Report of Data Analysis

Chen Yang

Student ID: 03700140

Chapter 1

Problem 1.1

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 ?

Solution

For the first question, the genders of two babies in two pictures are independent events. Thus the probability for the second baby to be a girl is trivial, which is

$$P(\text{The second child is a gril}) = \frac{1}{2}$$

For the varied case, we need to use Bayesian probability. We call the picture Jane shows is picture A and the other picture is picture B. The event that she shows picture A is called event X. If the child on picture A is a boy, we call this event A_0, otherwise we call it A_1. The notation is the same for picture B.

Then we can write down some probabilities we need in Bayesian formula

$$P(X|A_0B_0) = 0, P(A_0B_0) = \frac{1}{4}$$
$$P(X|A_1B_0) = 1, P(A_1B_0) = \frac{1}{4}$$
$$P(X|A_0B_1) = 0, P(A_0B_1) = \frac{1}{4}$$
$$P(X|A_1B_1) = \frac{1}{2}, P(A_1B_1) = \frac{1}{4}$$

the probability we want is

$$P(A_1B_1|X) = \frac{P(X|A_1B_1)P(A_1B_1)}{\sum_{i,j=0,1} P(X|A_iB_j)P(A_iB_j)} = \frac{1}{3}$$

Problem 1.2

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

Solution

We can consider the difference between the times that T and H appears. For example, if we choose n_h as the number of times that H appears and n_t as the number of times that T appears in a 10-flip experiment and the random variable we care about is $n = |n_h - n_t|$. Then the possible value of n is from 0 to 5 and the probabilities are respectively

$$P(n=0) = \left(\begin{array}{c} 10 \\ 5 \end{array} \right) (1/2)^{10} = 252 \cdot (1/2)^{10}$$

If $n > 0$,

$$P(n) = \left(\left(\begin{array}{c} 10 \\ n \end{array} \right) + \left(\begin{array}{c} 10 \\ 10-n \end{array} \right) \right) (1/2)^{10} = 2 \left(\begin{array}{c} 10 \\ n \end{array} \right) (1/2)^{10}$$

Problem 1.3

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 ?

Solution

Firstly, we call the event that the real energy of the particle we measure is E_0 as E_0 and the event that we measure energy E as E.

Then the probability we care about is $P(E_0|E)$. Use the Bayesian formula so we have

$$P(E_0|E) = \frac{P(E|E_0)P(E_0)}{\int_{-\infty}^{\infty} \mathrm{d}E_0 P(E|E_0)P(E_0)}$$

Suppose that $P(E_0)$ is a constant number (flat prior) so that it will cancel. Then

$$P(E_0|E) = \frac{P(E|E_0)}{\int_{-\infty}^{\infty} \mathrm{d}E_0 P(E|E_0)} \tag{1.1}$$

Since the resolution of the detector is 10%, $P(E|E_0)$, as a function of E, should be like

$$P(E|E_0) \begin{cases} \neq 0 & \text{if } 0.9E_0 < E < 1.1E_0, \\ = 0 & \text{otherwise.} \end{cases}$$

This means $P(E_0|E)$ should have the same property because in (1.1) the numerator is $P(E|E_0)$ and the denominator is a nonzero number, so $P(E_0|E)$, as a function of E_0 should be like

$$P(E_0|E) \begin{cases} \neq 0 & \text{if } \frac{10E}{11} < E_0 < \frac{10E}{9}, \\ = 0 & \text{otherwise.} \end{cases}$$

Thus the biggest true energy value range is $\left(\frac{10E}{11}, \frac{10E}{9}\right)$. For a certain energy range (E_1, E_2), the probability that the real energy value is in this range is

$$P = \int_{E_1}^{E_2} \mathrm{d}E_0\, P(E_0|E) = \frac{\int_{E_1}^{E_2} \mathrm{d}E_0 P(E|E_0)}{\int_{0.9E_0}^{1.1E_0} \mathrm{d}E_0 P(E|E_0)}$$

As for the representation for $P(E|E_0)$, it depends on the definition of the resolution. For example, one definition is using the FWHM as the resolution and the measured energy value satisfied Gaussian distribution. Then we can write the concrete formula for $P(E|E_0)$.

Problem 1.4

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

Solution

The key point of this problem is Bayesian probability.

Firstly, we give some names to the events we need.

$$\begin{gathered} \text{Mongolian swamp fever } \rightarrow \mathrm{M} \\ \text{Spots} \rightarrow \mathrm{S} \\ \text{Lethargic} \rightarrow \mathrm{L} \\ \text{Thirsty} \rightarrow \mathrm{T} \\ \text{Sneeze} \rightarrow \mathrm{N} \end{gathered}$$

For the opposite events use the bar notation. For example, the event that a person does not have Mongolian swamp fever is $\overline{\mathrm{M}}$. Then

$$\begin{aligned} P(S|M) &= 0.03 \qquad & P(S|\overline{M}) &= 0.03 \\ P(L|M) &= 0.1 \qquad & P(L|\overline{M}) &= 0.1 \\ P(T|M) &= 0.6 \qquad & P(T|\overline{M}) &= 0.02 \\ P(N|M) &= 0.2 \qquad & P(N|\overline{M}) &= 0.05 \end{aligned}$$

$$\begin{aligned} P(M) &= 0.0001 \\ P(S) &= P(S|M)P(M) + P(S|\overline{M})P(\overline{M}) \approx P(S|\overline{M}) = 0.03 \\ P(L) &= P(L|M)P(M) + P(L|\overline{M})P(\overline{M}) \approx P(L|\overline{M}) = 0.1 \\ P(T) &= P(T|M)P(M) + P(T|\overline{M})P(\overline{M}) \approx P(T|\overline{M}) = 0.02 \\ P(N) &= P(N|M)P(M) + P(N|\overline{M})P(\overline{M}) \approx P(N|\overline{M}) = 0.05 \end{aligned}$$

The probability of having Mongolian swamp fever if one has four of theses symptoms is

$$P(M|SLTN) = \frac{P(M)P(SLTN|M)}{P(SLTN)} = P(M)\frac{P(S|M)}{P(S)}\frac{P(L|M)}{P(L)}\frac{P(T|M)}{P(T)}\frac{P(N|M)}{P(N)}$$
$$= P(M)\frac{P(T|M)}{P(T|\overline{M})}\frac{P(N|M)}{P(N|\overline{M})} = 0.012$$

For the case that there are three out of four of these symptoms, actually what matters is whether or not the three includes thirsty or sneeze. If the three includes both thirsty and sneeze, the probability is

$$P = 0.0012$$

If the three only includes thirsty, the probability is

$$P = P(M)\frac{P(T|M)}{P(T|\overline{M})} = 0.003$$

If the three only includes sneeze, the probability is

$$P = P(M)\frac{P(N|M)}{P(N|\overline{M})} = 0.0004$$

Chapter 2

Problem 2.1

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

Solution

(a) Follow the definition of mean and standard deviation so we have

$$E[x] = \int_0^\infty xP(x)\,\mathrm{d}x = \int_0^\infty x^2 e^{-x}\,\mathrm{d}x = \int_0^\infty 2e^{-x}\,\mathrm{d}x = 2$$

$$E[x^2] = \int_0^\infty x^2 P(x)\,\mathrm{d}x = \int_0^\infty x^3 e^{-x}\,\mathrm{d}x = \int_0^\infty 6e^{-x}\,\mathrm{d}x = 6$$

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

The probability in the interval $[2-\sqrt{2}, 2+\sqrt{2}]$ is

$$P = \int_{2-\sqrt{2}}^{2+\sqrt{2}} P(x) = -(1+x)e^{-x}\Big|_{2-\sqrt{2}}^{2+\sqrt{2}} = 0.7375$$

(b) Calculate the cumulative function.

$$F(x) = \int_0^x te^{-t}\,\mathrm{d}t = 1-(1+x)e^{-x}$$

According to the definition of median, we have

$$F(x_{med}) = 1-(1+x_{med})e^{-x_{med}} = 0.5$$

Use some software, e.g. Mathematica, we can get

$$x_{med} = 1.68$$

Using the definition of central interval, we can calculate the boundary of the central interval as

$$\begin{aligned} F(x_1) &= (1-0.68)/2 = 0.16 \implies x_1 = 0.71 \\ F(x_2) &= (1+0.68)/2 = 0.84 \implies x_2 = 3.29 \end{aligned}$$

The central interval is $(0.71, 3.29)$.

(c) According to the definition of mode, we have

$$\left.\frac{\mathrm{d}P(x)}{\mathrm{d}x}\right|_{x=x_{mod}} = (1-x_{mod})e^{-x_{mod}} = 0 \implies x_{mod} = 1$$

As for smallest interval, we can write down the definitive equations as

$$\begin{gathered} x_1e^{-x_1} = x_2e^{-x_2} \\ (1+x_1)e^{-x_1} - (1+x_2)e^{-x_2} = e^{-x_1} - e^{-x_2} = 0.68 \end{gathered}$$

Use Mathematica so we can get the result is

$$x_1 = 0.27 \qquad x_2 = 2.50$$

so the smallest interval is (0.27,2.50).

Problem 2.2

Energy	Trials	Successes
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

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

Solution As for flat prior, we can find the formula of posterior probability in the lecture note

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

However, it is nearly impossible to calculate this formula analytically because the factorial term is too big. A good idea is to calculate it numerically.

The scheme is that firstly we separate $(0,1)$ into 100 bins equally and choose the middle posterior as the posterior of each bin. Then we can find the mode and calculate the integration much more easily. Since the factorial terms are the main obstacles, we only calculate the latter term i.e. $p^r(1-p)^{N-r}$ for each bin then use a factor to normalize the sum(integration) to 1. The algorithm for finding the mode and smallest interval is simple.

Under this scheme, we can use program to solve this problem easily and the results are

Energy	p_{mod}	Smallest interval
0.5	0.00	(0.00,0.02)
1.0	0.04	(0.02,0.06)
1.5	0.20	(0.16,0.24)
2.0	0.58	(0.53,0.63)
2.5	0.92	(0.89,0.95)
3.0	0.99	(0.98,0.99)
3.5	0.99	(0.99,1.00)
4.0	0.99	(0.99,1.00)

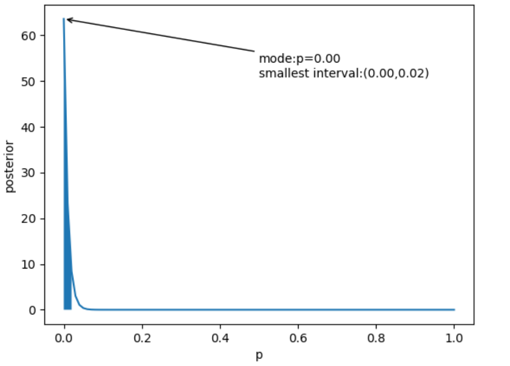

Figure 2.1: E=0.5

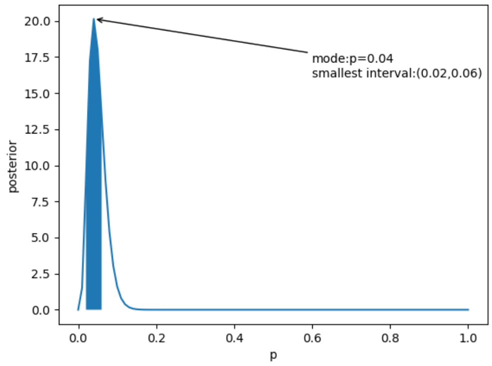

Figure 2.2: E=1.0

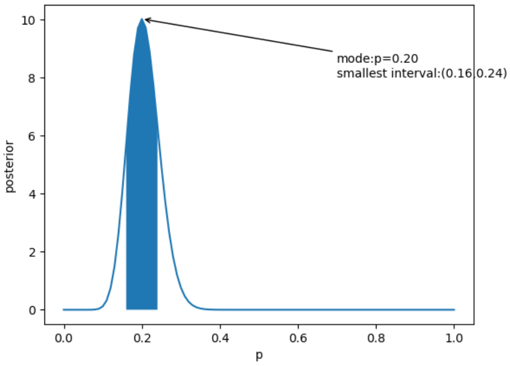

Figure 2.3: E=1.5

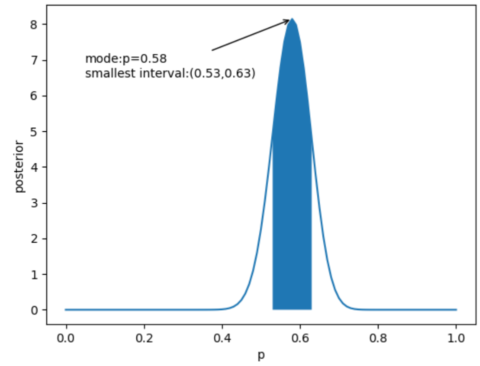

Figure 2.4: E=2.0

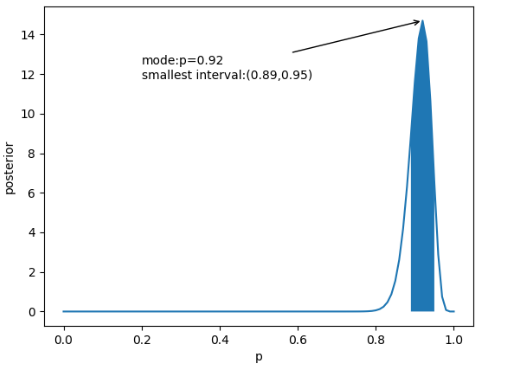

Figure 2.5: E=2.5

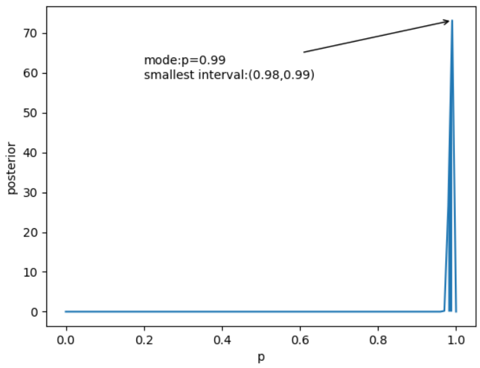

Figure 2.6: E=3.0

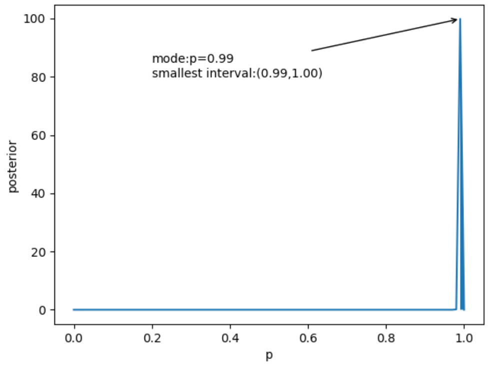

Figure 2.7: E=3.5

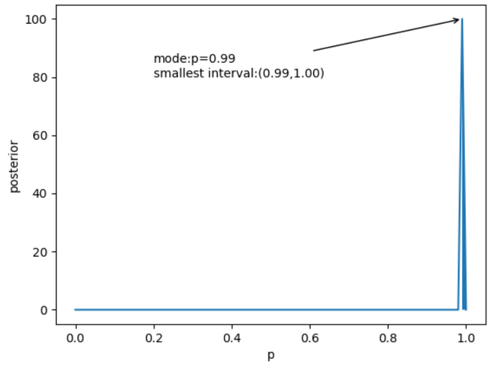

Figure 2.8: E=4.0

Figure note: The horizontal axis is the success probability and the vertical axis is the posterior probability. The blue shadows are the range of the smallest interval of each energy.

Problem 2.3

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

Solution

In the frequentist perspective, we need to construct the Neyman confidence level interval for $N = 100$ and $N = 1000$ binomial case. However, the problem only requires the confidence levels for certain energy, which means certain N and r, so we can only calculate these intervals for p instead of drawing the whole Neyman confidence level diagram.

According to the definition of 90% central interval, if p_0 is included in the central interval, it should satisfies that

$$\sum_{r=0}^{r_D} p(r|N, p_0) > 0.05 \text{ and } \sum_{r=r_D}^{N} p(r|N, p_0) > 0.05$$

at the same time. Under this condition, we can find the upper and lower bound of p easily by numerical method.

Fortunately, Python can calculate the combination number of 1000 quickly so that we do not need some special manipulation to deal with it. Then we can get the results as

Energy	r/N	90% Central Interval
0.5	0.000	(0.000,0.030)
1.0	0.040	(0.014,0.090)
1.5	0.200	(0.137,0.278)
2.0	0.580	(0.493,0.664)
2.5	0.920	(0.861,0.960)
3.0	0.987	(0.980,0.993)
3.5	0.995	(0.990,0.999)
4.0	0.998	(0.994,1.000)

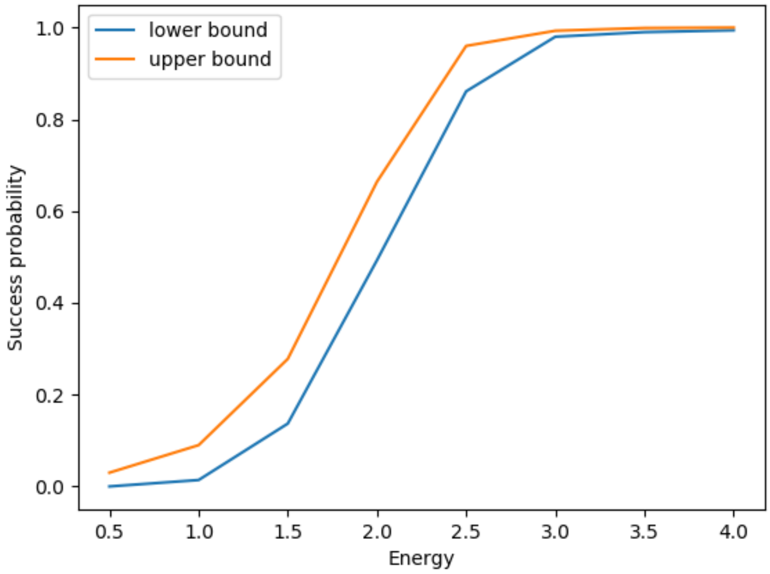

Figure 2.9: 90% CL interval for p in different energy

Problem 2.4

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

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

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

Solution It has been proved in the lecture note that if we have two data set N_1, r_1, N_2, r_2 and we choose the posterior of N_1, r_1 with flat prior as the new nonflat prior and use N_2, r_2 to calculate posterior, the result is

$$
\begin{aligned}
P(p|N_1, N_2, r_1, r_2) &= \frac{P(r_2|N_2, p)P(p|N_1, r_1)}{\int P(r_2|N_2, p)P(p|N_1, r_1)\,\mathrm{d}p} \\
&= \frac{(N_1 + N_2 + 1)!}{(r_1 + r_2)!(N_1 + N_2 - r_1 - r_2)!} p^{r_1+r_2}(1-p)^{N_1+N_2-r_1-r_2}
\end{aligned}
$$

Use this formula for n times so we can get

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

According to the lecture, the expectation value and variance for p are

$$
\begin{aligned}
E[p] &= \frac{r + 1}{N + 2} \\
V[p] &= \frac{E[p](1 - E[p])}{N + 3}
\end{aligned}
$$

Since the new posterior is in the same form with flat-prior case, we can use these two formulas directly. Their limit for $n \to \infty$ are

$$
\begin{aligned}
\lim_{n\to\infty} E[p] &= \lim_{n\to\infty} \frac{nr + 1}{nN + 2} = \frac{r}{N} \\
\lim_{n\to\infty} V[p] &= \lim_{n\to\infty} \frac{E[p](1 - E[p])}{N + 3} = \frac{r(N - r)}{N^2(N + 3)}
\end{aligned}
$$

Chapter 3

Problem 3.1

4. 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 peak?

Solution

(a) Since the function $f(x)$ is an even function, $xf(x)$ is an odd function of x. Thus, the mean value of x is 0 because the integral of $xf(x)$ is 0.

For the standard variance, calculate $E[x^2]$ first

$$
E[x^2] = \int_{-\infty}^{\infty} \frac{1}{2} x^2 e^{-|x|}\,\mathrm{d}x = \int_0^{\infty} x^2 e^{-x}\,\mathrm{d}x = \int_0^{\infty} 2e^{-x}\,\mathrm{d}x = 2
$$

Then, the standard variance is

$$
\sigma = \sqrt{V[x]} = \sqrt{E[x^2] - E[x]^2} = \sqrt{2} \approx 1.414
$$

(b) Calculate the FWHM first

$$\frac{1}{2}e^{-|x_{HW}|} = \frac{1}{4} \implies |x_{HW}| = \ln 2 \approx 0.6931$$

Obviously, the FWHM is much smaller than the standard deviation. To look more at this question, we can write down the formulas for these two variables

$$\sigma = \sqrt{\int_0^{\infty} x^2 f(x)\,\mathrm{d}x}$$

$$x_{HW} = f^{-1}\left(\frac{f(0)}{2}\right)$$

Actually there is no any certain relationship between these two variables but what about their relationship of size? For example, if the function has a certain FWHM x_0, we can easily find an even normalized function $f(x)$, which is

$$f(x) = \begin{cases} \frac{1}{2x_0} & \text{if} - x_0 < x < x_0, \\ \frac{1}{4x_0} & \text{if } x = \pm x_0, \\ 0 & \text{otherwise.} \end{cases}$$

so that $\sigma < x_{HW}$.

There is no certain relationship of size for σ and x_{HW} neither. However some conclusions can be given. If a function is convex, as the function in this problem, it is more possible for x_{HW} to be smaller than σ. On the contrary, for a concave function, it is more possible for σ to be smaller.

(c) The probability is

$$P = \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{2}e^{-|x|}\,\mathrm{d}x = \int_0^{\sqrt{2}} e^{-x}\,\mathrm{d}x \approx 0.7596$$

Problem 3.2

> 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 si the 68% confidence level interval for ν using the Neyman construction and the smallest interval definition?

Solution

(a) For flat prior, the posterior function is

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

To calculate the 95% probability lower limit, we need to find the value of ν for which the cumulative probability reaches 0.05. I.e.

$$F(\nu|n) = 0.05$$

$$F(\nu|n=9) = \int_0^{\nu} \frac{e^{-t}t^n}{n!}\,\mathrm{d}t = \frac{1}{9!}\int_0^{\nu} e^{-t}t^9 = 0.05$$

Use numerical method to find the root for ν so we have the lower limit is

$$\nu_{LM} = 5.42$$

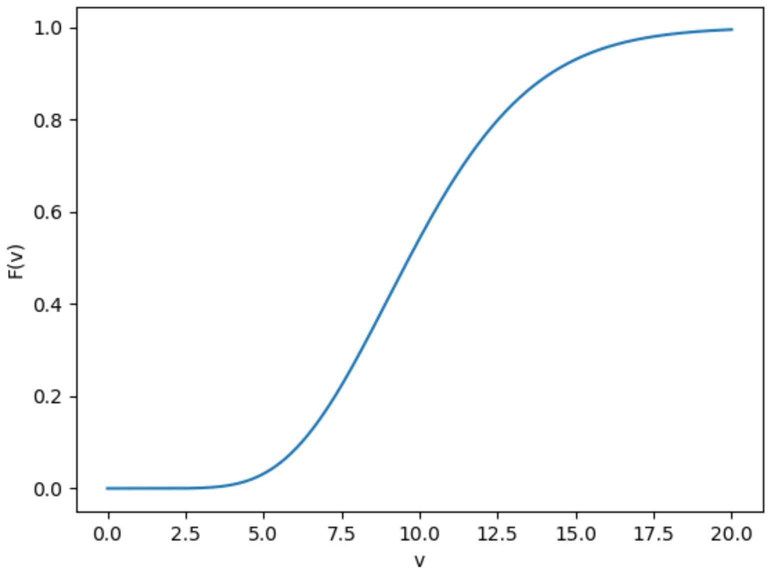

Figure 3.1: Cumulative Probability for ν

(b) Just follow the definition of Neyman confidence level interval and use numerical method to calculate them. Then we have

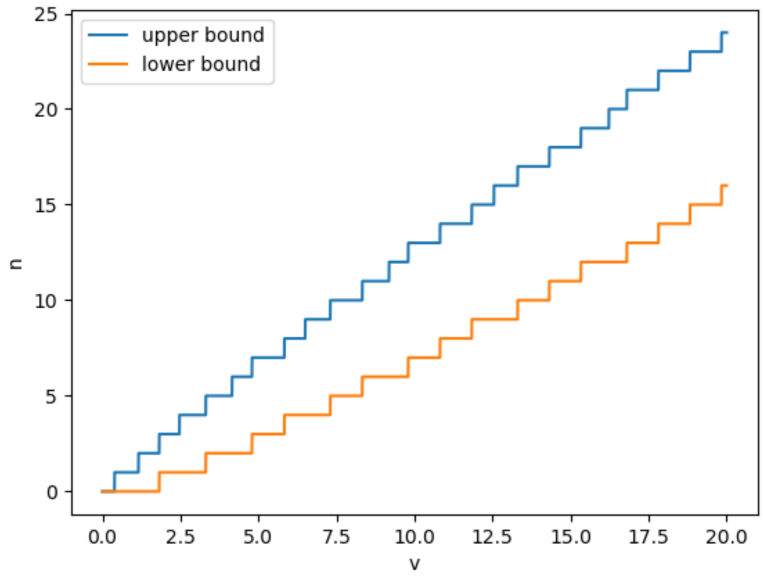

Figure 3.2: Neyman confidence level interval

and the 68% confidence level for ν is

$$(6.495, 13.301)$$

Problem 3.3

> 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.

Solution

(a) According to the definition of Feldman-Cousins confidence level, we need to find the value of $\hat{\mu}$, which maximizes $P(n|\hat{\mu})$ for certain n value. Since the background is 3.2, we have $\hat{\mu} \geq 3.2$. Hence, for $n \leq 3$, we choose $\hat{\mu} = 3.2$ and for $n > 3$, we choose $\hat{\mu} = n$.

Similar to calculate the smallest interval, we can draw the 68% Feldman-Cousins confidence level interval as

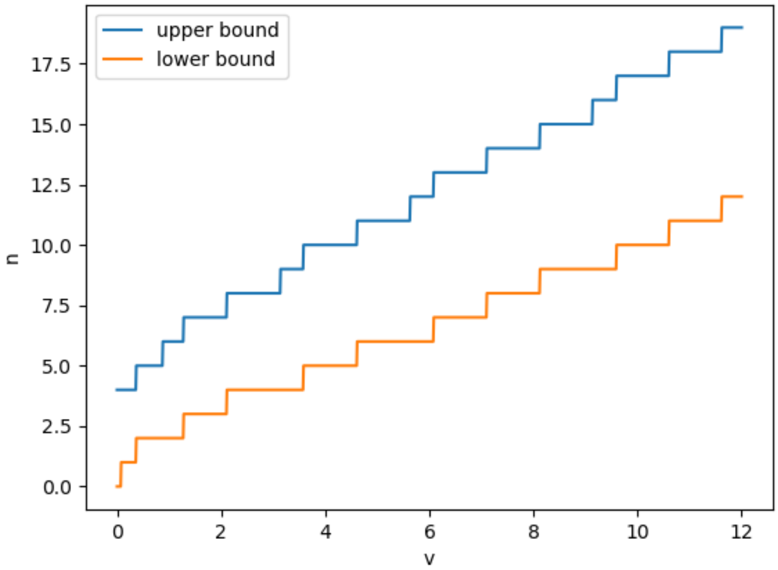

Figure 3.3: 68% Feldman-Cousins CL interval for $\lambda = 3.2$

(b) To compare with the Feldman-Cousins interval better, we choose the smallest interval when drawing Neyman CL interval. Actually, we do not need to calculate it again because we only need to revise a little to what we get in problem 7. We only need to shift the Neyman confidence level interval left for 3.2 in the x direction so we can get

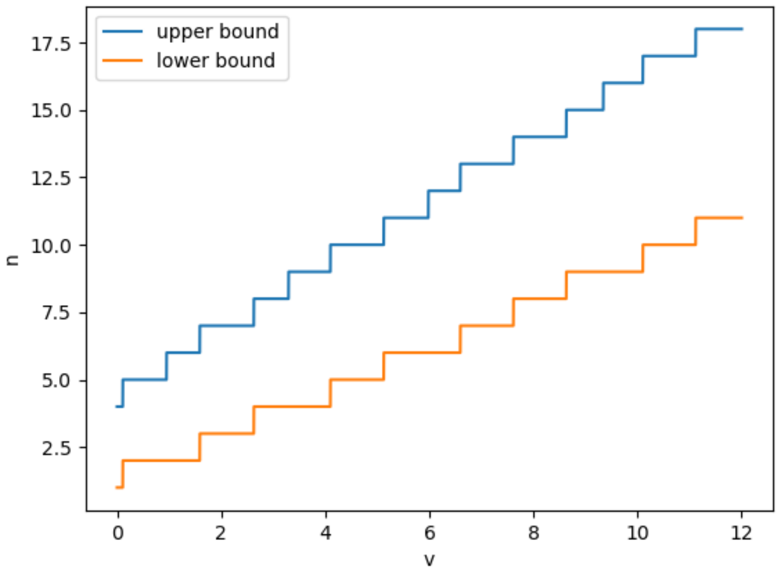

Figure 3.4: 68% Neyman CL interval for $\lambda = 3.2$

(c) As for credible interval for ν, we can get two answers from Feldman-Cousins interval and Neyman interval.

$$\begin{aligned} &(3.143, 9.588) && \text{for Feldman-Cousins interval} \\ &(3.295, 10.101) && \text{for Neyman interval} \end{aligned}$$

To see the difference between these two kinds of intervals better, we can draw these two intervals in one figure as

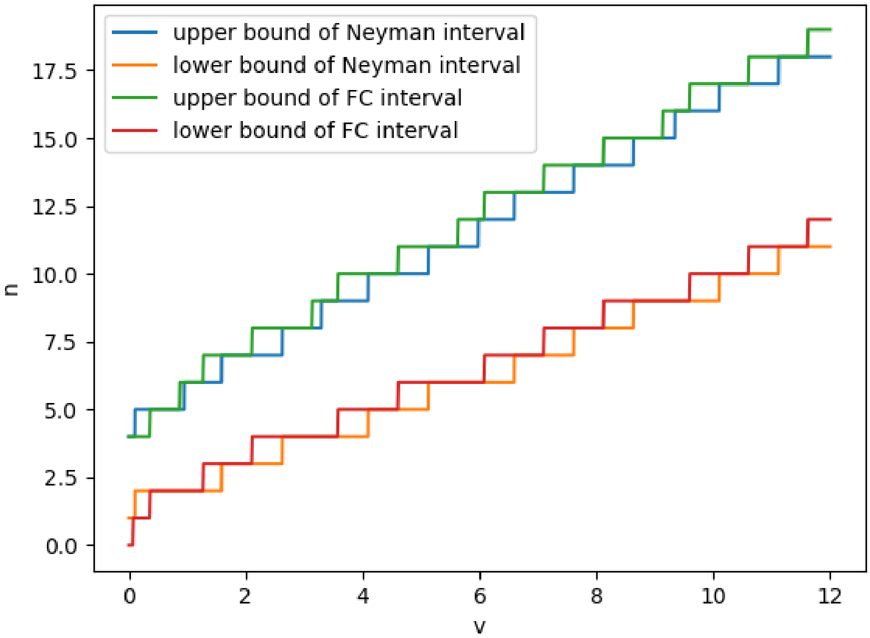

Figure 3.5: Neyman interval and FC interval for $\lambda = 3.2$

We can see for most of n, actually large value of n, the boundaries of Neyman interval are always higher than the boundaries of FC interval. However, when n is small and the confidence value of ν is smaller than 1, we can find that the boundaries of FC interval are higher.

The reason is that we consider the influence of background more in FC CL interval. When we calculate the r value for FC interval, for small n value, we actually can not choose the $\hat{\mu}$ which exactly maximizes $P(n|\hat{\mu})$. This means we magnify the weight of probability for small n value. This explains why for large n, the boundaries of FC interval are smaller.

Meanwhile, for very small ν and $n = 0, 1, 2, 3$, $P(n|\mu)$ is very close to $P(n|\hat{\mu})$ because now

$$\mu = \lambda + \nu = \hat{\mu} + \nu = 3.2 + \nu \approx 3.2 \qquad \text{When } \nu \text{ is small and } n = 0, 1, 2, 3$$

This means the r value for $n = 0, 1, 2, 3$ will be really close to 1 so the FC interval for small ν will prefer to include smaller n value. This cause the anomalous result for small ν value.

Problem 3.4

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 $x_i,\ i = 1, , n$. 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

$$\nu_j = \int_{\Delta_j} f(x|\lambda)\,\mathrm{d}x$$

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

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

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

When $K \to \infty$, ν_j and Δ will be very small and

$$\nu_j = \int_{\Delta_j} f(x|\lambda)\,\mathrm{d}x \approx f(x_j|\lambda)\Delta$$

where x_j is the x value for the infinitesimal bin Δ_j.

Since each bin will have either $n_j = 0$ or $n_j = 1$,

$$\text{if } n_j = 0,\ \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} = e^{-\nu_j} \approx 1$$

$$\text{if } n_j = 1,\ \frac{e^{-\nu_j}\nu_j^{n_j}}{n_j!} = e^{-\nu_j}\nu_j = f(x_j|\lambda)\Delta + O(\Delta)$$

so we have

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

Problem 3.5

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

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

Solution

In order to get the value of X, we firstly need to guarantee $N >= X$ and then choose X occurrences in the N to be 1 and the other to be 0. Translate this into mathematical formula so we have

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

Suppose that $i = N - X$ and we can rewrite the formula as

$$\begin{aligned}
P(X) &= \sum_{i=0}^{\infty} \frac{e^{-\nu}\nu^{i+X}}{i!X!} p^X (1-p)^i \\
&= \frac{e^{-\nu}(\nu p)^X}{X!} \sum_{i=0}^{\infty} \frac{[\nu(1-p)]^i}{i!} \\
&= \frac{e^{-\nu}(\nu p)^X}{X!} e^{\nu(1-p)} \\
&= \frac{e^{-\nu p}(\nu p)^X}{X!}
\end{aligned}$$

Chapter 4

Problem 4.1

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

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

   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

(a) Firstly, CLT is available for exponential distribution because no matter which order of moment, they are all finite for exponential distribution. Then, derive the expectation value of exponential distribution so that this is the expectation value for the mean of n samples

$$E[x] = \int_0^\infty \lambda e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$$

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

Then, the expectation value and standard variance for the mean value is

$$\mu = \frac{1}{\lambda} \qquad \sigma = \frac{1}{\lambda\sqrt{n}}$$

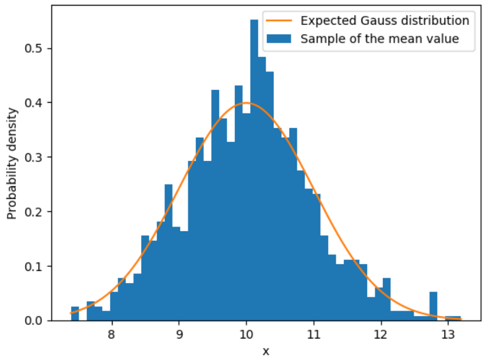

Figure 4.1: $\lambda = 0.1$, $n = 100$

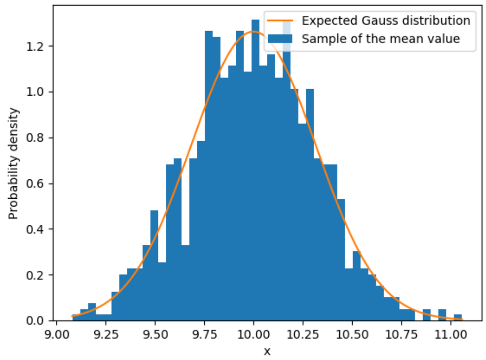

Figure 4.2: $\lambda = 0.1$, $n = 1000$

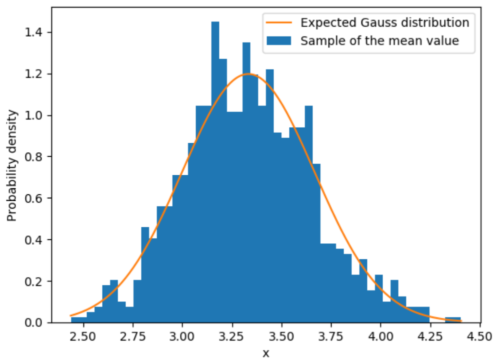

Figure 4.3: $\lambda = 0.3$, $n = 100$

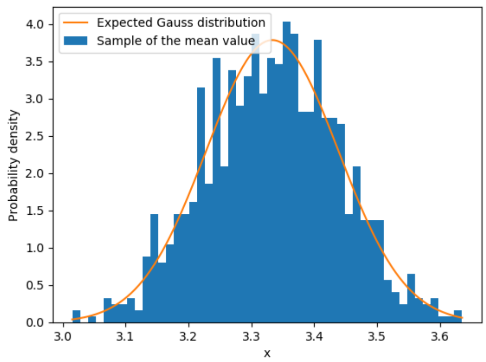

Figure 4.4: $\lambda = 0.3$, $n = 1000$

(b) CLT is not hold for Cauchy distribution because not all the moment of Cauthy distribution are finite. During our derivation of CLT, we have expanded the formula into the sum of all moments. Hence, if the high order moment is not finite, we can not drop the high order terms. In the pdf of Cauchy distribution, there is only a order-2 polynomial for x, which means that the moment whose order is bigger or equal than 2 will go infinity. Actually, the variance of Cauchy already goes infinity so we can not find a proper Gaussian to compare with the mean value sample of Cauchy distribution.

We can plot the $x_0 = 25$ and $\gamma = 3$ case as

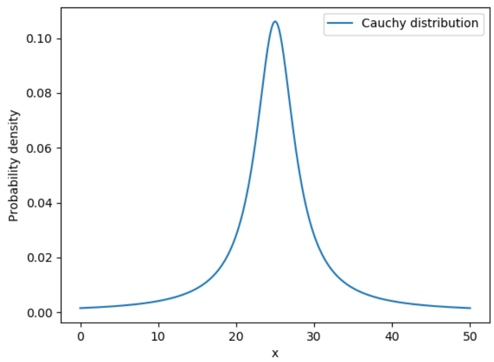

Figure 4.5: pdf of Cauchy distribution

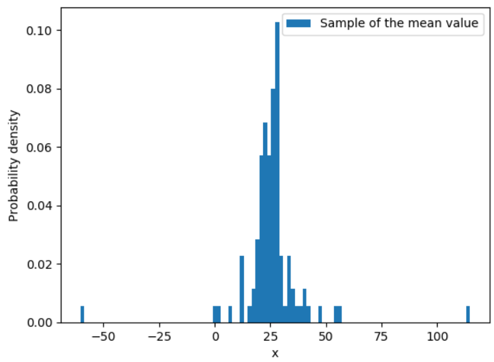

Figure 4.6: Sample of the mean value

We can find that there is a really big long tail in the Cauchy distribution and this causes that there are some really large mean value samples in Figure 4.6.

Problem 4.2

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

(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

(a) The bivariate Gauss function for $\mu_x = 0, \mu_y = 0, \sigma_x = 1, \sigma_y = 1, \rho_{xy} = 0$

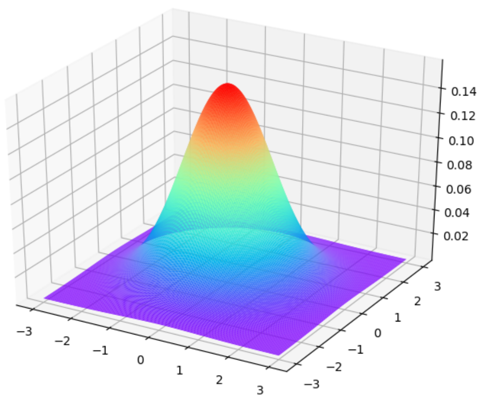

Figure 4.7: Bivariate Gauss function(a)

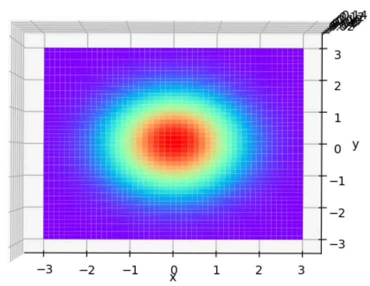

Figure 4.8: Bivariate Gauss function(a)

When $\rho = 0$, i.e. x and y are independent, in figure 4.8, it is a regularly symmetric circle.

(b) The bivariate Gauss function for $\mu_x = 1, \mu_y = 2, \sigma_x = 1, \sigma_y = 1, \rho_{xy} = 0.7$

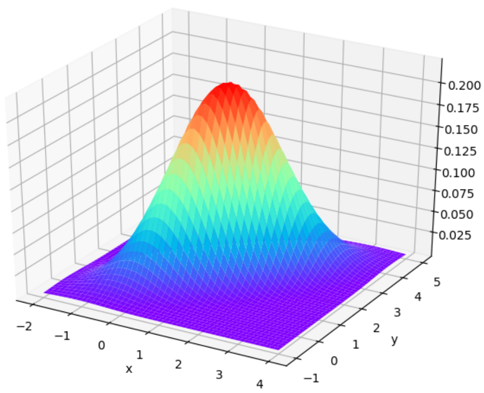

Figure 4.9: Bivariate Gauss function(b)

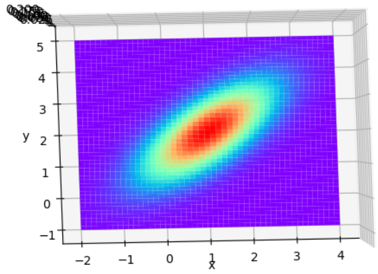

Figure 4.10: Bivariate Gauss function(b)

When $\rho = 0.7$, we can find in figure 4.10 there are ellipses looking down and the main axis is diagonal.

(c) The bivariate Gauss function for $\mu_x = 1, \mu_y = -2, \sigma_x = 1, \sigma_y = 2, \rho_{xy} = -0.7$

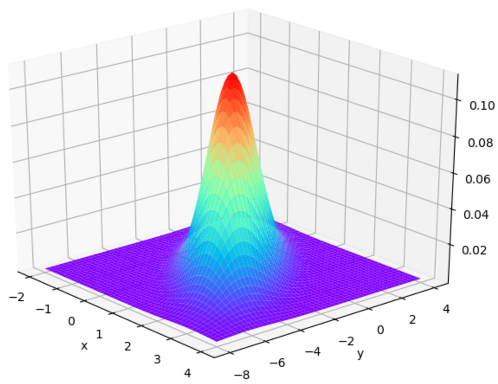

Figure 4.11: Bivariate Gauss function(c)

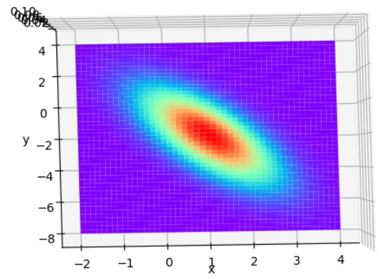

Figure 4.12: Bivariate Gauss function(c)

When $\rho = -0.7$, there are ellipses too but the main axis is anti-diagonal. These three cases just show clearly how the bivariate Gauss function is dependent on ρ.

Problem 4.3

12. Bivariate Gauss probability distribution

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

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

(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) In order to show the formula given is the pdf of bivariate Gauss probability distribution, two steps need to be done:

(1) Calculate the marginal probability distribution and show they satisfy the single Gauss distribution

(2) Calculate the covariance and show it is equal to $\rho\sigma_x\sigma_y$

Firstly, calculate the marginal probability distribution of x. Since it is symmetric for x and y, the calculation of marginal probability distribution is not needed.

$$P_x(x) = \int_{-\infty}^{\infty} P(x,y)\,\mathrm{d}y$$

Because

$$\frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y} = \left(\frac{y}{\sigma_y} - \rho\frac{x}{\sigma_x}\right)^2 - \rho^2\frac{x^2}{\sigma_x^2}$$

we have

$$\begin{aligned} P_x(x) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{x^2}{2\sigma_x^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y}{\sigma_y}-\rho\frac{x}{\sigma_x}\right)^2}\,\mathrm{d}y \\ &= \frac{1}{2\pi\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \int_{-\infty}^{\infty} e^{-t^2/2}\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \end{aligned}$$

This is exactly the form of Gauss distribution.

Then compute the covariance as

$$\mathrm{Cov}(x,y) = \iint_{-\infty}^{\infty} \frac{xy}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2\rho xy}{\sigma_x\sigma_y}\right)\right)\mathrm{d}x\,\mathrm{d}y$$

Suppose

$$t = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{y}{\sigma_y} - \rho\frac{x}{\sigma_x}\right), \qquad u = \frac{x}{\sigma_x}$$

Then

$$\mathrm{Cov}(x,y) = \frac{1}{2\pi}\iint (\sigma_x\sigma_y\sqrt{1-\rho^2}tu + \rho\sigma_x\sigma_y u^2)e^{-(u^2+t^2)/2}\,\mathrm{d}t\,\mathrm{d}u$$

$$= \frac{\rho\sigma_x\sigma_y}{2\pi}\int_{-\infty}^{\infty} u^2 e^{-u^2/2}\,\mathrm{d}u\int_{-\infty}^{\infty} e^{-t^2/2}\,\mathrm{d}t + \frac{\sigma_x\sigma_y\sqrt{1-\rho^2}}{2\pi}\int_{-\infty}^{\infty} ue^{-u^2/2}\,\mathrm{d}u\int_{-\infty}^{\infty} te^{-t^2/2}\,\mathrm{d}t$$

$$= \frac{\rho\sigma_x\sigma_y}{2\pi}\sqrt{2\pi}\sqrt{2\pi} = \rho\sigma_x\sigma_y$$

(b)

$$P_z(z) = \int_{-\infty}^{\infty} P(z+y,y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty}\mathrm{d}y\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\times$$

$$\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(z+y-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(z+(y-\mu_y)-\mu_x)y}{\sigma_x\sigma_y}\right)\right)$$

Having failed to compute this by hand for several times and got help from Mathematica, I get the answer is just what we expect in the text

$$P_z(z) = \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y)}}e^{-\frac{(z-\mu_x+\mu_y)^2}{2(\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y)}}$$

Problem 4.4

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 ?

Solution We can write down the formula of distribution function of y directly as

$$P(y) = \int_{-\infty}^{\infty} P(y|x)f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_y}e^{-\frac{(y-x)^2}{2\sigma_y^2}}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-x_0)^2}{2\sigma_x^2}}\,\mathrm{d}x$$

This formula can be regarded as the convolution of two function $f(x)$ and $g(x)$ where

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

i.e.

$$P(y) = f(x) * g(x)$$

Do Fourier transformation to both sides and we have

$$\begin{aligned}
\mathrm{F}(P(y)) =& \mathrm{F}(f(x))\mathrm{F}(g(x)) \\
=& e^{ikx_0 - \frac{k^2\sigma_x^2}{2}} e^{-\frac{k^2\sigma_y^2}{2}} \\
=& e^{ikx_0 - \frac{k^2(\sigma_x^2+\sigma_y^2)}{2}}
\end{aligned}$$

then,

$$P(y) = \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}} e^{-\frac{(y-x_0)^2}{2\sqrt{\sigma_x^2+\sigma_y^2}}}$$

Problem 4.5

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

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

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

(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

notice

The model in the question is a little strange because of the term "$\cos(\theta^2)$". From the point of view of expansion formula, I think it perhaps is $\cos^2(\theta)$ originally and there is some typo somehow. Consequently, during my solution I will change the form of the model into

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

(a) To find some posterior for parameters A, B, C, the most direct way is to build an joint probability distribution $P(A, B, C|\mathrm{data})$. However this is not practical here. The reason is when we use the Bayesian formula we have

$$\begin{aligned}
P(A, B, C|\mu, \sigma) &= \frac{P(\mu|A, B, C, \sigma)P_0(A, B, C)}{\iiint P(\mu|A, B, C, \sigma)P_0(A, B, C)\,\mathrm{d}A\,\mathrm{d}B\,\mathrm{d}C} \\
&= \frac{P(\mu|A, B, C, \sigma)}{\iiint P(\mu|A, B, C, \sigma)\,\mathrm{d}A\,\mathrm{d}B\,\mathrm{d}C}
\end{aligned}$$

the denominator will diverge. For the Gaussian function, it can only be integrated for once. Thus, we can only calculate the posterior for one variable and regard the other two parameters as constant number.

For example, to calculate the posterior of A, the Bayesian formula is

$$P(A|B,C,\mu,\sigma) = \frac{P(\mu|A,B,C,\sigma)}{\int P(\mu|A,B,C,\sigma)\,\mathrm{d}A}$$

Under this formula we can use the iteration formula for the expectation value and standard deviation.

$$\mu_A = \frac{\sum_i A_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$$

$$\frac{1}{\sigma_A^2} = \sum_i \frac{1}{\sigma_i^2}$$

Furthermore, the coefficients for B and C are not 1 so when using the iteration formula we need to divide the every expectation value and standard deviation by the coefficient. For instance

$$\sigma_{Bi} = \frac{\sigma_i}{\cos(\theta)}$$

Under this scheme, we can write down the coefficients and formula we need as

θ	30°	45°	90°	120°	150°
$\cos(\theta)$	0.866	0.707	0	-0.5	-0.866
$\cos^2(\theta)$	0.75	0.5	0	0.25	0.75
$\cos^4(\theta)$	0.5625	0.25	0	0.0625	0.5625
Cross Section	11	13	17	17	14
Error	1.5	1.0	2.0	2.0	1.5

$$\sigma_{exp} - A - \cos(\theta)B - \cos^2(\theta)C = 0$$

Then we can get the expectation value and standard deviation for A, B, C as

$$\mu_A = 13.651 - 0.244B - 0.515C \qquad \sigma_A = 0.647$$

$$\mu_B = 4.810 - 0.473A - 0.262C \qquad \sigma_B = 0.902$$

$$\mu_C = 20.762 - 1.605A - 0.421B \qquad \sigma_C = 1.143$$

We can write down the posterior easily with these expectation value and standard deviation.

(b) To reach the mode, we need

$$\mu_A = A \qquad \mu_B = B \qquad \mu_C = C$$

It is possible for this case and we can exactly use these three condition to get three linear equation of A, B, C so that we can solve them to get the values

$$\begin{aligned}
A + 0.244B + 0.515C &= 13.651 \\
0.473A + B + 0.262C &= 4.810 \\
1.605A + 0.421B + C &= 20.762
\end{aligned}$$

We can get

$$A = 17.331 \qquad B = -1.730 \qquad C = -6.326$$

Use these three value to calculate σ and compare them with the experiment value

θ	30°	45°	90°	120°	150°
Cross Section (Experiment)	11	13	17	17	14
Cross Section (Model)	11.08832	12.94489	17.331	16.6145	14.08468

We can see that for 90° and 120°, the deviation is a little larger than that of other degrees. In total, the fitting is not bad.

Chapter 5

Probelm 5.1

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

(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) Firstly, we need to do a rough estimate to E_0 and A in the Sigmoid function. Follow the method in the lecture we can choose $E_0 = 2.0$ and $\sigma_{E_0} = 0.5$. As for A, we have

$$\frac{A}{4} \approx 0.6 \Rightarrow A \approx 2.5$$

and we choose the standard variance of A is 0.5.

Then we can calculate the posterior function of E_0 and A using the formula

$$P(E_0, A|\{N\}, \{r\}) = \frac{P(\{r\}|\{N\}, E_0, A)P_0(E_0)P_0(A)}{\iint_{-\infty}^{\infty} P(\{r\}|\{N\}, E_0, A)P_0(E_0)P_0(A)\, \mathrm{d}A\, \mathrm{d}E_0}$$

where

$$P(\{r\}|\{N\}, E_0, A) = \prod_{i=1}^{8} C_{N_i}^{r_i} \epsilon(E_i|A, E_0)^{r_i}(1 - \epsilon(E_i|A, E_0))^{N_i - r_i}$$

It is not a very big trouble to calculate the posterior numerically in Python by making it into grids. Hence we can get the figure of the posterior as

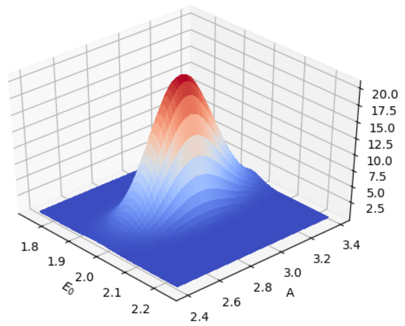

Figure 5.1: Posterior

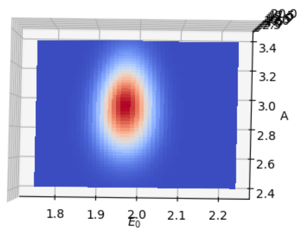

Figure 5.2: Posterior

According to the maximum value of the posterior, we can get the estimator of E_0 and A as

$$E_0 = 1.97 \qquad A = 2.93$$

Use these two values in Sigmoid function and compare with the experiment value we can get

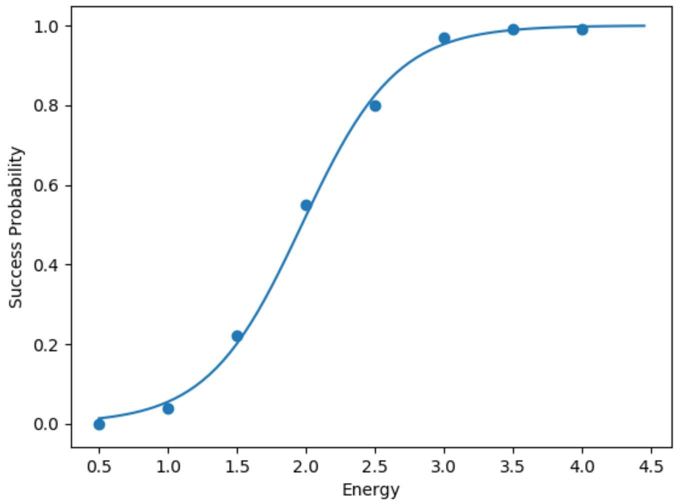

Figure 5.3: Sigmoid function and the experiment data

We can see that the fitting is good. Nearly all the data points are on the function.

(b) As for the test statistic, I just use the one in the lecture note.

$$\xi(\{r_i\}; A, E_0) = \prod_{i=1}^{8} C_{N_i}^{r_i} \epsilon(E_i|A, E_0)^{r_i} (1 - \epsilon(E_i|A, E_0))^{N_i - r_i}$$

There are not too much can be talked about because I just follow all steps in the lecture note rigorously. Finally we can get

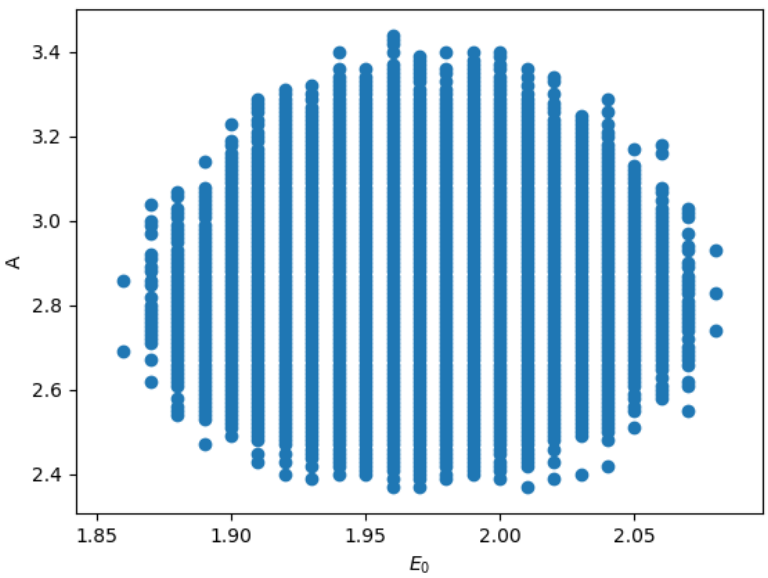

Figure 5.4: 68% CL interval

We can find that $A = 2.93$ and $E_0 = 1.97$, which we get in Bayesian analysis, is nearly in the center of the confidence level interval. This shows the consistence of these two methods.

Probelm 5.2

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

Solution

(a) Follow the same process of problem 5.1. Firstly, consider some estimator for E_0 and A. Since $r = 0$ when $E = 0.5$, we choose

$$E_0 = 0.5 \qquad \text{and} \qquad \sigma_{E_0} = 0.5$$

As for A, the success probability nearly reaches maximum when the energy is bigger than 2.5. Thus, we choose

$$A = \frac{\pi}{6} \qquad \text{and} \qquad \sigma_A = \frac{\pi}{6}$$

Then, use the same Bayesian formula to calculate the posterior we have

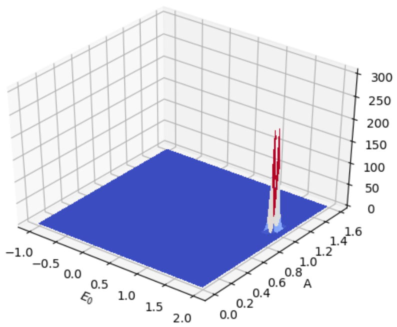

Figure 5.5: Posterior

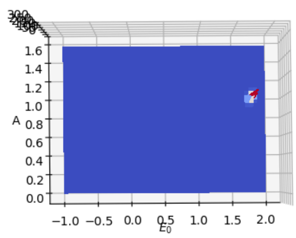

Figure 5.6: Posterior

Obviously, there is a huge deviation between the peak value and the estimated value. Try to find the maximum value of posterior we have

$$E_0 = 1.79 \qquad A = 1.03$$

Use these two values to fit for the success efficiency so we have

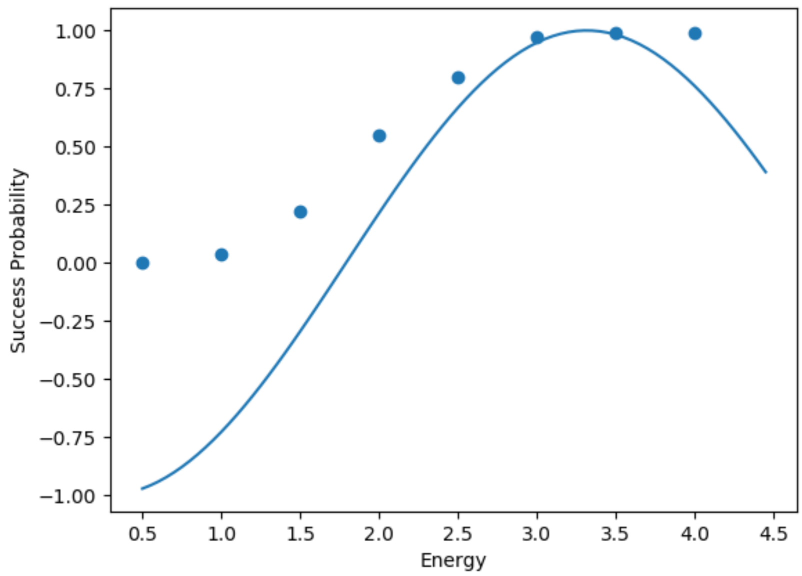

Figure 5.7: Model function and experiment data

We can see that the fitting is terrible and when the energy is small, there is a huge gap between the model value and experiment value.

(b) As to frequentist analysis, one problem is that it is not easy to find a effective region for A and E_0. The reason is we have to guarantee that the probability for binomial sample is positive. This condition require that the effective energy region has to be smaller than 0.5 and that A can not be too big otherwise $A(E - E_0)$ will be bigger than π.

Finally, I choose the grid region as

$$-1 < E_0 < 0.49 \qquad 0 < A < \pi/5$$

The result is that there is no 68% confidence level interval in this region. To look more at this result, I try to compare the test statistic of experiment data and that of generated data. Define a ratio which shows how many test statistic values of generated data are bigger than the test statistic of experiment data. Then we have

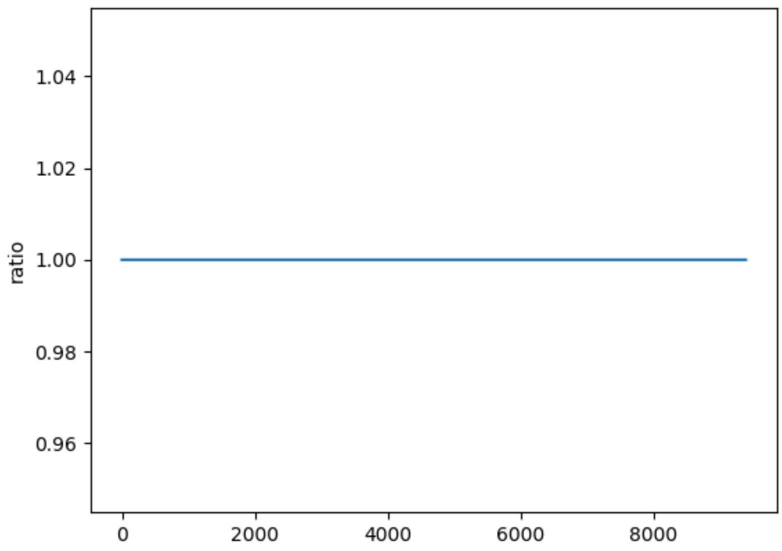

Figure 5.8: The ratio showing how many test statistic values of generated data are bigger than the test statistic of experiment data

We can see that this ratio is always 1, which means that the test statistic of data is always smaller that that of generated data. This is similar to the bad-model case in the lecture note.

(c) As we can see from the two kinds of methods, this model is not suitable for this question. We have two very obvious evidence

(1) The fitting in Bayesian method is terrible.

(2) In frequentist method, there is no confidence level interval and there is a huge gap between the test statistic of experiment data and generated data.

This just shows the importance of choice of model. When we plot the experiment data, we can easily see that sin is not a good choice for this model because we can not imagine how these data can be fitted to a sin function. The results of computation just prove this. Perhaps we can improve it by adding more parameters in sin function. However, it is always better to use as few parameters as possible to reach what we want. As John von Neumann once said:" With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

Consequently, choosing an appropriate model function is quite important. To do this, it is probably good to start with how the experiment data is curved on a line and use the properties of curve to find a suitable function.

Probelm 5.3

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

Solution The distribution function of canonical one-data-point χ^2 is

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

The mode is the easiest one because the distribution is a monotonously decaying function for χ^2. We can get the mode is

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

As to the mean and variance, we can of course follow the definition and compute the integrals. However, we can use some easier method.

The definition of χ^2 for single data point is

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

We can also regard the χ^2 as the square of a random variable which satisfied $\mathrm{N}(0,1)$, which means

$$\chi^2 = X^2 \qquad \text{where} X \sim \mathrm{N}(0,1)$$

Then, we can use the conclusions we have got in Gauss distribution

$$E[\chi^2] = E[X^2] = V[X] = 1$$

$$V[\chi^2] = E[X^4] - E[X^2]^2 = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - 1 = 2$$

Probelm 5.4

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:

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

(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

(a) First is to use chi-squared minimization fit. Just follow the process in the lecture note, we can calculate that

$$Y = \begin{pmatrix} 83.1375 \\ 54.48376 \\ 39.7710 \end{pmatrix} \qquad M = \begin{pmatrix} 1.1250 & 0.5906 & 0.3858 \\ 0.5906 & 0.3858 & 0.2820 \\ 0.3858 & 0.2820 & 0.2198 \end{pmatrix}$$

Then we can calculate the covariance matrix and estimator for A, B, C

$$\Sigma = M^{-1} = \begin{pmatrix} 15.2796 & -61.1145 & 51.5995 \\ -61.1145 & 286.2745 & -260.0619 \\ 51.5995 & -260.0619 & 247.6780 \end{pmatrix}$$

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} -7.2676 \\ 173.4671 \\ -28.8777 \end{pmatrix}$$

Plot the experiment data and model function together and we have

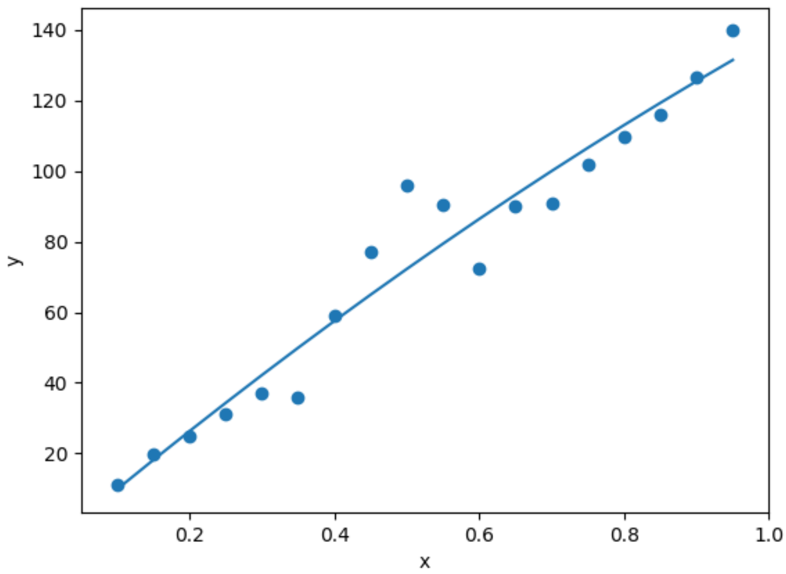

Figure 5.9: Fitting of model (a) by minimizing χ^2

As for calculating p-value, we can choose χ^2 as test statistic for convenient. The χ^2 for experiment data is

$$\chi^2_{exp} = 92.5065$$
$$p = \int_0^{92.5065} \frac{t^8 e^{-t/2}}{2^9 \times \Gamma(9)} \approx 1$$

This means this model is terrible for this set of data.

As for Bayesian analysis, we assume flat prior so that

$$p(A, B, C|\{y\}) = \frac{f(\{y\}|A, B, C)}{\iiint f(\{y\}|A, B, C)\, \mathrm{d}A\, \mathrm{d}B\, \mathrm{d}C} \propto f(\{y\}|A, B, C)$$
$$f(\{y\}|A, B, C) = \prod_i f(y_i|A, B, C) = \prod_i \frac{1}{4\sqrt{2\pi}} e^{-\frac{(y_i - f(x_i|A,B,C))^2}{2\sigma^2}} \propto e^{-\chi^2}$$

so what we need to do is to find a set of A, B, C value to maximize $f(A, B, C|\{y\})$. This means we need to minimize the χ^2, which is consistent with what we do before. Actually, with the help of former results for A, B, C, we can decide which grid region to be estimated much more easily and the estimator for A, B, C are

$$A = -7.2 \qquad B = 173.6 \qquad C = -29.0$$

(b) For model B, I have to say I can not solve it. Even though I followed the process about the general case of χ^2 minimization and calculate the Hessian matrix, the value of D, x_0, Γ are still very strange. Meanwhile, the fitting model is nearly invariant the Breit-Wigner term makes no difference. The value of χ^2 is still 92.5.

Then I also tried some global optimization methods like steepest descend and basin hoppin. However, it still doesn't work. I'm depressed about this. There are still many things I need to learn.

Last Set

Problem 6.1

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

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

(b) What is the maximum likelihood for p?

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

Solution Assume we have a data set $x_1, x_2, \ldots, x_n$ of this Bernoulli distribution. The the likelihood for p is

$$\mathcal{L}(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$\ln \mathcal{L}(p) = \sum_{i=1}^{n} x_i \ln p + (1-x_i)\ln(1-p)$$

We can define the sum of all measured values are $s = \sum_{i=1}^{n} x_i$. Then,

$$\ln \mathcal{L}(p) = s \ln p + (n-s)ln(1-p)$$

Find the maximum likelihood estimator for p

$$\left.\frac{\partial \ln \mathcal{L}(p)}{\partial p}\right|_{p=\hat{p}} = 0 \Rightarrow \frac{s}{\hat{p}} = \frac{n-s}{1-\hat{p}} \Rightarrow \hat{p} = \frac{s}{n}$$

Then, we can estimate the Fisher information as

$$\begin{aligned} I(p_0) = &- E\left[\frac{\partial^2 \ln P(x|p)}{\partial p^2}\right]\bigg|_{p_0} \approx -\frac{1}{n}\left.\frac{\partial^2 \ln \mathcal{L}(p)}{\partial p^2}\right|_{\hat{p}} \\ &= -\frac{1}{n}\left(\frac{n-s}{(1-\hat{p})^2} - \frac{s}{\hat{p}^2}\right) \end{aligned}$$

From the lecture note we have already known that the distribution of $\hat{p} - p_0$ is $\mathrm{N}(0, \frac{1}{nI(p_0)})$ so we can write down the distribution function as

$$P(\hat{p} - p_0) = \frac{1}{\sqrt{2\pi\left(\frac{s}{\hat{p}^2} - \frac{n-s}{(1-\hat{p})^2}\right)}} e^{-\frac{1}{2}\frac{(\hat{p}-p_0)^2}{\frac{s}{\hat{p}^2} - \frac{n-s}{(1-\hat{p})^2}}}$$

Problem 6.2

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

(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, $\hat{\lambda}$(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) From the lecture note, we have the formula for the estimator

$$\hat{\lambda} = \frac{N}{\sum_{i=1}^{N} x_i}$$

Firstly, we separately generate $n = 2, 10, 100$ values and take them into formula before to get an estimator of λ. Then we can get

$$\hat{\lambda}_{n=2} = 0.872 \qquad \hat{\lambda}_{n=10} = 0.903 \qquad \hat{\lambda}_{n=100} = 0.982$$

Repeat this process for 1000 times so that we can get the distribution of maximum likelihood estimator as

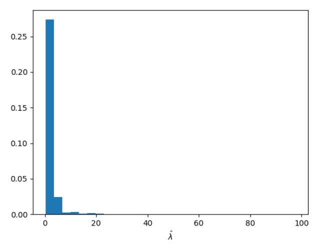

Figure 6.1: Distribution of estimator for n=2

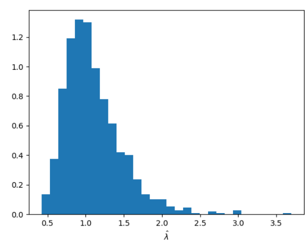

Figure 6.2: Distribution of estimator for n=10

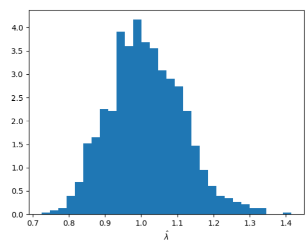

Figure 6.3: Distribution of estimator for n=100

(b) To get the expectation of LLN and CLT, we need to calculate the Fisher information first

$$I(\lambda_0) \approx -\frac{1}{n} \frac{\partial^2 \ln \mathcal{L}(\lambda)}{\partial \lambda^2}\bigg|_{\lambda_0}$$

Then the variance of $\hat{\lambda} - \lambda_0$ is

$$\sigma^2 = \frac{1}{nI(\lambda_0)} = \left(-\frac{\partial^2 \ln \mathcal{L}(\lambda)}{\partial \lambda^2}\bigg|_{\lambda_0} \right)^{-1} = \frac{\lambda_0^2}{n} = \frac{1}{n}$$

Then we can add the Gaussian distribution $\mathrm{N}(1, 1/n)$, which is the expectation of LLN and CLT, in $n = 10, 100$ case (I drop $n = 2$ case because it deviate from Gaussian function too much). Then we can get

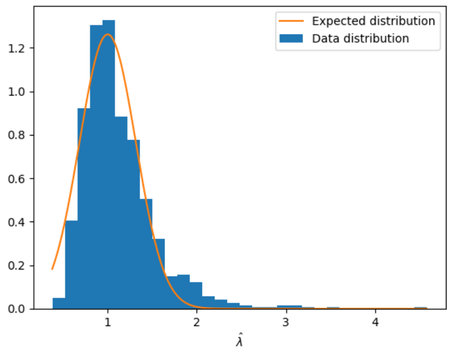

Figure 6.4: Distribution of estimator for n=10

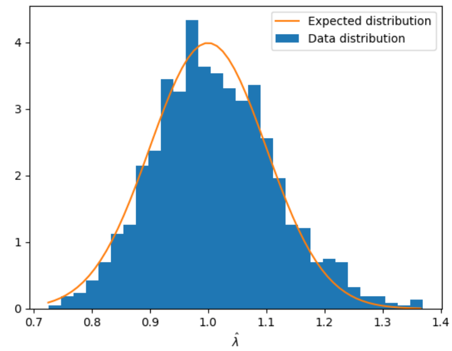

Figure 6.5: Distribution of estimator for n=100

We can see that the data distribution and expected distribution are approximately consistent and for $n = 100$, these two distribution fits better. Actually, the expected probability of estimator is the asymptotic behavior of the real distribution when $n \to \infty$ so this result is very natural.