UCSB, Physics 129AL, Computational Physics: Section Worksheet, Week 7A

Zihang Wang (UCSB), zihangwang@ucsb.edu

February 18, 2025

Section Participation and Submission Guidelines

Section attendance is required, but you do not need to complete all the work during the section. At each section, the TA will answer any questions that you might have, and you are encouraged to work with others and look for online resources during the section and outside of sections. Unless otherwise stated, the work will be due one week from the time of assignment. The TA will give you 1 point for each task completed. You can see your grades on Canvas.

We will use GitHub for section worksheet submissions. By the due date, you should have a single public repository on GitHub containing all the work you have done for the section work. Finally, upload a screenshot or a .txt file to Canvas with your GitHub username and repository name so the TA knows who you are and which repository you are using for the section.

Remember: talk to your fellow students, work together, and use GPTs. You will find it much easier than working alone. Good luck! All work should be done in the Docker container, and don't forget to commit it to Git!

Task 1: Statistical Inference on Biased Coins

Let's consider a problem similar to the one we discussed in class.

You are given three datasets, each containing 500 coin-flip outcomes represented as **Boolean** values (True means Head, and False means Tail). In this problem, your task is to investigate potential biases in the coin flips within these datasets. To **load the following datasets: dataset_1.json, dataset_2.json, dataset_3.json** from the website, you can use json.load().

a) Bayesian Inference

As discussed in class, for a given p, the probability that $N = 500$ tosses result in M heads and $N - M$ tails is given by the binomial distribution, with the probability mass function, namely the likelihood function:

$$P(M,N|p) = \binom{N}{M} p^M (1-p)^{N-M}.$$

Then the corresponding posterior is given by Bayes' theorem,

$$P(p|M,N) \sim \frac{1}{B(M,N)} P(M,N|p)P(p),$$

where,

$$B(M,N) = \int_0^1 dp\ p^M (1-p)^{N-M} = \frac{\Gamma(M+1)\Gamma(N-M+1)}{\Gamma(N+2)}.$$

Note that the above integral defines the Beta function, expressed in terms of the Gamma function (factorial, $\Gamma(n+1) = n!$).

First, we assume that our initial model for p is a uniform distribution (prior), i.e., $P(p) \sim 1/N$. With this prior distribution, apply the Bayesian inference methods discussed in class to calculate the likelihood functions for the above three datasets. Subsequently, create a similar plot to visualize the corresponding posterior distributions for each dataset. Additionally, numerically calculate the expectation value and variance of each posterior distribution. Use the fisher information matrix to calculate the variance at each batch. You can take a batch size of 50 when you iterate the prior and posterior.

b) Stirling's approximation

As we discussed in class, the frequentist inference assumes a fixed p_{true}. Maximum Likelihood Estimation (MLE) is a technique used to estimate the parameter p_{true} of a presumed probability distribution (in this case, binomial) based on observed data. It involves maximizing a likelihood function,

$$P(M,N|p) = \binom{N}{M} p^M (1-p)^{N-M}.$$

with the objective of making the observed data most likely under the assumed model. In practice, working with the natural logarithm of the likelihood function $P(M,N|p)$, known as the log-likelihood, is often more convenient,

$$\mathcal{P}(M,N|p) = \log[P(M,N|p)] = \log[\binom{N}{M}] + M\log(p) + (N-M)\log(1-p).$$

When N is large, we can approximate the factorial by Stirling's approximation,

$$\log(n!) \approx n\log(n) - n + \frac{1}{2}\log(2\pi n).$$

Numerically check the Stirling's approximation by compute both sides of the above equation and make a (2,1) plot that shows the following: 1) for factorial, vary N from 1 to 10 and calculate each value then make a scatter plot. Plot a **smooth curve** that shows the Stirling's formula (remember, the Stirling's formula takes real values) and the Gamma function $\Gamma(N+1)$. 2) plot the difference between the Stirling's formula and the Gamma function.

e) Bootstrapping

Bootstrapping belongs to the wider category of resampling methods and involves any test or metric that employs random sampling with replacement, effectively replicating the sampling process. This process involves a substantial number of iterations, and for each of these bootstrap samples, we compute its expectation value or other statistical properties. The distribution of these expectation values or other statistical properties then provides insights into the characteristics of the underlying distribution.

Bootstrap three datasets 100 times each with sample sizes of 5, 15, 40, 60, 90, 150, 210, 300, and 400, and create a (3,3) histogram (so there will be a total of three (3,3) histograms). Calculate the expectation value and variance for each and label the subplot title with the sample size. Plot the expectation values and variances of the above in comparison with the one you have calculated in **b)**.

Task2: Particle Decay

Unstable particles are emitted from a source and undergo decay at a distance x, a continuous real number following an exponential probability distribution with a characteristic length λ. Decay events can be detected only if they happen within a range extending from $x = 1$ to $x = \infty$, (a.u.). A total of N decay events are observed at positions $x_1, \ldots, x_N$.

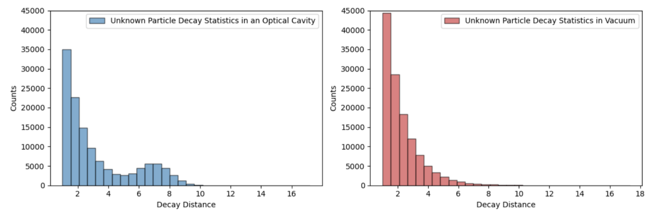
For a given decay parameter, λ, the probability of observing a particle at a distance x is given by an exponential function,

$$P(x \mid \lambda) = \frac{1}{Z(\lambda)} \frac{1}{\lambda} e^{-x/\lambda} \qquad 0 < x < \infty \tag{1}$$

where

$$Z(\lambda) = \int_0^\infty dx \frac{1}{\lambda} e^{-x/\lambda} = e^{-1/\lambda}.$$

You are given two datasets that record the decay distances of approximately 10^4 data points in both a vacuum and an optical cavity. In this question, you are asked to infer the values of the decay constants, λ, under both conditions. The measurement files are Vacuum_decay_dataset.json and Cavity_decay_dataset.json Here are snapshots of the data.



a) Unknown particle

You can observe that the optical cavity modifies a fraction of particles into a different type with distinct decay properties. What are the decay constants in both cases? How can you define the decay constant in the presence of an optical cavity? What additional structures have you observed? Let's assume the second particle type follows a Gaussian probability distribution function:

$$\mathcal{F}(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

What are the parameters μ, σ, λ that best fit the data? Find mean and variance of the associated with the fit parameter. What is the maximum likelihood estimator in this case? Calculate the fisher information matrix around the best fit parameter.

b) null hypothesis

Construct a null hypothesis and reject this null hypothesis with 95% confidence that there exists an additional decay contribution in the optical cavity. You do not reject the Gaussian structure.