



Properties of Shannon and Rényi entropies of the Poisson distribution as the functions of intensity parameter

Volodymyr Braiman^a , Anatoliy Malyarenko^b , Yuliya Mishura^{a,b,1} ,
Yevheniia Anastasiia Rudyk^a 

^aTaras Shevchenko National University of Kyiv,
64/13 Volodymyrska St., Kyiv, Ukraine
volodymyr.braiman@knu.ua; yuliyamishura@knu.ua
rudykyao@knu.ua

^bMälardalen University, 721 23 Västerås, Sweden
anatoliy.malyarenko@mdu.se

Received: February 7, 2024 / **Revised:** May 6, 2024 / **Published online:** July 1, 2024

Abstract. We consider two types of entropy, namely, Shannon and Rényi entropies of the Poisson distribution, and establish their properties as the functions of intensity parameter. More precisely, we prove that both entropies increase with intensity. While for Shannon entropy the proof is comparatively simple, for Rényi entropy, which depends on additional parameter $\alpha > 0$, we can characterize it as nontrivial. The proof is based on application of Karamata's inequality to the terms of Poisson distribution.

Keywords: Shannon entropy, Rényi entropy, Poisson distribution, Karamata's inequality.

1 Introduction

The concept and formulas for different types of entropy, which come mainly from information theory, are now widely used in many applications, in particular, in detecting DDoS attacks (see, for example, [2, 6]), in investigation of structure of the neural codes [17], network traffic [14], keyword extraction [21], stock market forecast modelling [12], and many others. Considering these applications, it is natural to assume that, for example, the distribution of the number of DDoS attacks and of some other related phenomena is Poisson with some fixed intensity at least over some fixed time interval. Considering two types of entropy, namely, Shannon and Rényi entropy, it is natural to assume that both entropies increase together with intensity parameter. To the best of our knowledge, this statement was not established rigorously. Moreover, if for Shannon entropy the proof is comparatively simple, for Rényi entropy, which depends on additional parameter $\alpha > 0$,

¹The author supported by The Swedish Foundation for Strategic Research (grant No. UKR22-0017) and the ToppForsk project No. 274410 of the Research Council of Norway with title STORM: Stochastics for Time-Space Risk Models.

we can characterize it as nontrivial. In order to get the proof, we apply Karamata’s inequality to the terms of Poisson distribution. The verification of conditions of this inequality also is nontrivial. It involves rearrangement of the terms of Poisson distribution into a nonincreasing sequence and exploring the monotonicity of partial sums of this sequence.

If to take a historical look, the concept of entropy for a random variable was introduced by Shannon [20] to characterize the irreducible complexity inherent in a specific form of randomness. Then Shannon entropy was generalized by Rényi entropy [19] by introducing an additional parameter $\alpha > 0$ that allows for a range of entropy measures. The presence of this parameter makes it difficult to accurately calculate the Rényi entropy for various distributions and study its behavior. A pleasant exception is the normal distribution. For it, many types of entropies are calculated exactly, and it can be noted that these entropies increase along with the variance; see [15]. Furthermore, if the decrease of the Rényi entropy with respect to the parameter α is a well-known fact, its convexity for discrete distributions depends on the distribution; see [4].

Taking into account the difficulty with exact computation of entropies, many attempts were devoted to the numerical calculation and approximation of entropies (see, e.g., [1–3, 5, 9, 10, 22]), while the limit of Shannon entropy of Poisson distribution was calculated, e.g., in [8].

As already mentioned, the purpose of this paper is to prove analytically the natural fact that the entropy of a Poisson distribution increases with the intensity of the distribution. We considered Shannon and Rényi entropy. The paper is constructed as follows. In Section 2, we consider increase and convexity of Shannon entropy of Poisson distribution as the function of intensity parameter. In Section 3 the increase in intensity of Rényi entropy for any $\alpha > 0$ is proved with the help of Karamata’s inequality. Some by-product inequalities are obtained in Section 4, while auxiliary statements are postponed to Appendix.

2 Analytical properties of Shannon entropy of Poisson distribution as the function of intensity parameter λ

Consider a discrete distribution $\{p_i, i \geq 1\}$. Its Shannon entropy is defined as

$$H_S(p_i, i \geq 1) = - \sum_{i \geq 1} p_i \log p_i.$$

In particular, we consider Poisson distribution with parameter λ :

$$\mathbf{P}\{\xi_\lambda = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N} \cup \{0\}.$$

Its Shannon entropy $H_S(\lambda)$ equals

$$H_S(\lambda) = - \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \log \frac{\lambda^k e^{-\lambda}}{k!} = -\lambda \log \frac{\lambda}{e} + e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k \log k!}{k!}. \quad (1)$$

It is natural to assume that Shannon entropy $H_S(\lambda)$ of Poisson distribution strictly increases with $\lambda \in (0, +\infty)$. We will prove this result, as well as the concavity property of $H_S(\lambda)$, in the next statement.

Theorem 1. *Shannon entropy $H_S(\lambda), \lambda \in (0, +\infty)$ is strictly increasing and concave in λ , its derivative in λ is strictly positive and decreasing.*

Proof. Obviously, we can differentiate the series (1) term by term for any $\lambda > 0$, and we get the equality

$$\begin{aligned}
 H'_S(\lambda) &= -\log \frac{\lambda}{e} - 1 - e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k \log k!}{k!} + e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-1} \log k!}{(k-1)!} \\
 &= -\log \lambda + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)!}{k!} - e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k \log k!}{k!} \\
 &= -\log \lambda + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!}. \tag{2}
 \end{aligned}$$

It is clear that both terms in the right-hand side of (2) are nonnegative for $\lambda \in (0, 1]$, and the second one is strictly positive, therefore $H'_S(\lambda) > 0$ for $\lambda \in (0, 1]$. So, it is necessary to prove that $H'_S(\lambda) > 0$ for $\lambda > 1$. Let us calculate

$$\begin{aligned}
 H''_S(\lambda) &= -\frac{1}{\lambda} - e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1} \log(k+1)}{(k-1)!} \\
 &= -\frac{1}{\lambda} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(k+2)}{k!} - e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} \\
 &= -\frac{1}{\lambda} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(1 + \frac{1}{k+1})}{k!} \\
 &< -\frac{1}{\lambda} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k+1)!} < -\frac{1}{\lambda} + e^{-\lambda} \frac{1}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} \\
 &< -\frac{1}{\lambda} + e^{-\lambda} \frac{1}{\lambda} e^{\lambda} = 0.
 \end{aligned}$$

So, $H''_S(\lambda) < 0$ for all $\lambda > 0$. Therefore $H'_S(\lambda)$ strictly decreases in λ , and it is sufficient to prove that

$$\lim_{\lambda \rightarrow \infty} H'_S(\lambda) \geq 0.$$

However,

$$\lim_{\lambda \rightarrow \infty} H'_S(\lambda) = \lim_{\lambda \rightarrow \infty} \log \lambda \left(e^{-\lambda} (\log \lambda)^{-1} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} - 1 \right),$$

and it is sufficient to establish that

$$\liminf_{\lambda \rightarrow \infty} e^{-\lambda} (\log \lambda)^{-1} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} \geq 1.$$

This inequality is proved in Lemma A1. Finally, we get that $H'_S(\lambda) > 0$ for all $\lambda \geq 0$ and $H''_S(\lambda) < 0$ for all $\lambda \geq 0$, whence the proof follows. \square

Remark 1. Strictly increasing property of H_S was also established in [7] by a bit different calculations and without convexity.

3 Analytical properties of Rényi entropy of Poisson distribution as the function of intensity parameter λ

Again, consider a discrete distribution $\{p_i, i \geq 1\}$. Its Rényi entropy is defined as

$$H_R^\alpha(p_i, i \geq 1) = \frac{1}{1-\alpha} \log \left(\sum_{i \geq 1} p_i^\alpha \right), \quad \alpha > 0, \alpha \neq 1,$$

and

$$H_R^\alpha(p_i, i \geq 1) \rightarrow H_S(p_i, i \geq 1)$$

as $\alpha \rightarrow 1$, where $H_S(p_i, i \geq 1)$ is a Shannon entropy. In the case of Poisson distribution,

$$H_R^\alpha(\lambda) = \frac{1}{1-\alpha} \log \left(e^{-\alpha\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k\alpha}}{(k!)^\alpha} \right), \quad \alpha > 0, \alpha \neq 1. \tag{3}$$

As for the Shannon entropy, our goal is to investigate the behaviour of Rényi entropy of Poisson distribution as the function of intensity λ . To be more precise, we wish to prove that for any fixed $\alpha > 0, \alpha \neq 1$, Rényi entropy of Poisson distribution increases in λ . Let us take into account equality (3) and consider two cases.

- Let $\alpha \in (0, 1)$. Then $1/(1-\alpha) > 0$, and since logarithm is strictly increasing, for $H_R^\alpha(\lambda)$ to increase in λ , the function

$$\psi(\alpha, \lambda) = e^{-\alpha\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda^k}{k!} \right)^\alpha \tag{4}$$

should increase in $\lambda \in (0, +\infty)$.

- Let $\alpha > 1$. Then $1/(1-\alpha) < 0$, and for $H_R^\alpha(\lambda)$ to increase in λ , functions $\log(\psi(\alpha, \lambda))$ and $\psi(\alpha, \lambda)$ should decrease in $\lambda \in (0, +\infty)$.

In Theorem 2 below, we will establish desired character of monotonicity of $\psi(\alpha, \lambda)$ in $\lambda \in (0, +\infty)$ for both cases $0 < \alpha < 1$ and $\alpha > 1$. To prove this result, we will apply Karamata's inequality (see Lemma A2 in Appendix). This inequality deals with

nonincreasing sequences, so it suggests to rearrange the terms of the Poisson distribution in nonincreasing order and study the properties of the resulting sequence.

Let, as before, $p_k(\lambda) = \mathbf{P}\{\xi_\lambda = k\} = (\lambda^k/k!)e^{-\lambda}$, $k \in \mathbb{N} \cup \{0\}$. Denote by $\{q_k(\lambda), k \geq 0\}$ the terms of the sequence $\{p_k(\lambda), k \geq 0\}$, which are rearranged in nonincreasing order. Note that the sequence $\{p_k(\lambda), k \geq 0\}$ may contain equal terms. In this case the sequence $\{q_k(\lambda), k \geq 0\}$ will contain equal terms as well.

Lemma 1. For each $n \geq 0$, the function $S_n(\lambda) = \sum_{k=0}^n q_k(\lambda)$ strictly decreases on $(0, +\infty)$.

Proof. Fix $n \geq 0$. Firstly, we will establish that for every $\lambda > 0$, the number $S_n(\lambda)$ is equal to the sum of some $n + 1$ consecutive terms of the initial sequence $\{p_k(\lambda), k \geq 0\}$. Note that $p_{k+1}(\lambda)/p_k(\lambda) = \lambda/(k + 1)$, thus $p_k(\lambda) < p_{k+1}(\lambda)$ for $k < \lambda - 1$, and $p_k(\lambda) > p_{k+1}(\lambda)$ for $k > \lambda - 1$. Therefore for every $0 \leq i_1 < i_2 < i_3$, we have $p_{i_2}(\lambda) > \min\{p_{i_1}(\lambda), p_{i_3}(\lambda)\}$. It follows that $q_0(\lambda), \dots, q_n(\lambda)$ are $n + 1$ consecutive terms of the sequence $\{p_k(\lambda), k \geq 0\}$. Thus $(q_0(\lambda), \dots, q_n(\lambda))$ is a permutation of terms $(p_\ell(\lambda), \dots, p_{\ell+n}(\lambda))$ for some $\ell = \ell(\lambda) \geq 0$, so $S_n(\lambda) = \sum_{k=0}^n q_k(\lambda) = \sum_{k=\ell}^{\ell+n} p_k(\lambda)$. Let us determine $\ell(\lambda)$. Put $s_n(m, \lambda) = \sum_{k=m}^{m+n} p_k(\lambda)$, $m \geq 0$. Clearly,

$$S_n(\lambda) = s_n(\ell, \lambda) = \max_{m \geq 0} s_n(m, \lambda).$$

Note that

$$\begin{aligned} & s_n(m + 1, \lambda) - s_n(m, \lambda) \\ &= \sum_{k=m+1}^{m+n+1} p_k(\lambda) - \sum_{k=m}^{m+n} p_k(\lambda) = p_{m+n+1}(\lambda) - p_m(\lambda) < 0 \end{aligned}$$

if and only if

$$\frac{p_{m+n+1}(\lambda)}{p_m(\lambda)} = \frac{\lambda^{n+1}m!}{(m + n + 1)!} = \frac{\lambda^{n+1}}{(m + 1) \dots (m + n + 1)} < 1,$$

i.e., $\lambda < c_m = ((m + 1) \dots (m + n + 1))^{1/(n+1)}$. It follows that $s_n(m + 1, \lambda) > s_n(m, \lambda)$ for $\lambda > c_m$ and $s_n(m + 1, \lambda) < s_n(m, \lambda)$ for $\lambda < c_m$. Therefore $\max_{m \geq 0} s_n(m, \lambda)$ is achieved at

$$\ell(\lambda) = \begin{cases} 0 & \text{for } 0 < \lambda \leq c_0, \\ m & \text{for } c_{m-1} \leq \lambda \leq c_m, m \geq 1 \end{cases}$$

(in particular, in the case $\lambda = c_k$, we have $s_n(k + 1, \lambda) = s_n(k, \lambda)$, so one can take either $\ell(\lambda) = k$ or $\ell(\lambda) = k + 1$).

For $0 < \lambda \leq c_0$, we have $S_n(\lambda) = \sum_{k=0}^n p_k(\lambda) = \sum_{k=0}^n (\lambda^k/k!)e^{-\lambda}$, so

$$\begin{aligned} S'_n(\lambda) &= \sum_{k=1}^n \frac{\lambda^{k-1}}{(k - 1)!} e^{-\lambda} - \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} e^{-\lambda} - \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda} \\ &= -\frac{\lambda^n}{n!} e^{-\lambda} < 0, \end{aligned}$$

and the function $S_n(\lambda)$ strictly decreases on $(0, c_0]$. For $c_{m-1} < \lambda \leq c_m$, $m \geq 1$, we have $S_n(\lambda) = \sum_{k=m}^{m+n} p_k(\lambda) = \sum_{k=m}^{m+n} (\lambda^k/k!)e^{-\lambda}$, so

$$\begin{aligned} S'_n(\lambda) &= \sum_{k=m}^{m+n} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} - \sum_{k=m}^{m+n} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=m-1}^{m+n-1} \frac{\lambda^k}{k!} e^{-\lambda} - \sum_{k=m}^{m+n} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} - \frac{\lambda^{m+n}}{(m+n)!} e^{-\lambda} < 0, \end{aligned}$$

which is equivalent to $\lambda^{n+1} > m(m+1) \cdots (m+n)$ or $\lambda > c_{m-1}$. Hence the function $S_n(\lambda)$ also strictly decreases on $[c_{m-1}, c_m]$ for each $m \geq 1$. Clearly, $c_m \rightarrow +\infty$ as $m \rightarrow \infty$. Thus $S_n(\lambda)$ decreases on

$$(0, c_0] \cup \bigcup_{m=1}^{\infty} [c_{m-1}, c_m] = (0, +\infty). \quad \square$$

Recall that $\psi(\alpha, \lambda) = \sum_{k=0}^{\infty} p_k^\alpha(\lambda)$, $\alpha > 0$, $\lambda > 0$.

Theorem 2. For every $0 < \alpha < 1$, the function $\psi(\alpha, \lambda)$ strictly increases as a function of λ on $(0, +\infty)$, while for every $\alpha > 1$, the function $\psi(\alpha, \lambda)$ strictly decreases as a function of λ on $(0, +\infty)$.

Proof. Put $r_n(\lambda) = \sum_{k=n+1}^{\infty} q_k(\lambda)$, $n \geq 0$. Note that $r_n(\lambda) \rightarrow 0$ as $n \rightarrow \infty$ as a remainder of a convergent series. It follows from the proof of Lemma 1 that for every $\lambda > 0$, there exists $N(\lambda) \in \mathbb{N}$ such that $(q_0(\lambda), \dots, q_n(\lambda))$ is a permutation of $(p_0(\lambda), \dots, p_n(\lambda))$ for each $n \geq N(\lambda)$. Hence $q_n(\lambda) = p_n(\lambda)$ for every $n \geq N(\lambda)$. Without loss of generality, let $N(\lambda) \geq 2\lambda$. Then $p_{k+1}(\lambda)/p_k(\lambda) = \lambda/(k+1) < 1/2$ for all $k \geq N(\lambda)$. Thus

$$r_n(\lambda) = \sum_{k=n+1}^{\infty} p_k(\lambda) \leq \sum_{k=n+1}^{\infty} \frac{p_n(\lambda)}{2^{k-n}} = p_n(\lambda) = q_n(\lambda), \quad n \geq N(\lambda).$$

Fix any $0 < \lambda_1 < \lambda_2$. For every $n \geq N = \max\{N(\lambda_1), N(\lambda_2)\}$, we have

- (i) $q_0(\lambda_i) \geq q_1(\lambda_i) \geq \dots \geq q_n(\lambda_i) \geq r_n(\lambda_i)$, $i = 1, 2$,
- (ii) $\sum_{k=0}^i q_k(\lambda_1) > \sum_{k=0}^i q_k(\lambda_2)$ for $0 \leq i \leq n$ by Lemma 1,
- (iii) $\sum_{k=0}^n q_k(\lambda_1) + r_n(\lambda_1) = \sum_{k=0}^n q_k(\lambda_2) + r_n(\lambda_2) = 1$,

i.e., $(q_0(\lambda_1), \dots, q_n(\lambda_1), r_n(\lambda_1))$ majorizes $(q_0(\lambda_2), \dots, q_n(\lambda_2), r_n(\lambda_2))$ (see Lemma A2 for the definition of majorization).

For $0 < \alpha < 1$, the function x^α , $x \in [0, 1]$, is concave, thus by Karamata's inequality

$$\sum_{k=0}^n q_k^\alpha(\lambda_1) + r_n^\alpha(\lambda_1) \leq \sum_{k=0}^n q_k^\alpha(\lambda_2) + r_n^\alpha(\lambda_2), \quad n \geq N. \quad (5)$$

Passing to the limit as $n \rightarrow \infty$, we obtain

$$\psi(\alpha, \lambda_1) = \sum_{k=0}^{\infty} q_k^\alpha(\lambda_1) \leq \sum_{k=0}^{\infty} q_k^\alpha(\lambda_2) = \psi(\alpha, \lambda_2), \quad (6)$$

therefore the function $\psi(\alpha, \lambda)$ increases as a function of λ on $(0, +\infty)$. Now we will strengthen inequalities (5) and (6) to ensure that $\psi(\alpha, \lambda)$ is strictly monotonic as a function of λ . Fix m such that $q_m(\lambda_1) > q_{m+1}(\lambda_1)$ (actually, one can take $m = 0$ or $m = 1$ because the sequence $\{q_k(\lambda_1), k \geq 0\}$ contains at most two copies of each term). Put $\tilde{q}_m(\lambda_1) = q_m(\lambda_1) - \delta, \tilde{q}_{m+1}(\lambda_1) = q_{m+1}(\lambda_1) + \delta$, where $\delta > 0$ is such that

$$\delta \leq \frac{1}{2}(q_m(\lambda_1) - q_{m+1}(\lambda_1))$$

and

$$\delta < \sum_{k=0}^m q_k(\lambda_1) - \sum_{k=0}^m q_k(\lambda_2).$$

Note that for every $n \geq \tilde{N} = \max\{N(\lambda_1), N(\lambda_2), m + 1\}$, conditions (i)–(iii) above remain valid if $q_m(\lambda_1), q_{m+1}(\lambda_1)$ are replaced with $\tilde{q}_m(\lambda_1), \tilde{q}_{m+1}(\lambda_1)$. Indeed, only inequalities $\tilde{q}_m(\lambda_1) \geq \tilde{q}_{m+1}(\lambda_1)$ in (i) and $\sum_{k=0}^{m-1} q_k(\lambda_1) + \tilde{q}_m(\lambda_1) > \sum_{k=0}^m q_k(\lambda_2)$ in (ii) are to be verified anew, and both of them hold by the choice of δ . It means that

$$(q_0(\lambda_1), \dots, q_{m-1}(\lambda_1), \tilde{q}_m(\lambda_1), \tilde{q}_{m+1}(\lambda_1), q_{m+2}(\lambda_1), \dots, q_n(\lambda_1), r_n(\lambda_1))$$

majorizes $(q_0(\lambda_2), \dots, q_n(\lambda_2), r_n(\lambda_2))$. Therefore inequalities (5) and (6) remain valid if $q_m(\lambda_1), q_{m+1}(\lambda_1)$ are replaced with $\tilde{q}_m(\lambda_1), \tilde{q}_{m+1}(\lambda_1)$. Thus

$$\sum_{k=0}^{\infty} q_k^\alpha(\lambda_1) + \tilde{q}_m^\alpha(\lambda_1) + \tilde{q}_{m+1}^\alpha(\lambda_1) - q_m^\alpha(\lambda_1) - q_{m+1}^\alpha(\lambda_1) \leq \sum_{k=0}^{\infty} q_k^\alpha(\lambda_2).$$

Since the function $x^\alpha, x \in [0, 1]$, is strictly concave on $[0, 1]$, we have

$$\tilde{q}_m^\alpha(\lambda_1) + \tilde{q}_{m+1}^\alpha(\lambda_1) > q_m^\alpha(\lambda_1) + q_{m+1}^\alpha(\lambda_1).$$

Hence

$$\psi(\alpha, \lambda_1) = \sum_{k=0}^{\infty} q_k^\alpha(\lambda_1) < \sum_{k=0}^{\infty} q_k^\alpha(\lambda_2) = \psi(\alpha, \lambda_2). \tag{7}$$

For $\alpha > 1$, the function $x^\alpha, x \in [0, 1]$, is convex, thus by Karamata’s inequality inequalities (5), (6), and (7) are reversed. □

Corollary 1. For every $\alpha > 0, \alpha \neq 1$, Rényi entropy $H_R^\alpha(\lambda)$ strictly increases as a function of λ on $(0, +\infty)$.

Remark 2. The case $\alpha = 2$ was treated in [18].

4 By-product inequalities

As can be seen from Section 3, the proof of increasing (decreasing) properties of function $\psi(\alpha, \lambda)$ did not use differentiation of $\psi(\alpha, \lambda)$ and the properties of its derivatives. However, function $\psi(\alpha, \lambda)$ can be differentiated term by term at any point $\alpha > 0$ in $\lambda > 0$,

and as a result, we shall get after some elementary transformations that for any $\alpha, \lambda > 0$,

$$\psi'_\lambda(\alpha, \lambda) = \alpha e^{-\alpha\lambda} \sum_{k=0}^{\infty} (k - \lambda) \frac{\lambda^{\alpha k - 1}}{(k!)^\alpha}.$$

Obviously, if $\alpha = 1$, then for any $\lambda > 0$, the derivative is zero because it differs by a strictly positive multiplier $\alpha e^{-\alpha\lambda}$ from the expression

$$R(\alpha, \lambda) = \sum_{k=0}^{\infty} (k - \lambda) \frac{\lambda^{\alpha k - 1}}{(k!)^\alpha},$$

and

$$R(1, \lambda) = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} - \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 0.$$

However, taking into account Theorem 2, we can establish some nontrivial inequalities.

Lemma 2. For any $\lambda > 0$ and $\alpha \in (0, 1)$,

$$\sum_{k=1}^{\infty} \frac{\lambda^{\alpha k - 1} k^{1 - \alpha}}{((k - 1)!)^\alpha} \geq \sum_{k=0}^{\infty} \frac{\lambda^{\alpha k}}{(k!)^\alpha}, \tag{8}$$

while for any $\lambda > 0$ and $\alpha > 1$,

$$\sum_{k=1}^{\infty} \frac{\lambda^{\alpha k - 1} k^{1 - \alpha}}{((k - 1)!)^\alpha} \leq \sum_{k=0}^{\infty} \frac{\lambda^{\alpha k}}{(k!)^\alpha}. \tag{9}$$

Proof. Both inequalities follow immediately from the fact that $R(\alpha, \lambda)$ and $\psi'_\lambda(\alpha, \lambda)$ are of the same sign, therefore it follows from Theorem 2 that $R(\alpha, \lambda) \geq 0$ for all $\lambda > 0$ and $\alpha \in (0, 1)$ and $R(\alpha, \lambda) \leq 0$ for all $\lambda > 0$ and $\alpha > 1$. \square

Remark 3. The nontrivial character of inequalities (8) and (9) is implied by the fact that we cannot compare the respective series term by term, and the relation between k th terms depends on whether the condition $\lambda > k$ or the inverse one is satisfied. The similar situation was with $\psi(\alpha, \lambda)$, but now, having Theorem 2 in hand, we do not need to analyze the series in more detail in order to compare them.

5 Graphical support of results

In this section, we present several plots and surfaces illustrating the behaviour of Rényi entropy as the function of parameters α and λ . Obviously, all of them confirm our theoretical results. First, we demonstrate the behaviour of the function $\psi(\alpha, \lambda)$ from (4).

We see from Fig. 1(a) that for fixed $\alpha \in (0, 1)$ (the values of α are chosen from 0.1 to 0.9 with interval 0.1), the function $\psi(\alpha, \lambda)$ strictly increases in $\lambda \in (0, +\infty)$. The same is confirmed by the surface on Fig. 1(b).

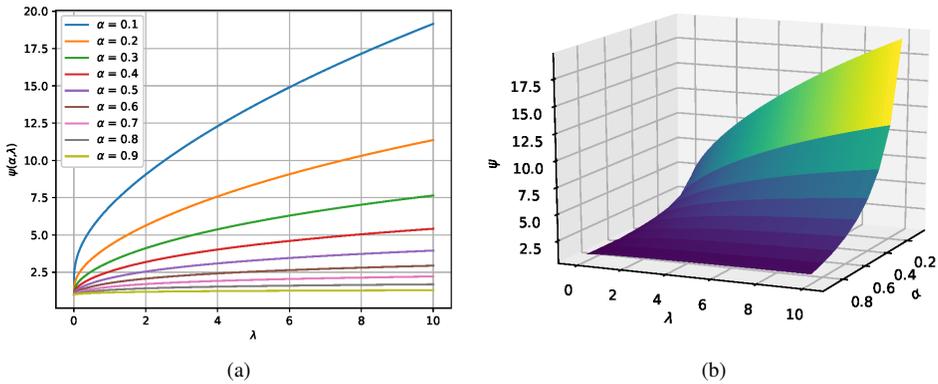


Figure 1. $\psi(\alpha, \lambda)$ is an increasing function of λ when α is fixed, $0 < \alpha < 1, \lambda > 0$.

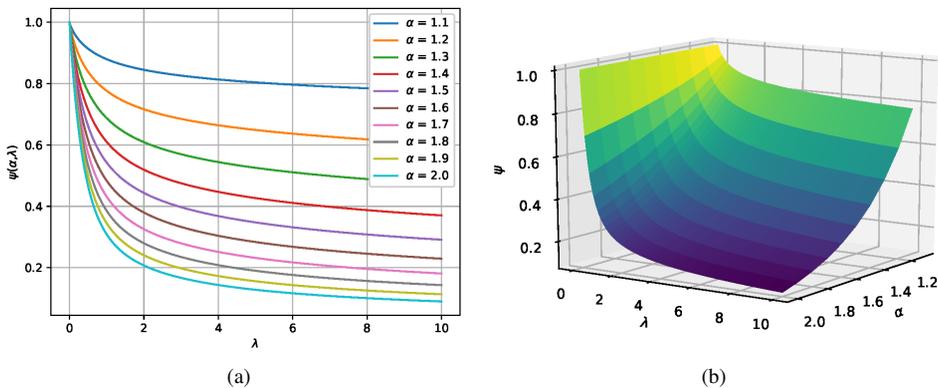


Figure 2. $\psi(\alpha, \lambda)$ is a decreasing function of λ when α is fixed, $\alpha > 1, \lambda > 0$.

Respectively, we see from Fig. 2(a) that for fixed $\alpha > 1$ (the values of α are chosen from 1.1 to 2.0 with interval 0.1), the function $\psi(\alpha, \lambda)$ strictly decreases in $\lambda \in (0, +\infty)$. The same is confirmed by the surface on Fig. 2(b).

Now, let us illustrate Lemma 2.

We see from Fig. 3(a) that for $\lambda > 0$ and $\alpha \in (0, 1)$ (the values of α are chosen from 0.1 to 0.9 with interval 0.1), $R(\alpha, \lambda) > 0$. The same is confirmed by the surface on Fig. 3(b).

Respectively, we see from Fig. 4(a) that for $\lambda > 0$ and $\alpha > 0$ (the values of α are chosen from 1.1 to 2.0 with interval 0.1), $R(\alpha, \lambda) < 0$. The same is confirmed by the surface on Fig. 4(b).

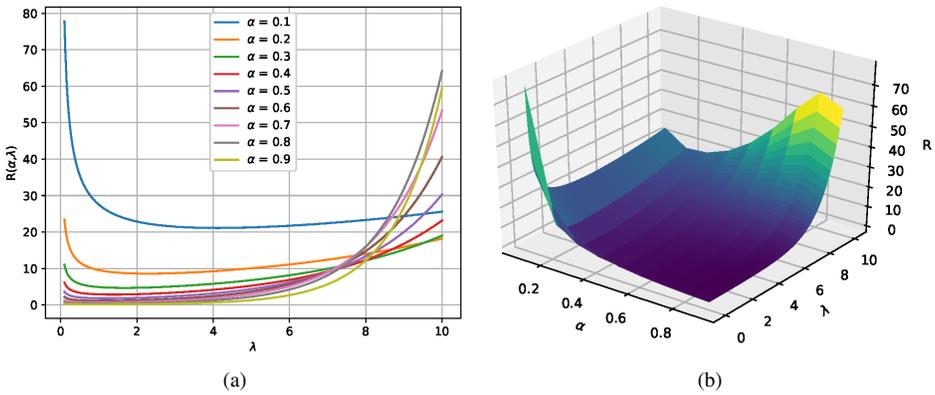


Figure 3. $R(\alpha, \lambda) > 0$ when $0 < \alpha < 1, \lambda > 0$.

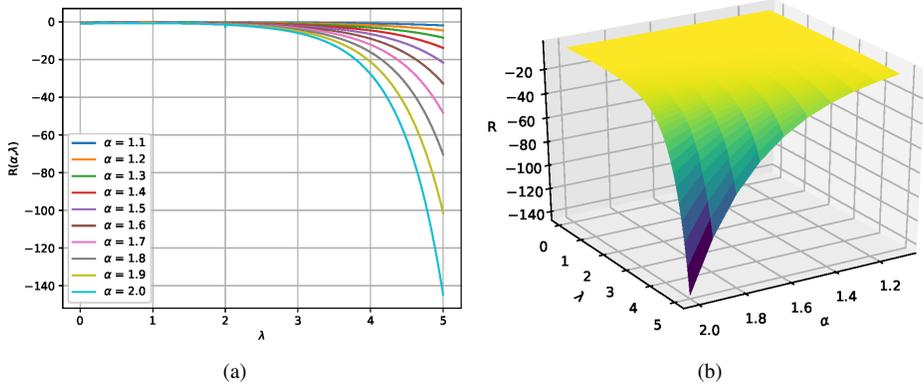


Figure 4. $R(\alpha, \lambda) < 0$ when $\alpha > 1, \lambda > 0$.

Appendix

Lemma A1. *The following relation holds:*

$$\liminf_{\lambda \rightarrow \infty} e^{-\lambda} (\log \lambda)^{-1} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} \geq 1.$$

Proof. It is sufficient to consider $\lambda > 4$. Let us rewrite the sum under the sign of a limit as follows:

$$\begin{aligned} S(\lambda) &:= \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} = \sum_{k=1}^{[\lambda/2]} \frac{\lambda^k \log(k+1)}{k!} + \sum_{k \geq [\lambda/2]+1} \frac{\lambda^k \log(k+1)}{k!} \\ &=: S_1(\lambda) + S_2(\lambda), \end{aligned}$$

where $[\lambda/2]$ is the floor function of $\lambda/2$ (i.e., the greatest integer less than or equal to $\lambda/2$). Obviously,

$$S_1(\lambda) \leq \log \left(\left[\frac{\lambda}{2} \right] + 1 \right) \sum_{k=1}^{[\lambda/2]} \frac{\lambda^k}{k!}.$$

Moreover, $\lambda^k/k!$ increases in $k = 1, \dots, [\lambda/2]$. Therefore

$$S_1(\lambda) \leq \log \left(\left[\frac{\lambda}{2} \right] + 1 \right) \frac{\lambda^{[\lambda/2]}}{[\frac{\lambda}{2}]!} \left[\frac{\lambda}{2} \right].$$

Now, according to two-sided Stirling’s approximation of factorial, for any $n > 1$,

$$\sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{1/(12n+1)} < n! < \sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{1/(12n)}.$$

So, we can bound $e^{-\lambda}(\log \lambda)^{-1}S_1(\lambda)$ as $\lambda \rightarrow \infty$:

$$\begin{aligned} 0 < e^{-\lambda}(\log \lambda)^{-1}S_1(\lambda) \\ &\leq \frac{e^{-\lambda}\lambda^{[\lambda/2]}}{[\frac{\lambda}{2}]!} \cdot \frac{\log([\frac{\lambda}{2}] + 1)}{\log \lambda} \left[\frac{\lambda}{2} \right] \leq \frac{e^{-\lambda}\lambda^{[\lambda/2]}e^{[\lambda/2]}[\frac{\lambda}{2}]}{\sqrt{2\pi}\sqrt{[\frac{\lambda}{2}]}[\frac{\lambda}{2}]^{[\lambda/2]}e^{1/(12[\lambda/2]+1)}}. \end{aligned}$$

Now shift the range of the values of λ under consideration to $\lambda > 42$ and note that for such λ , we have the bounds $[\lambda/2] \geq \lambda/2 - 1 \geq \lambda/2.1$. Therefore for $\lambda > 42$,

$$\begin{aligned} e^{-\lambda}(\log \lambda)^{-1}S_1(\lambda) &\leq \frac{e^{[\lambda/2]-\lambda}(2.1)^{[\lambda/2]}[\frac{\lambda}{2}]}{\sqrt{2\pi}\sqrt{[\frac{\lambda}{2}]}e^{1/(12[\lambda/2]+1)}} < \frac{(\frac{2.1}{e})^{[\lambda/2]}\sqrt{[\frac{\lambda}{2}]}}{\sqrt{2\pi}e^{1/(12[\lambda/2]+1)}} \\ &\xrightarrow{\lambda \rightarrow \infty} 0. \end{aligned} \tag{A.1}$$

Now consider

$$e^{-\lambda}(\log \lambda)^{-1}S_2(\lambda) \geq e^{-\lambda} \frac{\sum_{k=[\lambda/2]+1}^{\infty} \frac{\lambda^k}{k!} \log([\frac{\lambda}{2}] + 1)}{\log \lambda}.$$

Obviously, $\log([\lambda/2] + 1)\log \lambda \rightarrow 1$ as $\lambda \rightarrow \infty$. Therefore

$$\liminf_{\lambda \rightarrow \infty} e^{-\lambda}(\log \lambda)^{-1}S_2(\lambda) \geq \liminf_{\lambda \rightarrow \infty} \frac{\sum_{k=[\lambda/2]+1}^{\infty} \frac{\lambda^k}{k!}}{e^\lambda}. \tag{A.2}$$

Obviously, $e^\lambda = \sum_{k=0}^{\infty} \lambda^k/k!$, while it follows immediately that, similarly to previous calculations in (A.1),

$$\begin{aligned} 0 &\leq \liminf_{\lambda \rightarrow \infty} \frac{\sum_{k=1}^{[\lambda/2]} \frac{\lambda^k}{k!}}{e^\lambda} \leq \limsup_{\lambda \rightarrow \infty} \frac{\frac{\lambda^{[\lambda/2]}}{[\frac{\lambda}{2}]!} \left[\frac{\lambda}{2} \right]}{e^\lambda} \\ &\leq \limsup_{\lambda \rightarrow \infty} \frac{(\frac{2.1}{e})^{[\lambda/2]}\sqrt{[\frac{\lambda}{2}]}}{\sqrt{2\pi}e^{1/(12[\lambda/2]+1)}} = 0. \end{aligned} \tag{A.3}$$

Relation (A.3) implies that the following limit exists:

$$\lim_{\lambda \rightarrow \infty} \frac{\sum_{k=[\lambda/2]+1}^{\infty} \frac{\lambda^k}{k!}}{e^\lambda} = \lim_{\lambda \rightarrow \infty} \frac{e^\lambda - \sum_{k=0}^{[\lambda/2]} \frac{\lambda^k}{k!}}{e^\lambda} = 1. \tag{A.4}$$

The proof immediately follows from (A.1), (A.2), and (A.4). □

Lemma A2 [Karamata’s inequality]. (See [11, 13, 16].) *Let f be a convex function on an interval $I \subset \mathbb{R}$ and $a_1, \dots, a_n, b_1, \dots, b_n$ be numbers in I such that (a_1, \dots, a_n) majorizes (b_1, \dots, b_n) , i.e., the following conditions are fulfilled:*

- (i) $a_1 \geq a_2 \geq \dots \geq a_n, b_1 \geq b_2 \geq \dots \geq b_n$;
- (ii) $\sum_{k=1}^i a_k \geq \sum_{k=1}^i b_k$ for every $1 \leq i \leq n - 1$;
- (iii) $\sum_{k=1}^n a_k = \sum_{k=1}^n b_k$.

Then the inequality $\sum_{k=1}^n f(a_k) \geq \sum_{k=1}^n f(b_k)$ holds.

If f is a concave function on I , the inequality is reversed.

Author contributions. All Authors (V.B., A.M., Yu.M., and Ye.A.R.) have contributed as follows: methodology, V.B. and Yu.M.; formal analysis, V.B., A.M., and Yu.M.; software, Ye.A.R.; validation, A.M. and Yu.M.; writing – original draft preparation, V.B. and Yu.M.; writing – review and editing, V.B., A.M., Yu.M., and Ye.A.R. All authors have read and approved the published version of the manuscript.

Conflicts of interest. The authors declare no conflicts of interest.

Acknowledgment. We would like to thank Prof. Igor Shevchuk for the attention to our work and useful advices.

References

1. J. Acharya, A. Orlitsky, A.T. Suresh, H. Tyagi, Estimating Rényi entropy of discrete distributions, *IEEE Trans. Inf. Theory*, **63**(1):38–56, 2017, <https://doi.org/10.1109/TIT.2016.2620435>.
2. M.A. Aladaileh, M. Anbar, A.J. Hintaw, I.H. Hasbullah, A.A. Bahashwan, T.A. Al-Amiedy, D.R. Ibrahim, Effectiveness of an entropy-based approach for detecting low- and high-rate DDoS attacks against the SDN controller: Experimental analysis, *Appl. Sci.*, **13**(2):775, 2023, <https://doi.org/10.3390/app13020775>.
3. E. Archer, I.M. Park, J.W. Pillow, Bayesian entropy estimation for countable discrete distributions, *J. Mach. Learn. Res.*, **15**:2833–2868, 2014.
4. F. Buryak, Yu. Mishura, Convexity and robustness of the Rényi entropy, *Mod. Stoch., Theory Appl.*, **8**(3):387–412, 2021, <https://doi.org/10.15559/21-vmsta185>.
5. M. Cheraghchi, Expressions for the entropy of basic discrete distributions, *IEEE Trans. Inf. Theory*, **65**(7):3999–4009, 2019, <https://doi.org/10.1109/TIT.2019.2900716>.

6. J. David, C. Thomas, DDoS attack detection using fast entropy approach on flow-based network traffic, *Procedia Comput. Sci.*, **50**:30–36, 2015, <https://doi.org/https://doi.org/10.1016/j.procs.2015.04.007>.
7. N. Ebrahimi, E. Maasoumi, E.S. Soofi, Ordering univariate distributions by entropy and variance, *J. Econometrics*, **90**(2):317–336, 1999, [https://doi.org/10.1016/S0304-4076\(98\)00046-3](https://doi.org/10.1016/S0304-4076(98)00046-3).
8. R.J. Evans, J. Boersma, The entropy of a Poisson distribution (C. Robert Appledorn), *SIAM Rev.*, **30**(2):314–317, 1988, <https://doi.org/10.1137/1030059>.
9. Y. Han, J. Jiao, T. Weissman, Adaptive estimation of Shannon entropy, in *2015 IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, China, 2015, pp. 1372–1376, <https://doi.org/10.1109/ISIT.2015.7282680>.
10. J. Jiao, K. Venkat, Y. Han, T. Weissman, Minimax estimation of functionals of discrete distributions, *IEEE Trans. Inf. Theory*, **61**(5):2835–2885, 2015, <https://doi.org/10.1109/TIT.2015.2412945>.
11. Z. Kadelburg, D. Đukić, M. Lukić, I. Matić, Inequalities of Karamata, Schur and Muirhead, and some applications, *The Teaching of Mathematics*, **14**:31–45, 2005, <http://eudml.org/doc/256196>.
12. Y. Karaca, Y.-D. Zhang, K. Muhammad, Characterizing complexity and self-similarity based on fractal and entropy analyses for stock market forecast modelling, *Expert Syst. Appl.*, **144**:113098, 2020, <https://doi.org/10.1016/j.eswa.2019.113098>.
13. J. Karamata, Sur une inégalité relative aux fonctions convexes, *Publ. Inst. Math.*, **1**(1):145–147, 1932, <http://eudml.org/doc/254514>.
14. A. Lall, V. Sekar, M. Ogihara, J. Xu, H. Zhang, Data streaming algorithms for estimating entropy of network traffic, in *SIGMETRICS '06/Performance '06: Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems, Saint Malo, France, June 26–30, 2006*, ACM, New York, 2006, pp. 145–156, <https://doi.org/10.1145/1140277.1140295>.
15. A. Malyarenko, Yu. Mishura, K. Ralchenko, Ye.A. Rudyk, Properties of various entropies of Gaussian distribution and comparison of entropies of fractional processes, *Axioms*, **12**(11):1026, 2023, <https://doi.org/10.3390/axioms12111026>.
16. A.W. Marshall, I. Olkin, B.C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., Springer Ser. Stat., Springer, New York, 2011, <https://doi.org/10.1007/978-0-387-68276-1>.
17. L. Paninski, Estimation of entropy and mutual information, *Neural Comput.*, **15**(6):1191–1253, 2003, <https://doi.org/10.1162/089976603321780272>.
18. I. Raşa, Complete monotonicity of some entropies, *Period. Math. Hung.*, **75**(2):159–166, 2017, <https://doi.org/10.1007/s10998-016-0177-5>.
19. A. Rényi, On measures of entropy and information, in J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Held at the Statistical Laboratory University of California June 20–July 30, 1960. Vol. I: Contributions to the Theory of Statistics*, Univ. California Press, Los Angeles, CA, 1960, pp. 547–561.
20. C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**:379–423, 623–656, 1948, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

21. A. Singhal, D.K. Sharma, Keyword extraction using Rényi entropy: A statistical and domain independent method, in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, Vol. 1*, IEEE, 2021, pp. 1970–1975, <https://doi.org/10.1109/ICACCS51430.2021.9441909>.
22. Y. Wu, P. Yang, Minimax rates of entropy estimation on large alphabets via best polynomial approximation, *IEEE Trans. Inf. Theory*, **62**(6):3702–3720, 2016, <https://doi.org/10.1109/TIT.2016.2548468>.