

Article

# Asymptotic Distribution of Certain Types of Entropy under the Multinomial Law

Andrea A. Rey <sup>1</sup>, Alejandro C. Frery <sup>2,\*</sup>, Magdalena Lucini <sup>3</sup>, Juliana Gambini <sup>4</sup>, Eduarda T. C. Chagas <sup>5</sup>  
and Heitor S. Ramos <sup>5</sup>

<sup>1</sup> Signal and Image Processing Center, Universidad Tecnológica Nacional, Ciudad Autónoma de Buenos Aires C1179AAQ, Argentina

<sup>2</sup> School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6140, New Zealand

<sup>3</sup> Department of Mathematics, FaCENA, Universidad Nacional del Nordeste and CONICET, Corrientes W3400Ayy, Argentina

<sup>4</sup> CIDIA, Universidad Nacional de Hurlingham, Pcia. de Buenos Aires Argentina—CPSI Universidad Tecnológica Nacional, Buenos Aires C1041AAJ, Argentina

<sup>5</sup> Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil

\* Correspondence: alejandro.frery@vuw.ac.nz

**Simple Summary:** We obtain expressions for the asymptotic distributions of the Rényi and Tsallis of order  $q$  entropies, and Fisher information when computed on the maximum likelihood estimator of probabilities from multinomial random samples. We recall results related to the Shannon entropy. We build a test for comparing entropies of different types and categories.

**Abstract:** We obtain expressions for the asymptotic distributions of the Rényi and Tsallis of order  $q$  entropies and Fisher information when computed on the maximum likelihood estimator of probabilities from multinomial random samples. We verify that these asymptotic models, two of which (Tsallis and Fisher) are normal, describe well a variety of simulated data. In addition, we obtain test statistics for comparing (possibly different types of) entropies from two samples without requiring the same number of categories. Finally, we apply these tests to social survey data and verify that the results are consistent but more general than those obtained with a  $\chi^2$  test.

**Keywords:** multinomial distribution; entropy; asymptotic distributions; hypothesis tests



**Citation:** Rey, A.A.; Frery, A.C.; Lucini, M.; Gambini, J.; Chagas, E.T.C.; Ramos, H.S. Asymptotic Distribution of Certain Types of Entropy under the Multinomial Law. *Entropy* **2023**, *25*, 734. <https://doi.org/10.3390/e25050734>

Academic Editor: Takuya Yamano

Received: 14 February 2023

Revised: 24 March 2023

Accepted: 21 April 2023

Published: 28 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The multinomial distribution is an adequate model for describing how observations fall into categories. Quoting Johnson et al. [1], “The Multinomial distribution, like the Multivariate Normal distribution among the continuous multivariate distributions, consumed a sizable amount of the attention that numerous theoretical as well as applied researchers directed towards the area of discrete multivariate distributions.”

The entropy of a (multivariate, in our case) random variable is a substantial quantity. It quantifies the predictability of a system whose outputs can be described by such a model. Entropy has several definitions, both conceptual and mathematical. The concept of entropy originated as a way to relate a system’s energy and temperature [2]. The same concept was used to describe the number of ways the particles of a system can be arranged.

Entropy has been seldom studied as a random variable. Hutcheson [3] and Hutcheson and Shenton [4] discussed the exact expected value and variance of the Shannon entropy under the multinomial model. These works also provided approximate expressions that circumvent the numerical issues when using the exact value.

Jacquet and Szpankowski [5] studied high-quality analytic approximations of the Rényi entropy, of which the Shannon entropy is a particular case, under the binomial model. With the same approach, Cichoń and Gołębiewski [6] obtained expressions for

more general functionals that include the multinomial distribution. These works treat the entropy as a fixed quantity. Cook et al. [7] studied almost unbiased estimators of functions of the parameter of the binomial distribution. The authors extended those results to find an almost-unbiased estimator for the entropy under multinomial laws.

Chagas et al. [8] treated the Shannon entropy as a random variable. The authors obtained its asymptotic distribution when indexing by the maximum likelihood estimators of the proportions under the multinomial distribution. This result allowed the devising of unilateral and bilateral tests for comparing the entropy from two samples in a very general way. These tests do not require having the same number of categories.

In this work, our attention is directed toward the asymptotic distribution of other forms of entropy under the multinomial model. This allows the comparison of large samples throughout their entropies and, with this, they may have different numbers of classes. The comparison also allows using different types of entropy. We firstly apply the multivariate delta method and, in the case of the Rényi entropy, we transform the resulting multivariate normal distribution into that of the logarithm of the absolute value of a normally distributed random variable. Then, we provide the general expression of a test statistic that suits our needs.

This paper unfolds as follows. Section 2 recalls the main properties of the multinomial distribution and defines the four types of entropies we will study. In Section 3, we present the central results, i.e., the asymptotic distribution of those entropies. We describe the techniques we used and left for Appendix A.1 technical details. We validate our results with simulation studies in Section 4: we show the adequacy of the normal distribution as limit law for the entropies under three probability models of different support, considering various sample sizes. In Section 5, we show that these asymptotic properties lead to a helpful hypothesis test between samples with different categories. We conclude the article in Section 6. Appendix A.2 comments on applications that justify our choices of the number of categories and sample sizes in the simulation studies. Appendix A.3 discloses relevant computational information, including reproducibility.

## 2. Entropies and the Multinomial Distribution

Consider a series of  $n$  independent trials, where only one of  $k$  mutually exclusive events  $\pi^1, \pi^2, \dots, \pi^k$  must be observed in each one, with probability  $\mathbf{p} = \{p_1, p_2, \dots, p_k\}$  such that  $p_\ell \geq 0$  and  $\sum_{\ell=1}^k p_\ell = 1$ . Let  $\mathbf{N} = (N_1, N_2, \dots, N_k)$  be the random vector that counts the number of occurrences of the events  $\pi^1, \pi^2, \dots, \pi^k$  in the  $n$  trials, with  $N_\ell \geq 0$  and  $\sum_{\ell=1}^k N_\ell = n$ . A sample from  $\mathbf{N}$ , say  $\mathbf{n}$ , is a  $k$ -variate vector of integer values  $\mathbf{n} = (n_1, n_2, \dots, n_k)$ . Then, the joint distribution of  $\mathbf{N}$  is

$$\Pr(\mathbf{N} = \mathbf{n}) = \Pr(N_1 = n_1, N_2 = n_2, \dots, N_k = n_k) = n! \prod_{\ell=1}^k \frac{p_\ell^{n_\ell}}{n_\ell!}. \quad (1)$$

We denote this situation as  $\mathbf{N} \sim \text{Mult}(n, \mathbf{p})$ .

In practice, one does not know the true values of  $\mathbf{p}$ , the probabilities that index this multinomial distribution. Such values are estimated by computing  $\hat{p}_\ell$ , the proportion of times the class (category, event)  $\pi^\ell$  was observed among the  $k$  possible categories  $\boldsymbol{\pi} = \{\pi^1, \pi^2, \dots, \pi^k\}$  during the  $n$  trials. The maximum likelihood estimator for  $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$  is the random vector of proportions. This maximum likelihood estimator coincides with the intuitive estimator based on the distribution's first moments, and is the most frequently used in applications.

We study the distribution of several forms of entropy of the random vector  $\hat{\mathbf{p}}$  for fixed  $k$ . Notice that  $\hat{\mathbf{p}}$  is computed over a single  $k$ -variate measurement of random proportions corresponding to a single random sample from  $\mathbf{N} \sim \text{Mult}(n, \mathbf{p})$ . The asymptotic behaviors we derive hold for typical cases in which  $n \gg k$ .

The Shannon entropy measures the disorder or unpredictability of systems characterized by a probability distribution. On the one hand, the minimum Shannon value occurs

when there is complete knowledge about the system behavior and total confidence in predicting the following observation. On the other hand, when a uniform distribution describes the system’s behavior, that is, when all possibilities have the same probability of occurrence, the knowledge about the behavior of the data is minimal. In Chagas et al. [8], we studied the asymptotic distribution of the Shannon entropy. In this work, we extend those results to three other forms of entropy.

Other types of descriptors have been proposed in the literature to extract additional information not captured by the Shannon entropy. Tsallis [9] and Rényi [10], for instance, proposed parametric versions, which include the Shannon entropy.

Fisher information [11] is defined by an average logarithm derivative of a continuous probability density function. In the case of discrete densities, this measure can be approximated using differences of probabilities between consecutive distribution elements. While the Shannon entropy captures the degree of unpredictability of a system, the Fisher information is related to the rate of change of consecutive observations and, thus, quantifies small changes and perturbations.

Given a type of entropy  $H$ , we are interested in the distribution of  $H(\mathbf{p})$  when indexed by  $\hat{\mathbf{p}}$ , the maximum likelihood estimator of  $\mathbf{p}$ . Our problem then becomes finding the distribution of  $H(\hat{\mathbf{p}})$  for the following:

- The Shannon entropy

$$H_S(\hat{\mathbf{p}}) = - \sum_{\ell=1}^k \hat{p}_\ell \log \hat{p}_\ell, \tag{2}$$

- The Tsallis entropy with index  $q \in \mathbb{R} \setminus \{1\}$

$$H_T^q(\hat{\mathbf{p}}) = \sum_{\ell=1}^k \frac{\hat{p}_\ell - \hat{p}_\ell^q}{q - 1}, \tag{3}$$

- The Rényi entropy of order  $q \in \mathbb{R}^+ \setminus \{1\}$

$$H_R^q(\hat{\mathbf{p}}) = \frac{1}{1 - q} \log \sum_{\ell=1}^k \hat{p}_\ell^q, \tag{4}$$

- The Fisher information, also termed “Fisher Information Measure” in the literature, with renormalization coefficient  $F_0 = 4$

$$H_F(\hat{\mathbf{p}}) = F_0 \sum_{\ell=1}^{k-1} (\sqrt{\hat{p}_{\ell+1}} - \sqrt{\hat{p}_\ell})^2. \tag{5}$$

Among other possibilities, we used Equation (2.7) from Ref. [12].

### 3. Asymptotic Distributions of Entropies

The main results of this section are the asymptotic distributions of the Shannon (2), Tsallis of order  $q$  (3), and Rényi of order  $q$  (4) entropies, and Fisher information (5). These results are presented, respectively, in Equations (30)–(32) and (35). Notice that the Rényi entropy is not asymptotically normally distributed, while the other three are.

We recall the following theorems known respectively as the delta method and its multivariate version. We refer to Lehmann and Casella [13] for their proofs.

**Theorem 1.** *Let  $X_n$  be a sequence of independent and identically distributed random variables such that  $\sqrt{n}[X_n - \theta]$  converges in distribution to a  $\mathcal{N}(0, \sigma^2)$ . If  $\partial h / \partial \theta$  exists and does not vanish, then  $\sqrt{n}[h(X_n) - h(\theta)]$  converges in distribution to a  $\mathcal{N}(0, \sigma^2[\partial h / \partial \theta]^2)$ .*

**Theorem 2.** *Let  $\mathbf{X}_n = (X_{1n}, X_{2n}, \dots, X_{kn})$  be a sequence of independent and identically distributed vectors of random variables such that  $\sqrt{n}[X_{1n} - \theta_1, X_{2n} - \theta_2, \dots, X_{kn} - \theta_k]$  converges in*

distribution to a multivariate normal distribution  $\mathcal{N}_n(\mathbf{0}, \Sigma)$ , where  $\Sigma$  is the covariance matrix. Suppose that  $h_1, h_2, \dots, h_k$  are real functions continuously differentiable in a neighborhood of the parameter point  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$  and such that the matrix of partial derivatives  $B = (\partial h_\ell / \partial \theta_j)_{\ell, j=1}^k$  is non-singular in the mentioned neighborhood. Then, the following convergence in distribution holds:

$$\sqrt{n} [h_1(\mathbf{X}_n) - h_1(\boldsymbol{\theta}), h_2(\mathbf{X}_n) - h_2(\boldsymbol{\theta}), \dots, h_k(\mathbf{X}_n) - h_k(\boldsymbol{\theta})] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, B\Sigma B'),$$

where  $B'$  denotes the transpose of  $B$ .

Now, we focus on the case  $N \sim \text{Mult}(n, \mathbf{p})$ . Let  $\hat{\mathbf{p}} = N/n$  be the vector of sample proportions which coincides with the maximum likelihood estimator (MLE) of  $\mathbf{p}$  and  $\mathbf{Y}_n = \sqrt{n}(\hat{\mathbf{p}} - \mathbf{p})$ . Then

$$\mathbf{Y}_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{D}_p - \mathbf{p}\mathbf{p}'),$$

where  $\mathbf{D}_p = \text{Diag}(p_1, p_2, \dots, p_k)$ .

Let us explore the covariance matrix in this case:

$$\mathbf{D}_p - \mathbf{p}\mathbf{p}' = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_k \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix} \begin{pmatrix} p_1 & p_2 & \cdots & p_k \end{pmatrix} \tag{6}$$

$$= \begin{pmatrix} p_1 - p_1^2 & -p_1 p_2 & \cdots & -p_1 p_k \\ -p_2 p_1 & p_2 - p_2^2 & \cdots & -p_2 p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_k p_1 & -p_k p_2 & \cdots & p_k - p_k^2 \end{pmatrix} \tag{7}$$

It means that the covariance matrix  $\Sigma_{\mathbf{p}} \in \mathbb{R}^{k \times k}$  we are interested in is of the form

$$(\Sigma_{\mathbf{p}})_{\ell j} = \begin{cases} p_\ell(1 - p_\ell) & \text{if } \ell = j, \\ -p_\ell p_j & \text{if } \ell \neq j. \end{cases} \tag{8}$$

The above statements are generalized. In the following, we obtain new results for the Tsallis and Rényi entropies, and for the Fisher information. For the sake of completeness, we also include the results for the Shannon entropy.

In order to apply the delta method using Theorem 2, we consider the following functions:

$$h_\ell^S(p_1, p_2, \dots, p_k) = p_\ell \log p_\ell, \tag{9}$$

$$h_\ell^T(p_1, p_2, \dots, p_k) = p_\ell - p_\ell^q, \tag{10}$$

$$h_\ell^R(p_1, p_2, \dots, p_k) = p_\ell^q, \tag{11}$$

$$h_\ell^F(p_1, p_2, \dots, p_k) = (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})^2, \tag{12}$$

for  $\ell = 1, 2, \dots, k$  except for (12) that holds for  $\ell = 1, 2, \dots, k - 1$ . The assumptions are verified, and thus,

$$\frac{\partial h_\ell^S}{\partial p_\ell} = \log p_\ell + 1 \text{ and } \frac{\partial h_\ell^S}{\partial p_j} = 0 \text{ if } j \neq \ell, \tag{13}$$

$$\frac{\partial h_\ell^T}{\partial p_\ell} = 1 - qp_\ell^{q-1} \text{ and } \frac{\partial h_\ell^T}{\partial p_j} = 0 \text{ if } j \neq \ell, \tag{14}$$

$$\frac{\partial h_\ell^R}{\partial p_\ell} = qp_\ell^{q-1} \text{ and } \frac{\partial h_\ell^R}{\partial p_j} = 0 \text{ if } j \neq \ell, \tag{15}$$

$$\frac{\partial h_\ell^F}{\partial p_j} = \frac{\sqrt{p_{\ell+1}} - \sqrt{p_\ell}}{(-1)^{\ell+j-1}\sqrt{p_j}} \text{ if } j = \ell, \ell + 1 \text{ and } \frac{\partial h_\ell^F}{\partial p_j} = 0 \text{ if } j \neq \ell, \ell + 1. \tag{16}$$

Finally, we need the covariance matrix of the multivariate normal limit distribution, which is

$$\Sigma_p^{\Delta \mathcal{M}} = \left( \frac{\partial h_\ell^{\mathcal{M}}}{\partial p_j} \right)_{\ell,j=1}^k \Sigma_p \left( \frac{\partial h_\ell^{\mathcal{M}}}{\partial p_j} \right)_{\ell,j=1}^k, \tag{17}$$

where  $\mathcal{M} \in \{S, T, R, F\}$ . Since  $(\partial h_\ell^{\mathcal{M}} / \partial p_j)_{\ell,j=1}^k$  are diagonal matrices for  $\mathcal{M} \in \{S, T, R\}$ , we can use Equation (A1) to conclude that

$$(\Sigma_p^{\Delta S})_{\ell j} = \begin{cases} (p_\ell - p_\ell^2)(\log p_\ell + 1)^2 & \text{if } \ell = j, \\ -p_\ell p_j (\log p_\ell + 1)(\log p_j + 1) & \text{if } \ell \neq j; \end{cases} \tag{18}$$

$$(\Sigma_p^{\Delta T})_{\ell j} = \begin{cases} (p_\ell - p_\ell^2)(1 - qp_\ell^{q-1})^2 & \text{if } \ell = j, \\ -p_\ell p_j (1 - qp_\ell^{q-1})(1 - qp_j^{q-1}) & \text{if } \ell \neq j; \end{cases} \tag{19}$$

$$(\Sigma_p^{\Delta R})_{\ell j} = \begin{cases} q^2(p_\ell - p_\ell^2)p_\ell^{2(q-1)} & \text{if } \ell = j, \\ -q^2(p_\ell p_j)^q & \text{if } \ell \neq j. \end{cases} \tag{20}$$

In the case of  $\Sigma_p^{\Delta F}$ , from Equations (A3) and (A4) we have the following:

- For  $\ell, j = 1, 2, \dots, k - 2$  and  $\ell \neq j - 1, j, j + 1$ :

$$(\Sigma_p^{\Delta F})_{\ell j} = (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})(\sqrt{p_{j+1}} - \sqrt{p_j})(\sqrt{p_{\ell+1}p_j} + \sqrt{p_\ell p_{j+1}} - \sqrt{p_\ell p_j} - \sqrt{p_{\ell+1}p_{j+1}}). \tag{21}$$

- For  $\ell = 1, 2, \dots, k - 2$ :

$$(\Sigma_p^{\Delta F})_{\ell, \ell-1} = (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})(\sqrt{p_\ell} - \sqrt{p_{\ell-1}})(\sqrt{p_{\ell+1}p_{\ell-1}} + p_\ell - 1 - \sqrt{p_\ell p_{\ell-1}} - \sqrt{p_{\ell+1}p_\ell}). \tag{22}$$

- For  $\ell = 1, 2, \dots, k - 2$ :

$$(\Sigma_p^{\Delta F})_{\ell \ell} = (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})^2 (2\sqrt{p_\ell p_{\ell+1}} + 2 - p_\ell - p_{\ell+1}). \tag{23}$$

- For  $\ell = 1, 2, \dots, k - 2$ :

$$(\Sigma_p^{\Delta F})_{\ell, \ell+1} = (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})(\sqrt{p_{\ell+2}} - \sqrt{p_{\ell+1}})(p_{\ell+1} - 1 + \sqrt{p_\ell p_{\ell+2}} - \sqrt{p_\ell p_{\ell+1}} - \sqrt{p_{\ell+1}p_{\ell+2}}). \tag{24}$$

- For  $j = 1, 2, \dots, k - 2$ :

$$(\Sigma_p^{\Delta F})_{k-1, j} = (\sqrt{p_k} - \sqrt{p_{k-1}})(\sqrt{p_{j+1}} - \sqrt{p_j}) \left( \frac{p_k p_{j+1}}{\sqrt{p_{j+1}}} - \frac{p_{k-1} p_j}{\sqrt{p_j}} \right). \tag{25}$$

- Finally,

$$(\Sigma_p^{\Delta F})_{k-1, k-1} = (\sqrt{p_k} - \sqrt{p_{k-1}})^2 (1 - p_{k-1}). \tag{26}$$

Hence, we conclude that

$$\sqrt{n}[h_1^M(\hat{p}_1) - h_1^M(p_1), h_2^M(\hat{p}_2) - h_2^M(p_2), \dots, h_{k'}^M(\hat{p}_{k'}) - h_{k'}^M(p_{k'})] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma_p^{\Delta M}), \tag{27}$$

where  $\mathcal{M} \in \{S, T, R, F\}$  and  $k' = k$  in all cases except for the case of the Fisher information in which  $k' = k - 1$ . An equivalent expression is

$$\sqrt{n}[h_1^M(\hat{p}_1), h_2^M(\hat{p}_2), \dots, h_{k'}^M(\hat{p}_{k'})] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{n} \begin{pmatrix} h_1^M(p_1) \\ h_2^M(p_2) \\ \vdots \\ h_{k'}^M(p_{k'}) \end{pmatrix}, \Sigma_p^{\Delta M}\right). \tag{28}$$

If  $\mathbf{Y}$  is a vector of random variables such that  $\sqrt{n}\mathbf{Y} \xrightarrow{\mathcal{D}} \mathcal{N}(\sqrt{n}\boldsymbol{\mu}, \Sigma)$ , then it can be proved that  $E(\sqrt{n}\mathbf{Y}) \rightarrow \sqrt{n}\boldsymbol{\mu}$  and  $\text{Var}(\sqrt{n}\mathbf{Y}) \rightarrow \Sigma$ . Provided well-known properties, it holds that  $E(\mathbf{Y}) \rightarrow \boldsymbol{\mu}$  and  $\text{Var}(\mathbf{Y}) \rightarrow 1/n\Sigma$ . Applying this to (28),

$$[h_1^M(\hat{p}_1), h_2^M(\hat{p}_2), \dots, h_{k'}^M(\hat{p}_{k'})] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} h_1^M(p_1) \\ h_2^M(p_2) \\ \vdots \\ h_{k'}^M(p_{k'}) \end{pmatrix}, \frac{1}{n}\Sigma_p^{\Delta M}\right). \tag{29}$$

Now, using (29), we find the asymptotic distribution of (2)–(5). In order to do so, we need to know the distribution of the sum of  $k$  Gaussian random variables with different means and an arbitrary covariance matrix.

For any  $k$ -dimensional multivariate normal distribution  $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , with  $\boldsymbol{\mu} \in \mathbb{R}^k$  and covariance matrix  $\Sigma = (\sigma_{\ell j})$ , holds that the distribution of  $W = \mathbf{a}^T \mathbf{Z}$ , with  $\mathbf{a} \in \mathbb{R}^k$ , is  $\mathcal{N}(\mathbf{a}^T \boldsymbol{\mu}, \sum_{\ell=1}^k a_\ell^2 \sigma_{\ell\ell} + 2 \sum_{\ell=1}^{k-1} \sum_{j=\ell+1}^k a_\ell a_j \sigma_{\ell j})$ . Using the limit distribution presented in (29) and  $\mathbf{a} = (-1, -1, \dots, -1)$ , we directly have the asymptotic distribution of the Shannon entropy as follows:

$$H_S(\hat{\mathbf{p}}) = - \sum_{\ell=1}^k \hat{p}_\ell \log \hat{p}_\ell \xrightarrow{\mathcal{D}} \mathcal{N}\left(- \sum_{\ell=1}^k p_\ell \log p_\ell, \frac{1}{n} \sum_{\ell=1}^k p_\ell(1 - p_\ell)(\log p_\ell + 1)^2 - \frac{2}{n} \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k p_\ell p_j (\log p_\ell + 1)(\log p_j + 1)\right). \tag{30}$$

With similar arguments and  $\mathbf{a} = (1, 1, \dots, 1)$ , we obtain the asymptotic distribution for the Tsallis entropy of order  $q$ :

$$H_T^q(\hat{\mathbf{p}}) = \sum_{\ell=1}^k \frac{\hat{p}_\ell - \hat{p}_\ell^q}{q - 1} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sum_{\ell=1}^k \frac{p_\ell - p_\ell^q}{q - 1}, \sum_{\ell=1}^k \frac{(p_\ell - p_\ell^2)(1 - qp_\ell^{q-1})^2}{n(q - 1)^2} - 2 \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k \frac{p_\ell p_j (1 - qp_\ell^{q-1})(1 - qp_j^{q-1})}{n(q - 1)^2}\right). \tag{31}$$

The procedure is analogous for the Fisher information but with  $\mathbf{a} = (1, 1, \dots, 1) \in \mathbb{R}^{k-1}$ . Hence, it can be proved that

$$H_F(\hat{\mathbf{p}}) = F_0 \sum_{\ell=1}^{k-1} (\sqrt{\hat{p}_{\ell-1}} - \sqrt{\hat{p}_\ell})^2 \xrightarrow{\mathcal{D}} \mathcal{N}\left(F_0 \sum_{\ell=1}^{k-1} (\sqrt{p_{\ell-1}} - \sqrt{p_\ell})^2, \frac{F_0}{n}\Sigma^*\right), \tag{32}$$

where

$$\begin{aligned} \Sigma^* &= (\sqrt{p_k} - \sqrt{p_{k-1}})^2(1 - p_{k-1}) + \sum_{\ell=1}^{k-2} (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})^2(2\sqrt{p_\ell p_{\ell+1}} - p_\ell - p_{\ell+1} + 2) + \\ &2 \sum_{\ell=3}^{k-2} \sum_{j=1}^{\ell-2} \left[ (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})(\sqrt{p_{j+1}} - \sqrt{p_j})(\sqrt{p_{\ell+1}p_j} + \sqrt{p_\ell p_{j+1}} - \sqrt{p_\ell p_j} - \sqrt{p_{\ell+1}p_{j+1}}) \right] + \\ &2 \sum_{j=1}^{k-2} \left[ (\sqrt{p_k} - \sqrt{p_{k-1}})(\sqrt{p_{j+1}} - \sqrt{p_j})(p_k \sqrt{p_{j+1}} - p_{k-1} \sqrt{p_j}) \right] + \\ &2 \sum_{\ell=2}^{k-2} \left[ (\sqrt{p_{\ell+1}} - \sqrt{p_\ell})(\sqrt{p_\ell} - \sqrt{p_{\ell-1}})(\sqrt{p_{\ell+1}p_{\ell-1}} - \sqrt{p_\ell p_{\ell-1}} - \sqrt{p_{\ell+1}p_\ell} + p_\ell - 1) \right]. \end{aligned} \tag{33}$$

To obtain expression (33), we use the symmetry of the covariance matrix which implies that  $\sum_{\ell=1}^{k-1} \sum_{j=\ell+1}^k a_\ell a_j \sigma_{\ell j} = \sum_{\ell=2}^{k-1} \sum_{j=1}^{\ell-1} a_\ell a_j \sigma_{\ell j}$ . It is worth noticing that the expression of the covariance matrix for Fisher information is more complicated than the previously analyzed entropies since the matrix of partial derivatives is not diagonal in this case.

The case of Rényi entropy is different because, following the previous methodology, we can prove that

$$\sum_{\ell=1}^k \hat{p}_\ell^q \xrightarrow{\mathcal{D}} \mathcal{N} \left( \sum_{\ell=1}^k p_\ell^q, \frac{1}{n} \sum_{\ell=1}^k q^2(p_\ell - p_\ell^2) p_\ell^{2(q-1)} - \frac{2}{n} \sum_{\ell=1}^{k-1} \sum_{j=\ell+1}^k q^2(p_j p_\ell)^q \right). \tag{34}$$

Hence,

$$H_R^q(\hat{\mathbf{p}}) = \frac{1}{1-q} \log \sum_{\ell=1}^k \hat{p}_\ell^q \xrightarrow{\mathcal{D}} \mathcal{P}_R^q \tag{35}$$

where

$$\mathcal{P}_R^q(x) = \frac{1-q}{\sigma^* \sqrt{2\pi}} \exp[(1-q)x \log(k)] \exp \left\{ -\frac{1}{2} \left( \frac{\exp[(1-q)x \log(k)] - \mu^*}{\sigma^*} \right)^2 \right\}, \tag{36}$$

with  $\mu^* = \sum_{\ell=1}^k p_\ell^q$  and  $\sigma^* = n^{-1} \sum_{\ell=1}^k q^2(p_\ell - p_\ell^2) p_\ell^{2(q-1)} - 2n^{-1} \sum_{\ell=1}^{k-1} \sum_{j=\ell+1}^k q^2(p_\ell p_j)^q$ . Notice that this is not a normal distribution but the distribution of the logarithm of the absolute value of a normally distributed random variable.

Often, in practice, these entropies are scaled to be in  $[0, 1]$ ; these are called “normalized entropies”. The following modifications must be considered in the normalized versions of the entropies. For the normalized Shannon entropy, the asymptotic mean and variance are multiplied by  $1/\log k$  and  $1/(\log k)^2$ , respectively. In the case of the normalized Tsallis entropy, the asymptotic mean and variance are multiplied by  $(q-1)/(1-k^{1-q})$  and  $(q-1)^2/(1-k^{1-q})^2$ , respectively. Finally, the asymptotic distribution of the normalized Rényi entropy is  $\mathcal{P}_R^q(x) = \log k \mathcal{P}_R^q(x \log k)$ . Notice that normalized entropies do not depend on the logarithm basis. The Fisher information is, as defined in (5), already normalized.

#### 4. Analysis and Validation

In this section, we study the empirical distribution of the entropies computed from  $\hat{\mathbf{p}}$  under three models, four categories ( $k \in \{6, 24, 120, 720\}$ ), and three sample sizes ( $n \in \{10^2 k, 10^3 k, 10^4 k\}$ ) that depend on the number of categories. These choices of  $k$  and  $n$  are based on the values that appear in signal analysis with ordinal patterns; see details of this technique in Appendix A.2.

We considered the following probability functions  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ :

1. Linear:  $p_\ell = 2\ell/(k(k+1)), 1 \leq \ell \leq k$ .

2. One-Almost-Zero:  $p_\ell = 1/k$  for  $1 \leq \ell \leq k - 2$ ,  $p_{k-1} = \epsilon_0$ , and  $p_k = 2/k - \epsilon_0$  with  $\epsilon_0 = 2.220446 \times 10^{-16}$  (the smallest positive number for which, in our computer platform,  $1 + \epsilon_0 > 1$ ).
3. Half-and-Half:  $p_\ell = 1/k + \epsilon/k$  for  $1 \leq \ell \leq k/2$ , and  $p_\ell = 1/k - \epsilon/k$  for  $k/2 + 1 \leq \ell \leq k$ , with  $\epsilon \in \{0.1, 0.3, 0.5, 0.8\}$ .

These probability functions are illustrated, for  $k = 6$  and  $\epsilon = 0.3$ , in Figure 1. We studied the behavior of the Shannon entropy, the Rényi entropy with  $q \in \{1/3, 2/5\}$ , the Tsallis entropy with  $q \in \{1/2, 3/2\}$ , and the Fisher information computed on samples of sizes  $n \in \{10^2k, 10^3k, 10^4k\}$ . We used 300 independent samples (replicates).

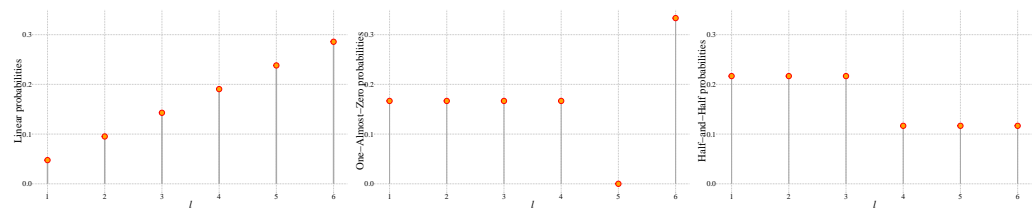


Figure 1. Linear, One-Almost-Zero, and Half-and-Half probability functions for  $k = 6$  and  $\epsilon = 0.3$ .

Although Equation (35) shows that the Rényi entropy is not asymptotically normal, we verified that its density is similar to that of a Gaussian distribution. With this in mind, we also checked of the normality of Rényi entropies. We used the Anderson–Darling test to verify the null hypothesis that the data follow a normal distribution. We chose this test because it uses the hypothesized distribution in calculating critical values. This test is more sensitive than other alternatives; see, for instance, the book by Lehman and Romano [14].

From Table 1, we notice that the Fisher information is the one that fails most times to pass the normality test at the 1%. The situation that appears with  $p$ -value = 0.0010 in the table has, in fact,  $p$ -value =  $9.606130 \times 10^{-3}$ ; the table shows rounded values. Figure 2 shows four of these cases, namely for  $k = 6$ ,  $n = 600$ , and  $\epsilon = 0.1, 0.3, 0.5, 0.8$ . We notice that the deviation from the normal hypothesis is more prevalent in both tails, being that the observations are larger than the theoretical quantiles.

Table 1. Situations for which the  $p$ -values of the Anderson–Darling test for the normality of samples of size 300 are less than 0.01 (“HF” stands for the Fisher information; “HaH” and “OAZ” are the Half-And-Half and One-Almost-Zero models).

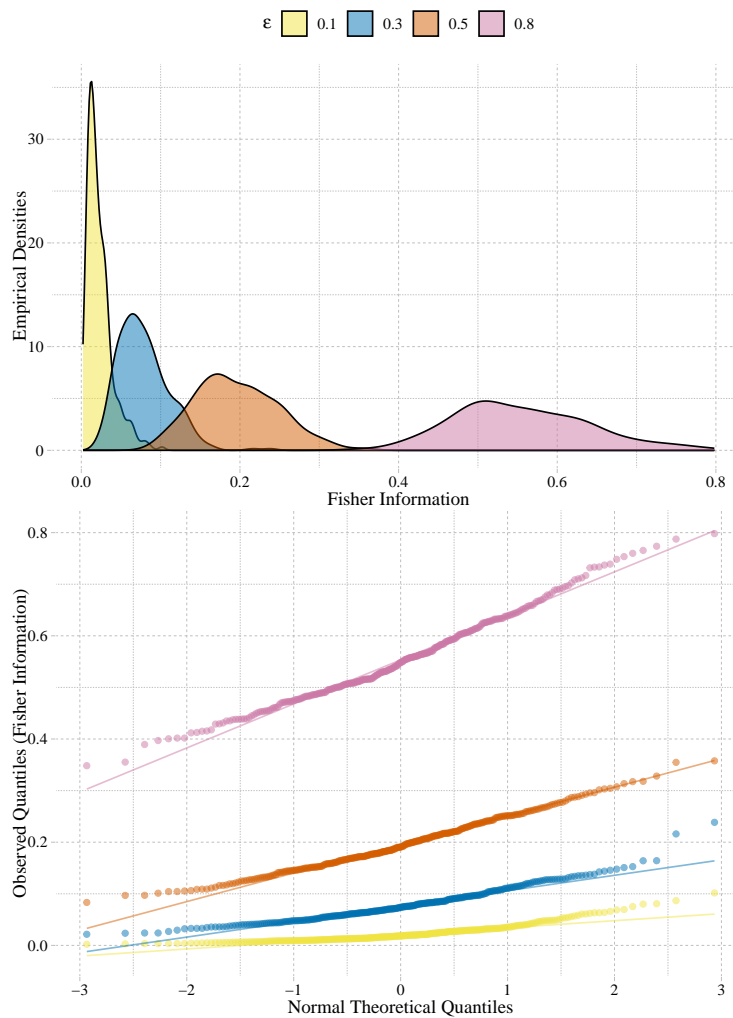
Type	Model	$\epsilon$	$k$	$n$	$p$ -Value
$H_F$	HaH	0.8	6	600	0.0030
$H_F$	HaH	0.1	6	600	0.0000
$H_F$	HaH	0.1	24	2400	0.0000
$H_F$	HaH	0.8	24	2400	0.0064
$H_F$	HaH	0.1	6	6000	0.0000
$H_F$	HaH	0.3	6	600	0.0000
$H_F$	HaH	0.1	120	120,000	0.0089
$H_F$	HaH	0.1	24	24000	0.0004
$H_F$	Linear	0	6	600	0.0000
$H_F$	Linear	0	24	2400	0.0000
$H_R^{1/3}$	HaH	0.1	6	600	0.0000
$H_R^{1/3}$	HaH	0.1	24	2400	0.0001
$H_R^{1/3}$	OAZ	0	24	2400	0.0000
$H_R^{2/3}$	HaH	0.1	6	600	0.0000
$H_R^{2/3}$	HaH	0.1	24	2400	0.0001
$H_R^{2/3}$	OAZ	0	24	2400	0.0000
$H_S$	HaH	0.1	6	600	0.0001
$H_S$	HaH	0.1	24	2400	0.0002



**Table 1.** Cont.

Type	Model	$\epsilon$	$k$	$n$	$p$ -Value
$H_S$	OAZ	0	24	2400	0.0000
$H_T^{1/2}$	HaH	0.1	6	600	0.0000
$H_T^{1/2}$	HaH	0.1	24	2400	0.0001
$H_T^{1/2}$	OAZ	0	24	2400	0.0000
$H_T^{3/2}$	HaH	0.1	6	600	0.0001
$H_T^{3/2}$	HaH	0.1	24	2400	0.0003
$H_T^{3/2}$	OAZ	0	24	2400	0.0000

The normality hypothesis was rejected at the 1% level by the Anderson–Darling test in only 24 out of 432 situations, showing that the asymptotic Gaussian model for the entropies is a good description for these data. Table 1 shows those situations.



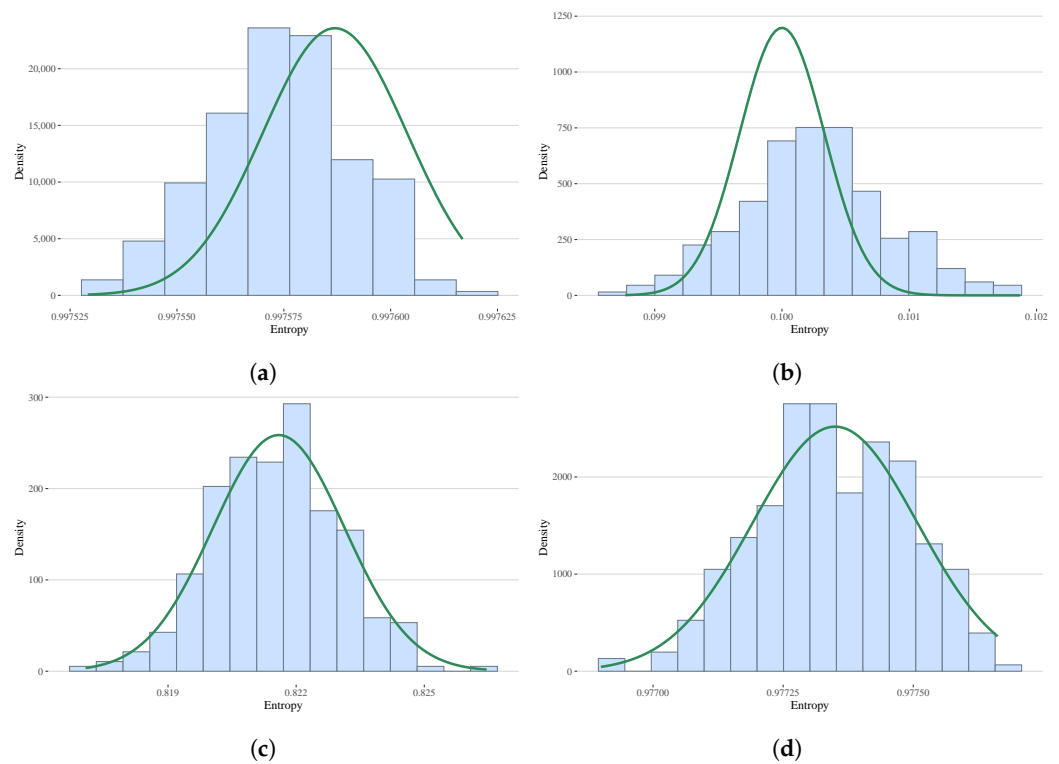
**Figure 2.** Empirical densities and normal QQ-plots of the Fisher information in situations that fail to pass the normality test at 1%.

With the aim to assess the goodness of fit of the asymptotic models, we applied the Kolmogorov–Smirnov test to fifty replicates of samples. Table 2 shows the results where the  $p$ -value of the test is at least equal to 0.05.

It is worth noticing that even in those cases where the  $p$ -value is lesser than 0.05, the asymptotic models are a good fit to the data as can be seen in several examples exhibited in Figure 3. The Fisher information shows the worst fitting. Additionally, notice in Figure 3d that, although the asymptotic distribution of the Rényi entropy is not normal, the probability density function is visually very close to the Gaussian model. We verified this similarity in all the cases we considered.

**Table 2.** Situations for which the  $p$ -values of the Kolmogorov–Smirnov test of samples of size 50 are larger than or equal to 0.05 (“HaH” and “OAZ” are the Half-And-Half and One-Almost-Zero models).

Type	Model	$\epsilon$	$k$	$n$	Type	Model	$\epsilon$	$k$	$n$						
$H_S$	HaH	0.1	6, 24	$10^3k, 10^4k$	$H_T^{3/2}$	HaH	0.1	6, 24	$10^3k, 10^4k$						
			120, 720	$10^4k$				0.3	6, 120	for all					
		24	$10^3k$	0.3			24		$10^3k$						
		720	$10^3k, 10^4k$				0.5	6, 24	for all						
		6, 24	for all	0.5				120, 720	$10^3k, 10^4k$						
	0.8	6	$10^2k, 10^3k$			Linear	0	6, 24, 120	for all						
	24, 720	for all	0	720				$10^3k, 10^4k$							
	120	$10^3k, 10^4k$		0				6, 24	for all						
	$H_F$	HaH						0.1	6	$10^3k$	$H_R^{1/3}$	HaH	0.3	6	$10^2k, 10^3k$
									0.3	6, 24				for all	0.5
6					$10^2k, 10^3k$	0.5	120	$10^4k$							
24			$10^3k, 10^4k$		0.8		6	$10^2k, 10^3k$							
120			$10^4k$	Linear		0	6	for all							
6		$10^2k, 10^3k$	0		24		$10^2k$								
0.8		24			for all		0	6	for all						
$H_T^{1/2}$		HaH			0.1			6	$10^3k, 10^4k$	$H_R^{2/3}$		HaH	0.3	6	for all
								0.3	6, 24					for all	0.5
				120, 720	$10^4k$	0.5		6	$10^2k, 10^3k$						
	6, 24		for all	0.8	6			$10^2k, 10^3k$							
	120, 720		$10^3k, 10^4k$		Linear	0	6	for all							
	6	$10^2k, 10^3k$	0	6			for all								
	24	$10^4k$		0			6	for all							
	6, 24	$10^3k, 10^4k$					0	6	$10^2k, 10^3k$						
	120, 720	$10^4k$						0	6		$10^2k, 10^3k$				
	6, 24	$10^3k, 10^4k$			0	6			$10^2k, 10^3k$						
120, 720	$10^4k$	0	6			$10^2k, 10^3k$									
0.1	6, 24		$10^3k, 10^4k$	0		6			$10^2k, 10^3k$						
120, 720	$10^4k$		0			6	$10^2k, 10^3k$								
6, 120	for all					0	6	$10^2k, 10^3k$							
24	$10^3k$				0		6	$10^2k, 10^3k$							
720	$10^3k, 10^4k$	0					6	$10^2k, 10^3k$							
6, 24	for all			0			6	$10^2k, 10^3k$							
120, 720	$10^3k, 10^4k$		0				6	$10^2k, 10^3k$							
6	$10^2k, 10^3k$					0	6	$10^2k, 10^3k$							
24	for all				0		6	$10^2k, 10^3k$							
120, 720	$10^3k, 10^4k$	0					6	$10^2k, 10^3k$							
0	for all			0			6	$10^2k, 10^3k$							
0	for all		0				6	$10^2k, 10^3k$							
6, 24	for all					0	6	$10^2k, 10^3k$							
120	$10^3k, 10^4k$				0		6	$10^2k, 10^3k$							
720	$10^4k$	0					6	$10^2k, 10^3k$							



**Figure 3.** Examples of cases where the null hypothesis of the Kolmogorov–Smirnov test is rejected. The histograms are computed with samples of size 300 using the Freedman–Diaconis rule [15], and the green lines are the asymptotic probability density functions. (a) Type:  $H_S$ , Model: OAZ,  $k = 120$ ,  $n = 10^4k$ ,  $p\text{-val} = 0.00202$ ; (b) Type:  $H_F$ , Model: OAZ,  $k = 120$ ,  $n = 10^4k$ ,  $p\text{-val} = 0.00013$ ; (c) Type:  $H_R^{1/3}$ , Model: Linear,  $k = 24$ ,  $n = 10^4k$ ,  $p\text{-val} \approx 0$ ; (d) Type:  $H_T^{1/2}$ , Model: HaH,  $\epsilon = 0.8$ ,  $k = 6$ ,  $n = 10^4k$ ,  $p\text{-val} = 0.04297$ .

### 5. Application

Inspired by an example from Agresti [16] (p. 200), we extracted data from the General Social Survey (GSS, a project of the independent research organization NORC at the University of Chicago, with principal funding from the National Science Foundation, available at <https://gss.norc.org/>. The data were downloaded on 24 December 2022). Table 3 shows the level of agreement to the assertion “Religious people are often too intolerant” as measured in three years.

**Table 3.** GSS data about religious intolerance.

Year	1998	2008	2018
STRONGLY AGREE	148	285	186
AGREE	429	602	496
NOT AGREE/DISAGREE	278	210	229
DISAGREE	275	196	181
STRONG DISAGREE	72	30	38
Total	1202	1323	1130

The  $p$ -values of pairwise  $\chi^2$  tests for the null hypotheses that the underlying probabilities are equal are

**1998 and 2008:**  $3.43 \times 10^{-22}$ ,

**1998 and 2018:**  $2.01 \times 10^{-8}$ ,

**2008 and 2018:**  $1.06 \times 10^{-3}$ .

On the one hand, these values attest that 1998 and 2008 and 1998 and 2018 are very different. On the other hand, although significant, the change between 2008 and 2018 is not so significant.

Table 4 shows the asymptotic mean and variance (in entropies normalized units) of the entropies of the proportions reported in Table 3.

**Table 4.** Asymptotic mean and variance of entropies.

	Mean			Variance		
	1998	2008	2018	1998	2008	2018
$H_S$	0.914	0.839	0.863	0.0000724	0.0001081	0.0001175
$H_R^{1/3}$	0.967	0.933	0.947	0.0001172	0.0002375	0.0002122
$H_R^{2/3}$	0.939	0.881	0.902	0.0000317	0.0000500	0.0000512
$H_T^{1/2}$	0.932	0.868	0.892	0.0000500	0.0000871	0.0000842
$H_T^{3/2}$	0.919	0.849	0.870	0.0000622	0.0001089	0.0001183
$H_F$	0.516	0.702	0.642	0.0008028	0.0011355	0.0011917

We perform the same hypothesis test with the asymptotic quantities presented in Table 4. Table 5 shows the  $p$ -values of the null hypothesis that the entropies are equal, using the test discussed by Chagas et al. [8] (Section 5):

$$p\text{-value} \approx 2 \left( 1 - \Phi \left( \frac{|H(\hat{p}_1) - H(\hat{p}_2)|}{\sqrt{\hat{\sigma}_{n_1, \hat{p}_1}^2 + \hat{\sigma}_{n_2, \hat{p}_2}^2}} \right) \right), \tag{37}$$

where  $\Phi$  is the cumulative distribution function of a standard normal random variable,  $H$  is any of the considered entropies computed with the observed proportions  $\hat{p}_i$ ,  $i = 1, 2$ , and  $\hat{\sigma}_{n_i, \hat{p}_i}^2$  is the corresponding sample asymptotic variance that takes into account the sample size  $n_i$ . Notice that the test based on entropies compares only these features, and not the underlying distribution.

**Table 5.**  $p$ -values of the hypothesis of equal entropies.

	1998–2008	1998–2018	2008–2018
$H_S$	0.0000000	0.0002606	0.1029
$H_R^{1/3}$	0.0705681	0.2520262	0.5316
$H_R^{2/3}$	0.0000000	0.0000478	0.0404
$H_T^{1/2}$	0.0000000	0.0004471	0.0690
$H_T^{3/2}$	0.0000001	0.0002316	0.1672
$H_F$	0.0000219	0.0045777	0.2118

The results in Table 5 are consistent with those provided by the  $\chi^2$  tests, i.e., the most significant differences arise between 1998 and 2008 and between 1998 and 2018. Moreover, the tests based on entropies do not reject the null hypothesis in the pair 2008–2018, except for Rényi entropy of order 2/3. The increased  $p$ -values are a consequence of the information reduction: whereas the  $\chi^2$  test compares count-by-count, those based on entropies compare two scalars.

In the second part of this application, we will illustrate the use of test statistics based on entropies for comparing samples with different categories. Situations like this may appear when applying alternative versions of the same questionnaire in a series of surveys.

We collapsed the categories of 1998 into three: “agreement” (by adding “strongly agree” and “agree”), “indifference” (“not agree/disagree”), and “disagreement” (by adding “disagree” and “strong disagree”). The resulting asymptotic mean entropies and asymptotic variances are shown in Table 6.

**Table 6.** Asymptotic mean and variance of entropies of the collapsed entries of 1998.

	Mean	Variance
$H_S$	0.955	0.0000678
$H_R^{1/3}$	0.985	0.0000171
$H_R^{2/3}$	0.970	0.0000083
$H_T^{1/2}$	0.971	0.0000281
$H_T^{3/2}$	0.949	0.0000896
$H_F$	0.192	0.0000678

Table 7 presents the  $p$ -values of the tests that verify the null hypothesis of the same entropy between the collapsed 1998 data (three categories), and 2008 and 2018 (five categories). These results agree with those presented in Table 5. Such an agreement suggests that, although the number of categories was reduced in 1998 from five to three, the tests based on entropies cope with the loss of information.

**Table 7.**  $p$ -values of the hypotheses of equal entropies using collapsed data in 1998.

	1998–2008	1998–2018
$H_S$	0.00000	0.0000
$H_R^{1/3}$	0.00115	0.0107
$H_R^{2/3}$	0.00000	0.0000
$H_T^{1/2}$	0.00000	0.0000
$H_T^{3/2}$	0.00000	0.0000
$H_F$	0.00000	0.0000

### 6. Conclusions

We presented expressions for the asymptotic distribution of the Rényi and Tsallis entropies of order  $q$ , and Fisher information. The Fisher information and the Tsallis and Shannon entropies have limit normal distribution with means and variances that depend on the underlying probability of patterns and the number of patterns. The Rényi entropy follows, asymptotically, a different distribution, cf. (35), but a Gaussian law can well approximate it. Those expressions pose no numerical challenges other than setting  $0 \log 0 \doteq 0$ . We verified that these asymptotic distributions are good models for data arising from both simulations with a variety of models and from the analysis of actual data.

On the one hand, the Fisher information is the one that fails more frequently to pass the Anderson–Darling normality tests. On the other hand, it does not provide evidence to reject the same hypothesis under the One-Almost-Zero model.

The distributions we present here can be used for building test statistics, as discussed by Chagas et al. [8]. Moreover, Equation (37) allows performing tests with mixed types of distributions, a situation that may appear in Internet of Things applications, in which, citing Borges et al. [17], one has to deal with “large time series data generated at different rates, of different types and magnitudes, possibly having issues concerning uncertainty, inconsistency, and incompleteness due to missing readings and sensor failures.”

**Author Contributions:** Conceptualization, A.A.R., A.C.F., J.G.; methodology, A.A.R., A.C.F., M.L., J.G., H.S.R.; software, A.A.R., A.C.F., M.L., J.G., E.T.C.C.; validation, A.A.R., A.C.F., M.L., J.G., E.T.C.C., H.S.R.; formal analysis, A.A.R., A.C.F., M.L., J.G., E.T.C.C., H.S.R.; investigation, A.A.R., A.C.F., M.L., J.G., E.T.C.C., H.S.R.; resources, A.C.F., H.S.R.; data curation, A.A.R., A.C.F., M.L., J.G., E.T.C.C., H.S.R.; writing—original draft preparation, A.A.R., A.C.F., M.L., J.G., E.T.C.C.; writing—review and editing, A.A.R., A.C.F., M.L., J.G., E.T.C.C., H.S.R.; visualization, A.A.R., A.C.F., M.L., J.G., E.T.C.C.; supervision, A.C.F., H.S.R.; project administration, A.A.R., A.C.F., J.G.; funding acquisition, A.C.F., H.S.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research and the APC were funded by Project 410695 from Victoria University of Wellington. It was also partially funded by Project 2020/05121-4 from Fundação de Amparo à

Pesquisa do Estado de São Paulo (FAPESP), and project APQ-00426-22 from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**Notation**

The following notation is used in this manuscript:

- $p$  vector of probabilities
- $p'$  the transpose of  $p$
- $\hat{p}$  an estimator of  $p$
- $N$  multivariate discrete random variable
- $n$  a sample from  $N$
- $H_S$  Shannon entropy
- $H_T^q$  Tsallis entropy of order  $q$
- $H_R^q$  Rényi entropy of order  $q$
- $H_F$  Fisher information measure
- $\Sigma$  covariance matrix

**Appendix A**

*Appendix A.1. Matrix Operations*

Consider the real matrix  $M \in \mathbb{R}^{k \times k}$ , and denote as  $M'$  its transpose. If  $\mathbf{D} = \text{Diag}(d_1, d_2, \dots, d_k) \in \mathbb{R}^{k \times k}$ , then

$$\begin{aligned}
 (\mathbf{DMD}')_{ij} &= \sum_{r=1}^k (\mathbf{DM})_{ir} \mathbf{D}'_{rj} = \sum_{r=1}^k \sum_{s=1}^k \mathbf{D}_{is} M_{sr} \mathbf{D}'_{rj} \\
 &= \mathbf{D}_{ii} M_{ij} \mathbf{D}_{jj} = \begin{cases} d_i^2 M_{ii} & \text{if } i = j, \\ d_i d_j M_{ij} & \text{if } i \neq j. \end{cases}
 \end{aligned} \tag{A1}$$

We consider now

$$B = \begin{pmatrix} b_{11} & b_{12} & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_{22} & b_{23} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & b_{k-1,k} & b_{kk} \\ 0 & 0 & 0 & 0 & \cdots & 0 & b_{kk} \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

Analogous to the computation in Equation (A1), it can be seen that

$$(BMB')_{ij} = \sum_{r=1}^k \sum_{s=1}^k B_{is} M_{sr} B'_{rj} = \sum_{r=1}^k \sum_{s=1}^k B_{is} M_{sr} B_{jr}. \tag{A2}$$

Due to the form of  $B$ , if  $i, j = 1, 2, \dots, k - 1$  then

$$\begin{aligned}
 (BMB')_{ij} &= \sum_{r=1}^k (B_{ii} M_{ir} + B_{i,i+1} M_{i+1,r}) B_{jr} \\
 &= (B_{ii} M_{ij} + B_{i,i+1} M_{i+1,j}) B_{jj} + (B_{ii} M_{i,j+1} + B_{i,i+1} M_{i+1,j+1}) B_{j,j+1}.
 \end{aligned} \tag{A3}$$

If  $i = k$ , replacing in Equation (A2),

$$(BMB')_{kj} = \sum_{r=1}^k B_{kk} M_{kr} B_{jr} = \begin{cases} B_{kk}(M_{kj}B_{jj} + M_{k,j+1}B_{j,j+1}) & \text{if } j \neq k, \\ B_{kk}^2 M_{kk} & \text{if } j = k. \end{cases} \quad (\text{A4})$$

### Appendix A.2. Ordinal Patterns

Symbolic data analysis [18] encompasses methods that study the statistical properties of data aggregated by criteria that meet some scientific question. Such methods have attracted lots of attention because they present competitive results in many data analysis applications [17,19,20].

Ordinal patterns [21] belong to this class of techniques. They impose low computational complexity and are inherently robust. This approach consists of constructing a set of symbolic ordinal patterns based on intrinsic data characteristics without any prior model. Ordinal patterns often reveal and quantify the underlying time series dynamics. In spite of their successful application to biomedicine, economics, mechanics and electronics engineering, image analysis and remote sensing, to name a few (see, for instance, Refs. [20,22,23]), little is known about the statistical properties of the features they induce. One of these features is entropy, in its several forms.

Signal analysis with ordinal patterns requires coding  $D$  observations into  $k = D!$  categories, in which  $D$  is typically small [8,20,24]. Motivated by these applications, we chose  $k \in \{6, 24, 120, 720\}$ , which allows checking results in various categories. Bear in mind that, when using ordinal patterns, the subsequent patterns are not independent and, thus, the multinomial distribution is an approximation.

### Appendix A.3. Computational Information

This article was written in Rmarkdown and is fully reproducible. We used RStudio version 2022.07.2 and R version 4.2.1. The code and data are available at <https://gitlab.ecs.vuw.ac.nz/freryal/asymptotic-distribution-of-various-types-of-entropy-under-the-multinomial-law>, accessed on 26 April 2023.

## References

- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Discrete Multivariate Distributions*; Wiley-Interscience: Hoboken, NJ, USA, 1997.
- Modis, T. Links between entropy, complexity, and the technological singularity. *Technol. Forecast. Soc. Chang.* **2022**, *176*, 121457. [CrossRef]
- Hutcheson, K. A test for comparing diversities based on the Shannon formula. *J. Theor. Biol.* **1970**, *29*, 151–154. [CrossRef] [PubMed]
- Hutcheson, K.; Shenton, L.R. Some moments of an estimate of Shannon's measure of information. *Commun. Stat. Theory Methods* **1974**, *3*, 89–94. [CrossRef]
- Jacquet, P.; Szpankowski, W. Entropy computations via analytic depoissonization. *IEEE Trans. Inf. Theory* **1999**, *45*, 1072–1081. [CrossRef]
- Cichoń, J.; Gołębiewski, Z. On Bernoulli Sums and Bernstein Polynomials. In Proceedings of the 23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms, Montreal, QC, Canada, 18–22 June 2012; pp. 179–190.
- Cook, G.W.; Kerridge, D.F.; Pryce, J.D. Estimations of Functions of a Binomial Parameter. *Sankhyā Indian J. Stat. Ser. A* **1974**, *36*, 443–448.
- Chagas, E.T.C.; Frery, A.C.; Gambini, J.; Lucini, M.M.; Ramos, H.S.; Rey, A.A. Statistical Properties of the Entropy from Ordinal Patterns. *Chaos Interdiscip. J. Nonlinear Sci.* **2022**, *32*, 113118. [CrossRef] [PubMed]
- Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
- Rényi, A. On Measures of Entropy and Information. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1961; Volume 1, pp. 547–561.
- Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2010.
- Sánchez-Moreno, P.; Yáñez, R.J.; Dehesa, J.S. Discrete densities and Fisher information. In Proceedings of the 14th International Conference on Difference Equations and Applications, Istanbul, Turkey, 19–23 October 2009; pp. 291–298.
- Lehmann, E.L.; Casella, G. *Theory of Point Estimation*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
- Lehman, E.L.; Romano, J.P. *Testing Statistical Hypothesis*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2005.
- Freedman, D.; Diaconis, P. On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **1981**, *57*, 453–476. [CrossRef]

16. Agresti, A. *An Introduction to Categorical Data Analysis*; Wiley-Interscience: Hoboken, NJ, USA, 2007.
17. Borges, J.B.; Ramos, H.S.; Loureiro, A.A.F. A Classification Strategy for Internet of Things Data Based on the Class Separability Analysis of Time Series Dynamics. *ACM Trans. Internet Things* **2022**, *3*, 1–30. [[CrossRef](#)]
18. Beranger, B.; Lin, H.; Sisson, S. New models for symbolic data analysis. *Adv. Data Anal. Classif.* **2022**, 1–41. [[CrossRef](#)]
19. Borges, J.B.; Medeiros, J.P.S.; Barbosa, L.P.A.; Ramos, H.S.; Loureiro, A.A. IoT Botnet Detection based on Anomalies of Multiscale Time Series Dynamics. *IEEE Trans. Knowl. Data Eng.* **2022**, *Early Access*. [[CrossRef](#)]
20. Chagas, E.T.C.; Frery, A.C.; Rosso, O.A.; Ramos, H.S. Analysis and Classification of SAR Textures using Information Theory. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 663–675. [[CrossRef](#)]
21. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)] [[PubMed](#)]
22. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **2012**, *14*, 1553–1577. [[CrossRef](#)]
23. Sigaki, H.Y.D.; Perc, M.; Ribeiro, H.V. History of art paintings through the lens of entropy and complexity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E8585–E8594. [[CrossRef](#)] [[PubMed](#)]
24. Chagas, E.T.C.; Queiroz-Oliveira, M.; Rosso, O.A.; Ramos, H.S.; Freitas, C.G.S.; Frery, A.C. White Noise Test from Ordinal Patterns in the Entropy-Complexity Plane. *Int. Stat. Rev.* **2022**, *90*, 374–396. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.