Homework 2, Statistical Mechanics: Concepts and applications 2019/20 ICFP Master (first year)

Botao Li, Valentina Ros, Victor Dagard, Werner Krauth

We will "soon" make a solution to this homework available. Please study in the meantime.

Please contact me if you find the exercise unclear and/or the solution unclear or wrong.

(Dated: October 14, 2019)

In lecture 02, we treated the maximum likelihood approach as one of the key methods for estimating the parameters of a distribution. Here we treat two examples. The second one was of great historical importance on the battlefields of WW2 although it proved necessary to go one step farther than we will do here.

I. POINT ESTIMATE FOR THE PARAMETERS OF A GAUSSIAN DISTRIBUTION

Suppose that we have n data points $x_1, x_2, x_3, , ..., x_n$ (these points are real numbers between $-\infty$ and ∞), and we know that they are drawn from a Gaussian distribution with unknown values of the variance σ^2 and the mean value $\langle x \rangle$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x - \langle x \rangle)^2/(2\sigma^2)). \tag{1}$$

What is the maximum likelihood estimator for the mean value $\langle x \rangle$ and variance σ^2 of this Gaussian distribution from the data?

Hint1 Remember (from the first lecture) that the likelihood function is given by $p(x_1)p(x_2)....p(x_n)$.

Hint2 If you use the log likelihood function, explain why this can be done.

Carefully explain your calculation.

The likelihood function is

$$\mathcal{L}(\mu, \sigma; \{X_i\}) \propto \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(X_i-\mu)^2/2\sigma^2}$$

Since log *function is monotonic, maximizing likelihood is equivalent to maximizing log likeli-*

hood. And,

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = 0$$
$$\hat{\mu} = \frac{1}{N}\sum_i X_i$$

and

$$\frac{\partial \log \mathcal{L}}{\partial \sigma} = 0$$
$$\hat{\sigma} = \frac{1}{N}\sum_i (X_i - \mu)^2$$

II. UNIFORM DISTRIBUTION

1. *Preparation*

What follows, a preparation for Section III, is described in Wasserman as a hard example "that many people find confusing". It will not confuse You!

2. *Application*

Suppose a uniform distribution between 0 and θ, and consider k samples drawn from this distribution.

- What is the likelihood function $L(\theta)$ given $x_1, \ldots x_k$? (Hint: suppose that "the probability to sample x_i" is $1/\theta$. The $1/\theta$ factor is "physically" clear). Plot $L(\theta)$.
- What is the maximum-likelihood extimator of θ given $x_1, \ldots, x_k$, that is, the samples?
- Comment your finding.

The likelihood function is

$$\mathcal{L} \propto \begin{cases} 0 & \exists X_i \notin [0, \theta] \\ 1/\theta^N & \forall X_i \in [0, \theta] \end{cases}$$

Likelihood is plotted in Fig. 1. θ is always positive, thus $[0, \theta]$ should cover all the data points

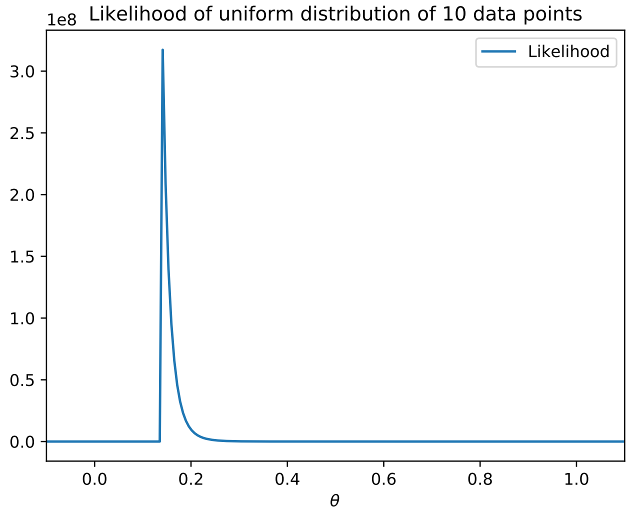

FIG. 1: Likelihood of uniform distribution.

in order to maximize the likelihood. And when $\forall X_i \in [0, \theta]$, the likelihood decrease with respect to θ. In order to maximize $\mathcal{L}$, $\hat{\theta}$ has to satisfy

$$\hat{\theta} = min(\{X_i\})$$

However, for an uniform distribution, $X_i < \theta$. Thus, the estimated value is always smaller than the real value of θ.

III. GERMAN TANK PROBLEM (FREQUENTIST APPROACH)

This example has been of considerable importance, first in WW2, then in the theory of statistics. It is the discrete version of Section II.

A. Preparation

Consider N balls numbered $1, 2, 3, ..., N$, and take k out them (urn problem without putting them back). What is the probability $p_N(m)$ that the largest of them is equal to m?

Hint0 How many ways are there to pick k (different) balls out of N?

Hint1 To solve this simple combinatorial problem, consider that m must be contained in $k, k+1, k+2, ...N$.

Hint2 Count the number of ways to pick $(k-1)$ balls so that they are smaller than m.

Carefully explain your calculation.

There are $\binom{N}{k}$ ways of selecting k balls out of N balls. The largest number being m is equivalent of having one ball being the number m and $k-1$ balls which have numbers smaller than m. If $m < k$, then there is no such cases. Thus, the probability is 0. When $m \geq k$, in order to reach the desired configuration, one of the balls has to be number m. And the number of cases of having $k-1$ balls which are smaller than m is $\binom{m-1}{k-1}$. Last but not least, m has to be smaller than N Thus,

$$p_N(m) = \begin{cases} 0 & m < k \text{ or } m > N \\ \binom{m-1}{k-1}/\binom{N}{k} & else \end{cases}$$

B. Application

From an urn with an unknown number N of balls (tanks), the following $k = 4$ balls were drawn (without putting them back):

$$1, 2, 4, 14$$

What is the maximum likelihood estimator of the total number N of balls (tanks) (based on the probability distribution of the sample maximum m, here 14) that are contained in the urn (destroyed tanks left on the battlefield)?

The likelihood function is

$$\mathcal{L} \propto \begin{cases} 0 & N < m \\ \binom{m-1}{k-1}/\binom{N}{k} & N \geq m \end{cases}$$

Thus, in order to maximize the likelihood function, $N \geq m$. The nominator of the likelihood function has nothing to do with N and can be ignored. When the value of N is increased by one, the likelihood function will be suppressed by $O(N/(N-k))$. Thus, $\hat{N} = m$.

The (disappointing) result of the maximum likelihood estimator (here in the famous "German tank problem") points to one of the limitations of the maximum likelihood method, namely that it presents a bias. Comment on this property. A trick allows to arrange the situation. In simple terms it consists in supposing that the mean of the intervals between the first ball and zero, the second and the first ball, the third and the second... etc is probably as large as the interval between the largest ball and N.

IV. GERMAN TANK PROBLEM (BAYESIAN APPROACH)

The Bayesian approach treats the total number N of the balls (tanks) as a random variable, and it has been much studied in the literature. But to start, simply write a program for $k = 4$ that samples N from a discretized exponential distribution with parameter λ. Then sample k *different balls* from this distribution, if possible.

A. Maximum

Numerically compute the probability distribution of all the N for which the largest of the $k = 4$ balls is equal to 14 (see previous example). Do this by sampling: Sample N, then sample $k = 4$ balls, and enter N into a histogram if the largest of the 4 balls is equal to 14.

Plot this distribution (histogram), its expectation and variance for different values of λ. For your convenience you find the Python program, already 95% written, on the course website. Modify it (to compute the expectation and variance as a function of λ), and run it.

B. Total sample

Numerically compute the probability distribution of all the N for which the $k = 4$ balls exactly equal $1, 2, 4, 14$. Plot this distribution (histogram), its mean value and variance for different values of λ. Do these distribution differ (empirically) from the ones in Section IV A? For your convenience, the Python program on the course website already contains the crucial modification. Is the outcome different from the one of the maximum version (Section IV A)?