HomeWork 2 : Few simple problems

This homework involves works on a computer, with the language of your choice. Present your result with graphs, plots, and data. Jupyter notebooks are a good option.

1 Sampling π

We come back to the problem of estimating π by sampling uniformly points in the square between $-1 < x < 1$ and $-1 < y < 1$, and counting how many hits are inside the unit circle. For each points $i = 1 \dots N$, we define the random variable $S_i = 0$ if the point is outside the circle, and $S_i = 4$ if the point is inside the circle.

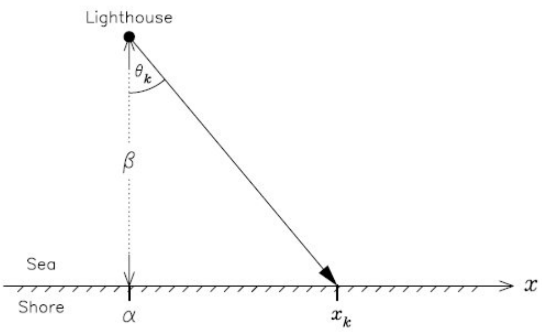
1. If each point i is sampled uniformly in the square, what is the probability p that $S_i = 4$? What are the mean m and the variance Δ of the distribution $P(S_i)$?
2. If we are given N perfectly random independent points, show that $\hat{m}$ and $\hat{\Delta}$

$$\hat{m} = \frac{1}{N} \sum S_i \qquad \hat{\Delta} = \frac{1}{N-1} \sum S_i^2 - \hat{m}^2$$

   are unbiased estimates of the mean and variance of $P(S)$.
3. What is the variance of the estimator $\hat{m}$? Deduce the value of the typical error made by this estimator when using N points to compute π.
4. Implement the sampling strategy where each points is selected uniformly in the square. Make a plot to make sure you are doing it well. For different values of N, compute 1000 times different estimates of π. Check that the variance of your estimation is indeed close to the one predicted in Q.3. Compute the probability to make an error larger 0.01 and compare with the various bounds we discussed in the course.

2 Find the lighthouse

A lighthouse is somewhere off a piece of straight coastline at a position α along the shore and a distance β out at sea. It emits a series of short highly collimated flashes at random intervals and hence at random azimuths. These pulses are intercepted on the coast by photo-detectors that record only the fact that a flash has occurred, but not the angle from which it came. N flashes have been recorded so far at positions $\{x_k\}$. Where is the lighthouse ?



1. Show that the probability to observe an event at value x_k is given by

$$P(x_k; \alpha, \beta) = \frac{\beta}{\pi\left[\beta^2 + (x_k - \alpha)^2\right]}$$

   This is known as a Cauchy or Lorentz distribution.

2. Let us sample from it so that we can have some synthetic data to work with. Generate some synthetic data using thhe values $\alpha = 30.0$ and $\beta = 10.0$. For instance $N = 10$ points (most language do have a generator of random number from a Cauchy distribution).
3. Assume that the value of β is known. Plot the likelihood (as a function of your estimated α), together with a point indicating the the Maximum Likelihood estimate, for different values of N between 10 and 1000.
4. Bonus question ; when sampling the Cauchy distribution, does the mean coincides with the mode (i.e. the maximum) of the posterior ? Why is that ? Will they coincide in the $N \to \infty$ limit ?

3 Statistical inference and Maximum Likelihood

We shall consider the problem discussed in our lecture, where a system emits particle with a half-length decay λ, which we detect if $1 < \lambda < 20$. The (density) probability to observe such a decay is thus

$$\begin{aligned} P_\lambda(x_i) &= \frac{e^{-x/\lambda}}{Z(\lambda)} \text{ if } 1 < x < 20 \\ P_\lambda(x_i) &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

1. Compute $Z(\lambda)$ such that the probability is normalized. What the probability $P_\lambda(\{x\})$ to observe a set of n events $(\{x\})$? Write a program that output n such observation sampled from the probability distribution (1).
2. We choose the true value to be $\lambda^* = 10$. Generate $n = 10$ observations. Plot the likelihood as a function of λ and see how it is peaked arround the true value. Repeat for $n = 100$ and $n = 1000$.
3. We now assume that we are given a set of n observations, without being told the true value of λ. We consider the maximum likelihood estimator

   $$\hat{\lambda}_{\mathrm{ML}}(\{x\}) = \mathrm{argmax}_\lambda P_\lambda(\{x\}) \tag{2}$$

   and we shall define the squared error as $\mathrm{SE} = (\hat{\lambda}_{ML}(\{x\}) - \lambda)^2$.

   Create some data set with $n = 10, 100, 1000$ for different values of λ and see how the ML estimator performs. Note that finding the maximizer $\hat{\lambda}_{ML}$ can be done numerically
4. If we average of many realizations of this process we can obtain the mean square error $MSE(\lambda, \hat{\lambda}, n)$, which is thus a function of n, λ and of the estimator $\hat{\lambda}$. Compute, for instance for $n = 10000$, the curve $MSE(\lambda, \hat{\lambda}_{\mathrm{ML}}, n = 10000)$.

   How does this curve compare with the Cramers-Rao bound $\mathrm{MSE}(\hat{\lambda}) \geq \frac{1}{I(\lambda)}$, where $I(\lambda)$ is the total Fisher information. Is the ML estimator unbiased ?