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

Botao Li, Valentina Ros, Victor Dagard, Werner Krauth
Tutorial exercises

I. STATISTICAL INFERENCE

1. **Maximum Likelihood Method for the Bernoulli distribution.** Suppose we have a coin which falls heads up with probability p. Let $X_i = X_1, ..., X_N$ represent the outcome of the i^{th} flip ($x = 1$ for heads and $x = 0$ for tails). X has a Bernoulli distribution with PDF

$$\pi(x; p) = p^x (1-p)^{1-x} \quad \text{for} \quad x = 0, 1. \tag{1}$$

   (a) Estimate p using the MLE.

   (b) Find a 95% confidence interval for p.

   (c) Let $\tau = e^p$. Find the MLE for τ.

   HINT: Use one of the properties of the MLE.

: *(a) Calculate the likelihood function, take its log, and find the value of p that maximizes the log.*

$$L_N(p) = \prod_{i=1}^{N} \pi(X_i; \theta) = \prod_{i=1}^{N} p^{X_i} (1-p)^{1-X_i}$$

$$\log L_N(p) = \left(\sum_{i=1}^{N} X_i \right) \log p + \left(N - \sum_{i=1}^{N} X_i \right) \log(1-p)$$

$$\frac{\partial \log L_N(p)}{\partial p}\Big|_{p=p_{mle}} = \frac{\sum_{i=1}^{N} X_i}{p} + \frac{N - \sum_{i=1}^{N} X_i}{1-p}\Big|_{p=p_{mle}} = 0$$

$$p_{mle} \equiv \hat{p} = \sum_{i=1}^{N} X_i / N = \bar{X}$$

(b) Approximate 95% confidence interval is given by

$$\left[\hat{p} - \sqrt{\frac{1}{I_N(\hat{p})}}, \hat{p} + \sqrt{\frac{1}{I_N(\hat{p})}} \right]$$

where $I_N(\hat{p})$ is Fisher information evaluated at the MLE $\hat{p}$.

$$I_N(\hat{p}) = -\frac{\partial^2 \log L_N(p)}{\partial p^2}\Big|_{p=\hat{p}}$$

$$\frac{\partial^2 \log L_N(p)}{\partial p^2} = \frac{-N\bar{X}}{p^2} - \frac{N(1-\bar{X})}{(1-p)^2}$$

Therefore the confidence interval is given by

$$\left[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\right]$$

(c) Use the fact that MLE is ***reparametrization invariant:***

$$\hat{\tau} = e^{\hat{p}} = e^{\bar{X}}$$

2. **Bootstrap.**

We consider a sample of $N = 2^n$ numbers with unknown distribution from a population of $M >> N$ numbers

$$x_1, x_2, \ldots, x_N \tag{2}$$

We wonder which is the minimal value that the random variables x_j can assume inside the population and would like to use bootstrap to compute the variance of the minimum.

To our great surprise the numbers in the sample have the simple form 2^j, where $j = 1, 2, \ldots, n+1$

$$x_i \in \{2^j\}_{j\in\{1,\ldots,n+1\}}\,, \tag{3}$$

and their multiplicity is given by $m(j) = 2^{n-j}$ for $j \leq n$ and 1 for $j = n+1$. With this information we can predict the outcome of the bootstrap sampling and directly estimate the empirical distribution of the minimum

$$\mathcal{P}(\min\{x_i^*\} = 2^j) \tag{4}$$

of the bootstrap realization $\{x_i^*\} \in \{x_1, \ldots, x_N\}$.

(a) Determine $\mathcal{P}(\min\{x_i^*\} = 2^j)$ and the bootstrap variance $\mathrm{Var}_{boot}^{\min} =< \min\{x_i^*\}^2 > - < \min\{x_i^*\} >^2$ of the minimum.
Hint: Note that the the probability that the minimum of $\{x_i^*\}$ is 2^j with some j is given by

$$\mathcal{P}(\min\{x_i^*\} = 2^j) = \mathcal{P}(\min\{x_i^*\} > 2^{(j-1)}) - \mathcal{P}(\min\{x_i^*\} > 2^j) \tag{5}$$

and that the probability that a number x_i^* in the sample is equal to 2^j is given by its multiplicity

$$\mathcal{P}(2^j \in \{x_i^*\}) = \frac{m(j)}{N} \tag{6}$$

Finally remember the geometric sum

$$\sum_{i=1}^{j} a^i = \frac{a^{j+1} - a}{a - 1} \tag{7}$$

(b) Approximate the expression assuming N large (if N is large, also n is large). You should get $\mathrm{Var}_{boot}^{\min} = 2^{2-N}$.

(c) A bootstrap $1-\alpha$ confidence interval is given by

$$\mathcal{P}\left(\min\{x_i\} \in \left[2\min\{x_i\} - \mu_{1-\alpha/2}\ ,\ 2\min\{x_i\} - \mu_{\alpha/2}\right]\right) = 1-\alpha \tag{8}$$

Compute the bootstrap $1-\alpha$ confidence interval for the minimum of the sample $\min\{x_i\} = 2$. With which confidence $1-\alpha$ can we state that the minimal value is equal to 2?

Hint: Find $\mu_{1-\alpha/2}$ and $\mu_{\alpha/2}$ using

$$\mathcal{P}(\min[\{x_i^*\}] \le \mu_{1-\alpha/2}) = 1 - \frac{\alpha}{2} \qquad \mathcal{P}(\min[\{x_i^*\}] \le \mu_{\alpha/2}) = \frac{\alpha}{2}\,. \tag{9}$$

and the previous result for $\mathcal{P}(\min\{x_i^*\} > 2^j)$.

: *The bootstrap method uses simulation for estimating errors and computing confidence intervals. However, due to the simple properties of the sample, we can calculate them analytically.*

First of all, we must compute the probability that, picking a random bootstrap realization $\{x_1^*, x_2^*, \ldots, x_N^*\}$ (with $x_i^* \in \{x_1, x_2, \ldots, x_N\}$), this has a given minimal value. (We note that multiplying this probability by the number of realizations B of the simulation gives the average number of configurations with given minimum.) Since

$$\mathcal{P}(\min\{x_i^*\} = 2^j) = \mathcal{P}(\min\{x_i^*\} > 2^{(j-1)}) - \mathcal{P}(\min\{x^*\} > 2^j) \tag{10}$$

and

$$\mathcal{P}(\min\{x_i^*\} > 2^j) = \left(1 - \sum_{i\le j} \mathcal{P}(2^i \in \{x_k^*\})\right)^N = \left(1 - \sum_{i\le j} \frac{m(i)}{N}\right)^N = \begin{cases} 2^{-jN} & j \le n \\ 0 & j = n+1\,. \end{cases}, \tag{11}$$

we have

$$\mathcal{P}(\min\{x_i^*\} = 2^j) = \left(1 - \frac{1}{N}\sum_{i<j} m(i)\right)^N - \left(1 - \frac{1}{N}\sum_{i\le j} m(i)\right)^N = \begin{cases} 2^{-jN}(2^N-1) & j \le n \\ 2^{-nN} & j = n+1\,. \end{cases} \tag{12}$$

The bootstrap variance of the minimum is

$$\mathrm{Var}_{boot}^{\min} = (2^N-1)\sum_{j=1}^{n} 2^{2j}2^{-jN} + 2^{2(n+1)}2^{-nN} - \left((2^N-1)\sum_{j=1}^{n} 2^j 2^{-jN} + 2^{(n+1)}2^{-nN}\right)^2 =$$
$$(1-2^{-N})\frac{2^2 - 2^{2(n+1)-nN}}{1-2^{2-N}} + 2^{2(n+1)}2^{-nN} - \left((1-2^{-N})\frac{2-2^{(n+1)-nN}}{1-2^{1-N}} + 2^{(n+1)}2^{-nN}\right)^2 \xrightarrow{\ll N} 2^{2-N} \tag{13}$$

A confidence interval can be obtained using the bootstrap estimation of the probability distribution. A bootstrap $1-\alpha$ confidence interval is given by

$$\left[2\min\{x\} - \mu_{1-\alpha/2}, 2\min\{x\} - \mu_{\alpha/2}\right] \tag{14}$$

where

$$\mathcal{P}(\min\{x^*\} \le \mu_{1-\alpha/2}) = 1 - \frac{\alpha}{2} \tag{15}$$

and

$$\mathcal{P}(\min\{x^*\} \le \mu_{\alpha/2}) = \frac{\alpha}{2}\,, \tag{16}$$

see below for the derivation. The interpretation of the confidence interval is as follows (see Wasserman Sec. 6.3.2): if one (i) takes (ideally) infinitely many samples $x_1, \cdots, x_N$ *from the population of size* M*, (ii) computes the corresponding value of* $\min(x_i)$*, (iii) computes the quantities* $\mu_{1-\alpha/2}, \mu_{\alpha/2}$ *from the bootstrap distribution, then one finds that* $(1-\alpha)\%$ *of the intervals contains the minimum of the population.*
From (11) *and using that* $\mathcal{P}(\min\{x^*\} \le \mu_{\alpha/2}) = 1 - \mathcal{P}(\min\{x^*\} > \mu_{\alpha/2})$ *it then follows (we denote with* $\lfloor \ldots \rfloor$ *the floor function)*

$$\mu_{1-\alpha/2} = 2^{\left\lfloor \frac{1}{N}\log_2 \frac{2}{\alpha} \right\rfloor} \tag{17}$$

$$\mu_{\alpha/2} = 2^{\left\lfloor \frac{1}{N}\log_2 \frac{2}{2-\alpha} \right\rfloor} \tag{18}$$

Finally, the bootstrap $1-\alpha$ *confidence interval is*

$$\left[2\min\{x\} - 2^{\left\lfloor \frac{1}{N}\log_2 \frac{2}{\alpha} \right\rfloor}, 2\min\{x\} - 2^{\left\lfloor \frac{1}{N}\log_2 \frac{2}{2-\alpha} \right\rfloor}\right] \tag{19}$$

If N *is large and we set* $\alpha = 2^{1-N}$ *the confidence interval become*

$$\sim [2, 4 - 2^{1/N}] \sim [2, 3] \tag{20}$$

While if $\alpha > 2^{1-N}$ *the lower bound is larger than the upper bound coming from the sample, i.e. 2. Since the minimum of the population is always less or equal then the minimum of the sample, the confidence interval has to include our minimum* 2*. Therefore with confidence level* $1 - 2^{1-N}$ *the minimum is* 2*.*

Derivation of the confidence interval. *Let* $T(x_1, \cdots, x_N)$ *be the statistics we are interested in (in the exercise,* $T(x_1, \cdots, x_N) = \min x_i$*), and let* $\hat{t}$ *be the value that it takes on the given sample. In principle one wants to find an interval* $[a, b]$ *such that:*

$$P\left(T \in [a, b]\right) = 1 - \alpha, \tag{21}$$

which is ensured if we set:

$$\begin{aligned} P\left(T \le a\right) &= \frac{\alpha}{2} \\ P\left(T \ge b\right) &= \frac{\alpha}{2}, \end{aligned} \tag{22}$$

or equivalently:

$$\begin{aligned} P\left(\hat{t} - T \ge \hat{t} - a\right) &= \frac{\alpha}{2} \\ P\left(\hat{t} - T \le \hat{t} - b\right) &= \frac{\alpha}{2}. \end{aligned} \tag{23}$$

In order to solve these equations for a, b, one should know the distribution of T. The idea of bootstrap is to replace the unknown distribution of T with the distribution constructed over the sample, so that $T \to \hat{t}$; at the same time, the bootstrap realizations $\{x_i^*\}$ give different realizations t^*, that we can use to build a statistics for $\hat{t}$. Therefore in the equations above we substitute the variable with unknown distribution $\hat{t} - T$ with its bootstrap approximation $\hat{t} - T \to t^* - \hat{t}$. This gives

$$
\begin{aligned}
P\left(\hat{t} - T \geq \hat{t} - a\right) &\approx P\left(t^* - \hat{t} \geq \hat{t} - a\right) = \frac{\alpha}{2} \\
P\left(\hat{t} - T \leq \hat{t} - b\right) &\approx P\left(t^* - \hat{t} \leq \hat{t} - b\right) = \frac{\alpha}{2}.
\end{aligned}
\tag{24}
$$

Then

$$
\begin{aligned}
P\left(t^* \leq 2\hat{t} - a\right) &= 1 - \frac{\alpha}{2} \\
P\left(t^* \leq 2\hat{t} - b\right) &= \frac{\alpha}{2},
\end{aligned}
\tag{25}
$$

and we can identify $\mu_{1-\frac{\alpha}{2}} = 2\hat{t} - a$, $\mu_{\frac{\alpha}{2}} = 2\hat{t} - b$.

3. **Bayesian inference.** In ideal gases of non-relativistic particles the speed v is described by the Maxwell-Boltzmann distribution:

$$
\pi_{\mathrm{MB}}(v|m, kT) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 e^{-\frac{mv^2}{2kT}}\,. \tag{26}
$$

We would like to infer the mass of the particles from a small sample $\{v\}$ consisting of n measurements of the velocities, taken at a given temperature kT.

(a) Construct a prior $\pi_{prior}(m|\pi_{\mathrm{MB}}, kT)$ that encodes our knowledge of the Maxwell-Boltzmann distribution and of the temperature; try to construct a prior that is invariant under reparametrizations, that is to say a prior independent of the particular functions of v that have been measured in the experiment.

(b) Estimate the mean and the variance of the mass using Bayesian inference.

: *If we rescale the velocity by a factor k, $v = \alpha u$, the distribution of u is given by*

$$
\pi(u|m, kT) = \left(\frac{\alpha^2 m}{2\pi kT}\right)^{3/2} 4\pi u^2 e^{-\frac{m\alpha^2 u^2}{2kT}}\,. \tag{27}
$$

This is still a Maxwell-Boltzmann distribution with a new mass $\alpha^2 m$. A sensible prior must be independent of the particular units of measurement used in the experiment, therefore this rescaling should map the prior on to itself:

$$
\pi(m|\pi_{\mathrm{MB}}, kT)dm = \pi(\alpha^2 m|\pi_{\mathrm{MB}}, kT)d(\alpha^2 m)\,. \tag{28}
$$

This means

$$
\frac{\pi(\alpha^2 m|\pi_{\mathrm{MB}}, kT)}{\pi(m|\pi_{\mathrm{MB}}, kT)} = \frac{1}{\alpha^2}\,, \tag{29}
$$

that is to say

$$\pi(m|\pi_{MB}, kT) \sim \frac{1}{m}\,. \tag{30}$$

This function is not normalizable, but, as we are about to see, the posterior probability will be. The posterior probability, defined as

$$\pi(m|\{v\}, \pi_{\rm MB}, kT) = \frac{\pi(m|\pi_{MB}, kT)\left(\prod_{i=1}^{n}\pi(v_i|m, kT)\right)}{\int_0^\infty dm\pi(m|\pi_{MB}, kT)\left(\prod_{i=1}^{n}\pi(v_i|m, kT)\right)} \tag{31}$$

(where we used Bayes's theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$*) is given by*

$$\pi(m|\{v\}, \pi_{\rm MB}, kT) = \left(\frac{\sum_{i=1}^{n} v_i^2}{2kT}\right)^{\frac{3n}{2}} \frac{m^{3n/2-1}}{\Gamma(\frac{3n}{2})} e^{-\frac{m}{2kT}\sum_{i=1}^{n} v_i^2}\,. \tag{32}$$

Here we introduced the Γ *function*

$$\Gamma(n+1) = \int_0^\infty \mathrm{d}k k^n e^{-k}\,, \tag{33}$$

which for integer n *is equal to the factorial:* $\Gamma(n+1) = n!$*. A simple computation then gives the average mass*

$$\langle m\rangle = \frac{2kT}{\sum_{i=1}^{n} v_i^2}\frac{\Gamma(\frac{3n}{2}+1)}{\Gamma(\frac{3n}{2})} = \frac{3kT}{\frac{1}{n}\sum_{i=1}^{n} v_i^2} = \frac{3kT}{v_{\rm rms}^2}\,. \tag{34}$$

and the variance

$$\mathrm{Var}[m] = \langle m^2\rangle - \langle m\rangle^2 = \frac{3}{2n}\left(\frac{2kT}{\frac{1}{n}\sum_{i=1}^{n} v_i^2}\right)^2 = \frac{2}{3n}\langle m\rangle^2\,. \tag{35}$$