Homework 1, Statistical Mechanics: Concepts and applications
2019/20 ICFP Master (first year)

Botao Li, Valentina Ros, Victor Dagard, Werner Krauth

A solution to this homework will be made available on 16 September 2019. Please study in the meantime. Please contact Botao Li if you find the exercise unclear and the solution unclear or wrong.

(Dated: 8 September 2019)

I. CHEBYCHEV INEQUALITY: VARIATIONS ON A THEME

The Chebychev inequality:

$$P(|\xi - E(\xi)| > \epsilon) \leq \frac{\text{Var}(\xi)}{\epsilon^2} \tag{1}$$

is one of the fundamental achievements in probability theory, and it is of great importance in statistics. In eq. (1), $\text{Var}(\xi)$ denotes the variance of the distribution and $E(\xi)$ its expectation (mean value).

1. State all the conditions on the probability distribution f_ξ for the Chebychev inequality to hold. For example, does it hold for discrete distributions (a sum of δ functions), for distributions with infinite expectation yet finite variance, etc?

   The only condition is that the distribution has a finite variance. Note that this implies that the mean value is also finite.

2. Review the proof of Chebychev's inequality given in Lecture 1. Can the Chebychev inequality be "sharp", that is:

   (a) Can there be a distribution f_ξ where, for some ϵ, one has

   $$P(|\xi - E(\xi)| > \epsilon) = \frac{\text{Var}(\xi)}{\epsilon^2} \tag{2}$$

   (note the "=" sign in eq. (2) instead of the "$\leq$" in eq. (1)). If so, construct this probability distribution. Otherwise, explain why this is not possible.

   Equation 2 is satisfied when f_ξ consists of delta function, located at ϵ, $-\epsilon$, or both.

   (b) Can there be a distribution f, where the inequality of eq. (1) is sharp for *all* real ϵ? If so, construct this probability distribution. Otherwise, explain why this is not possible.

It is impossible. Recall the proof of Chebychev inequality, equality is achieved only if $f_\xi = 0$ when $\xi \neq \pm\epsilon$. Thus, if a distribution satisfies equation 2 when $\epsilon = \epsilon_1$, equation 2 is not satisfied for $\{\epsilon | \epsilon \neq \epsilon_1\}$.

II. RÉNYI'S FORMULA FOR THE SUM OF UNIFORM RANDOM NUMBERS, VARIATIONS

In tutorial 1, you derived Rényi's formula for the sum of uniform random numbers between -1 and 1:

$$f_n(x) = \begin{cases} \frac{1}{2^n(n-1)!} \sum_{k=0}^{\left[\frac{n+x}{2}\right]} (-1)^k \binom{n}{k} (n + x - 2k)^{n-1} & \text{for } |x| < n \\ 0 & \text{else} \end{cases} \tag{3}$$

1. Compute the variance of the distribution of eq. (3) for $n = 1$, that is for uniform random numbers between -1 and 1.

   $Var(\xi) = 1/3$

2. Compute the variance of Rényi's distribution for general n (Hint: this can be computed in 1 minute, if you use a result presented in the lecture).

   Since these random numbers are independent, $Var(\xi) = n/3$

3. Implement eq. (3) in a computer program for general n. For your convenience, you will find such a computer program on the course website. This program also computes $P(\xi > \epsilon)$. Download this program and run it (in Python 2, or you can modify it so that it runs in Python 3). Notice that you may change the value of n in this program.

4. Modify the program (plot) so that it compares $P_n(\xi > \epsilon)$ to the upper limit given by the Chebychev inequality (Attention: you may modify Chebychev's inequality to take into account that $f_n(x)$ is symmetric around $x = 0$). Comment.

5. Modify the program (plot) so that it compares $P_n(\xi > \epsilon)$ to the Cantelli inequality:

$$P(\xi - E(\xi) > \epsilon) \leq \frac{\text{Var}(\xi)}{\text{Var}(\xi) + \epsilon^2} \tag{4}$$

   (note that there are now no absolute values). Comment.

6. Modify the program so that it compares $P(\xi > \epsilon)$ to Hoeffding's inequality. Hoeffding's inequality considers a probability distribution with zero expectation and $a_i \leq \xi_i < b_i$ (we will later take constant bounds a and b, but in fact, they may depend on i). For every $t > 0$, it states:

$$P(\sum_{i=1}^{n} \xi_i \geq \epsilon) \leq \exp(-t\epsilon) \prod_{i=1}^{n} \exp\left[t^2 (b_i - a_i)^2/8\right]. \tag{5}$$

Is Hoeffding's inequality *always* sharper than the Chebychev inequality, that is, is Hoeffding with the best value of t better than Chebychev for all ϵ? What is the asymptotic behavior for $\epsilon \to \infty$ behavior of Hoeffding's inequality, and why does it satisfy such a stringent bound if the Chebychev inequality does not achieve it? Return a plot that contains, next to $f_n(x)$ and its integral $P_n(\xi > \epsilon)$, the comparison with Chebychev, Cantelli, and Hoeffding.

Answer for 3, 4, 5, 6:

As plotted is figure 1, all the inequalities are satisfied. For Hoeffding's inequality, the value of t is tuned point-wise so that each point has the tightest constraint. Cantelli's inequality is the best at the center of the distribution. Hoeffding's inequality provides the tightest constraint at the tail. And Chebychev's inequality outperforms the other two in between. When $\epsilon \to \infty$, Hoeffding's inequality decays like a Gaussian distribution. Hoeffding's inequality utilize the fact that it constrains the distribution of the sum of bounded random variables, which results in a faster decay compared with the other two inequality, which make no assumption about the random variable they constrain.

III. LÉVY DISTRIBUTIONS, TWO SIMPLE DEMONSTRATIONS

In lecture 1 and tutorial 1, we discussed and derived Lévy distributions: Universal (stable) distributions that have infinite variance. A good example for producing such random variables is from uniform random numbers between 0 and 1, $\text{ran}(0,1)$ taken to a power $-1 < \gamma < -0.5$. Such random numbers are distributed according to a distribution

$$f_\xi(x) = \begin{cases} \frac{\alpha}{x^{1+\alpha}} & \text{for } 1 < x < \infty \\ 0 & \text{else} \end{cases} \tag{6}$$

where $\alpha = -1/\gamma$ (you may check this by doing a histogram, and read up on this in SMAC book).

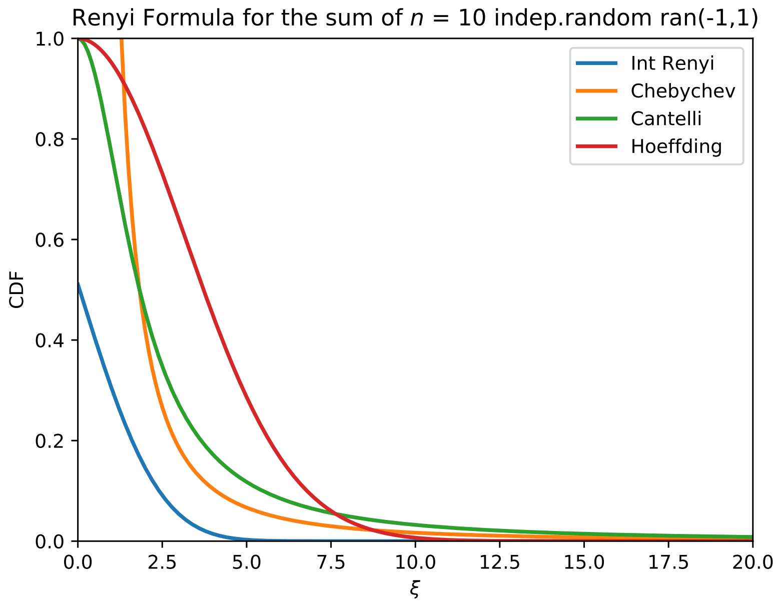

FIG. 1: Rényi's distribution and inequalities

1. Is the probability distribution of eq. (6) normalized for $\gamma = -0.8$ (that is $\alpha = 1.25$), is it normalized for $\gamma = -0.2$ (that is $\alpha = 5$)?

   Both of them are normalized

2. What is the expectation of the probability distribution for the above two cases, and what is the variance?

   For $\alpha = 1.25$, $E(\xi) = 5$, $Var(\xi)$ does not exist. For $\alpha = 5$, $E(\xi) = 5/4$, $Var(\xi) = 5/3$.

3. Write a (two-line) computer program for generating the sum of 1000 random numbers with $\gamma = -0.2$, and plot the empirical histogram of this distribution (that is, generate 1000 times the sum of 1000 such random numbers. Interpret what you observe. For your convenience, you may find a closely related program on the course website. Modify it so that it solves the problem at hand, and adapt the range in the drawing routine. Produce output and discuss it.

   $\alpha = 5$, $Var(\xi) = 5/3$. The distribution of the sum is Gaussian, as shown in figure 2, due to the central limit theorem.

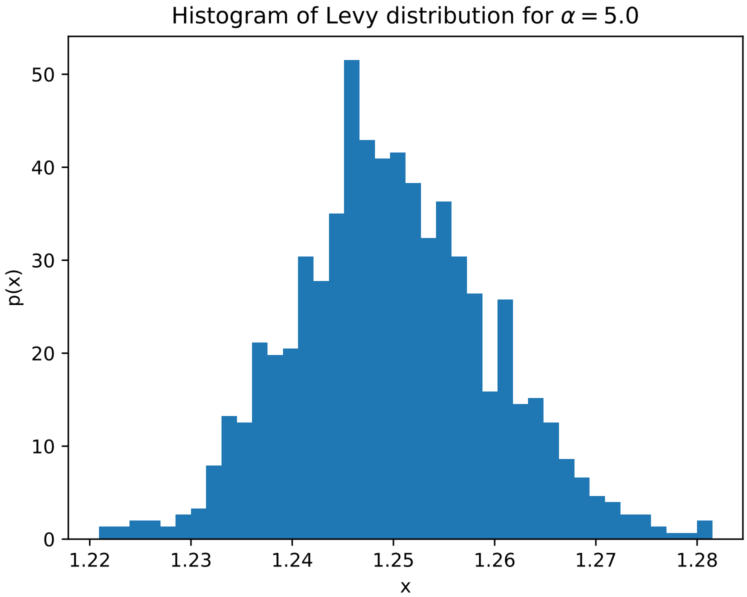

FIG. 2: $\alpha = 5$

4. Write a (two-line) computer program for generating the sum of 1000 random numbers with $\gamma = -0.8$, and plot the empirical histogram of this distribution. Interpret what you observe. For your convenience, please take the closely related program from the course website. Modify it so that it solves the problem at hand, and adapt the range in the drawing routine. Produce output and discuss it.

$\alpha = 1.25$, $Var(\xi)$ does not exist. The central limit theorem no longer applies.

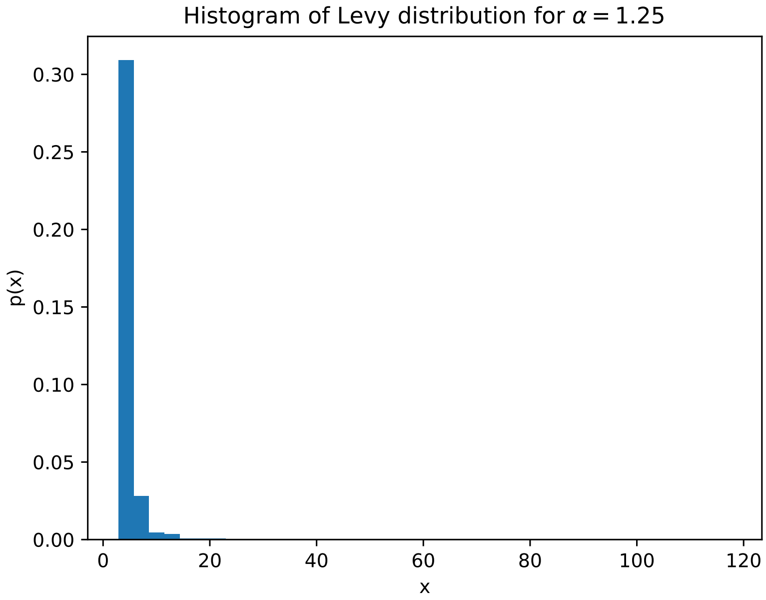

FIG. 3: $\alpha = 1.25$