Data Analysis WS17/18: Exercise solutions

Irina Ilicheva, 03663403

5. 4. 2018

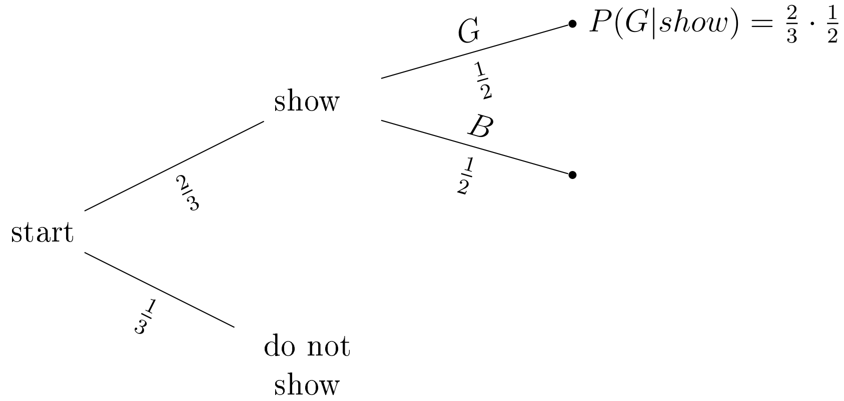
Contents

1 Introduction to Probabilistic Reasoning

1.1 Ex.1

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?

First scenario: both possibilities to have a male or a female child do not depend on each other, therefore $P(G) = P(B) = 1/2$.



Second scenario: Jane could have 3 children sets: two girls (GG), one girl and one boy (GB) or two boys (BB). There is at least one girl in two of three sets, therefore $P(show) = 2/3$. Then we have $P(G) = P(B) = 1/2$ as described above. The total possibility that the second child is also a girl is $P(G|show) = 2/3 \cdot 1/2 = 1/3$

1.2 Ex.2

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

Probability of the data is the likelihood that the measured data will have a specific attribute. This attribute has to be well-defined to avoid misunderstanding, for example (6H and 6T) is not the same outcome as (HHHTTTHHHTTT).

1.3 Ex.3

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?

Insufficient information. To form a conclusion, the probability distribution of the particle detector is necessary.

1.4 Ex.4

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)

Abbreviations: Sp = spots, L = lethargy, T = thirst and Sn = sneezes. Use Bayes theorem

$$P(SF|\{Sp,L,T,Sn\}) = \frac{P(\{Sp,L,T,Sn\}|SF)\cdot P(SF)}{P(\{Sp,L,T,Sn\})} \tag{1}$$

where the probability to have swamp fever with all four symptoms

$$P(\{Sp,L,T,Sn\}|SF) = P(Sp|SF)\cdot P(L|SF)\cdot P(T|SF)\cdot P(Sn|SF) \tag{2}$$

and the probability to have all four symptoms is the weighted sum of the probabilities to have all four symptoms with and without swamp fewer

$$\begin{aligned} P(\{Sp,L,T,Sn\}) = P(SF)\cdot P(\{Sp,L,T,Sn\}|SF) &+ \\ + P(\overline{SF})\cdot P(\{Sp,L,T,Sn\}|\overline{SF}) \end{aligned} \tag{3}$$

Insert (2) and (3) in (1) and use given numbers:

$$P(SF|\{Sp, L, T, Sn\}) =$$

$$= \frac{1 \cdot 1 \cdot 0.6 \cdot 0.2 \cdot 10^{-4}}{10^{-4} \cdot 1 \cdot 1 \cdot 0.6 \cdot 0.2 + (1 - 10^{-4}) \cdot 0.03 \cdot 0.1 \cdot 0.02 \cdot 0.05} \approx 80\% \quad (4)$$

Other probabilities are calculated analogously:

$$P(SF|\{Sp, L, T\}) \approx 50\%$$
$$P(SF|\{Sp, T, Sn\}) \approx 28.57\%$$
$$P(SF|\{Sp, L, Sn\}) \approx 11.77\%$$
$$P(SF|\{L, T, Sn\}) \approx 10.72\%$$

2 Binomial and Multinomial Distribution

2.1 Ex. 8

For the following function

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

(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

(a) Mean value $\overline{x}$ can be defined as

$$\overline{x} = E[x] = \int_0^\infty x \cdot P(x)dx$$

Standard deviation σ is defined as

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

Integrals are evaluated using integration by parts (I.b.P.):

$$\overline{x} = E[x] = \int_0^\infty x^2 e^{-x} dx \overset{\text{I.b.P.}}{=} 2$$

$$\sigma = \sqrt{\int_0^\infty x^3 e^{-x} dx - E[x]^2} \overset{\text{I.b.P.}}{=} \sqrt{6 - 2^2} = \sqrt{2}$$

The probability content in the interval $(\overline{x} - \sigma, \overline{x} + \sigma)$:

$$\int_{\overline{x}-\sigma}^{\overline{x}+\sigma} P(x)dx \overset{\text{I.b.P.}}{=} 0.7375$$

(b) Median $\mathrm{x}_{\mathrm{med}}$ is the value of x for which the cumulative probability $F(x_{\mathrm{med}})$ reaches 50 %

$$F(x_{\mathrm{med}}) = \int_0^{x_{\mathrm{med}}} P(x)dx = -e^{-x_{\mathrm{med}}}(1 + x_{\mathrm{med}}) + 1 = \frac{1}{2}$$

This condition is satisfied only when $x_{\text{med}} \approx 1.6784$

A so-called central interval $\mathcal{O}^C_{1-\alpha}$ is an interval $[a, b]$ with $\overline{x} = (b-a)/2$ that contains at least $1-\alpha$ probability:

$$P(r < a) <= \alpha/2 \quad \text{and} \quad P(r > b) <= \alpha/2 \tag{5}$$

In order to evaluate the 68% confidence interval α was set to 0.32.

$$\int_0^a P(x)dx = 0.16 \quad \text{and} \quad \int_b^\infty P(x)dx = 0.16$$

Solving for a and b gives

$$\mathcal{O}^C_{0.68} = [0.7120, 3.2885]$$

(c) Mode x^* is the value of x where $P(x)$ has a maximum. $P(x)$ is a product of one linearly increasing function and one exponentially decreasing function, therefore P(x) has one extremum, which is a maximum. The extremum of a function can be found by taking a derivative of the function and setting it to zero

$$P'(x^*) = e^{-x^*}(1 + x^*) = 0 \quad \Longrightarrow \quad x^* = 1$$

A so-called smallest interval $\mathcal{O}^S_{1-\alpha}$ is the shortest interval $[a, b]$ that contains at least $1-\alpha$ probability. In a unimodal distribution the smallest interval is defined as

$$\int_a^b P(x)dx = 1 - \alpha \qquad P(a) = P(b) \qquad a < x^* < b \tag{6}$$

The smallest interval could not be found analytically. However, it is still possible to determine the SI numerically. A short Python script was written for this purpose.

The working principle of the script (See Fig. 1):

(i) Define the variable h. At the beginning its value is $h = P(x^*)$ - the maximal value of $P(x)$.

(ii) Find a and b such that $P(a) = P(b) \approx h$. When we calculate something numerically, we can not completely evade uncertainty, therefore the absolute numeric tolerance of 0.0001 is used. In the first iteration both a and b are equal to x^*, but as the script runs a decreases and b increases. a and b are used as the integration limits.

(iii) Integrate $P(x)$ over $[a, b]$ and check if the value of the integral is larger than $(1 - \alpha)/100$. If so, the present values of a and b are final and are displayed. If not, h is lowered by 0.0001 and steps 1. to 4. are performed anew.

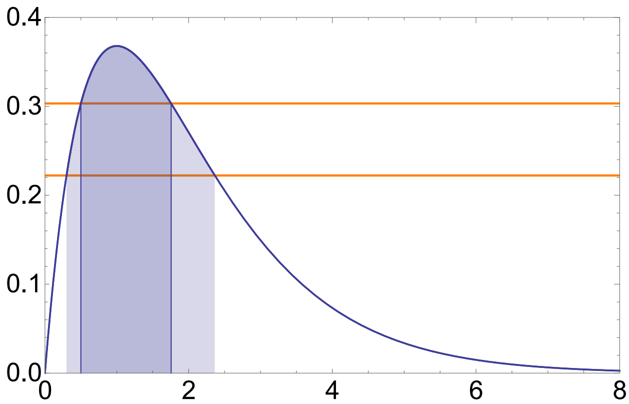


Figure 1: A graph of the $P(x)$: h is denoted by orange lines, a and b are denoted by blue vertical lines. With each iteration h decreases, which makes the integral of $P(x)$ between a and b larger

The script gives following result:

$$\mathcal{O}^S_{0.68} = [0.27, 2.492]$$

2.2 Ex. 10

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

Table 1

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.

The posterior is the binomial distribution:

$$P(r|N,p) = \binom{N}{r} p^r (1-p)^{N-r} \tag{7}$$

Some posteriors with different N and p values are shown in Fig. 2.

The mode of each distribution p^* is computed numerically. The working principle of this computation is:

(i) Define the range of $p \in [0,1]$ with the step size Δp that is small enough. Set the value of p^* to zero.

(ii) Check if $P(r|N,p^*+\Delta p) > P(r|N,p^*)$. If so, set $p^* = p^* + \Delta p$.

(iii) Do (ii) until $p = 1$ is reached.

The 68% smallest intervals for p were computed similar to Exercise 2.1. Both mode and 68% smallest interval for each distribution are presented on the Fig. 3. The smallest intervals corresponding to energies of 3.0 eV, 3.5 eV and 4.0 eV are very small (sic!) because in these cases the number of measurements N is significantly larger than in other cases.

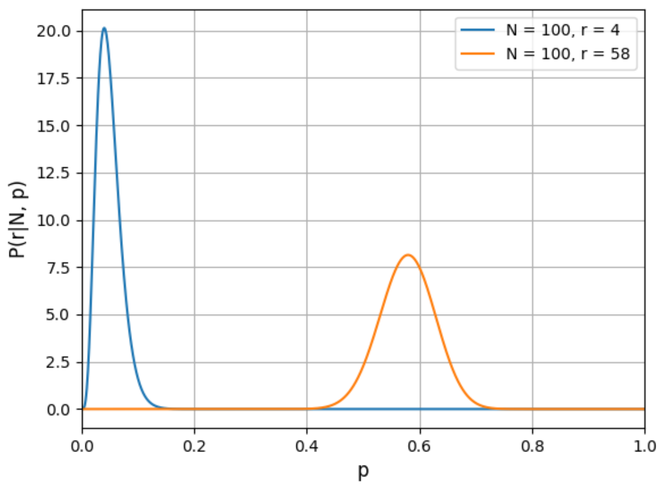


Figure 2: Some posteriors. They are all unimodal and differ in the mode value, maximum value and FWHM

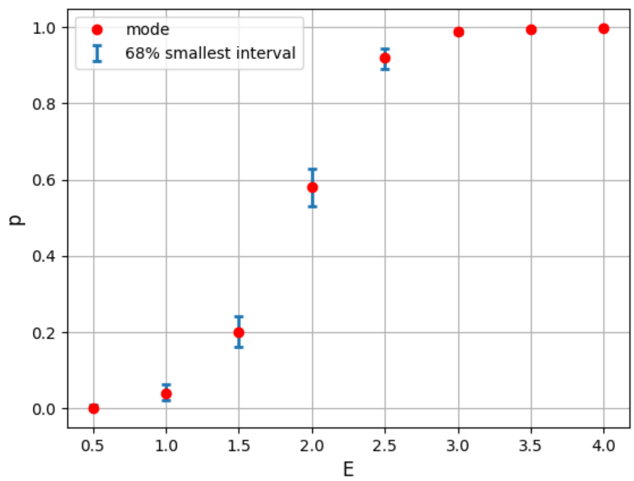


Figure 3: Modes and smallest intervals for p

2.3 Ex. 11

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

For the frequentist analysis one has to construct the Neyman Confidence Level using the central interval definition as in Eq. 5. Again, α was set to 0.32 in order to find 68% confidence interval. The Neyman bands are built by computing a and b for several values of $p \in [0,1]$ with step size of $\Delta p = 0.001$. Obtained Neyman bands are shown in Fig. 4. The central interval corresponding a certain r is the interval of p between a and b bands for a given r.

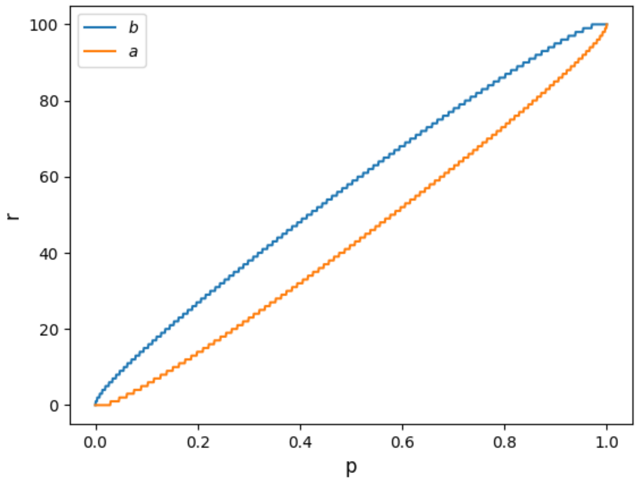


Figure 4: 90% Confidence bands

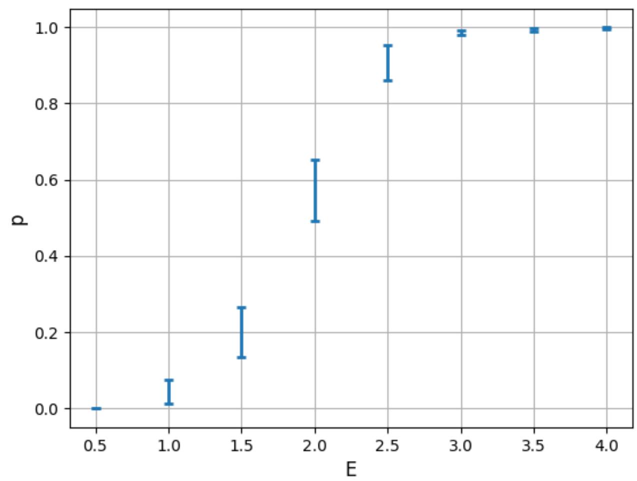


Figure 5: 90% CL intervals

The 90% CL intervals for p are better visualized on the Fig. 5. For each energy E the respective value of r from the Tab. 1 was taken and their central intervals calculated.

2.4 Ex. 13

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

$$P_n(p|r,N) = \frac{(nN+1)!}{(nr)!(nN-nr)!} p^{nr}(1-p)^{n(N-r)} : \tag{8}$$

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

First, find $P_1(p|r,N)$:

$$P_1(p|r,N) = \frac{P(r|N,p)P_0(p)}{\int P(r|N,p)P_0(p)dp}$$

with the binomial distribution $P(r|N,p)$ (Eq. 7) and the flat prior $P_0(p) = 1$

$$P_1(p|r,N) = \frac{\binom{N}{r}p^r(1-p)^{N-r}}{\int \binom{N}{r}p^r(1-p)^{N-r}dp} \overset{\beta \text{ function}}{=} \frac{(N+1)!}{r!(N-r)!}p^r(1-p)^{(N-r)}$$

Then, find $P_2(p|r,N)$:

$$P_2(p|r,N) = \frac{P(r|N,p)P_1(r|N,p)}{\int P(r|N,p)P_1(r|N,p)dp} = \frac{p^{2r}(1-p)^{2N-2r}}{\int p^{2r}(1-p)^{2N-2r}dp}$$
$$\overset{\beta \text{ function}}{=} \frac{(2N+1)!}{(2r)!(2N-2r)!}p^{2r}(1-p)^{(2N-2r)}$$

The result (Eq. 8) can be received via mathematical induction.

The expectation value and the variance:

$$E[p] = \frac{r+1}{N+2} \Longrightarrow E[p]_n = \frac{nr+1}{nN+2}$$
$$\Longrightarrow \lim_{n\to\infty} E_n[p] = \frac{r}{N}$$

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

3 Poisson Distribution

3.1 Ex. 4

Consider the function $P(x) = \frac{1}{2}e^{-|x|} \quad -\infty \leq x < \infty$
(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?

(a)

$$E[x] = \int_{-\infty}^{\infty} x \cdot P(x)dx = \frac{1}{2}\left[\int_{-\infty}^{0} x \cdot e^{x}dx + \int_{0}^{\infty} x \cdot e^{-x}dx\right] \overset{\text{I.b.P.}}{=} 0$$

(b)

$$V[x] = E[x^2] - E[x]^2 = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{2}e^{-|x|}dx \overset{\text{I.b.P.}}{=} 2 \Longrightarrow \sigma = \sqrt{2}$$

The maximal value of $P(x)$ is $\frac{1}{2}$, therefore its half maximum value is $\frac{1}{4}$.

$$\frac{1}{4} \overset{!}{=} \frac{1}{2}e^{-|x|} \Longrightarrow x = \pm\ln(2)$$

therefore

$$\text{FWHM} = 2\ln(2) < \sigma$$

(c) The peak of $P(x)$ is obviously at $x^* = 0$, therefore the probability content in $[-\sigma, \sigma]$:

$$\int_{-\sigma}^{\sigma} P(x) = \frac{1}{2}\left[\int_{-\infty}^{0} e^{x}dx + \int_{0}^{\infty} e^{-x}dx\right] = 0.7569$$

3.2 Ex. 7

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?

Consider a Poisson distribution $P(\nu|n)$ with $n = 9$.

(a) The 95 % probability lower limit on the Poisson expectation value ν if found using repeated numerical computing of the value $P(\nu|9)$ for $\nu \in [0, x]$ where x is high enough. Once $P(\nu|9)$ reaches 0.05, the program stops and gives the last used ν For this computation $x = 9$ is used due to $P(\nu|9)$ being unimodal and $\nu^* = 9$ being the mode. The step size for ν is 0.001.
The program gives

$$\nu_{\text{lowlim}} = 5.444$$

The Poisson probability distribution, the cumulative probability and the 95 % probability lower limit on ν are visualized in the Figure 6

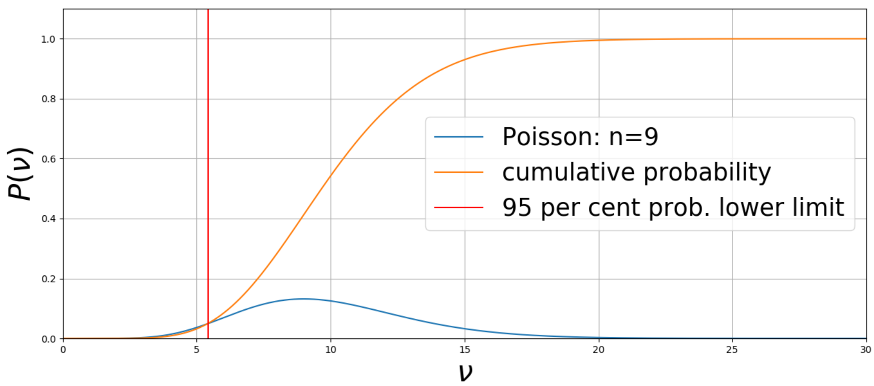


Figure 6

(b) Neyman construction and 68% confidence level interval visualization is done as described in Section 2.2. Smallest intervals are found as described in Section 2.1.

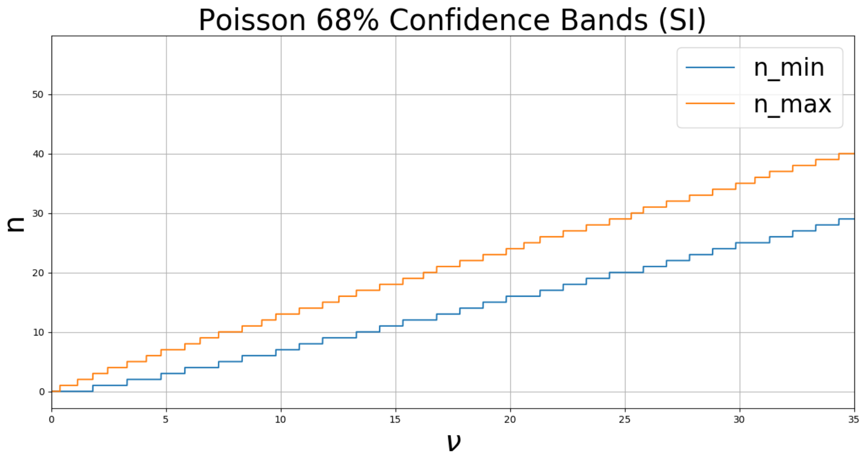


Figure 7

Note that the Poisson confidence bands diverge as opposed to the binomial confidence bands.

3.3 Ex. 8

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

(a)

(b) First, the Neyman 68% CL bands (with smallest interval definition) corresponding to each $\mu = \nu + \lambda$ are calculated. They have the same form as the confidence bands from the previous exercise (Fig. 7). Due to $\nu = \mu - \lambda$ and $\lambda = 3.2$ the confidence bands corresponding to ν are the bands from the Fig. 7 that are shifted by 3.2 downwards. Note that ν can never be negative because the event rate is always positive. All values of ν that would be negative are set to zero. The new CL bands are presented on the Fig. 8.

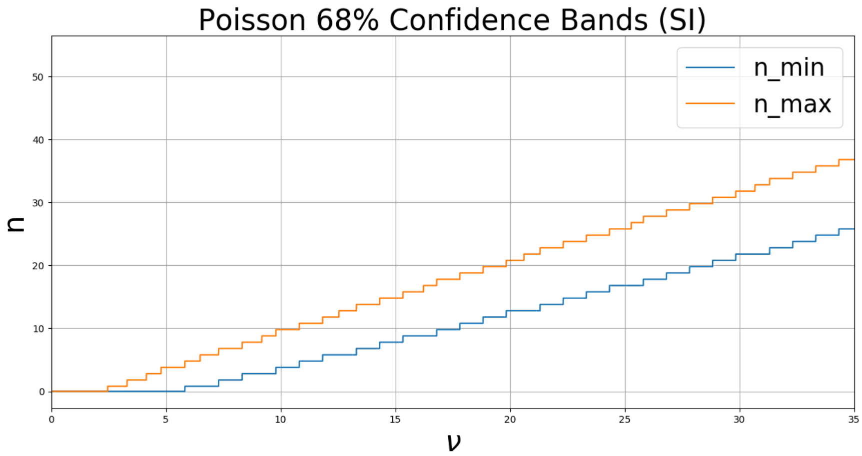


Figure 8: Poisson CL bands with known background of $\lambda = 3.2$ events

In this case there are significantly more empty intervals. It is fine in the Neyman procedure. If one does not want to see any empty intervals, one should use the Feldman-Cousins construction.

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

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

Let us use the smallest interval definition. The smallest 68% Credible interval for ν can be found by applying the script described in Ex. 82.1 on the posterior $P(\nu|3.2, 9)$.

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

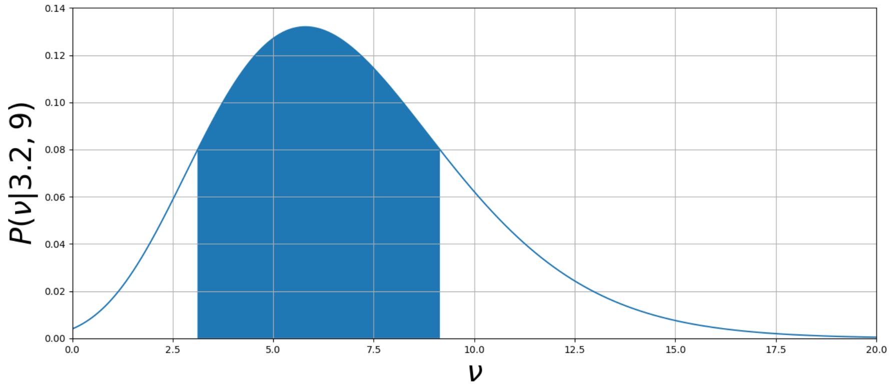


Figure 9: Bayesian posterior as the function of ν with the 68% Credible interval (filled)

Both posterior and the Credible interval are shown of Fig. 9.

3.4 Ex. 13

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 with $i = 1, ..., n$. The unbinned likelihood is defined as the product of the densities

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

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

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

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

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

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.

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

$$
\begin{aligned}
\lim_{K\to\infty}\prod_{j=1}^{K}\frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} &\overset{\text{(i)}}{=} \lim_{K\to\infty}\prod_{j=1}^{K} e^{-\nu_j}\nu_j^{n_j} \overset{\text{(ii)}}{=} \prod_{j=1}^{\infty} e^{-f_j\Delta}(f_j\Delta)^{n_j} \\
&\overset{\text{(iii)}}{\propto} \prod_{j=1}^{\infty}(1-f_j\Delta)\cdot(f_j\Delta)^{n_j} = \prod_{j=1}^{\infty}(f_j\Delta)^{n_j} - \underbrace{(f_j\Delta)^{n_j+1}}_{\text{negligible}} \\
&= \prod_{i=1}^{n} f_j\Delta = \prod_{i=1}^{n} f(x_i|\lambda)\Delta
\end{aligned}
$$

3.5 Ex. 16

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

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

Show that the probability distribution is given by

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

Use $N = X + M$ where X and M stand for number of hits and misses respectively. Bayes:

$$P(x) = \sum_{N}^{\infty} P(X|N)P(N) = \sum_{N=X}^{\infty} P(X|N)P(N) \tag{9}$$

Both sums are equivalent because $P(X) = 0$ for $N < X$ (only the "misses" are considered).
In our case we have a fixed number of occurrences N. Each occurrence either happens ("hit") or does not happen ("miss") with their respective probabilities p and $1-p$. Therefore one can use the binomial distribution for

$$P(X|N) = \binom{N}{X} p^X (1-p)^{N-X}$$

$P(N)$ is defined in the exercise:

$$P(N) = P(N|\nu) = \frac{\nu^N}{N!}e^{-\nu}$$

Plug both distributions in Eq. 9 and use $M = N - X$:

$$\begin{aligned}
P(X) &= \sum_{N=H}^{\infty} \frac{N!}{X!(N-X)!}p^X(1-p)^{N-X} \cdot \frac{\nu^N}{N!}e^{-\nu} \\
&= \frac{p^X e^{-\nu}}{X!} \sum_{N=H}^{\infty} \frac{(1-p)^{N-X}}{(N-X)!}\nu^X \\
&= \frac{\nu^X p^X e^{-\nu}}{X!} \sum_{M=0}^{\infty} \frac{((1-p)\nu)^M}{M!} \\
&= \frac{(\nu p)^X}{X!}e^{-\nu p}
\end{aligned}$$

4 Gaussian Probability Distribution Function

4.1 Ex. 8

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

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

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

$$x = -\frac{\ln U}{\lambda} \tag{10}$$

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

$$x = \gamma \tan(\pi U - \pi/2) + x_0 \tag{11}$$

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.

(a) The average x in the $p(x)$ distribution:

$$E[x] = \int_0^\infty x \cdot p(x)dx = \frac{1}{\lambda}$$

Its variance:

$$V[x] = \sigma^2 = E[x^2] - E[x]^2 = \frac{1}{\lambda^2}$$

At first n samples are taken using Eq. 10 and the random.random() Python function. Then the means of these n samples is calculated ans stored. These two steps are repeated 10000 times in order to gather enough means for the plots. The following parameters of a Gauss distribution are expected in this case:

$$\mu = E[x] = \frac{1}{\lambda} \qquad \text{and} \qquad \sigma = \frac{1}{\lambda\sqrt{n}} \tag{12}$$

Let us see how the distribution of means changes with the increasing number of samples n. The distributions in the Fig. 10 are normalized in order to enable the comparison of the distribution and a corresponding Gaussian curve.

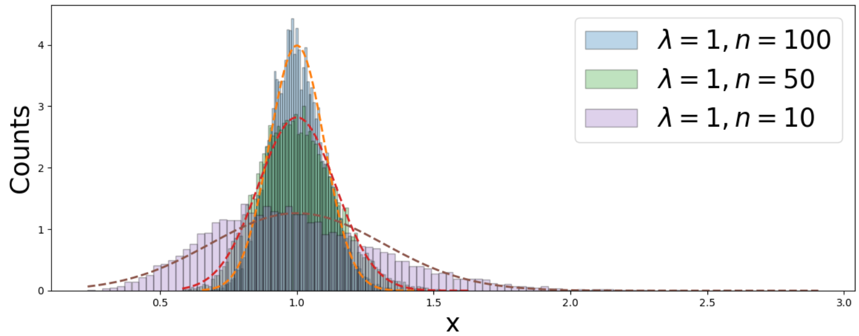


Figure 10: Distributions built with different sample quantity n and Gaussian curves with μ and σ according to Eq. 12

It is evident that:

- the distribution narrows with increasing sample quantity
- the Gaussian curves with μ and σ according to Eq. 12 comply with their respective distributions.

Now let us look how the distributions behave when λ is changed. Three distributions with different λ are shown on the Fig. 11.

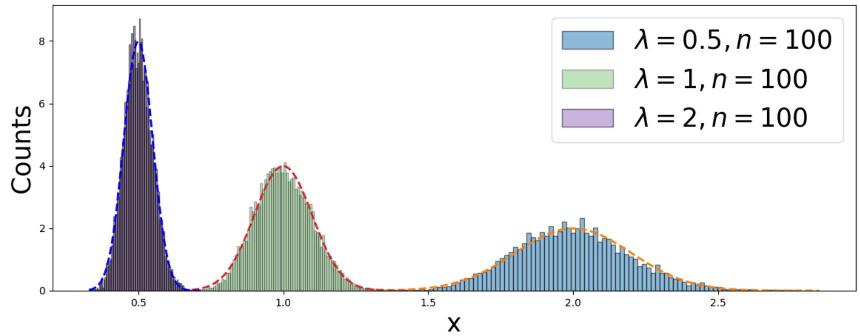


Figure 11: Distributions built with different parameters λ and Gaussian curves with μ and σ according to Eq. 12

Again, the Gaussian curves with μ and σ according to Eq. 12 comply with the distributions of the means. This shows that even given enough experiments, the averages of x tend towards the normal (Gaussian) distribution despite the fact that the individual x are not normally distributed. Therefore, CLT is valid in this case.

(b) Firstly, let us plot the Cauchy distribution:

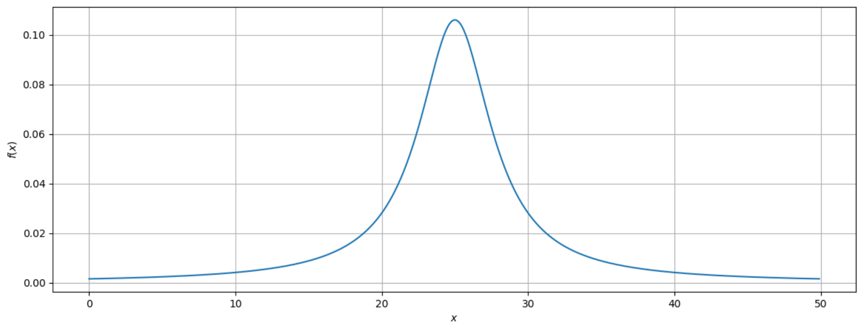


Figure 12: The Cauchy distribution according to Eq. 12 with $x_0 = 25$ and $\gamma = 3$

One can see that the distribution has very long "tails". It means that the random x picked using the Eq. 11 could be anywhere on the X axis. This phenomenon can be better visualized if we plot both "pick functions", the Eq.s 10 and 11:

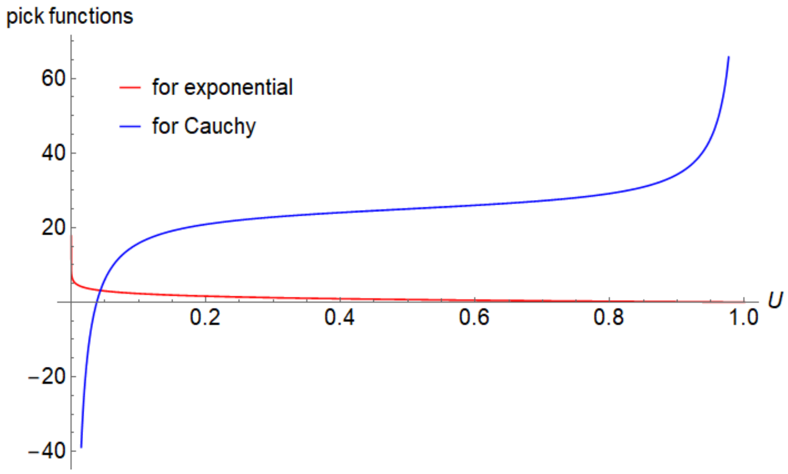


Figure 13: The pick functions for random values of x from the exponential and Cauchy distributions

The value of the pick function for the exponential distribution is limited, therefore the obtained x values are highly unlikely to differ from each other. The pick function for the Cauchy distribution diverges at $U \to 0$ and $U \to \infty$, thus making the distribution of x flatter and more dependent on the random quantity U than in previous case.

It is also worth noting, that the Cauchy distribution does not have a formal average value:

$$E[x] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \cdots \overset{\omega = \mathrm{x} - \mathrm{x}_0}{=} \underbrace{\int_{-\infty}^{\infty} \frac{\omega}{\omega^2 + \gamma^2} d\omega}_{\text{diverges}} + \int_{-\infty}^{\infty} \frac{X_0}{\omega^2 + \gamma^2} d\omega$$

Therefore it is not possible to calculate the predicted values of mu and σ in this case.

Fig. 14 shows the distribution of x and the "distribution" of the average of 100 picked values of x. The distribution of x values was calculated many times in order to get something that looks relatively nice. For example, in many attempts some of the x values were larger than ten thousand although the bulk of the x values resided around thirty, thus making the histogram uncomfortable to look at.

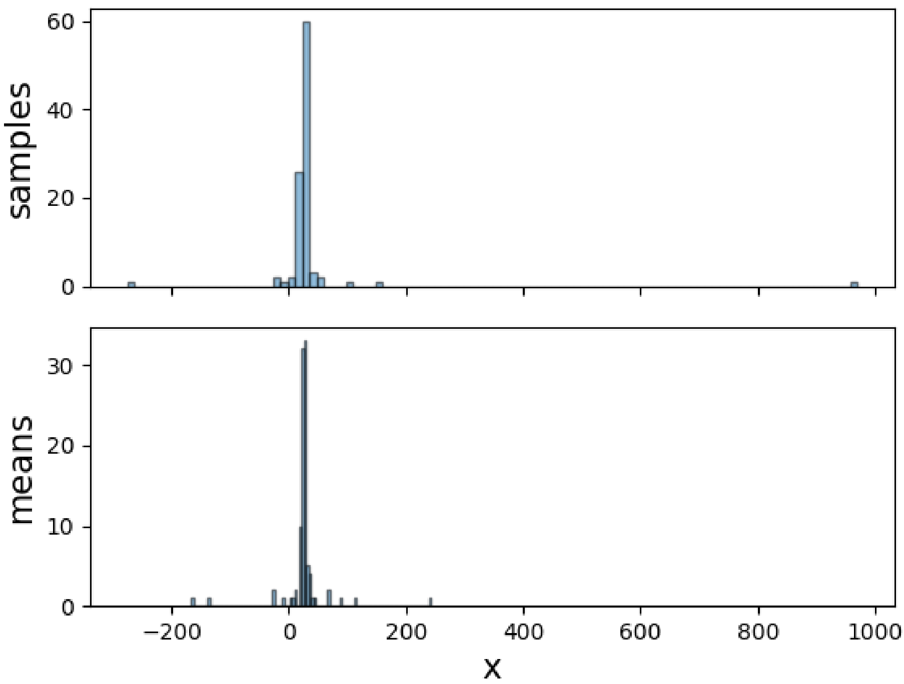
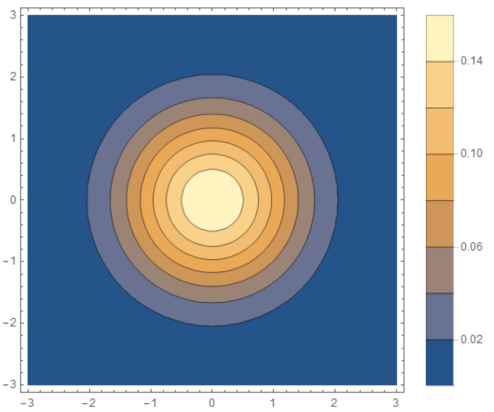


Figure 14: The distributions of x and $\overline{x}$ according to Cauchy distribution

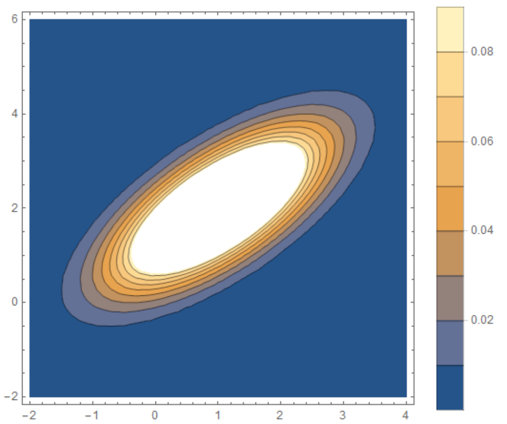
It is evident that it is not possible to fit the distribution of means with the Gauss curve, because the distribution is way too narrow. Together with the fact, that the Cauchy distribution does not have an average (and therefore the variance) and thus the parameters μ and σ do not exist, it is clear that in this case the CLT in NOT valid.

4.2 Ex. 11

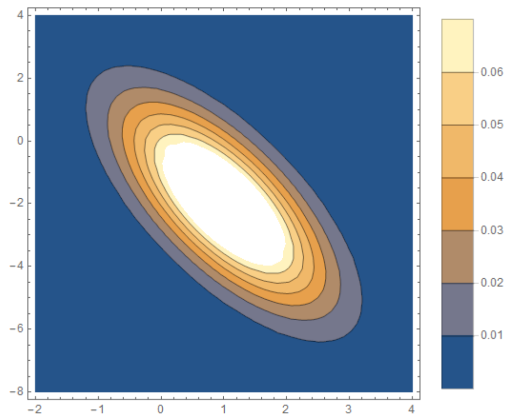
With a plotting program, draw contours of the bivariate Gauss function (see next exercise for the definition of the function) for the following parameters:
(a) $\mu_x = 0, \quad \mu_y = 0, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho_{xy} = 0$
(b) $\mu_x = 1, \quad \mu_y = 2, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho_{xy} = 0.7$
(c) $\mu_x = 1, \quad \mu_y = -2, \quad \sigma_x = 1, \quad \sigma_y = 2, \quad \rho_{xy} = -0.7$



(a)



(b)



(c)

Figure 15

4.3 Ex. 12

Bivariate Gauss probability distribution
(a) Show that the pdf can be written in the form

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

(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

$$\mu_z = \mu_x - \mu_y$$
$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

(a) The multivariate Gaussian distribution is defined as

$$f(x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right) \tag{13}$$

with the quantity of variables N and covariance matrix Σ.

$$\Sigma = \begin{pmatrix} cov(x_1,x_1) & cov(x_1,x_2) & \dots & cov(x_1,x_N) \\ cov(x_2,x_1) & \ddots & & \dots \\ \vdots & & & \\ cov(x_N,x_1) & & & cov(x_N,x_N) \end{pmatrix}$$

Use $cov(x_i,x_j) = \rho_{ij}\sigma_i\sigma_j$ and $N=2$ to get:

$$|\Sigma| = \begin{vmatrix} \overbrace{\rho_{xx}}^{=1}\sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \underbrace{\overset{=1}{\rho_{yy}}}\sigma_y^2 \end{vmatrix} = (1-\rho_{xy})\sigma_x^2\sigma_y^2$$

The inverse matrix of Σ:

$$\Sigma^{-1} = \frac{1}{(1-\rho_{xy})\sigma_x^2\sigma_y^2}\begin{pmatrix} \sigma_y^2 & -\rho_{xy}\sigma_x\sigma_y \\ -\rho_{xy}\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}$$

With the values calculated above and $(\vec{x}-\vec{\mu}) = (x-\mu_x, y-\mu_y)^T$ the exponent of the Eq. 13 is

$$\Big(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\Big) = \dots = -\frac{1}{2(1-\rho_{xy}^2)\sigma_x^2\sigma_y^2} \cdot \Big((x-\mu_x)^2\sigma_y^2 - 2\rho\sigma_x\sigma_y(x-\mu_x)(y-\mu_y) + \sigma_x^2(y-\mu_y)^2\Big)$$

By inserting $|\Sigma|$ and the exponent in Eq. 13 we get

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

(b)

$$z = x - y \Longrightarrow y = x - z$$

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

Substitute the exponent with function $\beta' x^2 - \gamma' x + \delta'$ with terms

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

The terms without an apostrophe are useful for taking the coefficient outside:

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

Now the pdf takes the form:

$$
\begin{aligned}
P(z) &= A\int_{-\infty}^{+\infty} \exp(-\beta x^2 + \gamma x - \delta)dx \\
&= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta x^2 + \gamma x)dx \\
&= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta x(x - \frac{\gamma x}{\beta})dx \\
&= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta(x - \frac{\gamma}{2\beta})(x + \frac{\gamma}{2\beta}))dx \\
&= A\exp(-\delta)\int_{-\infty}^{+\infty} \exp(-\beta(x^2 - \frac{\gamma^2}{4\beta^2}))dx \\
&= A\exp(-\delta + \frac{\gamma^2}{4\beta^2})\int_{-\infty}^{+\infty} \exp(-\beta x^2)dx \\
&= A\sqrt{\frac{\pi}{\beta}}\exp(-\delta + \frac{\gamma^2}{4\beta^2})
\end{aligned}
$$

In the fourth line the property of the integration over $\mathbb{R}$ is used and in the last line the Gauss integral is used.

The back substitution of β, γ and δ yields the Gaussian probability distribution with

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

4.4 Ex. 13

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

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

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

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

What is the predicted distribution for the observed quantity y?

The characteristic function of $f(x)$ is defined as

$$\Phi_{f(x)}(k) = \int_{-\infty}^{\infty} e^{ikx} f(x) dx \tag{14}$$

Calculate the characteristic functions for $f(x)$ with $\omega = x - x_0$:

$$\begin{aligned}
\Phi_{f(x)}(k) &= \int_{-\infty}^{\infty} e^{ikx} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-x_0)^2}{2\sigma_x^2}} dx = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} e^{ik\omega} \cdot e^{ikx_0} \cdot e^{-\frac{\omega^2}{2\sigma_x}} d\omega \\
&= \frac{e^{ikx_0}}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{2\sigma_x}} \cdot [\cos(k\omega) + i \sin(k\omega)] \, d\omega \\
&= \frac{e^{ikx_0}}{\sqrt{2\pi}\sigma_x} \left[\int_{-\infty}^{\infty} e^{-\frac{\omega^2}{2\sigma_x}} \cos(k\omega) d\omega + i \underbrace{\int_{-\infty}^{\infty} e^{-\frac{\omega^2}{2\sigma_x}} \sin(k\omega) d\omega}_{= 0 \text{ , integrand is an odd function}} \right] \\
&= \exp\left[ikx_0 - \frac{k^2\sigma_x^2}{2} \right]
\end{aligned}$$

In the last step the integral $\int_0^\infty e^{-ax^2} \cos(bx) dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} e^{-b^2/4a}$ is used. Due to the integrand being even, this value can be multiplied by two to calculate the integral over the interval $(-\infty, \infty)$.

The characteristic functions for $P(y|x)$ is calculated analogously with $\omega' = y - x$:

$$\Phi_{P(y|x)}(k) = \int_{-\infty}^{\infty} e^{iky} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-x)^2}{2\sigma_y^2}} dy = \exp\left[ikx - \frac{k^2\sigma_y^2}{2} \right]$$

For the characteristic function of predicted distribution for quantity y calculate the product of both recently calculated functions:

$$\Phi_{f(x)}(k) \cdot \Phi_{P(y|x)}(k) = \exp\left[ik(x + x_0) - \frac{k^2}{2}(\sigma_x^2 + \sigma_y^2) \right]$$

The product has the same form as the characteristic function of the Gaussian distribution, therefore the predicted distribution for quantity y is again a Gaussian distribution.

4.5 Ex. 14

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

Table 2

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

$$\sigma(\theta) = A + B\cos(\theta) + C\cos(\theta^2) \tag{15}$$

(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 ?

(a) The posterior probability for the unknown parameter a with observed data y_i at the points x_i is derived by Bayes Theorem

$$P(a|y_i) = \frac{P(y_i|a)P_0(a)}{\int P(y_i|a)P_0(a)da} \quad \text{with} \quad P(y_i|a) \stackrel{\text{def}}{=} P(y_i|f(x_i|a))$$

For N data points we get:

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

Assuming flat priors, our posterior probability becomes:

$$P(a|y) = \prod_{i=1}^{N} \frac{P(y_i|a)P_0(a)}{\int P(y_i|a)P_0(a)da} \stackrel{\text{Flat prior}}{=} \prod_{i=1}^{N} \frac{P(y_i|a)}{\int P(y_i|a)da}$$
$$= \prod_{i=1}^{N} \frac{P(y_i|f(x_i|a))}{\int P(y_i|f(x_i|a))da} = \prod_{i=1}^{N} \frac{\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(y_i-\sigma(\theta_i))^2}{2\sigma_i^2}\right)}{\int P(y_i|f(x_i|a)da}$$

In this exercise the cross sections σ_i are used as y_i and the angles θ_i are used as x_i.

(b) The derived posterior probability is now used to determine the values A, B and C of the Eq. 15. To determine these parameters, the above-mentioned posterior probability needs to be maximized. The posterior is calculated numerically and the mode values A^*, B^* and C^* are found

$$A^* = 15.356, B^* = -1.076, C^* = -2.568$$

The given model (Eq. 15) with the mode values is fitted to data points in Fig 16. The fit curve does not go beyond the error bars, therefore it can be considered a good fit.

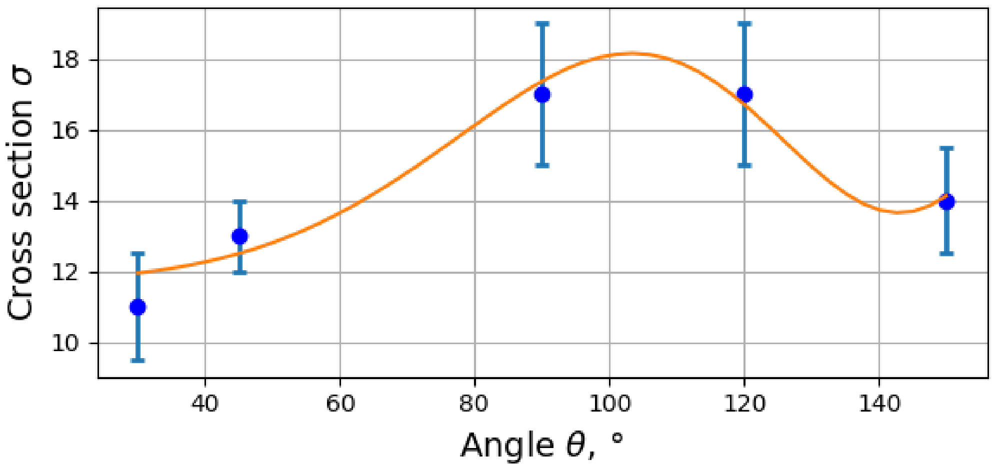


Figure 16: Data points and the fit (Eq. 15) with calculated parameters A^*, B^* and C^*.

5 Model Fitting and Model selection

5.1 Ex. 1

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

Table 3

(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).

(a) The Sigmoid function $S(E)$ that is used as a model is defined as

$$S(E) = \frac{1}{1 + \exp^{-A(E-E_0)}}$$

The posterior probability distribution for the parameters (A, E_0) can be calculated using the Bayes formula

$$P(A, E_0|\{r_i\}, \{N_i\}) = \frac{\overbrace{P(\{r_i\}, \{N_i\}|A, E_0)}^{=\mathcal{L}} \cdot P_0(A) \cdot P_0(E_0)}{\int P(\{r_i\}, \{N_i\}|A, E_0) \cdot P_0(A) \cdot P_0(E_0) dA dE_0}$$

with the likelihood $\mathcal{L}$ defined as

$$\mathcal{L} = \prod_{i=1}^{8} \binom{N_i}{r_i} S(E_i)^{r_i} (1 - S(E_i))^{N_i - r_i}$$

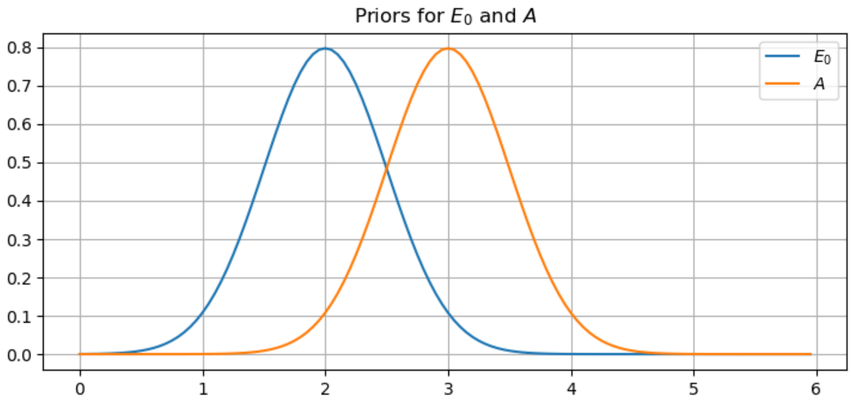


Figure 17: Assumed Gaussian priors for E_0 and A

where N_i, r_i are given in the Tab. 3.

Assume Gaussian priors $P_0(E_0)$ and $P_0(A)$ that are depicted on the Fig. 17.

To visualize the posterior probability distribution for, a grid of the (A, E_0) values with a certain step size can be used. For each point on the grid, the likelihood $\mathcal{L}$ and then the posterior $P(A, E_0|\{r_i\}, \{N_i\})$ are calculated. For the numerical calculation of the integral in the denominator the following approximation is used:

$$P(\{r_i\}, \{N_i\}|A, E_0) \cdot P_0(A) \cdot P_0(E_0) dA dE_0 \approx \mathcal{L} \cdot P_0(A) \cdot P_0(E_0) \cdot A_{\text{step}} E_{0,\text{step}}$$

where $A_{\text{step}}, E_{0,\text{step}}$ denote the step size of the A and E_0 on the grid, respectively. The value of the posterior is then stored.
For this calculation, the following grid is used:

A_{start}	2.5
A_{end}	3.5
$E_{0,\text{start}}$	1.0
$E_{0,\text{end}}$	3.0
no. of steps	100

Table 4: Grid parameters

Fig. 18 shows the posterior probability distribution function for A and E_0. The 3D surface shows the values of pdf and the contours below visualize its thickness. The mode of the pdf is at $(A^*, E_0^*) =$

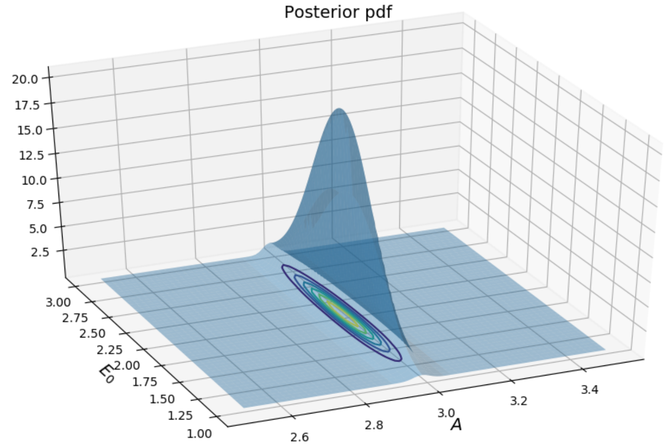


Figure 18: Posterior pdf for A and E_0

$(3.005, 1.970)$.
The Sigmoid function with $A = A^*$ and $E_0 = E_0^*$ fits the data nicely, as shown on the Fig. 19.

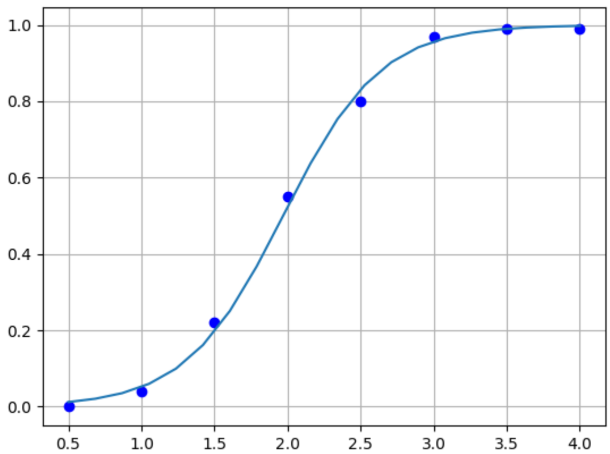


Figure 19: Data points and the Sigmoid function with the parameters A and E_0 calculated above. The curve fits the data nicely

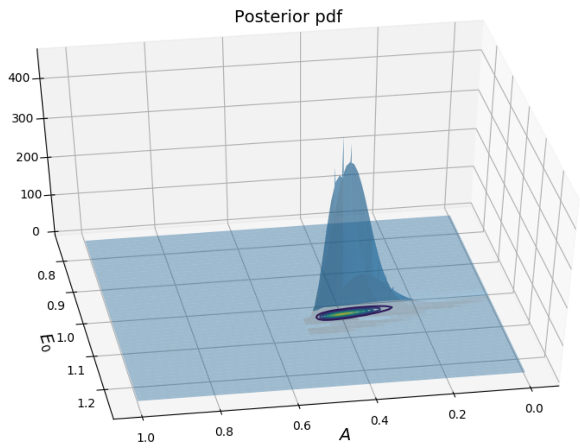


Figure 20: Posterior pdf for A and E_0

(b)

5.2 Ex. 2

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

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

(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

(a) This exercise is done analogously to Ex.1a with one exception: here the function $\epsilon(E)$ is used as a model. The priors $P_0(A)$ and $P_0(E_0)$ are assumed to be the same as before.

Fig. 20 shows the posterior probability distribution function for A and E_0. In this case the pdf looks much more "ragged" than before. The mode values are $(A^*, E_0^*) = (0.576, 0.952)$.
The $\epsilon(E)$ function with $A = A^*$ and $E_0 = E_0^*$ is not good for the data, as shown on the Fig. 21.

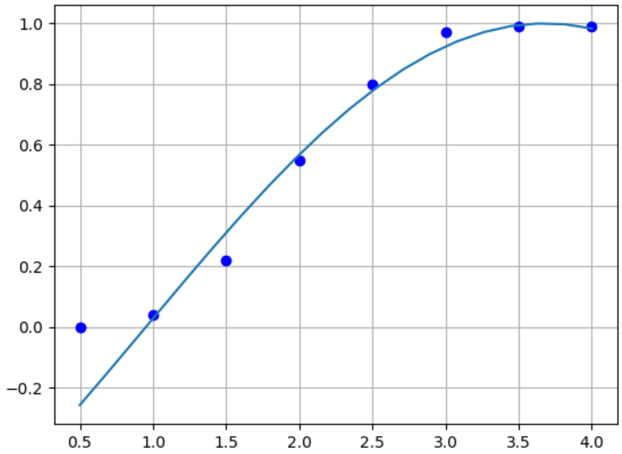


Figure 21: Data points and the $\epsilon(E)$ function with the parameters A and E_0 calculated above. Thus function does not suit the data

(b)

(c)

5.3 Ex. 3

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

χ^2 is defined as

$$\chi^2 = \sum_i \left(\frac{y_i - f(x_i|\lambda)}{\sigma_i} \right)^2$$

with $\sigma_i = \sigma(x_i|\lambda)$. For one data point:

$$\chi^2 = \left(\frac{y - f(x|\lambda)}{\sigma} \right)^2$$

Assume that the model function $f(x_i|\lambda)$ predicts the measurement values at points x and that the measurements are distributed according to the Gaussian pdf

$$P(y) = \mathcal{G}(y|f(x|\lambda), \sigma(x|\lambda))$$

Due to χ^2 being a function of y, the following holds:

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

The factor 2 is there because both positive and negative values of $(y - f(x|\lambda))$ contribute to the same χ^2.
Calculate the derivative of χ^2:

$$|\frac{d\chi^2}{dy}| = \frac{2}{\sigma} \cdot \left(\frac{y - f(x|\lambda)}{\sigma}\right) = \frac{2\sqrt{\chi^2}}{\sigma}$$

Therefore the probability density function for χ^2 is

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

First, find the average $\overline{\chi^2}$. For convenience, χ^2 is denoted as x.

$$\begin{aligned}
\overline{\chi^2} = \overline{x} = \int_0^\infty x \cdot P(x)dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{x} e^{-\frac{1}{2}\chi^2} dx \\
\overset{\text{I.b.P.}}{=} \frac{1}{\sqrt{2\pi}} \left\{ (0-0) + \int_0^\infty \frac{1}{\sqrt{x}} e^{-\frac{1}{2}x} dx \right\} \\
= \int_0^\infty \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}x} dx = \int_0^\infty P(x)dx = 1
\end{aligned}$$

The variance is calculated similar to the mean using the integration by parts (I.b.P.) two times

$$V[x] = \int_0^\infty x^2 \cdot P(x)dx - \left(\int_0^\infty x \cdot P(x)dx\right)^2 = \cdots = 2$$

The mode of the χ^2 distribution is

$$\chi^{2^*} = 0$$

due to the pdf being a product of an exponentially decaying function with the finite values at $\chi^2 \in [0, \infty]$ and a function with a positive pole at $\chi^2 = 0$.

5.4 Ex. 8

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:

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

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

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

(c) 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?

(d) 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	0.9	0.95
y	90.3	72.2	89.9	91.0	102.0	109.7	116.0	126.6	139.8

Table 5

6 Maximum Likelihood Estimators

6.1 Ex. 1

The family of Bernoulli distributions have the probability density

$$P(x|p) = p^x(1-p)^{1-x} \tag{16}$$

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

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

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

(a) The Fisher information:

$$\begin{aligned} I(p) &= -\mathrm{E}\left[\frac{\partial^2 \ln P(x|p)}{\partial p^2}\right] = -\mathrm{E}\left[-\frac{x}{p^2} - \frac{1-x}{(1-p)^2}\right] \\ &= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{(1-p)} \\ &= \frac{1}{p(1-p)} \end{aligned}$$

(b) We need to find the $\hat{p}$ that maximizes the likelihood $\mathcal{L}(p)$. The value $\hat{p}$ where the likelihood function $\ln(\mathcal{L}(p))$ has a maximum is also the value of p that maximizes the likelihood function $\mathcal{L}(p)$ itself.

$$\mathcal{L}(p) = p^x(1-p)^{1-x} \Longrightarrow \ln(\mathcal{L}(p)) = x\ln(p) + (1-x)\ln(1-p)$$

$$\partial_p \ln(\mathcal{L}(p))|_{\hat{p}} = \partial_p \mathcal{L}(p)|_{\hat{p}} \stackrel{!}{=} 0$$

$$\partial_p \ln(\mathcal{L}(p)) = \frac{x}{p} - \frac{1-x}{1-p} = 0 \Longrightarrow x(1-p) - p + px = 0$$

For 1 experiment:

$$\hat{p} = x$$

For n experiments ($\mathcal{L}_n'(p) = 0$):

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

6.2 Ex. 2

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

(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, λ (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.

(a) After generating n values of x by using $x = -\ln U$, where U is a uniformly distributed random number between (0,1) and plotting them, one can see that the distribution tends towards the exponential distribution $P(x|p)$, see Fig. 22.

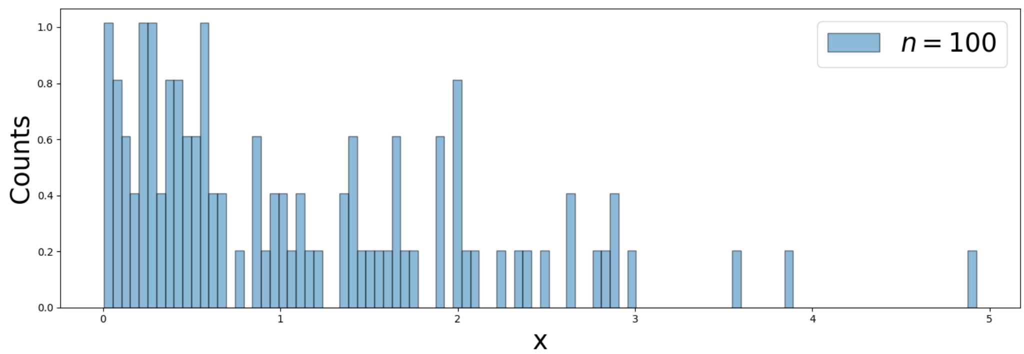


Figure 22: Hundred values generated from $P(x|\lambda)$

In order to determine the MLE estimator the results of $n_{\text{exp}} = 1000$ experiments for values $n = 2, 10, 100$ are calculated and plotted it in Fig. 23.

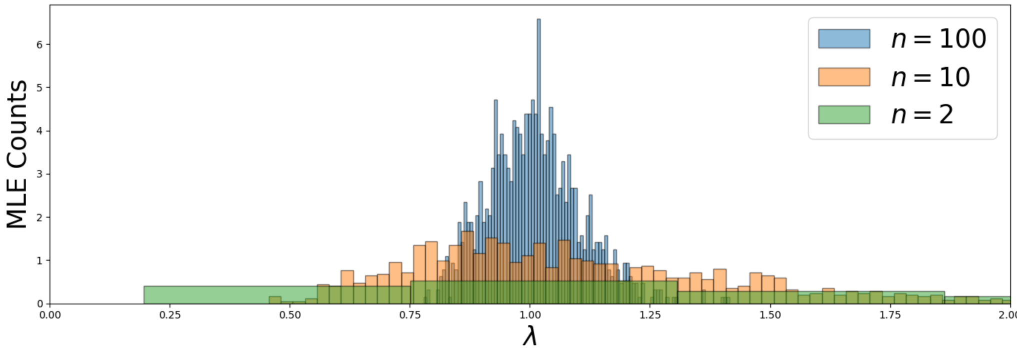


Figure 23: Distributions of MLE estimator for $n_{\text{exp}} = 1000$ repetitions of $n = 2$, and $n = 10$ and $n = 100$, respectively

For the experiments with average value of $n = 2$ and $n = 10$ 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 is similar to a Gauss distribution with expectation value of 1.

(b) A Gaussian distribution for the MLE is expected.

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

$$\begin{aligned}\mathcal{L}_n &= \frac{1}{n}\sum_{i=1}^{n} \log(P(x_i|\lambda)) = \frac{1}{n}\sum_{i=1}^{n}(\log\lambda + \log(e^{-\lambda x}))\\ &= \log\lambda + \frac{1}{n}\sum_{i=1}^{n}(-\lambda x_i) = \log\lambda - \bar{x}\end{aligned}$$

$$\partial_\lambda \mathcal{L} = \frac{1}{\lambda} - \bar{x} \overset{!}{=} 0 \qquad \Longrightarrow \qquad \hat{\lambda} = \frac{1}{\bar{x}}$$

In case of a large number of experiments:

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

The standard deviation σ is proportional to $\frac{1}{n}$. Both the distribution (Gauss) and the theoretical values (expectation value and standard deviation of λ) are supported by the results above.

List of Figures