expected distribution for $\hat{p} - p_0$:

$$\frac{f(a)-f(b)}{a-b} = f'(c) \qquad c \in [a,b]$$
$$a \hat{=} \hat{p} \qquad b \hat{=} p_0 \qquad f(p) \hat{=} L'_n(p) \qquad L'_n(\hat{p}) = 0$$
$$\frac{L'_n(\hat{p}) - L'_n(p_0)}{\hat{p} - p_0} = L''_n(p_1) \qquad p_1 \in [\hat{p}, p_0] \quad \rightarrow \quad \hat{p} - p_0 = \frac{L'_n(p_0)}{L''_n(p_1)}$$
$$L(p) = E[lnP]|_{p_0} \qquad \text{LNN:} \quad L_n \rightarrow L$$
$$L'_n(p_0) = L'_n(p_0) - \underbrace{L'(p_0)}_{=0} = \frac{1}{n}\sum_{i=1}^{n}(\partial_p P|_{p_0} - E[\partial_p lnP|_{p_0}])$$

Assumption: $\partial_p P|_{p_0}$ is well behaving, we can apply CLT $\rightarrow L'(p_0) \propto N(0|\sigma)$

$$L''_n(p) = \frac{1}{n}\sum_{i=1}^{n}\frac{-p^2+2px-x}{p^2(1-p)^2} \qquad n \rightarrow \infty: \quad LNN: E[\frac{-p^2+2px-x}{p^2(1-p)^2}]|_{p_0} = -I(p_0)$$
$$Var[(\frac{L'_n(p_0)}{L''_n(p_1)})^2] = E[\frac{(L'_n(p_0))^2}{nI(p_0)^2}] = \frac{\overbrace{[(L'_n(p_0))^2]}^{=I(p_0)}}{nI(p_0)^2} = \frac{1}{nI(p_0)} = Var[\frac{L'_n}{L''_n}]$$
$$\hat{p} - p_0 = N(0, \frac{1}{\sqrt{nI(p_0)}})$$

6.2 Exercise 2

Problem

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

(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.

Solution

(a) Generating n values of x by using $x = -\ln U$, where U is a uniformly distributed random number between (0,1) and plotting them, we see that our distribution tends towards the exponential distribution $P(x|p)$, see figure 20. For finding our MLE estimator, we determined the results of $n_{experiments} = 1000$ experiments for values $n = 2, 10, 100$ and plotted it in figure 21. For the experiments with average value of $n = 2andn = 10$ we cannot recognize a Gauss distribution, but we see that the expectation value of the MLE estimator tends towards 1. For $n = 100$ the distribution of the MLE estimator clearly goes to a Gauss distribution with expectation value of 1.

(b) Comparison with the theory

For the distribution we would expect a Gaussian distribution with

$$P(x|\lambda) = \lambda e^{-\lambda x} \qquad L = \prod_{i=1}^{n} P(x_i|\lambda)$$

$$L_n = \frac{1}{n}\sum_{i=1}^{n} log(P(x_i|\lambda)) = \frac{1}{n}\sum_{i=1}^{n}(log\lambda + log(e^{-\lambda x})) = log\lambda + \frac{1}{n}\sum_{i=1}^{n}(-\lambda x_i) = log\lambda - \bar{x}$$

$$\partial_\lambda L = \frac{1}{\lambda} - \bar{x} \stackrel{!}{=} 0 \qquad \rightarrow \qquad \hat{\lambda} = \frac{1}{\bar{x}}$$

$$\text{here: } \lim_{n\to\infty:} \qquad x_n = \bar{x} = \int_0^\infty x e^{-x} dx = 1 \qquad \rightarrow \hat{\lambda} = 1$$

and a standard deviation $\sigma \propto \frac{1}{n}$. Both the distribution (Gauss) and the theoretical values (expectation value and standard deviation of λ) are supported by the results.

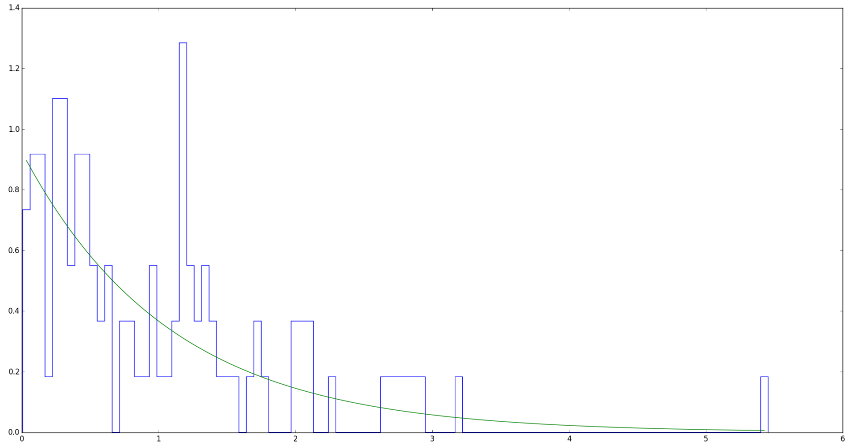

Figure 20: Distribution of $n = 100$ random values generated by $x = -\ln U$ where U with fit. The exponential probability is already recognizable.

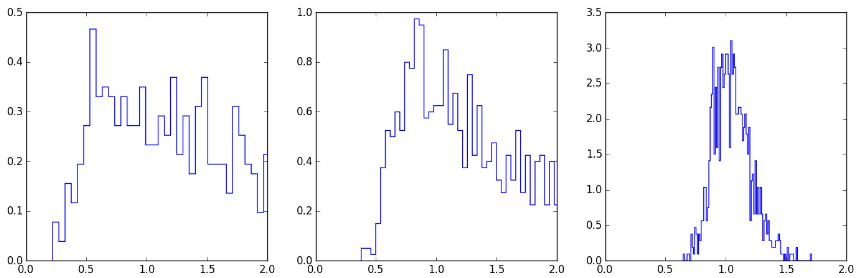

Figure 21: Distributions of MLE estimator $\hat{\sigma}$ for $n_{experiment} = 1000$ repetitions $n = 2$, and $n = 10$ and $n = 100$, respectively

Appendix for 5.8

import numpy as np import matplotlib.pyplot as plt import random import math

#lam=0.2 M=10000

def P(x): return lam*math.exp(-lam*x)

def Gauss(x,mu,sigma): return 1.0/(math.sqrt(2.0*math.pi)*sigma)*math.exp(-0.5*((x-mu)/sigma)**2)

def getRandomDist(lam): return - math.log(random.random())/lam

def getMeans(lam,n):

dist=np.zeros(n) mean=np.zeros(M) for i in range(M): mean[i]=0.0; for j in range(n): dist[j]=getRandomDist(lam) mean[i]+=dist[j] mean[i]=mean[i]/n return mean

x=np.arange(0,10,0.01) y=np.zeros(len(x)) density=1.

fig, plots = plt.subplots(nrows=3,ncols=3, figsize=(10, 8))

for i,plot_rows in enumerate(plots): for j,plot in enumerate(plot_rows): n=10**i lam=1.0/((j)**2+1) sigma=1.0/(lam*math.sqrt(n)) mu=1.0/lam plot.hist(getMeans(lam,n), 50, range=[0,9]) #plot.legend(['test']) plot.set_title("n=" + str(n) + "
$frac1\lambda$=" + str(mu))

x=np.arange(0,9,0.01) y=np.zeros(len(x)) for k in range(len(y)): y[k]=Gauss(x[k],mu,sigma); plot.plot(x,y) #plot.xlim(0,10)

plt.tight_layout()

plt.savefig('plot_ex04_08a.pdf', dpi=300, bbox_inches='tight')

plt.show()