Tutorial 2, Statistical Mechanics: Concepts and applications
2018/19 ICFP Master (first year)

Victor Dagard, Werner Krauth
Tutorial exercises

I. SOLUTIONS: 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.

2. **Bootstrap.**

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

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

   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 $j = 1, 2, \dots, n+1$

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

   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 $\{x_i^*\} \in \{x_1, \dots, x_N\}$.

   (a) Determine $\mathcal{P}(\min\{x_i^*\} = 2^j)$ and the bootstrap variance $\text{Var}_{boot}^{\min} =< \min\{x_i^*\}^2 > - < \min\{x_i^*\} >^2$ 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

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

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

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

$$\pi_{\mathrm{MB}}(v|m, kT) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 e^{-\frac{mv^2}{2kT}}\,. \tag{10}$$

We would like to infer the mass of the particles from a small sample $\{v\}$ consisting of n measuraments 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.