

Convexity and robustness of the Rényi entropy

Filipp Buryak*, Yuliya Mishura

*Kyiv National Taras Shevchenko University, Faculty of Mechanics and Mathematics,
Department of Probability Theory, Statistics and Actuarial Mathematics,
Volodymyrska 64, 01601 Kyiv, Ukraine*

filippburyak2000@gmail.com (F. Buryak), myus@univ.kiev.ua (Yu. Mishura)

Received: 27 May 2021, Revised: 23 June 2021, Accepted: 24 June 2021,
Published online: 26 July 2021

Abstract We study convexity properties of the Rényi entropy as function of $\alpha > 0$ on finite alphabets. We also describe robustness of the Rényi entropy on finite alphabets, and it turns out that the rate of respective convergence depends on initial alphabet. We establish convergence of the disturbed entropy when the initial distribution is uniform but the number of events increases to ∞ and prove that the limit of Rényi entropy of the binomial distribution is equal to Rényi entropy of the Poisson distribution.

Keywords Discrete distribution, Rényi entropy, convexity

2010 MSC 60E05, 94A17

1 Introduction

Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be a probability space supporting all distributions considered below. For any $N \geq 1$ introduce the family of discrete distributions $p = (p_1, p_2, \dots, p_N)$ with probabilities

$$p_i \geq 0, 1 \leq i \leq N, N \geq 1, p_1 + \dots + p_N = 1.$$

*Corresponding author.

In the present paper we investigate some properties of the Rényi entropy, which was proposed by Rényi in [14],

$$\mathcal{H}_\alpha(p) = \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right), \quad \alpha > 0, \quad \alpha \neq 1,$$

including its limit value as $\alpha \rightarrow 1$, i.e., the Shannon entropy

$$\mathcal{H}(p) = - \sum_{k=1}^N p_k \log(p_k).$$

Due to this continuity, it is possible to put $\mathcal{H}_1(p) = \mathcal{H}(p)$. We consider the Rényi entropy as a functional of various parameters. The first approach is to fix the distribution and consider $\mathcal{H}_\alpha(p)$ as the function of $\alpha > 0$. Some of the properties of $\mathcal{H}_\alpha(p)$ as the function of $\alpha > 0$ are well known. In particular, it is known that $\mathcal{H}_\alpha(p)$ is continuous and nonincreasing in $\alpha \in (0, \infty)$, $\lim_{\alpha \rightarrow 0+} \mathcal{H}_\alpha(p) = \log m$, where m is the number of nonzero probabilities, and $\lim_{\alpha \rightarrow +\infty} \mathcal{H}_\alpha(p) = -\log \max_k p_k$. However, for the reader's convenience, we provide the short proofs of this and some other simple statements in the Appendix. One can see that these properties of the entropy itself and its first derivative are common for all finite distributions. Also, it is known that Rényi entropy is Schur concave as a function of a discrete distribution, that is

$$(p_i - p_j) \left(\frac{\partial \mathcal{H}_\alpha(p)}{\partial p_i} - \frac{\partial \mathcal{H}_\alpha(p)}{\partial p_j} \right) \leq 0, \quad i \neq j.$$

Some additional results such as lower bounds on the difference in the Rényi entropy for distributions defined on countable alphabets could be found in [12]. Those results usually use the Rényi divergence of order α of a distribution P from a distribution Q

$$D_\alpha(P||Q) = \frac{1}{\alpha-1} \log \left(\sum_{i=1}^N \frac{p_i^\alpha}{q_i^{\alpha-1}} \right),$$

which is very similar to the Kullback–Leibler divergence. Some most important properties of the Rényi divergences were reviewed and extended in [7]. Rényi divergences for most commonly used univariate continuous distributions could be found in [8]. The Rényi entropy and divergence is widely used in majorization theory [6, 15], statistics [4, 13], information theory [12, 7, 1] and many other fields. Boundedness of the Rényi entropy was shown in [5] for discrete log-concave distributions depending on its variance. There are other operational definitions of the Rényi entropy given in [11], which are used in practice. It is also used in analysis of financial time series. As it is stated in [16], the Rényi entropy can deal effectively with heavy-tailed distributions and reflect a short-range characteristics of financial time series. So, in some sense, entropy can be connected to the long- and short-range dependence of the selected stochastic processes and allows to compare the memory inherent to various processes. Other methods of comparing the memory of the processes are proposed, e.g., in [3] and [2]. What is more, the Rényi entropy is used in physics. For it to be

physically meaningful as thermostistical quantity it should not change drastically if the probability distribution is slightly changed. It is important that experimental uncertainty in determining the distribution function not cause entropy to diverge. In [9] it is shown that the Rényi entropy is uniformly continuous for probabilities on finite sets. In our paper we go on and find the rate of convergence. In the present paper we restrict ourselves to the standard Rényi entropy and go a step ahead in comparison of standard properties, namely, we investigate convexity of the Rényi entropy with the help of the second derivative. It turned out that from this point of view, the situation is much more interesting and uncertain in comparison with the behavior of the first derivative, and crucially depends on the distribution. One might say that all the standard guesses are wrong. Of course, the second derivative is continuous (evidently, it simply means that it is continuous at 1 because at all other points the continuity is obvious), but then the surprises begin. If the second derivative starts with a positive value at zero, it can either remain positive or have inflection points, depending on the distribution. If it starts from the negative value, it can have the first inflection point both before 1 and after 1, depending on the distribution, too (point 1 is interesting as some crucial point for entropy, so, we compare the value of inflection points with it). The value of the second derivative at zero is bounded from below but unbounded from above. Due to the over-complexity of some expressions, which defied analytical consideration, we propose several illustrations performed by numerical methods. We investigate robustness of the Rényi entropy w.r.t. the distribution, and it turns out that the rate of respective convergence depends on initial distribution, too. Further, we establish convergence of the disturbed entropy when the initial distribution is uniform but the number of events increases to ∞ , and prove that the limit of Rényi entropy of the binomial distribution is equal to entropy of the Poisson distribution. It was previously proved in [10] that the Shannon entropy of the binomial distribution is increasing to entropy of the Poisson distribution. Our proof of this particular fact is simpler because uses only Lebesgue's dominated convergence theorem. The paper is organized as follows. Section 2 is devoted to the convexity properties of the Rényi entropy, Section 3 describes robustness of the Rényi entropy, and Section A contains some auxiliary results.

2 Convexity of the Rényi entropy

To start, we consider the general properties of the 2nd derivative of the Rényi entropy.

2.1 The form and the continuity of the 2nd derivative

Let us denote $S_i(\alpha) = \sum_{k=1}^N p_k^\alpha \log^i p_k$, $i = 0, 1, 2, 3$. Denote also $f(\alpha) = \log(\sum_{k=1}^N p_k^\alpha)$. Obviously, the function $f \in C^\infty(\mathbb{R}^+)$, and its first three derivatives equal

$$f'(\alpha) = \frac{S_1(\alpha)}{S_0(\alpha)}, \quad f''(\alpha) = \frac{S_2(\alpha)S_0(\alpha) - S_1^2(\alpha)}{S_0^2(\alpha)},$$

$$f'''(\alpha) = \frac{S_3(\alpha)S_0^2(\alpha) - 3S_2(\alpha)S_1(\alpha)S_0(\alpha) + 2S_1^3(\alpha)}{S_0^3(\alpha)}.$$

In particular, if one considers the random variable ξ taking values $\log p_k$ with probability p_k , then

$$\begin{aligned} f'(1) &= E(\xi) < 0, \quad f''(1) = E(\xi^2) - (E(\xi))^2 > 0, \\ f'''(1) &= E(\xi^3) - 3E(\xi^2)E(\xi) + 2(E(\xi))^3, \end{aligned} \tag{1}$$

and the sign of $f'''(1)$ is not fixed (as we can see below, it can be both + and -).

Lemma 1. *Let $p_k \neq 0$ for all $1 \leq k \leq N$. Then*

(i) *The 2nd derivative $\mathcal{H}_\alpha''(p)$ equals*

$$\mathcal{H}_\alpha''(p) = -\frac{1}{(1-\alpha)^3} \left(\sum_{k=1}^N ((1-\alpha)q'_k(\alpha) + 2q_k(\alpha)) \log \frac{q_k(\alpha)}{p_k} \right), \tag{2}$$

where

$$q_k(\alpha) = \frac{p_k^\alpha}{\sum_{k=1}^N p_k^\alpha}.$$

(b) *The 2nd derivative $\mathcal{H}_\alpha''(p)$ can be also presented as*

$$\mathcal{H}_\alpha''(p) = -\frac{1}{3} f'''(\theta) \tag{3}$$

for some $0 < \theta < \alpha$.

(ii) *The 2nd derivative $\mathcal{H}_\alpha''(p)$ is continuous on \mathbb{R}^+ if we put*

$$\mathcal{H}_1''(p) = -\frac{1}{3} f'''(1) = -\frac{1}{3} (E(\xi^3) - 3E(\xi^2)E(\xi) + 2(E(\xi))^3).$$

Proof. Equality (2) is a result of direct calculations. Concerning equality (3), we can present $\mathcal{H}_\alpha(p)$ as

$$\mathcal{H}_\alpha(p) = \frac{f(\alpha) - f(1)}{1-\alpha},$$

therefore, $-\mathcal{H}_\alpha(p)$ is a slope function for f . Taking successive derivatives, we get from the standard Taylor formula that

$$\mathcal{H}'_\alpha(p) = \frac{f'(\alpha)(1-\alpha) + f(\alpha)}{(1-\alpha)^2} = -\frac{1}{2} f''(\eta),$$

and

$$\mathcal{H}''_\alpha(p) = \frac{f''(\alpha)(1-\alpha)^2 + 2f'(\alpha)}{(1-\alpha) + 2f(\alpha)} = -\frac{1}{3} f'''(\theta),$$

where $\eta, \theta \in (1, \alpha)$. If $\alpha \rightarrow 1$, then both η and θ tend to 1. Taking into account (1), we immediately get both equality (3) and statement (ii). □

2.2 Behavior of the 2nd derivative at the origin

Let us consider the starting point for the 2nd derivative, i.e., the behavior of $\mathcal{H}_\alpha''(p)$ at zero as a function of a distribution vector p . Analyzing (2), we see that $\mathcal{H}_\alpha''(p)$ as function of α is continuous in 0. Moreover,

$$q_k(0) = 1/N, \quad q'_k(0) = \frac{\log p_k}{N} - \frac{\sum_{k=1}^N \log p_k}{N^2},$$

so we can present $\mathcal{H}_\alpha''(p)$ as

$$\begin{aligned} \mathcal{H}_0''(p) &= - \sum_{k=1}^N \left(\frac{1}{N} \log p_k - \frac{1}{N^2} \sum_{i=1}^N \log p_i + \frac{2}{N} \right) \log \frac{1}{N p_k} \\ &= \sum_{k=1}^N \left(\frac{1}{N} \log p_k - \frac{1}{N^2} \sum_{i=1}^N \log p_i + \frac{2}{N} \right) (\log N + \log p_k) \\ &= 2 \log N + \frac{1}{N} \sum_{k=1}^N (\log p_k)^2 - \frac{1}{N^2} \left(\sum_{k=1}^N \log p_k \right)^2 + \frac{2}{N} \sum_{k=1}^N \log p_k. \end{aligned}$$

Now we are interested in the sign of $\mathcal{H}_\alpha''(p)$. It is very simple to give an example of a distribution for which $\mathcal{H}_\alpha''(p) > 0$. One of such examples is given in Figure 1. Negative $\mathcal{H}_\alpha''(p)$ is also possible, however, at this moment we prefer to start with a more general result.

Lemma 2. *If some probability vector p is a point of local extremum of $\mathcal{H}_\alpha''(p)$ then either $p = p(\text{uniform}) = (\frac{1}{N}, \dots, \frac{1}{N})$ or it contains two different probabilities.*

Proof. Let us formulate the necessary conditions for $\mathcal{H}_0''(p)$ to have a local extremum at some point. Taking into account the limitation $\sum_{k=1}^N p_k = 1$, these conditions have the form

$$\begin{cases} 2 \log N + \frac{1}{N} \sum_{k=1}^N (\log p_k)^2 - \frac{1}{N^2} \left(\sum_{k=1}^N \log p_k \right)^2 + \frac{2}{N} \sum_{k=1}^N \log p_k \longrightarrow \text{extr} \\ \sum_{k=1}^N p_k = 1. \end{cases}$$

We write the Lagrangian function

$$\begin{aligned} L &= \lambda_0 \left(2 \log N + \frac{1}{N} \sum_{k=1}^N (\log p_k)^2 - \frac{1}{N^2} \left(\sum_{k=1}^N \log p_k \right)^2 + \frac{2}{N} \sum_{k=1}^N \log p_k \right) \\ &\quad + \lambda \left(\sum_{k=1}^N p_k - 1 \right). \end{aligned}$$

If some p is an extremum point then there exist λ_0 and λ such that $\lambda_0^2 + \lambda^2 \neq 0$ and $\frac{\partial L}{\partial p_i}(p) = 0$ for all $1 \leq i \leq N$, i.e.,

$$\frac{\partial L}{\partial p_i} = \lambda_0 \left(\frac{2}{N p_i} \log p_i - \frac{2}{N^2 p_i} \left(\sum_{k=1}^N \log p_k \right) + \frac{2}{N p_i} \right) + \lambda = 0.$$

If $\lambda_0 = 0$ then $\lambda = 0$. However, $\lambda_0^2 + \lambda^2 \neq 0$, therefore we can put $\lambda_0 = 1$. Then

$$-\lambda p_i = \frac{2}{N} \log p_i - \frac{2}{N^2} \left(\sum_{k=1}^N \log p_k \right) + \frac{2}{N}.$$

Taking the sum of these equalities we get that $\lambda = -2$, whence

$$p_i - \frac{1}{N} \log p_i = \frac{1}{N} - \frac{1}{N^2} \left(\sum_{k=1}^N \log p_k \right). \tag{4}$$

So, if the distribution vector p is an extremum point then $p_1 - \frac{1}{N} \log p_1 = \dots = p_N - \frac{1}{N} \log p_N$. Let us take a look at the continuous function $f(x) = x - \frac{1}{N} \log x$, $x \in (0, 1)$. Its derivative equals

$$f'(x) = 1 - \frac{1}{Nx} = 0 \Leftrightarrow x = \frac{1}{N}, \quad \text{sign}(f'(x)) = \text{sign}\left(x - \frac{1}{N}\right),$$

$$\lim_{x \rightarrow 0+} f(x) = +\infty, \quad \lim_{x \rightarrow +1} f(x) = 1.$$

So, $f(x)$ has its global minimum at point $x = \frac{1}{N}$, and for any $f(\frac{1}{N}) < y \leq 1$ there exist two points, $x' \neq x''$, $x', x'' \in (0, 1)$ such that $f(x') = f(x'') = y$. Thus, if $\mathcal{H}_0''(p)$ achieves local extremum at vector p , then it contains no more than two different probabilities. Obviously, it can be $p = p(\text{uniform}) = (\frac{1}{N}, \dots, \frac{1}{N})$. \square

Remark 1. Note that $\mathcal{H}_0''(p(\text{uniform})) = 0$. Therefore, in order to find the distribution for which $\mathcal{H}_0''(p) < 0$, let us consider the distribution vector that contains only two different probabilities p_0, q_0 such that:

$$\begin{cases} p_0 - q_0 = \frac{1}{N} (\log p_0 - \log q_0), \\ kp_0 + (N - k)q_0 = 1, \end{cases} \tag{5}$$

where $N, k \in \mathbb{N}$, $N > k$ and $p_0, q_0 \in (0, 1)$.

Lemma 3. *Let p be the distribution vector satisfying (5). Then $\mathcal{H}_0''(p) < 0$.*

Proof. First, we will show that $\mathcal{H}_0''(p)$ is nonpositive. For that we rewrite $\mathcal{H}_0''(p)$ in terms of p_0 and q_0 :

$$\begin{aligned} \mathcal{H}_0''(p) &= 2 \log N + \frac{1}{N} \left(k(\log p_0)^2 + (N - k)(\log q_0)^2 \right) \\ &\quad - \frac{1}{N^2} (k \log p_0 + (N - k) \log q_0)^2 + \frac{2}{N} (k \log p_0 + (N - k) \log q_0) \\ &= 2 \log N + \frac{k(N - k)}{N^2} \left((\log p_0)^2 - 2 \log p_0 \log q_0 + (\log q_0)^2 \right) \\ &\quad + \frac{2k}{N} (\log p_0 - \log q_0) + 2 \log q_0 \\ &= 2 \log Nq_0 + k(N - k)(p_0 - q_0)^2 + 2k(p_0 - q_0). \end{aligned}$$

We know that $kp_0 + (N - k)q_0 = 1$, whence $k = \frac{Nq_0 - 1}{q_0 - p_0}$, and $N - k = \frac{1 - Np_0}{q_0 - p_0}$. Then

$$\begin{aligned} \mathcal{H}_0''(p) &= 2 \log Nq_0 + (1 - Nq_0)(Np_0 - 1) + 2(1 - Nq_0) \\ &= 2 \log Nq_0 + N(p_0 - q_0) + 1 - N^2 p_0 q_0 \\ &= \log(Nq_0)^2 + \log \frac{p_0}{q_0} + 1 - N^2 p_0 q_0 = \log N^2 p_0 q_0 - N^2 p_0 q_0 + 1. \end{aligned}$$

Note that $\log x - x + 1 < 0$ for $x > 0, x \neq 1$. We want to show that under conditions (5) $N^2 p_0 q_0$ cannot be equal to 1. Suppose that $N^2 p_0 q_0 = 1$. Then it follows from (5) that

$$\frac{k}{N^2} + (N - k)q_0^2 = q_0.$$

It means that q_0 and p_0 are algebraic numbers. Thus, their difference $p_0 - q_0$ is also algebraic. On the other hand, by the Lindemann–Weierstrass theorem $\frac{1}{N}(\log p_0 - \log q_0)$ is transcendental number, which contradicts (5). So $N^2 p_0 q_0 \neq 1$ and $\mathcal{H}_0''(p) < 0$. □

Theorem 1. *For any $n > 2$ there exists $N \geq n$ and a probability vector $p = (p_1, \dots, p_N)$ such that $\mathcal{H}_0''(p) < 0$.*

Proof. Consider a distribution vector p that satisfies conditions (5). From Lemma 3 we know that $\mathcal{H}_0''(p) < 0$. Now we want to show that there exist an arbitrarily large $N \in \mathbb{N}$ and a distribution vector p of length N that satisfy those conditions. For that we denote

$$x = Np_0, \quad y = Nq_0, \quad r = \frac{k}{N} = \frac{y - 1}{y - x}.$$

Then $0 < x < 1 < y$ and $r < 1$ and $x - y = \log x - \log y$. The function $x - \log x$ is decreasing on $(0, 1)$, is increasing on $(1, +\infty)$ and is equal to 1 at point 1. Let $y = y(x)$ be the implicit function defined by $x - y = \log x - \log y$. By that we get the 1-to-1 correspondence from $x \in (0, 1)$ to $y \in (1, +\infty)$. We also have function the $r(x) = \frac{y(x) - 1}{y(x) - x}$. If we find $x' \in (0, 1)$ such that $r' = r(x')$ is rational then we could pick $N, k \in \mathbb{N}$ such that $r' = \frac{k}{N}$ and get q distribution vector p satisfying (5) with $p_0 = \frac{x}{N}, q_0 = \frac{y}{N}$. However, we will not look for such x' , but will just show that they exist. To do that, observe that $y(x)$ is a continuous function of x and so is the function $r(x) = \frac{y(x) - 1}{y(x) - x}$. What is more,

$$y(x) \rightarrow +\infty, \quad x \rightarrow 0+ \text{ so } r(x) \rightarrow 1, \quad x \rightarrow 0+.$$

Let us fix $x_0 \in (0, 1), r(x_0) < 1$. Then for any $r' \in (r(x_0), 1)$ there exists $x' \in (0, x_0)$ such that $r(x') = r'$. By taking $r' \in \mathbb{Q}$ we get that there exists x' such that $\frac{k}{N} < 1$ and is rational. Finally, we want to show that N can be arbitrarily large. For that simply observe that $\frac{k}{N} = r'$ so as $r' \rightarrow 1-$ we get that $N \rightarrow +\infty$. □

Lemma 4. *Let N be fixed. Then $\mathcal{H}_0''(p)$ as the function of vector p is bounded from below and is unbounded from above.*

Proof. Recall that $\mathcal{H}_0''(p) = 0$ on the uniform distribution and exclude this case from the further consideration. In order to simplify the notations, we denote $x_k = \log p_k$, and let

$$S_N := N(\mathcal{H}_0''(p) - 2 \log N) = \sum_{k=1}^N (x_k)^2 - \frac{1}{N} \left(\sum_{k=1}^N x_k \right)^2 + 2 \sum_{k=1}^N x_k.$$

Note that there exists $n \leq N - 1$ such that

$$x_1 < \log \frac{1}{N}, \dots, x_n < \log \frac{1}{N}, \quad x_{n+1} \geq \log \frac{1}{N}, \dots, x_N \geq \log \frac{1}{N}.$$

Further, denote the rectangle $A = [\log \frac{1}{N}; 0]^{N-n} \subset \mathbb{R}^{N-n}$, and let

$$S_{N,1} = \sum_{k=1}^n x_k, \quad S_{N,2} = \sum_{k=n+1}^N x_k.$$

Let us establish that $\mathcal{H}_0''(p)$ is bounded from below. In this connection, rewrite S_N as

$$S_N = \sum_{k=1}^n x_k^2 + \sum_{k=n+1}^N x_k^2 - \frac{1}{N} \left((S_{N,1})^2 + 2S_{N,1}S_{N,2} + (S_{N,2})^2 \right) + 2S_{N,1} + 2S_{N,2}.$$

By the Cauchy–Schwarz inequality we have

$$\left(\sum_{k=1}^n x_k \right)^2 \leq n \sum_{k=1}^n x_k^2, \quad \left(\sum_{k=n+1}^N x_k \right)^2 \leq (N - n) \sum_{k=n+1}^N x_k^2.$$

Therefore

$$\begin{aligned} S_N &\geq \left(1 - \frac{n}{N}\right) \sum_{k=1}^n x_k^2 + \frac{n}{N} \sum_{k=n+1}^N x_k^2 - \frac{2}{N} S_{N,1} S_{N,2} + 2S_{N,1} + 2S_{N,2} \\ &= \sum_{k=1}^n \left(\left(1 - \frac{n}{N}\right) x_k^2 + x_k \left(2 - \frac{2}{N} S_{N,2}\right) \right) + \frac{n}{N} \sum_{k=n+1}^N x_k^2 + 2S_{N,2} \\ &= \frac{1}{N} \sum_{k=1}^n \left((N - n) x_k^2 + 2x_k (N - S_{N,2}) \right) + \frac{n}{N} \sum_{k=n+1}^N x_k^2 + 2S_{N,2}. \end{aligned}$$

There exists $M > 0$ such that for every $n \leq N - 1$ we have $|S_{N,2}| \leq M$ because A is compact and $S_{N,2}$ is continuous on A . Obviously, $\frac{n}{N} \sum_{k=n+1}^N x_k^2 \geq 0$. Finally, for every $1 \leq k \leq n$ we have that $(N - n) x_k^2 + 2x_k (N - S_{N,2})$ is bounded from below by the value $-\frac{(N - S_{N,2})^2}{N - n} \geq -N^2 - M^2$. Resuming, we get that S_N is bounded from below, and consequently $\mathcal{H}_0''(p)$ is bounded from below for fixed N .

Now we want to establish that $\mathcal{H}_0''(p)$ is not bounded from above. In this connection, let $\varepsilon > 0$, and let us consider the distribution of the form $p_1 = \varepsilon, p_2 = \dots = p_N = \frac{1-\varepsilon}{N-1}$. Then we have

$$\begin{aligned} \mathcal{H}_0''(p) &= 2 \log N + \frac{1}{N} \sum_{k=1}^N (\log p_k)^2 - \frac{1}{N^2} \left(\sum_{k=1}^N \log p_k \right)^2 + \frac{2}{N} \sum_{k=1}^N \log p_k \\ &= 2 \log N + \frac{N-1}{N} \left(\log \frac{1-\varepsilon}{N-1} \right)^2 + \frac{1}{N} (\log \varepsilon)^2 \\ &\quad - \frac{1}{N^2} \left((N-1) \log \frac{1-\varepsilon}{N-1} + \log \varepsilon \right)^2 + \frac{2(N-1)}{N} \log \frac{1-\varepsilon}{N-1} + \frac{2}{N} \log \varepsilon \\ &= \left(\frac{1}{N} - \frac{1}{N^2} \right) (\log \varepsilon)^2 + \left(\frac{2}{N} - \frac{2(N-1)}{N^2} \log \frac{1-\varepsilon}{N-1} \right) \log \varepsilon + 2 \log N \\ &\quad + \left(\frac{N-1}{N} - \frac{(N-1)^2}{N^2} \right) \left(\log \frac{1-\varepsilon}{N-1} \right)^2 \\ &\quad + \frac{2(N-1)}{N} \log \frac{1-\varepsilon}{N-1} \rightarrow +\infty, \varepsilon \rightarrow 0+. \end{aligned}$$

□

2.3 Superposition of entropy that is convex

Now we establish that the superposition of entropy with some decreasing function is convex. Namely, we shall consider the function

$$\mathcal{G}_\beta(p) = -\mathcal{H}_{1+\frac{1}{\beta}}(p) = \beta \log \left(\sum_{k=1}^N p_k^{1+1/\beta} \right), \beta > 0, \tag{6}$$

and prove its convexity. Because now we consider the tools that do not include differentiation, we can assume that some probabilities are zero. In order to prove the convexity, we start with the following simple and known result whose proof is added for the reader's convenience.

Lemma 5. *For any measure space $(\mathcal{X}, \Sigma, \mu)$ and any measurable $f \in L^p(\mathcal{X}, \Sigma, \mu)$ for some interval $p \in [a, b]$, $\|f\|_p = \|f\|_{L^p(\mathcal{X}, \Sigma, \mu)}$ is log-convex as a function of $1/p$ on this interval.*

Proof. For any $p_2, p_1 > 0$ and $\theta \in (0, 1)$, denote $p = (\theta/p_1 + (1-\theta)/p_2)^{-1}$ and observe that

$$\theta p/p_1 + (1-\theta)p/p_2 = 1.$$

Therefore, by the Hölder inequality

$$\begin{aligned} \|f\|_p^p &= \int_{\mathcal{X}} |f(x)|^{\theta p} \cdot |f(x)|^{(1-\theta)p} \mu(dx) \\ &\leq \left(\int_{\mathcal{X}} |f(x)|^{p_1} \mu(dx) \right)^{\theta p/p_1} \left(\int_{\mathcal{X}} |f(x)|^{p_2} \mu(dx) \right)^{(1-\theta)p/p_2}, \end{aligned}$$

whence

$$\log \|f\|_p \leq \theta \log \|f\|_{p_1} + (1 - \theta) \log \|f\|_{p_2},$$

as required. □

Corollary 1. For any probability vector $p = (p_k, 1 \leq k \leq N)$, the function $\mathcal{G}_\beta(p)$, $\beta > 0$, is convex.

Proof. It follows from Lemma 5 by setting $\mathcal{X} = \{1, \dots, N\}$, $\mu(A) = \sum_{k \in A} p_k$, $f(k) = p_k, k \in \mathcal{X}$. □

Remark 2. It follows immediately from (6) that for the function

$$\mathcal{G}_\beta(p) = \beta \log \sum_{k=1}^N p_k^{1+1/\beta}, \quad \beta > 0,$$

$\mathcal{H}_\alpha(p) = \mathcal{G}_{\frac{1}{\alpha-1}}(p)$. For $\alpha > 1$, $\frac{1}{\alpha-1}$ is convex. If there be such p that $G.(p)$ be nondecreasing on an interval, then $\mathcal{G}_{\frac{1}{\alpha-1}}(p)$ be convex on that interval and $\mathcal{H}_\alpha(p)$ be convex, too. However,

$$\begin{aligned} \mathcal{G}'_\beta(p) &= \log \sum_{k=1}^N p_k^{1+1/\beta} - \frac{1}{\beta} \frac{\sum_{k=1}^N p_k^{1+1/\beta} \log p_k}{\sum_{k=1}^N p_k^{1+1/\beta}} \\ &= - \sum_{k=1}^N \frac{p_k^{1+1/\beta}}{\sum_{k=1}^N p_k^{1+1/\beta}} \log \frac{p_k^{1/\beta}}{\sum_{k=1}^N p_k^{1+1/\beta}} \leq 0. \end{aligned}$$

In some sense, this is a reason why we cannot say something definite concerning the 2nd derivative of entropy either on the whole semiaxes or even in the interval $[1, +\infty)$.

2.4 Graphs of $\mathcal{H}_\alpha(p)$ and its second derivative for several probability distributions

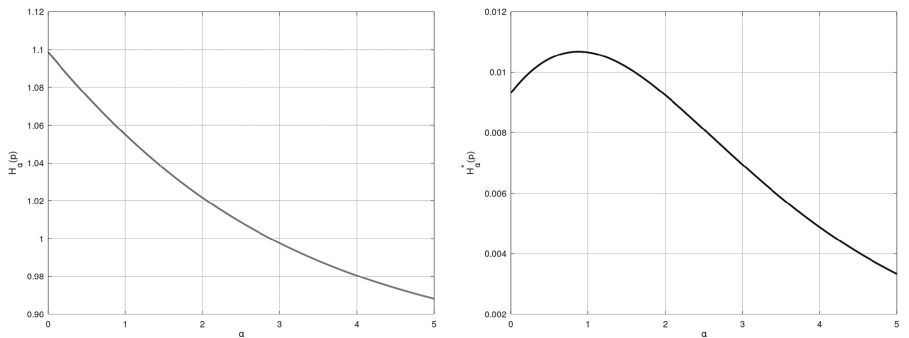


Fig. 1. Graphs of $\mathcal{H}_\alpha(p)$ and $\mathcal{H}''_\alpha(p)$, where $p_1 = p_2 = 0.4, p_3 = 0.2$. Here $\mathcal{H}_\alpha(p)$ is convex as a function of $\alpha > 0$

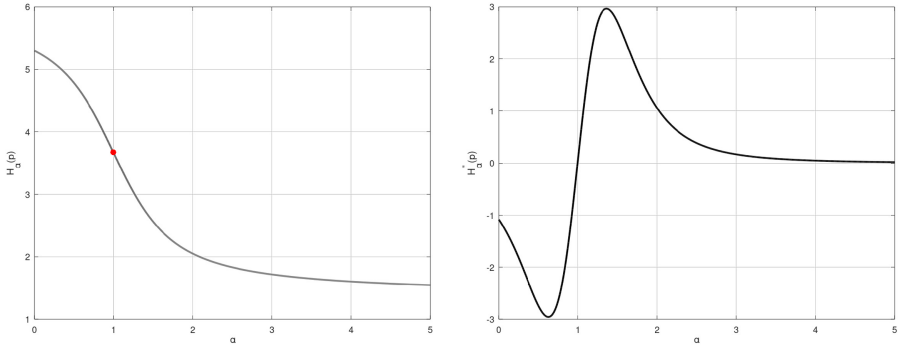


Fig. 2. Graphs of $\mathcal{H}_\alpha(p)$ and $\mathcal{H}''_\alpha(p)$, where $p_1 = \dots = p_{198} = \frac{1}{400}$, $p_{199} = p_{200} = \frac{101}{400}$. Dot is the point where $\mathcal{H}''_\alpha(p) = 0$ and this point is $\alpha = 0.99422$

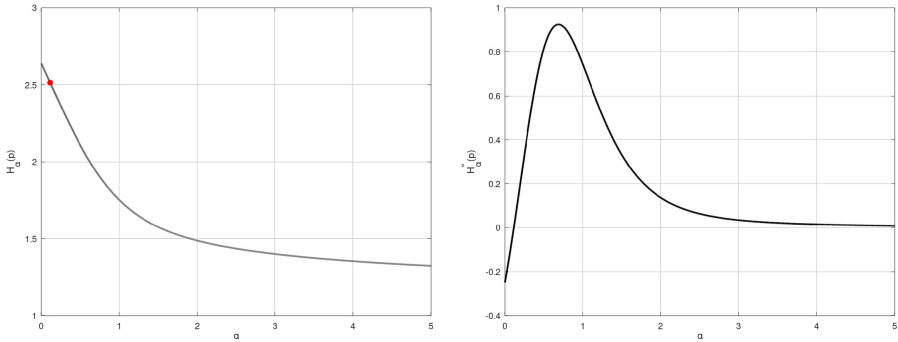


Fig. 3. Graphs of $\mathcal{H}_\alpha(p)$ and $\mathcal{H}''_\alpha(p)$, where $p_1 = \dots = p_{10} = 0.01$, $p_{11} = p_{12} = 0.15$, $p_{13} = p_{14} = 0.3$. Here the second derivative becomes positive long before point 1 (at point 0.11318)

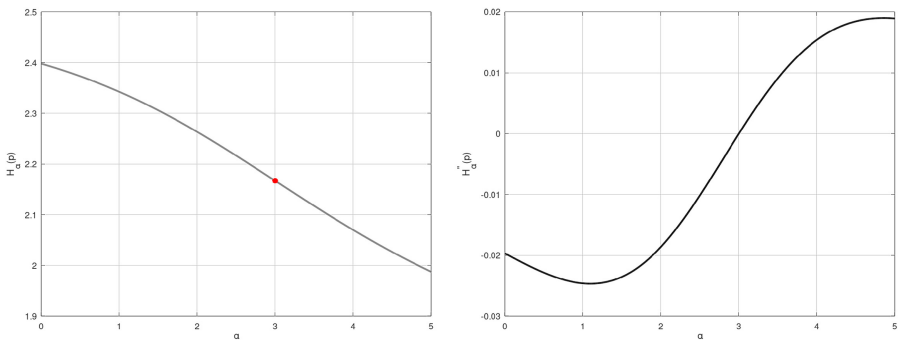


Fig. 4. Graphs of $\mathcal{H}_\alpha(p)$ and $\mathcal{H}''_\alpha(p)$, where $p_1 = \dots = p_{10} = 0.08$, $p_{11} = 0.2$. Here the second derivative becomes positive after point 1 (at point 2.9997)

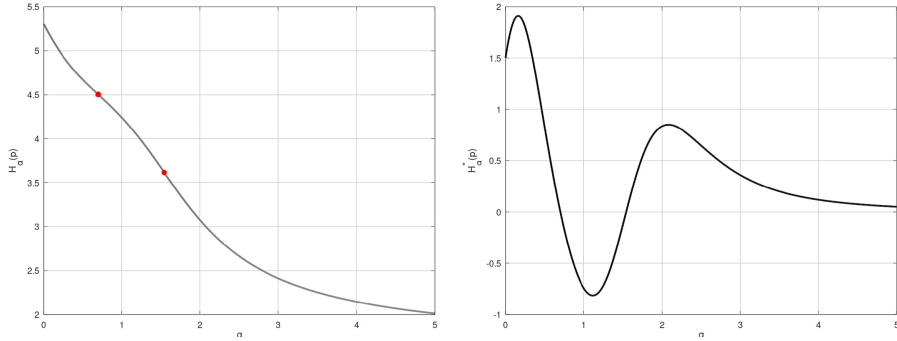


Fig. 5. Graphs of $\mathcal{H}_\alpha(p)$ and $\mathcal{H}_\alpha''(p)$, where $p_1 = \dots = p_{100} = 0.0001$, $p_{101} = \dots = p_{200} = 0.0079$, $p_{201} = 0.2$. Here the second derivative has two zeros

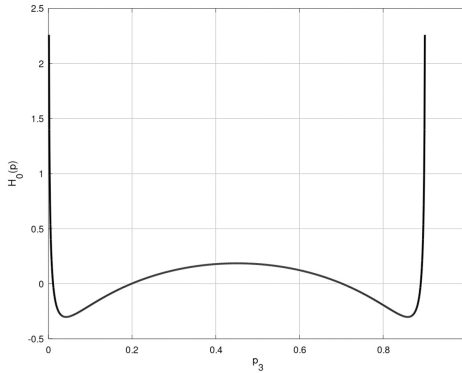


Fig. 6. Graph of $\mathcal{H}_\alpha(p)$, where $p_1 = p_2 = 0.05$, p_3 is changing from 0 to 0.9 and $p_4 = 1 - p_1 - p_2 - p_3$

3 Robustness of the Rényi entropy

Now we study the asymptotic behavior of the Rényi entropy depending on the behavior of the involved probabilities. The first problem is the stability of the entropy w.r.t. involved probabilities and the rate of its convergence to the limit value when probabilities tend to their limit value with the fixed rate.

3.1 Rate of convergence of the disturbed entropy when the initial distribution is arbitrary but fixed

Let us look at distributions that are “near” some fixed distribution $p = (p_k, 1 \leq k \leq N)$ and construct the approximate distribution $p(\epsilon) = (p_k(\epsilon), 1 \leq k \leq N)$ as follows. We can assume that some probabilities are zero, and we shall see that this assumption influences the rate of convergence of the Rényi entropy to the limit value. So, let $0 \leq N_1 < N$ be a number of zero probabilities, and for them we consider approximate values of the form $p_k(\epsilon) = c_k \epsilon$, $0 \leq c_k \leq 1$, $1 \leq k \leq N_1$. Further, let $N_2 = N - N_1 \geq 1$ be a number of nonzero probabilities, and for them we consider

approximate values of the form $p_k(\epsilon) = p_k + c_k\epsilon$, $|c_k| \leq 1$, $N_1 + 1 \leq k \leq N$, where $c_1 + \dots + c_N = 0$ and $\epsilon \leq \min_{N_1+1 \leq k \leq N} p_k$. Assume also that there exists $k \leq N$ such that $c_k \neq 0$, otherwise $\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) = 0$. So, we disturb the initial probabilities linearly in ϵ with different weights whose sum should necessarily be zero. These assumptions imply that $0 \leq p_k(\epsilon) \leq 1$ and $p_1(\epsilon) + \dots + p_N(\epsilon) = 1$. Now we want to find out how entropy of the disturbed distribution will differ from the initial entropy, depending on parameters ϵ , N and α . We start with $\alpha = 1$.

Theorem 2. *Let the number N and coefficients c_1, \dots, c_N be fixed, and let $\alpha = 1$. We have three different situations:*

(i) *Let $N_1 \geq 1$ and there exists $k \leq N_1$ such that $c_k \neq 0$. Then*

$$\mathcal{H}_1(p) - \mathcal{H}_1(p(\epsilon)) \sim \epsilon \log \epsilon \sum_{k=1}^{N_1} c_k, \quad \epsilon \rightarrow 0.$$

(ii) *Let for all $k \leq N_1$ $c_k = 0$ and $\sum_{k=N_1+1}^N c_k \log p_k \neq 0$. Then*

$$\mathcal{H}_1(p) - \mathcal{H}_1(p(\epsilon)) \sim \epsilon \sum_{k=N_1+1}^N c_k \log p_k, \quad \epsilon \rightarrow 0.$$

(iii) *Let for all $k \leq N_1$ $c_k = 0$ and $\sum_{k=N_1+1}^N c_k \log p_k = 0$. Then*

$$\mathcal{H}_1(p) - \mathcal{H}_1(p(\epsilon)) \sim \frac{\epsilon^2}{2} \sum_{k=N_1+1}^N \frac{c_k^2}{p_k}, \quad \epsilon \rightarrow 0.$$

Proof. First of all, we will find the asymptotic behavior of two auxiliary functions as $\epsilon \rightarrow 0$. First, let $0 \leq c_k \leq 1$. Then

$$c_k \epsilon \log(c_k \epsilon) = c_k \epsilon \log \epsilon + c_k \epsilon \log c_k = c_k \epsilon \log \epsilon + o(\epsilon \log \epsilon), \quad \epsilon \rightarrow 0.$$

Second, let $p_k > 0$, $|c_k| \leq 1$. Taking into account the Taylor expansion of logarithm

$$\log(1 + x) = x - \frac{x^2}{2} + o(x^2), \quad x \rightarrow 0,$$

we can write:

$$\begin{aligned} (p_k + c_k \epsilon) \log(p_k + c_k \epsilon) - p_k \log p_k &= c_k \epsilon \log p_k + (p_k + c_k \epsilon) \log(1 + c_k p_k^{-1} \epsilon) \\ &= c_k \epsilon \log p_k + (p_k + c_k \epsilon) \left(c_k p_k^{-1} \epsilon - \frac{1}{2} (c_k p_k^{-1} \epsilon)^2 + o(\epsilon^2) \right) \\ &= \epsilon (c_k \log p_k + c_k) + \epsilon^2 \left(c_k^2 p_k^{-1} - \frac{1}{2} c_k^2 p_k^{-1} \right) + o(\epsilon^2) \\ &= \epsilon (c_k \log p_k + c_k) + \frac{c_k^2 \epsilon^2}{2 p_k} + o(\epsilon^2), \quad \epsilon \rightarrow 0. \end{aligned} \tag{7}$$

In particular, we immediately get from (3.1) that

$$(p_k + c_k \varepsilon) \log(p_k + c_k \varepsilon) - p_k \log p_k = o(\varepsilon \log \varepsilon), \quad \varepsilon \rightarrow 0,$$

and

$$(p_k + c_k \varepsilon) \log(p_k + c_k \varepsilon) - p_k \log p_k = \varepsilon(c_k \log p_k + c_k) + o(\varepsilon), \quad \varepsilon \rightarrow 0$$

Now simply observe the following.

$$\begin{aligned} (i) \quad & \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_1(p) - \mathcal{H}_1(p(\varepsilon))}{\varepsilon \log \varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon \log \varepsilon} \sum_{k=1}^{N_1} c_k \varepsilon \log c_k \varepsilon \\ & \quad + \frac{1}{\varepsilon \log \varepsilon} \sum_{k=N_1+1}^N ((p_k + c_k \varepsilon) \log(p_k + c_k \varepsilon) - p_k \log p_k) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon \log \varepsilon} \left(\sum_{k=1}^{N_1} (c_k \varepsilon \log \varepsilon + o(\varepsilon \log \varepsilon)) + \sum_{k=N_1+1}^N o(\varepsilon \log \varepsilon) \right) \\ &= \sum_{k=1}^{N_1} c_k. \end{aligned}$$

(ii) Since for any $k \leq N_1$ we have that $c_k = 0$ and the total sum $c_1 + \dots + c_N = 0$ then $c_{N_1+1} + \dots + c_N = 0$. Furthermore, in this case

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_1(p) - \mathcal{H}_1(p(\varepsilon))}{\varepsilon} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{k=N_1+1}^N ((p_k + c_k \varepsilon) \log(p_k + c_k \varepsilon) - p_k \log p_k) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sum_{k=N_1+1}^N (\varepsilon(c_k \log p_k + c_k) + o(\varepsilon)) \\ &= \sum_{k=N_1+1}^N (c_k \log p_k + c_k) = \sum_{k=N_1+1}^N c_k \log p_k. \end{aligned}$$

(iii) In this case we have the following relations:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_1(p) - \mathcal{H}_1(p(\varepsilon))}{\varepsilon^2} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \sum_{k=N_1+1}^N ((p_k + c_k \varepsilon) \log(p_k + c_k \varepsilon) - p_k \log p_k) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \sum_{k=N_1+1}^N (\varepsilon(c_k \log p_k + c_k) + \frac{c_k^2 \varepsilon^2}{2p_k} + o(\varepsilon^2)) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \sum_{k=N_1+1}^N \left(\frac{c_k^2 \varepsilon^2}{2p_k} + o(\varepsilon^2) \right) = \frac{1}{2} \sum_{k=N_1+1}^N \frac{c_k^2}{p_k}. \end{aligned}$$

Theorem is proved. □

Now we proceed with $\alpha < 1$.

Theorem 3. *Let the number N and coefficients c_1, \dots, c_N be fixed, and let $\alpha < 1$. Then we have three different situations:*

(i) *Let $N_1 \geq 1$ and there exists $k \leq N_1$ such that $c_k \neq 0$. Then*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\epsilon^\alpha}{\alpha - 1} \left(\sum_{k=1}^{N_1} c_k^\alpha \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(ii) *Let for all $k \leq N_1$ $c_k = 0$ and $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \neq 0$. Then*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\alpha \epsilon}{\alpha - 1} \left(\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(iii) *Let for all $k \leq N_1$ $c_k = 0$ and $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$. Then*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\alpha \epsilon^2}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

Proof. Similarly to proof of Theorem 2, we start with several asymptotic relations as $\epsilon \rightarrow 0$. Namely, let $p_k > 0, |c_k| \leq 1$. Taking into account the Taylor expansion of $(1 + x)^\alpha$ that has the form

$$(1 + x)^\alpha = 1 + \alpha x + o(x), \quad x \rightarrow 0,$$

we can write:

$$\begin{aligned} \alpha c_k (p_k + c_k \epsilon)^{\alpha-1} &= \alpha c_k p_k^{\alpha-1} (1 + c_k p_k^{-1} \epsilon)^{\alpha-1} \\ &= \alpha c_k p_k^{\alpha-1} (1 + (\alpha - 1) c_k p_k^{-1} \epsilon + o(\epsilon)) \\ &= \alpha c_k p_k^{\alpha-1} + \alpha(\alpha - 1) c_k^2 p_k^{\alpha-2} \epsilon + o(\epsilon), \quad \epsilon \rightarrow 0. \end{aligned} \tag{8}$$

As a consequence, we get the following asymptotic relations:

$$\alpha c_k (p_k + c_k \epsilon)^{\alpha-1} = o(\epsilon^{\alpha-1}), \quad \epsilon \rightarrow 0, \tag{9}$$

and

$$\alpha c_k (p_k + c_k \epsilon)^{\alpha-1} = \alpha c_k p_k^{\alpha-1} + o(1), \quad \epsilon \rightarrow 0$$

(i) Applying L'Hospital's rule, we get:

$$\lim_{\epsilon \rightarrow 0} \frac{H_\alpha(p) - H_\alpha(p(\epsilon))}{\epsilon^\alpha} = \lim_{\epsilon \rightarrow 0} \frac{1}{(\alpha - 1)\epsilon^\alpha} \log \left(\sum_{k=1}^{N_1} (c_k \epsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \epsilon)^\alpha \right)$$

$$\begin{aligned}
& - \frac{1}{(\alpha - 1)\varepsilon^\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{\alpha \varepsilon^{\alpha-1}} \\
& \times \frac{\sum_{k=1}^{N_1} \alpha c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=1}^{N_1} c_k^\alpha + \varepsilon^{1-\alpha} \sum_{k=N_1+1}^N o(\varepsilon^{\alpha-1})}{\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{1}{\alpha - 1} \left(\sum_{k=1}^{N_1} c_k^\alpha \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}.
\end{aligned}$$

(ii) In this case we can transform the value under a limit as follows:

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon} \\
& = \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon} \left(\log \left(\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) - \log \left(\sum_{k=1}^N p_k^\alpha \right) \right) \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=N_1+1}^N (\alpha c_k p_k^{\alpha-1} + o(1))}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{\alpha}{\alpha - 1} \left(\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}.
\end{aligned}$$

(iii) Finally, in the 3rd case,

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon^2} \\
& = \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon^2} \left(\log \left(\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) - \log \left(\sum_{k=1}^N p_k^\alpha \right) \right) \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \frac{\sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \frac{\sum_{k=N_1+1}^N (\alpha c_k p_k^{\alpha-1} + \alpha(\alpha - 1)c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
& = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \frac{\sum_{k=N_1+1}^N (\alpha(\alpha - 1)c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha}
\end{aligned}$$

$$= \frac{\alpha}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}.$$

Theorem is proved. □

Now we conclude with $\alpha > 1$. In this case, five different asymptotics are possible.

Theorem 4. *Let the number N and coefficients c_1, \dots, c_N be fixed, and let $\alpha > 1$. Then five different situations are possible:*

(i) *Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \neq 0$. Then for any $N_1 \geq 0$ and $\alpha > 1$, we have that*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\alpha\epsilon}{\alpha - 1} \left(\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(ii) *Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $N_1 \geq 1$, and there exists $k \leq N_1$ such that $c_k \neq 0$. Then for $\alpha < 2$ it holds that*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\epsilon^\alpha}{\alpha - 1} \left(\sum_{k=1}^{N_1} c_k^\alpha \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(iii) *Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $N_1 \geq 0$ and for all $k \leq N_1$ we have that $c_k = 0$. Then for $\alpha < 2$ it holds that*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\alpha\epsilon^2}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(iv) *Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $\alpha = 2$. Then for any $N_1 \geq 0$ and c_k , $k \leq N_1$, we have that*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \epsilon^2 \left(\sum_{k=1}^{N_1} c_k^2 \right) \left(\sum_{k=N_1+1}^N p_k^2 \right)^{-1}, \quad \epsilon \rightarrow 0.$$

(v) *Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $\alpha > 2$. Then for any $N_1 \geq 0$ and c_k , $k \leq N_1$, we have that*

$$\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\epsilon)) \sim \frac{\alpha\epsilon^2}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}, \quad \epsilon \rightarrow 0.$$

Proof. As in the proof of Theorem 3, we shall use expansions (8) and (9). The main tool will be L'Hospital's rule.

(i) Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \neq 0$. Then for any $N_1 \geq 0$ and $\alpha > 1$, we have the following relations:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon} \log \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) \\ &\quad - \frac{1}{(\alpha - 1)\varepsilon} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \frac{1}{\alpha - 1} \\ &\quad \times \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=1}^{N_1} \alpha c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\ &= \frac{\alpha}{\alpha - 1} \left(\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}. \end{aligned}$$

(ii) Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $N_1 \geq 1$, and there exists $k \leq N_1$ such that $c_k \neq 0$. Then for $\alpha < 2$ we have that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon^\alpha} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon^\alpha} \log \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) \\ &\quad - \frac{1}{(\alpha - 1)\varepsilon^\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{\alpha \varepsilon^{\alpha-1}} \\ &\quad \times \frac{\sum_{k=1}^{N_1} \alpha c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\ &= \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=1}^{N_1} c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N ((\alpha - 1)c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{\varepsilon^{\alpha-1} \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right)} \\ &= \frac{1}{\alpha - 1} \left(\sum_{k=1}^{N_1} c_k^\alpha \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}. \end{aligned}$$

(iii) Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $N_1 \geq 0$ and for all $k \leq N_1$ we have that $c_k = 0$. Then for $\alpha < 2$ it holds that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon^2} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon^2} \left(\log \left(\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) - \log \left(\sum_{k=1}^N p_k^\alpha \right) \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \frac{\sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\
 &= \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=N_1+1}^N (\alpha c_k p_k^{\alpha-1} + \alpha(\alpha - 1) c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{2\varepsilon \left(\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right)} \\
 &= \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=N_1+1}^N (\alpha(\alpha - 1) c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{2\varepsilon \left(\sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right)} \\
 &= \frac{\alpha}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}.
 \end{aligned}$$

(iv) Obviously, in the case $\alpha = 2$ we have the simple value of the entropy:

$$\mathcal{H}_2(p) = -\log \left(\sum_{k=1}^N p_k^2 \right).$$

Therefore, if $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $\alpha = 2$, then for any $N_1 \geq 0$ and c_k , $k \leq N_1$, we have that

$$\begin{aligned}
 \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_2(p) - \mathcal{H}_2(p(\varepsilon))}{\varepsilon^2} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \log \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^2 + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^2 \right) \\
 &\quad - \frac{1}{\varepsilon^2} \log \left(\sum_{k=1}^N p_k^2 \right) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \\
 &\quad \times \frac{\sum_{k=1}^{N_1} 2c_k^2 \varepsilon + \sum_{k=N_1+1}^N 2c_k (p_k + c_k \varepsilon)}{\sum_{k=1}^{N_1} (c_k \varepsilon)^2 + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^2} \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=1}^{N_1} c_k^2 + \sum_{k=N_1+1}^N c_k^2}{\sum_{k=1}^{N_1} (c_k \varepsilon)^2 + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^2} \\
 &= \left(\sum_{k=1}^N c_k^2 \right) \left(\sum_{k=N_1+1}^N p_k^2 \right)^{-1}.
 \end{aligned}$$

(v) Let $\sum_{k=N_1+1}^N c_k p_k^{\alpha-1} = 0$, $\alpha > 2$. Then for any $N_1 \geq 0$ and c_k , $k \leq N_1$, we have that

$$\begin{aligned}
 &\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}_\alpha(p) - \mathcal{H}_\alpha(p(\varepsilon))}{\varepsilon^2} \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{(\alpha - 1)\varepsilon^2} \log \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right) \\
 &\quad - \frac{1}{(\alpha - 1)\varepsilon^2} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon}
 \end{aligned}$$

$$\begin{aligned} & \times \frac{\sum_{k=1}^{N_1} \alpha c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N \alpha c_k (p_k + c_k \varepsilon)^{\alpha-1}}{\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha} \\ &= \frac{1}{\alpha - 1} \lim_{\varepsilon \rightarrow 0} \frac{\sum_{k=1}^{N_1} \alpha c_k^\alpha \varepsilon^{\alpha-1} + \sum_{k=N_1+1}^N (\alpha(\alpha - 1) c_k^2 p_k^{\alpha-2} \varepsilon + o(\varepsilon))}{2\varepsilon \left(\sum_{k=1}^{N_1} (c_k \varepsilon)^\alpha + \sum_{k=N_1+1}^N (p_k + c_k \varepsilon)^\alpha \right)} \\ &= \frac{\alpha}{2} \left(\sum_{k=N_1+1}^N c_k^2 p_k^{\alpha-2} \right) \left(\sum_{k=N_1+1}^N p_k^\alpha \right)^{-1}. \end{aligned}$$

Theorem is proved. □

3.2 Convergence of the disturbed entropy when the initial distribution is uniform but the number of events increases to ∞

The second problem is to establish conditions of stability of the entropy of uniform distribution when the number of events tends to ∞ . Let $N > 1$, $p_N(\text{uni}) = (\frac{1}{N}, \dots, \frac{1}{N})$ be a vector of uniform distribution with N possible states, $\varepsilon = \varepsilon(N) \leq \frac{1}{N}$, and $\{c_{kN}; N \geq 1, 1 \leq k \leq N\}$ be a family of fixed numbers (not all zero) such that $|c_{kN}| \leq 1$ and $\sum_{k=1}^N c_{kN} = 0$. Note that for any $N \geq 1$ there are strictly positive numbers c_{kN} for some k and consider the disturbed distribution vector $p'_N = (\frac{1}{N} + c_{1N}\varepsilon, \dots, \frac{1}{N} + c_{NN}\varepsilon)$.

Theorem 5. *Let $\varepsilon(N) = o(\frac{1}{N})$, $N \rightarrow \infty$. Then*

$$\mathcal{H}_\alpha(p_N) - \mathcal{H}_\alpha(p'_N) \rightarrow 0, \quad N \rightarrow \infty.$$

Proof. We know that $N\varepsilon \rightarrow 0$, as $N \rightarrow \infty$, and the family of numbers $\{c_{kn}; n \geq 1, 1 \leq k \leq n\}$ is bounded. Therefore the values

$$\sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \rightarrow 1, \quad \inf_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \rightarrow 1, \quad N \rightarrow \infty,$$

as the functions of N , and for every $N \geq 1$, $\sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \geq 1$. Recall that the function $x \log x$ is increasing for $x \geq 1$, and $x \log x \leq 0$ for $0 < x < 1$. Moreover, Rényi entropy is maximal on the uniform distribution. As a consequence of all these observations and assumptions we get that

$$\begin{aligned} 0 \leq \mathcal{H}_1(p_N) - \mathcal{H}_1(p'_N) &= \frac{1}{N} \sum_{k=1}^N (1 + Nc_{kN}\varepsilon) \log(1 + Nc_{kN}\varepsilon) \\ &\leq \frac{1}{N} \sum_{k=1}^N \sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \log \sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \\ &= \sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \log \sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \\ &\rightarrow 0, \quad N \rightarrow \infty. \end{aligned}$$

Let $\alpha > 1$. Then

$$\begin{aligned} 0 \leq \mathcal{H}_\alpha(p_N) - \mathcal{H}_\alpha(p'_N) &= \frac{1}{\alpha - 1} \log \left(\frac{1}{N} \sum_{k=1}^N (1 + Nc_{kN}\varepsilon)^\alpha \right) \\ &\leq \frac{1}{\alpha - 1} \log \left(\frac{1}{N} \sum_{k=1}^N \left(\sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \right)^\alpha \right) \\ &= \frac{\alpha}{\alpha - 1} \log \left(\sup_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \right) \rightarrow 0, \quad N \rightarrow \infty. \end{aligned}$$

Similarly, for $0 < \alpha < 1$ we produce the transformations

$$\begin{aligned} 0 \leq \mathcal{H}_\alpha(p_N) - \mathcal{H}_\alpha(p'_N) &= \frac{1}{\alpha - 1} \log \left(\frac{1}{N} \sum_{k=1}^N (1 + Nc_{kN}\varepsilon)^\alpha \right) \\ &\leq \frac{1}{\alpha - 1} \log \left(\frac{1}{N} \sum_{k=1}^N \left(\inf_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \right)^\alpha \right) \\ &= \frac{\alpha}{\alpha - 1} \log \left(\inf_{n \geq 1, 1 \leq k \leq n} (1 + Nc_{kn}\varepsilon) \right) \rightarrow 0, \quad N \rightarrow \infty, \end{aligned}$$

and the proof follows. □

3.3 Binomial and Poisson distributions

In this section we look at convergence of Rényi entropy of the binomial distribution to Rényi entropy of the Poisson distribution.

Theorem 6. *Let $\lambda > 0$ be fixed. For any $\alpha > 0$*

$$\lim_{n \rightarrow \infty} \mathcal{H}_\alpha \left(B \left(n, \frac{\lambda}{n} \right) \right) = \mathcal{H}_\alpha(\text{Poi}(\lambda)).$$

Proof. First, let $\alpha = 1$. We will find and regroup entropies of the binomial and Poisson distributions.

$$\begin{aligned} \mathcal{H}_1(B(n, p)) &= - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \log \left(\binom{n}{k} p^k (1-p)^{n-k} \right) \\ &= - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \log \binom{n}{k} \\ &\quad - n(p \log p + (1-p) \log(1-p)) \\ &= - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} (\log n! - \log k! - \log(n-k)!) + np \log n \\ &\quad - np \log np - n \log(1-p) + np \log(1-p) \\ &= np \log(1-p) - n \log(1-p) - np \log np + np \log n \end{aligned}$$

$$\begin{aligned}
 & - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} (\log n! - \log k! - \log(n-k)!). \\
 \mathcal{H}_1(\text{Poi}(\lambda)) &= - \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \log \left(e^{-\lambda} \frac{\lambda^k}{k!} \right) = \lambda - \lambda \log \lambda + \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \log k!
 \end{aligned}$$

We want to show componentwise convergence of entropies. For that let us take $np = \lambda$ and observe that

$$\begin{aligned}
 np \log(1-p) &= \lambda \log(1-p) \rightarrow \lambda \log 1 = 0, \quad n \rightarrow \infty, \quad p \rightarrow 0. \\
 -n \log(1-p) &= \log \left(1 - \frac{\lambda}{n} \right)^{-n} \rightarrow \lambda, \quad n \rightarrow \infty, \quad p \rightarrow 0. \\
 -np \log np &= -\lambda \log \lambda, \\
 np \log n - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} (\log n! - \log k! - \log(n-k)!) & \\
 &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} k \log n \\
 &\quad - \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} (\log n! - \log k! - \log(n-k)!) \\
 &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} (\log n^k - \log n! + \log k! + \log(n-k)!) \\
 &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \left(\log \frac{n^k (n-k)!}{n!} + \log k! \right).
 \end{aligned}$$

It is well known that $\frac{\log(x)}{x} \leq 1, x > 0$. Using this fact, we get the following representation:

$$\begin{aligned}
 \binom{n}{k} p^k (1-p)^{n-k} \log \frac{n^k (n-k)!}{n!} &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n} \right)^k \left(1 - \frac{\lambda}{n} \right)^{n-k} \log \frac{n^k (n-k)!}{n!} \\
 &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n} \right)^{n-k} \frac{n!}{n^k (n-k)!} \log \frac{n^k (n-k)!}{n!} \\
 &\leq \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n} \right)^{n-k} \leq \frac{\lambda^k}{k!}.
 \end{aligned}$$

For the second part of the sum simply observe that

$$\begin{aligned}
 \binom{n}{k} p^k (1-p)^{n-k} \log k! &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n} \right)^k \left(1 - \frac{\lambda}{n} \right)^{n-k} \log k! \\
 &= \frac{\lambda^k}{k!} \log k! \left(1 - \frac{\lambda}{n} \right)^{n-k} \frac{n!}{(n-k)!n^k}
 \end{aligned}$$

$$\leq \frac{\lambda^k}{k!} \log k!$$

As $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} (1 + \log k!) < \infty$, by Lebesgue's dominated convergence theorem:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \left(\log \frac{n^k (n-k)!}{n!} + \log k! \right) \\ &= \sum_{k=0}^{\infty} \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} \left(\log \frac{n^k (n-k)!}{n!} + \log k! \right) \\ &= \sum_{k=0}^{\infty} \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n} \right)^{n-k} \frac{n!}{(n-k)! n^k} \left(\log \frac{n^k (n-k)!}{n!} + \log k! \right) \\ &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \log k! \end{aligned}$$

Finally, we get that

$$\lim_{n \rightarrow \infty} \mathcal{H}_1 \left(B \left(n, \frac{\lambda}{n} \right) \right) = \mathcal{H}_1(\text{Poi}(\lambda)).$$

For $\alpha \neq 1$ we have:

$$\begin{aligned} \mathcal{H}_\alpha(\text{binomial}) &= \frac{1}{1-\alpha} \log \sum_{k=0}^n \left(\binom{n}{k} p^k (1-p)^{n-k} \right)^\alpha, \\ \mathcal{H}_\alpha(\text{poisson}) &= \frac{1}{1-\alpha} \log \sum_{k=0}^{+\infty} \left(e^{-\lambda} \frac{\lambda^k}{k!} \right)^\alpha. \end{aligned}$$

Thus, to show that

$$\lim_{n \rightarrow \infty} \mathcal{H}_\alpha \left(B \left(n, \frac{\lambda}{n} \right) \right) = \mathcal{H}_\alpha(\text{Poi}(\lambda)),$$

it is enough to show the convergence of sums, which follows from Lebesgue's dominated convergence theorem and

$$\left(\binom{n}{k} p^k (1-p)^{n-k} \right)^\alpha \leq \left(\frac{\lambda^k}{k!} \right)^\alpha, \quad \sum_{k=0}^{+\infty} \left(\frac{\lambda^k}{k!} \right)^\alpha < +\infty.$$

□

A Appendix

We let $0 \log 0 = 0$ by continuity and prove several auxiliary results. Stating these three lemmas, we assume that $p_i \geq 0, 1 \leq i \leq N$ are fixed.

Lemma 6. $H_\alpha(p) \rightarrow H(p)$, $\alpha \rightarrow 1$.

Proof. Using L'Hospital's rule, we get the following relations:

$$\begin{aligned} \lim_{\alpha \rightarrow 1} H_\alpha(p) &= \lim_{\alpha \rightarrow 1} \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \lim_{\alpha \rightarrow 1} \frac{1}{-1} \frac{\sum_{k=1}^N p_k^\alpha \log(p_k)}{\sum_{k=1}^N p_k^\alpha} \\ &= - \sum_{k=1}^N p_k \log(p_k) = H(p). \end{aligned}$$

□

Let $\mathcal{H}_1(p) := \mathcal{H}(p)$ (Shannon entropy), and so $H_\alpha(p)$ is defined for all $\alpha > 0$ and is continuous in α .

Lemma 7. $H_\alpha(p)$ is nonincreasing in $\alpha > 0$.

Proof. Indeed,

$$\begin{aligned} \frac{\partial H_\alpha(p)}{\partial \alpha} &= \frac{1}{(1-\alpha)^2} \log \left(\sum_{i=1}^N p_i^\alpha \right) + \frac{1}{1-\alpha} \frac{\sum_{k=1}^N p_k^\alpha \log p_k}{\sum_{k=1}^N p_k^\alpha} \\ &= \frac{1}{(1-\alpha)^2} \sum_{k=1}^N p_k^\alpha \left(\log \left(\sum_{i=1}^N p_i^\alpha \right) + \log p_k^{1-\alpha} \right) \\ &= \frac{-1}{(1-\alpha)^2} \sum_{k=1}^N \frac{p_k^\alpha}{\sum_{i=1}^N p_i^\alpha} \log \left(\frac{p_k^\alpha}{\sum_{i=1}^N p_i^\alpha} \frac{1}{p_k} \right). \end{aligned}$$

Let $q_k = \frac{p_k^\alpha}{\sum_{i=1}^N p_i^\alpha}$. Then

$$\frac{\partial H_\alpha(p)}{\partial \alpha} = \frac{-1}{(1-\alpha)^2} \sum_{k=1}^N q_k \log \frac{q_k}{p_k} \leq 0.$$

The fact that $\mathcal{H}_\alpha(p) \leq \mathcal{H}_1(p) \leq \mathcal{H}_\beta(p)$, where $0 < \beta < 1 < \alpha$ follows from Lemma 6. □

Lemma 8. $\mathcal{H}_\alpha(p) \leq \log N$ and it reaches maximum when distribution is uniform.

Proof. Let $1 \leq m \leq N$ be the number of nonzero probabilities. Then we have

$$\lim_{\alpha \rightarrow 0+} \mathcal{H}_\alpha(p) = \lim_{\alpha \rightarrow 0+} \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \log m \leq \log N.$$

So $\mathcal{H}_\alpha(p) \leq \log N$ due to Lemma 7. For the second part, put $p_1 = \dots = p_N = \frac{1}{N}$.

$$\mathcal{H}_1(p) = - \sum_{k=1}^N \frac{1}{N} \log \left(\frac{1}{N} \right) = \log N.$$

$$\mathcal{H}_\alpha(p) = \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N \frac{1}{N^\alpha} \right) = \frac{1}{1-\alpha} \log \left(\frac{N}{N^\alpha} \right) = \log N.$$

□

Remark 3. Let $1 \leq m \leq N$ be the number of nonzero probabilities and without loss of generality let $p_k < p_1 = \dots = p_{N_1}$ for every $N_1 + 1 \leq k \leq N$. Then we can also define

$$\mathcal{H}_0(p) := \lim_{\alpha \rightarrow 0^+} \mathcal{H}_\alpha(p) = \lim_{\alpha \rightarrow 0^+} \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \log m.$$

$$\begin{aligned} \mathcal{H}_\infty(p) &:= \lim_{\alpha \rightarrow +\infty} \mathcal{H}_\alpha(p) = \lim_{\alpha \rightarrow +\infty} \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right) = \\ &= \lim_{\alpha \rightarrow +\infty} - \frac{\sum_{k=1}^N p_k^\alpha \log p_k}{\sum_{k=1}^N p_k^\alpha} \\ &= \lim_{\alpha \rightarrow +\infty} - \frac{N_1 \log p_1 + \sum_{k=N_1+1}^N \left(\frac{p_k}{p_1}\right)^\alpha \log p_k}{N_1 + \sum_{k=N_1+1}^N \left(\frac{p_k}{p_1}\right)^\alpha} \\ &= -\log p_1. \end{aligned}$$

References

- [1] Acharya, J., Orlitsky, A., Suresh, A.T., Tyagi, H.: The Complexity of Estimating Rényi Entropy. Proceedings of the 2015 Annual ACM-SIAM Symposium on Discrete Algorithms (2015). Book Code:PRDA15. MR3451148. <https://doi.org/10.1137/1.9781611973730.124>
- [2] Banna, O., Buryak, F., Yu, M.: Distance from fractional Brownian motion with associated Hurst index $0 < H < 1/2$ to the subspaces of Gaussian martingales involving power integrands with an arbitrary positive exponent. Mod. Stoch. Theory Appl. **7**, pp. 191–202 (2020). MR4120614. <https://doi.org/10.15559/20-vmsta156>
- [3] Banna, O., Mishura, Yu., Ralchenko, K., Shklyar, S.: Fractional Brownian Motion. Approximations and Projections. Wiley-ISTE, London (2019). 288 p.
- [4] Bégin, L., Germain, P., Laviolette, F., Roy, J.-F.: PAC-Bayesian Bounds based on the Rényi Divergence. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, vol. 51, pp. 435–444. PMLR (2016).
- [5] Bobkov, S.G., Marsiglietti, A., Melbourne, J.: Concentration functions and entropy bounds for discrete log-concave distributions. arXiv:2007.11030v1 [math.PR]
- [6] Erven, T., Harremoës, P.: Rényi Divergence and majorization. In: IEEE Transactions on Information Theory (ISIT), INSPEC 11434178 (2010).
- [7] Erven, T., Harremoës, P.: Rényi Divergence and Kullback-Leibler Divergence. IEEE Trans. Inf. Theory **60**, 3797–3820 (2014). MR3225930. <https://doi.org/10.1109/TIT.2014.2320500>

- [8] Gil, M., Alajaji, F., Linder, T.: Rényi divergence measures for commonly used univariate continuous distributions. *Inf. Sci.* **249**, 124–131 (2013). [MR3105467](#). <https://doi.org/10.1016/j.ins.2013.06.018>
- [9] Hanel, R., Thurner, S., Tsallis, C.: On the robustness of q-expectation values and Rényi entropy. *Europhys. Lett. Assoc.* **852** (2009).
- [10] Harremoës, P.: Binomial and Poisson Distributions as Maximum Entropy Distributions. *IEEE Trans. Inf. Theory* **475**, 2039–2041 (2001). [MR1842536](#). <https://doi.org/10.1109/18.930936>
- [11] Harremoës, P.: Interpretations of Rényi entropies and divergences. *Physica A* (2006). [MR2223335](#). <https://doi.org/10.1016/j.physa.2006.01.012>
- [12] Ho, S.-W., Verdú, S.: Convexity/concavity of Rényi entropy and α -mutual information. In: *IEEE International Symposium on Information Theory (ISIT)*, pp. 745–749 (2015). <https://doi.org/10.1109/ISIT.2015.7282554>
- [13] Lenzi, E.K., Mendes, R.S., Silva, L.R.: Statistical mechanics based on Rényi entropy. *Physica A* **280**, 337–345 (2000). [MR1758523](#). [https://doi.org/10.1016/S0378-4371\(00\)00007-8](https://doi.org/10.1016/S0378-4371(00)00007-8)
- [14] Rényi, A.: On Measures of Entropy and Information. In: *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, pp. 547–561. The Regents of University of the California, Berkeley, CA, USA (1961). [MR0132570](#)
- [15] Sason, I.: Tight Bounds on the Rényi Entropy via Majorization with Applications to Guessing and Compression. *Entropy* **20**(12), 896 (2018).
- [16] Yang, Y., Li, J., Yang, Y.: Multiscale multifractal multiproperty analysis of financial time series based on Rényi entropy. *Int. J. Mod. Phys. C* **28**(2) (2017). [MR3284253](#). <https://doi.org/10.1016/j.physa.2014.11.009>