DRAFT NISTIR 8139

Identifying Uniformity with Entropy and Divergence

Dmitry A. Cousin

National Institute of Standards and Technology
U.S. Department of Commerce

DRAFT NISTIR 8139

Identifying Uniformity with Entropy and Divergence

Dmitry A. Cousin
Computer Security Division
Information Technology Laboratory

February 2017

National Institute of Standards and Technology
Kent Rochford, Acting NIST Director and Under Secretary of Commerce for Standards and Technology

National Institute of Standards and Technology Internal Report 8139
30 pages (February 2017)

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

There may be references in this publication to other publications currently under development by NIST in accordance with its assigned statutory responsibilities. The information in this publication, including concepts and methodologies, may be used by federal agencies even before the completion of such companion publications. Thus, until each publication is completed, current requirements, guidelines, and procedures, where they exist, remain operative. For planning and transition purposes, federal agencies may wish to closely follow the development of these new publications by NIST.

Organizations are encouraged to review all draft publications during public comment periods and provide feedback to NIST. Many NIST cybersecurity publications, other than the ones noted above, are available at http://csrc.nist.gov/publications.

Public comment period: *February 2, 2017* through *March 9, 2017*

National Institute of Standards and Technology
Attn: Computer Security Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 8930) Gaithersburg, MD 20899-8930
Email: Comments-IR-8139@nist.gov

All comments are subject to release under the Freedom of Information Act (FOIA).

Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof of concept implementations, and technical analyses to advance the development and productive use of information technology. ITL's responsibilities include the development of management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of other than national security-related information in federal information systems.

Abstract

Entropy models are frequently utilized in tests identifying either qualities of randomness or randomness uniformity of formal and/or observed distributions. The NIST special publications SP 800-22 [1] and SP 800-90 (A, B, & C) [2, 3, 4] discuss tests and methods leveraging both Shannon and min entropies. Shannon and min entropies represent two particular cases of Renyi entropy, which is a more general one parameter entropy model. Renyi entropy insightfully unifies Hartley, Shannon, collision, and min entropies and belongs to the class of one parameter entropy models, such as entropies named after Havrda-Charvat-Daroczy, Tsallis, Abe, and Kaniadakis. Renyi entropy along with the other members of the one parameter entropy models class can be in turn viewed as a case of the Sharma-Mittal entropy, which is a bi-parametric generalized entropy model. This NIST Internal Report (NISTIR) focuses on using Renyi and Tsallis entropy and divergence models to analyze similarities and differences between probability distributions of interest. The report introduces extensions for the traditional uniformity identification and measurement techniques that were proposed in the NIST special publications SP 800-22 and SP 800-90.

Keywords

Renyi entropy; Tsallis entropy; Sharma-Mittal entropy; randomness; uniformity identification

Note to Reviewers

Some basic understanding of probability theory and familiarity with mathematical formalisms would allow readers to get deeper understanding of the topics discussed. The document is structured to make possible independent reading and understanding of each section. Small examples are presented to illustrate the observations presented in the document.

Table of Contents

List of Appendices

List of Figures

List of Tables

1 Introduction

Entropy based distribution evaluation techniques are of particularly high interest in the fields of cryptography, crypto analysis, statistical security, and security automation. NIST special publications 800-22 [1] and 800-90 (A, B, & C) [2, 3, 4] use Shannon [5] and min entropy [6] models. Both Shannon and min entropies present particular cases of Renyi entropy [7]. This publication focuses on the similarities, differences, and relationships between Shannon, min and generalized (Renyi, Tsallis, etc.) entropies. In addition to entropy models, the distribution divergences may come into play, when comparing multiple distribution models to each other.

Determining and evaluating empirical distributions from observations is an important problem in theoretical and applied computer security in particular and computer science in general. In the field of cryptography or crypto-analytics evaluating min or Shannon entropies is considered sufficient for accepting or discarding a particular model of randomness quality. Though in more general applications in security automation and artificial intelligence, the ability to distinguish two distributions on the basis of more than Shannon or min entropy leads to the questions: "How **uniform** is the distribution produced by the model examined?" and "How to pick a **more uniform** model out of a set of models?" This report illustrates a few ways and approaches of answering the above questions.

2 Finite Discrete Probability Distributions and Δ_n Simplexes

Consider a finite discrete random variable $\boldsymbol{\xi}$ defined over a probability space $\langle \boldsymbol{\Omega}, \mathfrak{B}, \boldsymbol{p}() \rangle$, where $\boldsymbol{\Omega}$ is the universal set of events, $\mathfrak{B}$ is a Borel algebra over the events of $\boldsymbol{\Omega}$, and $\boldsymbol{p}()$ is the probability metric defined as a mapping of all elements of the algebra $\mathfrak{B}$ to the real interval [0,1]. Assume that random variable $\boldsymbol{\xi}$ is taking values $\xi(\omega_i)$ on the countable set of events $\boldsymbol{\omega} = \{\omega_i|_{i=1}^n\}$ from $\mathfrak{B}$. The probability distribution for $\boldsymbol{\xi}$ can be used as either a set $\boldsymbol{p}$ or a vector $\overline{\boldsymbol{p}}$ of probabilities. The choices and structures of the set $\boldsymbol{\Omega}$ and of the algebra $\mathfrak{B}$ may present research interest, but will not be focus of this report. This report will mostly focus on the structure and interrelations of the probability values defined by the metric $\boldsymbol{p}()$ from $\langle \boldsymbol{\Omega}, \mathfrak{B}, \boldsymbol{p}() \rangle$, while neglecting the underlying structures of $\boldsymbol{\Omega}$, $\mathfrak{B}$, and the values $\xi(\omega_i)$.

At the adopted level of abstraction, we will assume that a distribution family or simply a distribution is defined if a set of values $\boldsymbol{p} = \{p_i|_{i=1}^n\}$ or an n-dimensional vector $\overline{\boldsymbol{p}} \triangleq (p_i|_{i=1}^n)$, correspondingly, is known. Any of the equivalent notations will be used for representing the given distributions as a set: $\boldsymbol{p} \stackrel{\text{def}}{=} [p_1, p_2, \dots, p_n] = \{p_i|_{i=1}^n\} = \{p(\xi_i)|_{i=1}^n\} = \{p(x = \xi_i)|_{i=1}^n\}$, which will be used in cases when the order of probability values is not important. For the cases, when the order of the values matters the vector notation will be used $\overline{\boldsymbol{p}} \stackrel{\text{def}}{=} (p_1, p_2, \dots, p_n) = (p(\xi_i)|_{i=1}^n) = (p(\xi_i)|_{i=1}^n) = (p(x = \xi_i)|_{i=1}^n)$.

Note, that an arbitrary finite discrete distribution of cardinality n with the probabilities $p(\xi_i) \geq 0$, can always be thought of as a vector $\overline{\boldsymbol{p}} \triangleq (p(\xi_i)|_{i=1}^n)$ from an n-dimensional vector space with the additivity property $\sum_{i=1}^n p(\xi_i) = 1$. The additivity to 1 hots to a simple geometrical interpretation of the expression $\sum_{i=1}^n p(\xi_i) = 1$.

Consider a space of all possible real-valued n-element partitions $\{p(\xi_i) \geq 0|_{i=1}^{n}\}$ with the constraint $\sum_{i=1}^{n} p(\xi_i) = 1$ of the n-dimensional probability space. Under these constraints the additivity condition can be represented as: $\sum_{i=1}^{n} p(\xi_i) - 1 = 0$, which is the canonical hyperplane equation in n-dimensional space. The vector orthogonal to the hyperplane is $\bar{n} = (1, \ldots \text{n times} \ldots, 1)$ and the resulting hyperplane intersects each of the axes $p(\xi_i)$ at the following n points: $(\{p(\xi_1) = 1\}, 0 \ldots 0), (0, \{\text{p}(\xi_2) = 1\}, 0 \ldots 0), \ldots (0, .. 0, \{\text{p}(\xi_n) = 1\})$. The same linear expression explains the (n-1) dimensionality of this set, which is one less than n - the dimension of the original space in consideration. This happens because the relation of any arbitrarily selected j[th] dimension with the other (n-1) dimensions can be expressed as follows:

$$\text{p}(\xi_j) = 1 - \sum_{i \neq j, i=1}^{n} \text{p}(\xi_i) \qquad (\mathbf{e2.1})$$

The earlier hyperplane expression rewritten in the form of expression ($\mathbf{e2.1}$) demonstrates that any probability dimension, regardless of the dimension selection, is linearly dependent on the other n-1 dimensions. This structure, containing all possible discrete finite probability distributions of dimension n is usually called simplex, denoted as Δ_n, and defined as follows:

$$\Delta_n \stackrel{\text{def}}{=} \left\{ (p_1, p_2, \ldots, p_n) | p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, n > 1 \right\} \qquad (\mathbf{e2.2})$$

Figure 1 demonstrates the two simplexes for the cases of n=2 and n=3. The illustrations below represent 2D domain subspace for binomial and 3D domain subspace for simplest multinomial distribution families:

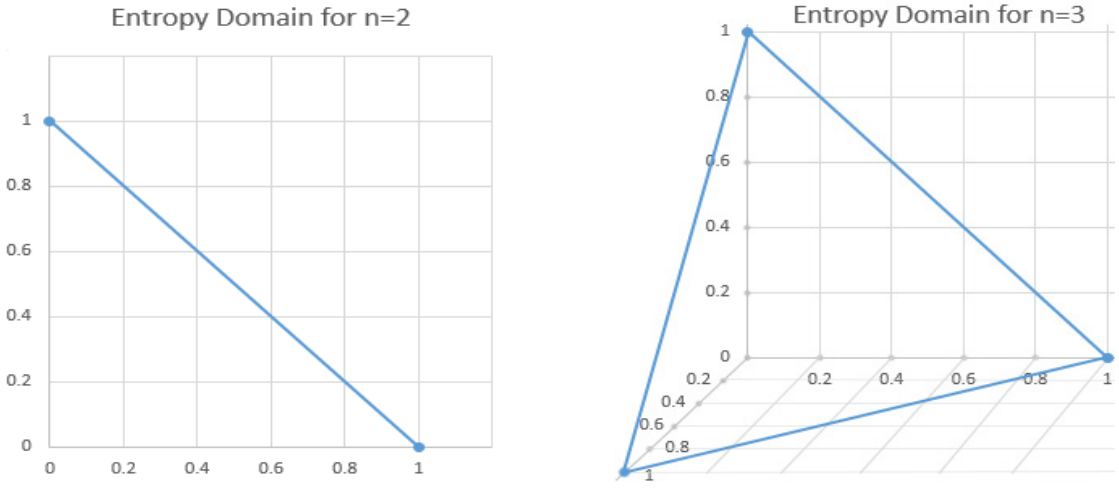


Figure 1: 2D and 3D Simplexes for all Possible Binomial and 3 Elements Multinomial Probability Distributions

With the established terminology an entropy can be defined as a function $H: \Delta_n \to \mathfrak{R}$ reflecting an n-dimensional simplex Δ_n into a set of real numbers. Simply speaking, an entropy function would yield a single (usually real) value for a given distribution that can be identified as a point in the simplex of the distribution corresponding dimensionality or higher.

3 Renyi Generalization of the Shannon, Hartley and Min Entropies

For a given probability distribution, the Hartley, Shannon, collision and min entropies map the given distribution to a real number. The idea of generalizing the Hartley, Shannon, collision, and min entropies was presented by Alfred Renyi [7] more than 50 years ago. The generalization proposed by Renyi establishes correspondence of a given probability distribution $\boldsymbol{p}(\boldsymbol{\xi})$ and a parameter α value to a real number. In this notation the Renyi entropy denoted as $H_\alpha(\boldsymbol{\xi})$ for a given parameter α can be defined as follows:

$$H_\alpha(\boldsymbol{\xi}) \stackrel{\text{def}}{=} \frac{1}{1-\alpha}\log_\beta\left(\sum_{i=1}^{n} p(\xi_i)^\alpha\right); \qquad (\mathbf{e3.1})$$

For a particular random variable $\boldsymbol{\xi}$ defined with the corresponding probabilities $\boldsymbol{p} = (p_1, \dots, p_n)$ in the simplex Δ_n, a fixed α (and a particular logarithm base β) $H_\alpha(\boldsymbol{\xi})$ would simply produce a scalar real value. Following calculus methods one could analyze the changes of the resulting value as a function of varying random variables $\boldsymbol{\xi}$ with the corresponding probability distributions $\boldsymbol{p} = (p_1, \dots, p_n)$ in the simplex Δ_n and/or the parameter α. It's useful to note, that for different values of the parameter $\alpha > 0$, the resulting Renyi entropy values for the same given distribution are usually different from each other, because Renyi entropy is a non-increasing function of α. Detailed analysis of the Renyi entropy as the function of parameter α establishes that Renyi entropy can be viewed a generalization of Hartley, Shannon, collision, and min entropies at the values of $\alpha = 0$, $\alpha = 1$, $\alpha = 2$, and $\alpha \to \infty$ correspondingly [6].

Strictly speaking, if one considers only discrete distributions, then the Renyi entropy can be seen as an implicit mapping of the given discrete distribution family generated by any permutations of the given distribution probabilities and a given parameter α to a real number. Renyi entropy is not a bijection in the strict sense. Though for each of the distribution family (i.e. all distributions that can be obtained by permutation of the given probabilities) and the same parameter α the corresponding Renyi entropy value will remain **analytically the same** because of the commutativity and associativity of the sum. This property of entropy value permutation invariance is rather important for practical applications. When working with arbitrary distributions the permutation invariance property of Renyi entropy permits use of safer computation techniques by sorting distribution probabilities in ascending order, which allows to minimize the rounding precision loss during computation.

Considering the properties of logarithms, powers and generalized p-norms the following identities can be easily established:

$$H_\alpha(\boldsymbol{\xi}) = \frac{\alpha}{1-\alpha}\log_\beta\left(\sqrt[\alpha]{\sum_{i=1}^{n}[p(\xi_i)]^\alpha}\right) = \frac{\alpha}{1-\alpha}\log_\beta(\|\mathbf{P}\|_\alpha); \qquad (\mathbf{e3.2})$$

$$H_\alpha(\boldsymbol{\xi}) = \frac{1}{1-\alpha}\log_\beta\left(\sum_{i=1}^{n} p(\xi_i)^\alpha\right) = -\log_\beta\left(\sqrt[\alpha-1]{\sum_{i=1}^{n} p(\xi_i)[p(\xi_i)]^{\alpha-1}}\right); \qquad (\mathbf{e3.3})$$

$$H_\alpha(\boldsymbol{\xi}) = \frac{1}{1-\alpha}\log_\beta\left(\sum_{i=1}^{n} p(\xi_i)^\alpha\right) = \log_\beta\left(\frac{1}{\sqrt[\alpha-1]{\sum_{i=1}^{n} \mathrm{p}(\xi_i)[\mathrm{p}(\xi_i)]^{\alpha-1}}}\right); \qquad (\mathbf{e3.4})$$

As in the case of Shannon entropy the logarithm bases β in $(\mathbf{e3.1}) - (\mathbf{e3.4})$ can be easily changed to desired value with a multiplicative adjustment factor. The two rightmost expressions in the identity expression $(\mathbf{e3.2})$ can be interpreted as the Renyi entropy $H_\alpha(\boldsymbol{\xi})$ is the quantity proportional to a logarithm of $\|\mathbf{P}\|_\alpha$. Where $\|\mathbf{P}\|_\alpha$ notation is used to represent a "p-norm" (or rather α-norm), also known as a Minkowski generalized norm. The non-negativity of each particular probability of the i[th] random variable value $p(\xi_i) \geq 0$ makes the expression $\sqrt[\alpha]{\sum_{i=1}^{n}[p(\xi_i)]^\alpha}$ an α-norm. Otherwise, the middle part of the expression $(\mathbf{e3.2})$ would require component-wise absolute values by the p-norm (Minkowski norm) definition.

3.1 Generalized Weighted Averages and Entropy

Generalized weighted mean (GWM) $\mathrm{M}_q(\bar{x}, \bar{w})$ of the vector of values $\bar{x} = (x_i|_{i=1}^n)$ and vector of the corresponding weights $\bar{w} = (w_i|_{i=1}^n)$ is defined as follows:

$$\mathrm{M}_q(\bar{x}, \bar{w}) \stackrel{\text{def}}{=} \sqrt[q]{\sum_{i=1}^{n} x_i^q w_i} \qquad (\mathbf{e3.5})$$

At the parameter q tends to zero ($q \to 0$) the GWM tends to the following expression:

$$\lim_{q\to 0} \mathrm{M}_q(\bar{x}, \bar{w}) = \prod_{i=1}^{n} x_i^{(w_i)} \qquad (\mathbf{e3.6})$$

Thus, the expressions $(\mathbf{e3.3})$ and $(\mathbf{e3.4})$ can be viewed as a negative logarithm of a GWM $(\mathbf{e3.3})$ or positive logarithm of the GWM's inverse $(\mathbf{e3.4})$. Considering expressions $(\mathbf{e3.5})$-$(\mathbf{e3.6})$ and that any finite discrete probability distribution can be thought of as a vector (ordered tuple) of fixed dimension as follows: $\bar{\boldsymbol{p}} = (p(\xi_i)|_{i=1}^n) = (p(x = \xi_i)|_{i=1}^n)$, one can easily obtain the limit for the expression $(\mathbf{e3.3})$ at $\alpha \to 1$ by noticing that the $\mathrm{M}_{\alpha-1}(\bar{p}, \bar{p})$ tends to $\lim_{q\to 0} \mathrm{M}_q(\bar{p}, \bar{p})$, hence the expression for $H_{\alpha\to 1}(\boldsymbol{\xi})$ will take the Shannon entropy form[5], which is easily demonstrated by the following chain of identities:

$$H_{\alpha\to 1}(\boldsymbol{\xi}) = -\log_\beta\left(\mathrm{M}_{\alpha-1}(\bar{p}, \bar{p})\right); = -\log_\beta\left(\prod_{i=1}^{n} p(\xi_i)^{\mathrm{p}(\xi_i)}\right) = -\sum_{i=1}^{n} p(\xi_i)\log_\beta p(\xi_i) \quad (\mathbf{e3.7})$$

More detailed additional analysis of the expressions $(\mathbf{e3.3})$-$(\mathbf{e3.4})$ can be found in [6].

4 Renyi Entropy α-Spectrum

Mathematical analysis of the analytical extension of the Renyi entropy as a function of α and the corresponding analysis of the function behavior around $\alpha \to 1$ in [6] demonstrate that the Shannon entropy [5] as well as Hartley and min entropies are just particular cases of the Renyi entropy at $\alpha = 1$, $\alpha = 0$, and $\alpha \to \infty$ correspondingly. For the case of $\alpha = 1$ it also converges to the expression $(\mathbf{e3}.\mathbf{7})$ value derived from the GWM properties above.

Shannon entropy can be computed as follows [5, 8, 9, 10] and $(\mathbf{e3}.\mathbf{7})$:

$$H_{\alpha=1}(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H_1(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H^S(\boldsymbol{\xi}) = \sum_{i=1}^{n} p(\xi_i) \log_\beta \left(\frac{1}{p(\xi_i)}\right) = -\sum_{i=1}^{n} p(\xi_i) \log_\beta p(\xi_i) \qquad (\mathbf{e4}.\mathbf{1})$$

Hartley entropy [5] is the of Renyi entropy value at $\alpha \to 0$ and it can be computed as follows:

$$H_{\alpha=0}(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H_0(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H_{max}(\boldsymbol{\xi}) = \log_\beta |\{p(\xi_i)|_{i=1}^{n} : p(\xi_i) > 0\}| = \log_\beta(n), \qquad (\mathbf{e4}.\mathbf{2})$$

Where notation $|\{a_i|_{i=1}^{n} : set\ elements\ condition\}|$ is used to present the cardinality of the set $\{a_i\}$, interpreted as the count of the set elements, that satisfy the set condition. In the case of the discrete distribution (considered in this report) the condition is that the probability $p(\xi_i)$ is positive and the cardinality of the set is n. The meaning of the value given by Hartley entropy is the logarithm of the non-zero elements count in the given distribution.

Min entropy can be computed as follows:

$$H_{\alpha\to\infty}(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H_\infty(\boldsymbol{\xi}) \stackrel{\text{def}}{=} H_{\min}(\boldsymbol{\xi}) = \min_{\forall i=1,n} \{-\log_\beta p(\xi_i)\} = -\log_\beta \max_{\forall i=1,n} \{p(\xi_i)\} \qquad (\mathbf{e4}.\mathbf{3})$$

The meaning of the min entropy is the logarithm of the most likely event's probability in the given distribution.

Theoretically, either expression $(\mathbf{e3}.\mathbf{1})$ or the interpretations $(\mathbf{e3}.\mathbf{2})$-$(\mathbf{e3}.\mathbf{4})$ can be used to compute Renyi entropy spectrum analytically over the whole range of $\alpha \in [0, \infty)$. Unfortunately, despite analytically computable behavior of the $H_\alpha(\boldsymbol{\xi})$ over $\alpha \in [0, \infty)$ the reality of the computer mathematics kicks in and the direct usage of the expressions $(\mathbf{e3}.\mathbf{1})$-$(\mathbf{e3}.\mathbf{4})$ can produce overflows, precision loss, or rounding errors as a result of computing the series needed to compute fractional powers, radicals, and logarithms called for in the expressions $(\mathbf{e3}.\mathbf{1})$-$(\mathbf{e3}.\mathbf{4})$. Since the series convergence is analytically guaranteed, $H_\alpha(\boldsymbol{\xi})$ is continuous over $\alpha \in [0, \infty)$, and the values and corresponding expressions that $H_\alpha(\boldsymbol{\xi})$ tends to are known, when approaching the special points $\alpha = 1$, $\alpha = 0$, and $\alpha \to \infty$ it is usually practical to compute $H_\alpha(\boldsymbol{\xi})$ by using the expressions $(\mathbf{e4}.\mathbf{1})$, $(\mathbf{e4}.\mathbf{2})$, and $(\mathbf{e4}.\mathbf{3})$ correspondingly.

When comparing $(\mathbf{e4}.\mathbf{2})$ and $(\mathbf{e4}.\mathbf{3})$ and assuming random variable $\boldsymbol{\xi}$ distributed according to the discrete uniform distribution U_n with the cardinality of the values $|\{p(\xi_i)|_{i=1}^{n}\}| = n$ and the probability of n equally probable events $p(\xi_i) \underset{\{\forall i=1,n\}}{=} \frac{1}{n}$, one can easily notice the following property, keeping in mind that $H_\alpha(\boldsymbol{\xi})$ is monotonously non-increasing function of α [6]:

$$H_0(\boldsymbol{\xi}) = \log_\beta(n) = -\log_\beta\left(\frac{1}{n}\right) = -\log_\beta \max_{\forall i=1,n}\left\{p(\xi_i) \underset{\{\forall i=1,n\}}{=} \frac{1}{n}\right\} = H_\infty(\boldsymbol{\xi}) \qquad (\mathbf{e4.4})$$

The expression $(\mathbf{e4.4})$ analytically conveys identical $H_\alpha(\boldsymbol{\xi})$ values at both ends of the possible range of the α-values.

When considering values $H_\alpha(\boldsymbol{\xi})$ for a random variable $\boldsymbol{\xi}$ over the whole range of parameter values $\alpha \in [0,\infty)$ the aggregated mapping of $\alpha \in [0,\infty)$ onto the set $\{H_\alpha(\boldsymbol{\xi})\} \subset [0,\infty)$ is called the Renyi entropy α-spectrum for the random variable $\boldsymbol{\xi}$ or for the distribution $\boldsymbol{p}(\boldsymbol{\xi})$. The random variable values only implicitly identify probabilities, hence it should be noticed that the Renyi entropy α-spectrum, strictly speaking, is uniquely defined by the probability distribution $\boldsymbol{p}(\boldsymbol{\xi})$.

4.1 The 3 Distributions with Matching Hartley and 2 with Matching Min Entropies

Compute Shannon, Hartley and min entropies for the distributions of cardinality 16:

- $\boldsymbol{\xi_{16}}$: uniform $\bar{p}(\boldsymbol{\xi_{16}}) \sim U_{16} = \bar{p}(\boldsymbol{\xi_{16}}) = \left\{\frac{1}{16}; \frac{1}{16}; \ldots; \frac{1}{16}\right\}$
- $\boldsymbol{\zeta_{16}}$: $\bar{p}(\boldsymbol{\zeta_{16}}) = \left\{\frac{1}{10}; \frac{3}{50}; \ldots; \frac{3}{50}\right\}$,
- $\boldsymbol{\eta_{16}}$: $\bar{p}(\boldsymbol{\eta_{16}}) = \left\{\frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{20} \ldots \frac{1}{20}\right\}$

Shannon entropies $(\mathbf{e4.1})$ for logarithm base $\beta = 2$ are easily computed as follows:

$$H^R_{\alpha=1}(\boldsymbol{\xi_{16}}) = -\sum_{i=1}^{16} p(\xi_i)\log_\beta p(\xi_i) = -\sum_{i=1}^{16} \frac{1}{16}\log_2\frac{1}{16} = -\frac{16*(-4)}{16} = \mathbf{4}$$

$$H^R_{\alpha=1}(\boldsymbol{\zeta_{16}}) = -\left(\frac{1}{10}\log_2\frac{1}{10} + \sum_{i=1}^{15}\frac{3}{50}\log_2\frac{3}{50}\right) \approx \mathbf{3.98519713}$$

$$H^R_{\alpha=1}(\boldsymbol{\eta_{16}}) = -\left(\frac{4}{10}\log_2\frac{1}{10} + \frac{12}{20}\log_2\frac{3}{50}\right) \approx \mathbf{3.921928095}$$

Because all distributions have 16 positive (non-zero probability) elements, the Hartley entropy for all 3 distributions according to $(\mathbf{e4.2})$ can be computed as follows**:**

$$H^R_{\alpha=0}(\boldsymbol{\xi_{16}}) = H^R_{\alpha=0}(\boldsymbol{\zeta_{16}}) = H^R_{\alpha=0}(\boldsymbol{\eta_{16}}) = \log_\beta(16) = \{\text{assuming } \beta = 2\} = \mathbf{4}$$

Applying the expression $(\mathbf{e4.3})$ the min entropies can be computed as follows:

For ξ_{16} the min entropy is $H^R_{\alpha\to\infty}(\boldsymbol{\xi_{16}}) = = -\log_\beta \max_{\forall i=1,16}\{p(\xi_i)\} = -log_2\left(\frac{1}{16}\right) = \mathbf{4}$

As $\max_{\forall i, p(\boldsymbol{\zeta_i})}\left\{\frac{1}{10}; \frac{3}{50}; \ldots; \frac{3}{50}\right\} = \max_{\forall i, p(\boldsymbol{\eta_i})}\left\{\frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{20} \ldots \frac{1}{20}\right\} = \frac{1}{10} \Rightarrow H^R_{\alpha\to\infty}(\boldsymbol{\zeta_{16}}) = H^R_{\alpha\to\infty}(\boldsymbol{\eta_{16}})$

$H^R_{\alpha\to\infty}(\boldsymbol{\zeta_{16}}) = H^R_{\alpha\to\infty}(\boldsymbol{\eta_{16}}) = -log_2\left(\frac{1}{10}\right) \approx \mathbf{3.321928095}$

4.2 Renyi Entropy α-Spectrum for Uniform and non-Uniform Distributions

Let's consider the fact that $H_\alpha(\boldsymbol{\xi})$ is a non-increasing as a function of α [6]. In the light of this, the expression $(\mathbf{e4.4})$ gives us a noteworthy property of the full α-spectrum of Renyi entropy $H_\alpha(\boldsymbol{\xi})$ over $\alpha \in [0,\infty)$ for the given $\boldsymbol{\xi}$, when random variable $\boldsymbol{\xi}$ is distributed according to the discrete uniform distribution with n values U_n, yielding $H_\alpha(\boldsymbol{\xi})$ as a constant function of α with the value:

$$H_\alpha(\boldsymbol{\xi}\sim U_n) \underset{\{\forall\alpha\in[0,\infty)\}}{=} \log_\beta(n) = -\log_\beta\left(\frac{1}{n}\right) \qquad (\mathbf{e4.5})$$

In validating random number generators and other sources of high informational entropy the problem of identifying the best source or quantifying the quality of the random sequence by sampling can be frequently reduced to identifying which sampled sequences approximate the discrete uniform distribution U_n as close as possible.

To illustrate the idea of using the expressions $(\mathbf{e3.1})$ and $(\mathbf{e4.1}) - (\mathbf{e4.4})$ for the uniformity testing we present the α-spectrum of Renyi entropy behavior on a few synthetic data sets drawn from the distributions presented in the following table:

Table 1: The Data for α-Spectrum of Renyi Entropy Illustration

ξ_1~ U_{16}		ξ_2~ U_{16} ±30%		ξ_3 ~ U_{16}±60%		ξ_4~ N(16,5)		ξ_5~ N(16,1)	
Counts	{p(ξ_i)}	Counts	{p(ξ_i)}	Counts	{p(ξ_i)}	Counts	{p(ξ_i)}	Counts	{p(ξ_i)}
100	0.06250	70.68681	0.04315	40.51917	0.02736	1.34990	0.000600	3.6709662E-48	5.3764423E-51
100	0.06250	71.01797	0.04335	41.85771	0.02826	2.55513	0.001136	7.7935368E-42	1.1414298E-44
100	0.06250	71.29977	0.04352	43.95745	0.02968	4.66119	0.002072	6.1171644E-36	8.9591076E-39
100	0.06250	71.67738	0.04375	44.59065	0.03011	8.19754	0.003643	1.7764821E-30	2.6018092E-33
100	0.06250	73.77597	0.04504	45.41941	0.03067	13.90345	0.006179	1.9106596E-25	2.7983235E-28
100	0.06250	74.06512	0.04521	45.55111	0.03076	22.75013	0.010111	7.6198530E-21	1.1159923E-23
100	0.06250	74.72694	0.04562	52.69224	0.03558	35.93032	0.015969	1.1285884E-16	1.6529137E-19
100	0.06250	113.01434	0.06899	52.73876	0.03561	54.79929	0.024355	6.2209606E-13	9.1111260E-16
100	0.06250	125.83834	0.07682	56.47474	0.03813	80.75666	0.035891	1.2798125E-09	1.8743943E-12
100	0.06250	125.88873	0.07685	142.07325	0.09593	115.06967	0.051141	9.8658765E-07	1.4449415E-09
100	0.06250	126.57872	0.07727	145.50200	0.09825	158.65525	0.070512	2.8665157E-04	4.1982561E-07
100	0.06250	126.78433	0.07739	149.58985	0.10101	211.85540	0.094156	3.1671242E-02	4.6385228E-05
100	0.06250	127.15765	0.07762	150.91161	0.10190	274.25312	0.121887	1.3498980E+00	1.9770405E-03
100	0.06250	127.86247	0.07805	151.57614	0.10235	344.57826	0.153142	2.2750132E+01	3.3319504E-02
100	0.06250	128.34226	0.07835	158.09023	0.10675	420.74029	0.186991	1.5865525E+02	2.3236412E-01
100	0.06250	129.43859	0.07901	159.46121	0.10767	500.00000	0.222217	5.0000000E+02	7.3229253E-01

The data in the Table 1: The Data for α-Spectrum of Renyi Entropy Illustration were synthesized as follows:

- $\bar{p}(\boldsymbol{\xi_1}) = \{p(\xi_i)|_{i=1}^{16}\}$ – all counts are the same taking probabilities of $p(\xi_i) \underset{\{\forall i=1,16\}}{=} \frac{1}{16}$

- $\bar{p}(\boldsymbol{\xi_2}) = \{p(\xi_i)|_{i=1}^{16}\}$ – all counts are taking randomized probabilities of up to 30% of the original uniform probability $p(\xi_i)_{\{\forall i=1,16\}} = \frac{1}{16} \pm \mathbf{rnd}\left[0, 30\% \text{ of } \frac{1}{16}\right]$. The columns are sorted in increasing order to emphasize discrepancy with the normal distribution samples.
- $\bar{p}(\boldsymbol{\xi_3}) = \{p(\xi_i)|_{i=1}^{16}\}$ – all counts are taking randomized probabilities of up to 60% of the original uniform probability $p(\xi_i)_{\{\forall i=1,16\}} = \frac{1}{16} \pm \mathbf{rnd}\left[0, 60\% \text{ of } \frac{1}{16}\right]$. The columns are sorted in increasing order to make discrepancy with the normal distribution samples more obvious.
- $\bar{p}(\boldsymbol{\xi_4}) = \{p(\xi_i)|_{i=1}^{16}\}$ – all counts are computed as thousand times the normal distribution with mean a $= 16$ and deviation $\sigma = 5$ at arguments $x = [1,16]$.
- $\bar{p}(\boldsymbol{\xi_5}) = \{p(\xi_i)|_{i=1}^{16}\}$ – all counts are computed as thousand times the normal distribution with mean a $= 16$ and deviation $\sigma = 1$ at arguments $x = [1,16]$.

The resulting data sampling are charted as the following histograms:

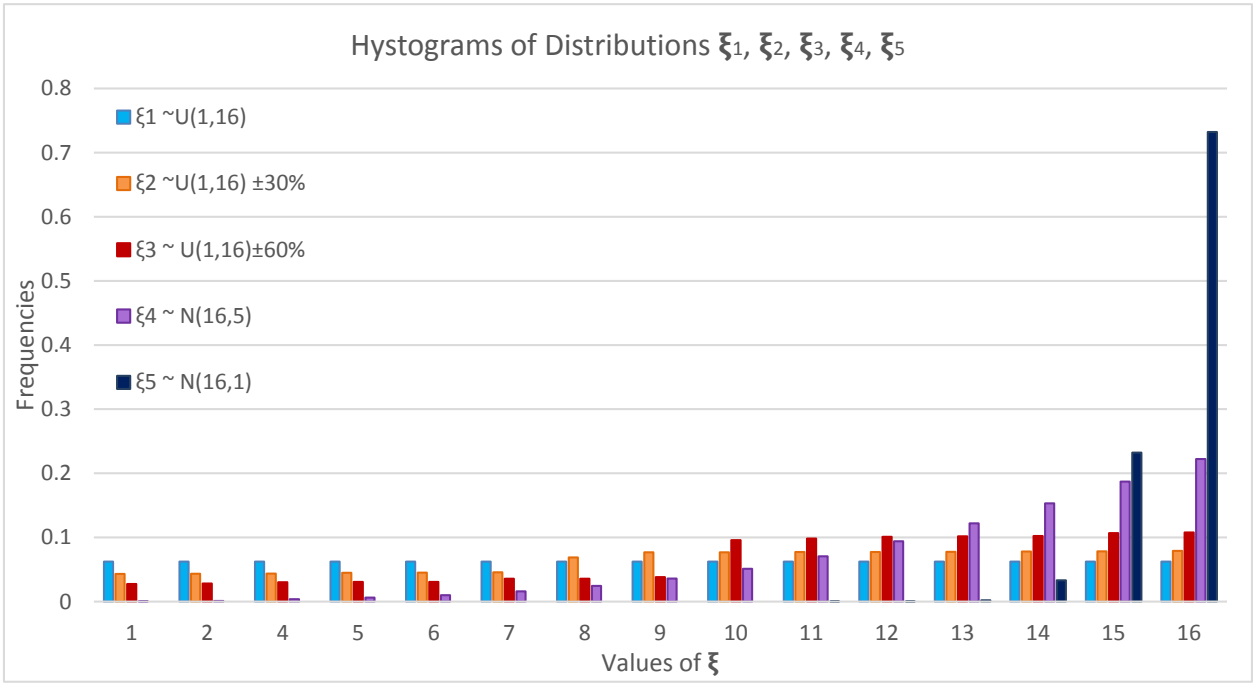


Figure 2: The Histograms of the Synthetic Data Set

The α-spectrum of Renyi entropy for the five discrete synthetic distribution samples $\boldsymbol{\xi_1}$, $\boldsymbol{\xi_2}$, $\boldsymbol{\xi_3}$, $\boldsymbol{\xi_4}$, and $\boldsymbol{\xi_5}$ described and charted above was computed on a selected number of points, yielding the α-spectrum of Renyi entropy lines $H_\alpha(\boldsymbol{\xi_1})$, $H_\alpha(\boldsymbol{\xi_2})$, $H_\alpha(\boldsymbol{\xi_3})$, $H_\alpha(\boldsymbol{\xi_4})$, and $H_\alpha(\boldsymbol{\xi_5})$. The lines were plotted on the same system of coordinates at the [11] same scale resulting in the following chart:

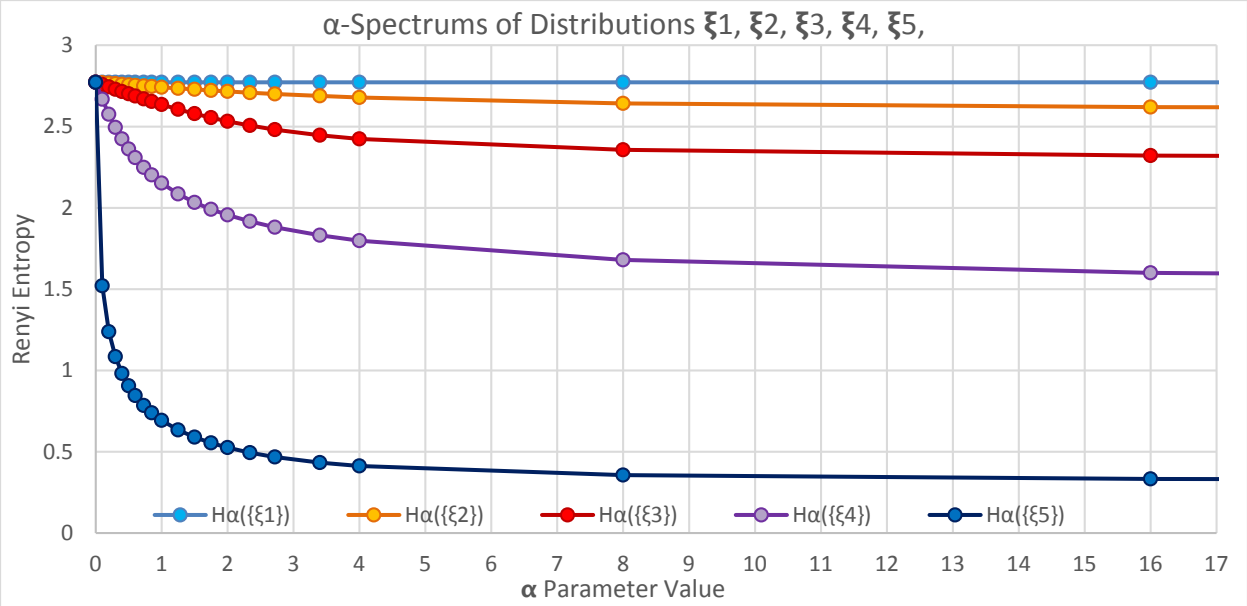


Figure 3: Renyi Entropy α-Spectrum for the Five Discrete Synthetic Distributions ξ1, ξ2, ξ3, ξ4, and ξ5

The obvious observation is that the $H_\alpha(\boldsymbol{\xi_1})$ (horizontal teal colored line) at the top of the chart in Figure 3 corresponds to the sample of the discrete uniform distribution as was predicted by the expressions $(\mathbf{e4.4})$and $(\mathbf{e4.5})$.

An intuitively visual interpretation of the Renyi entropy α-spectrum can be deduced from the line curvature of the two entropy α-spectrum graphs built for $H_\alpha(\boldsymbol{\xi_2})$ and $H_\alpha(\boldsymbol{\xi_3})$. The two weakened uniform distributions: $\boldsymbol{\xi_2}$ with up to $\pm$30%, and $\boldsymbol{\xi_3}$ with up to $\pm$60% randomly introduced inconsistency with uniformity (orange and red lines correspondingly) have general curvatures increasing with the size of the introduced inconsistency. Overall, the more Renyi entropy α-Spectrum graph of the given distribution deviates from the uniform distribution graph (horizontal line) the less uniform quality the distribution in question has.

5 Renyi Divergence α-Spectrum and Renyi Entropy

In order to determine how different are two distributions (or the samples drawn from distributions) Kulback-Leibler divergence is frequently used in the classical information theory. Renyi divergence [6], [11] generalizes the classical notion of the Kulback-Leibler divergence and can be defined for the case of finite discrete distributions and $\alpha > 0$ as follows:

$$D_\alpha(\boldsymbol{\xi} \parallel \boldsymbol{\zeta}) \stackrel{\text{def}}{=} \frac{1}{\alpha - 1} \log_\beta \sum_{i=1}^{n} p(\xi_i)^\alpha p(\zeta_i)^{1-\alpha} \qquad (\mathbf{e5.1})$$

Similarly, to the relation established for the Hartley, Shannon, and min entropies the Renyi divergence between the two distributions (or samples drawn from distributions) can be continuously extended for the cases of $\alpha \to 0$, $\alpha \to 1$, and $\alpha \to \infty$.

An interesting relation can be noticed by analyzing the Renyi divergence of an arbitrary

distribution from a discrete uniform random variable $\boldsymbol{\zeta}$ of *n* outcomes defined as $\boldsymbol{\zeta}{\sim}U_n = \left\{\frac{1}{n}, \frac{1}{n}, \frac{1}{n} \dots, \frac{1}{n}\right\}$. If we describe the arbitrary distribution of $\boldsymbol{\xi}$ as $\boldsymbol{\xi}{\sim}P = \{p(\xi_1) = p_1, p(\xi_2) = p_2, \dots, p(\xi_n) = p_n\}$ the expression $(\mathbf{e5.1})$ can yield the following identities:

$$D_\alpha(\boldsymbol{\xi}{\sim}P \parallel U_n) = \frac{1}{\alpha - 1}\log_\beta \sum_{i=1}^{n}\left[\left(\frac{1}{n}\right)^{1-\alpha} {p_i}^\alpha\right] = \frac{1}{\alpha - 1}\left(\log_\beta \left(\frac{1}{n}\right)^{1-\alpha} + \log_\beta \sum_{i=1}^{n} {p_i}^\alpha\right) \quad (\mathbf{e5.2})$$

By opening the brackets in $(\mathbf{e5.2})$ and using $(\mathbf{e3.1})$, $(\mathbf{e4.4})$, and $(\mathbf{e4.5})$ we obtain the following:

$$D_\alpha(P \parallel U_n) = \log_\beta(n) - \frac{1}{1-\alpha}\log_\beta \sum_{i=1}^{n} {p_i}^\alpha = \log_\beta(n) - H_\alpha(P) = H_\alpha(U_n) - H_\alpha(P) \quad (\mathbf{e5.3})$$

The expression $(\mathbf{e5.3})$ analytically proves the observation from Figure 3. The expression $(\mathbf{e5.3})$ also hints that a discrete uniform distribution U_n of n events is indeed a very special distribution because the Renyi divergence can be computed for any value of the parameter $\alpha \in [0, \infty)$ by using Renyi entropy of the discrete uniform distribution of equal cardinality instead of the whole expression $(\mathbf{e5.1})$.

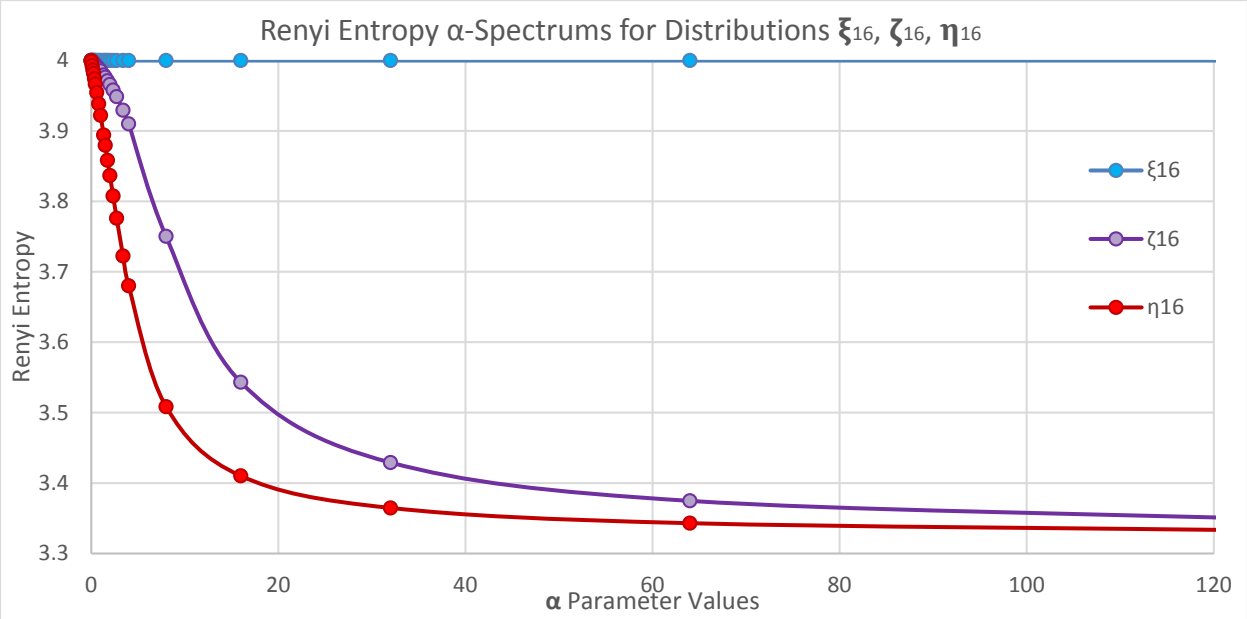
The expression $(\mathbf{e5.3})$ was analytically derived without any assumptions on the possible values of the parameter α and due to that should hold for $\forall\alpha \in [0, \infty)$. Though in practice, as noticed before, while computing $D_\alpha(P \parallel U_n)$ for α approaching the special for $H_\alpha(P)$ points $\alpha = 1$, $\alpha = 0$, and $\alpha \to \infty$ it may be computationally safer to use expressions $(\mathbf{e4.1})$, $(\mathbf{e4.2})$, and $(\mathbf{e4.3})$ correspondingly instead of relying on the analytical convergence.

5.1 Plotting Renyi Entropy and Divergence for the Distributions of Example 4.1

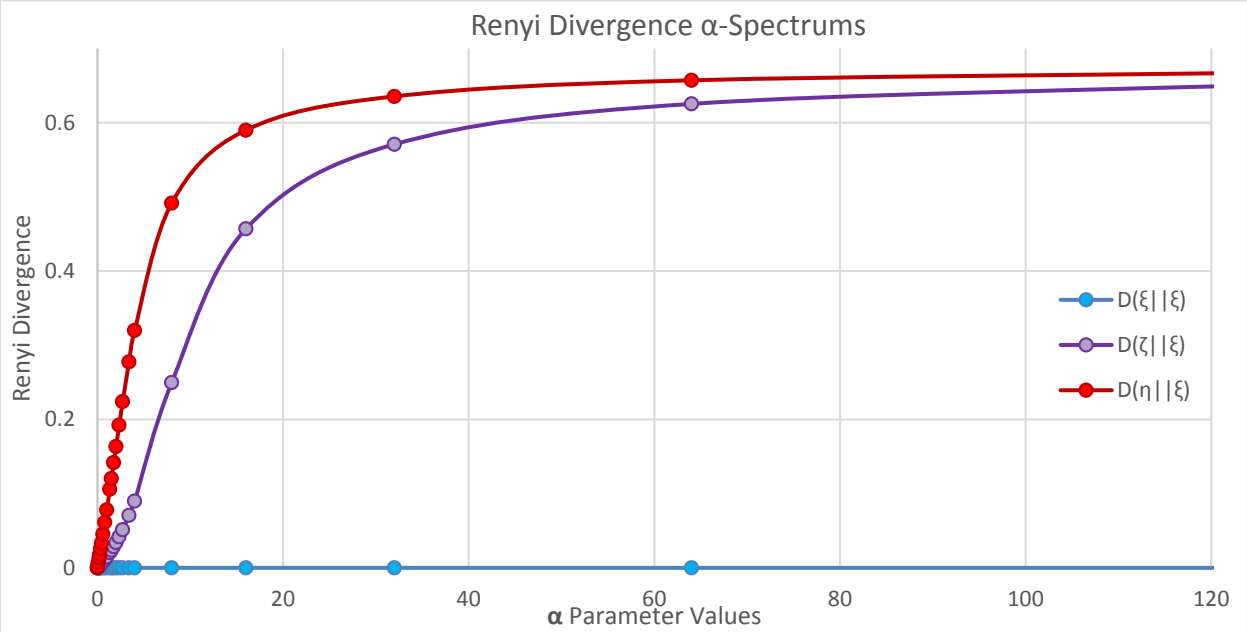
Renyi entropy (including Shannon, Hartley and min) and Renyi divergence α-spectrums are shown for the following discrete finite distributions of the events space cardinality 16, assuming the logarithm base $\beta = 2$:

- $\boldsymbol{\xi_{16}}$: uniform $\bar{p}(\boldsymbol{\xi_{16}}){\sim}U_{16} = \bar{p}(\boldsymbol{\xi_{16}}) = \left\{\frac{1}{16}; \frac{1}{16}; \dots; \frac{1}{16}\right\}$
- $\boldsymbol{\zeta_{16}}$: $\bar{p}(\boldsymbol{\zeta_{16}}) = \left\{\frac{1}{10}; \frac{3}{50}; \dots; \frac{3}{50}\right\}$,
- $\boldsymbol{\eta_{16}}$: $\bar{p}(\boldsymbol{\eta_{16}}) = \left\{\frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{20} \dots \frac{1}{20}\right\}$

As was established in the Example 4.1 all distributions $\boldsymbol{\xi_{16}}$, $\boldsymbol{\zeta_{16}}$, and $\boldsymbol{\eta_{16}}$ have the same Hartley entropy: $H^R_{\alpha=0}(\boldsymbol{\xi_{16}}) = H^R_{\alpha=0}(\boldsymbol{\zeta_{16}}) = H^R_{\alpha=0}(\boldsymbol{\eta_{16}}) = \log_2(16) = \mathbf{4}$. The two distributions $\boldsymbol{\zeta_{16}}$, and $\boldsymbol{\eta_{16}}$ have the same min entropy: $H^{\mathrm{R}}_{\alpha\to\infty}(\boldsymbol{\zeta}_{16}) = H^{\mathrm{R}}_{\alpha\to\infty}(\boldsymbol{\eta}_{16}) = -log_2\left(\frac{1}{10}\right) \approx \mathbf{3.321928095}$. By computing the Renyi entropy values for some of the intermediate values of the parameter α the following entropy chart will result:



For the Renyi divergence α-spectrum the expression (**e5. 3**) would yield similarly looking chart, with the top and bottom swapped and uniform distribution divergence from itself is zero.



These entropy and divergence charts for the revisited Example 4.2 hint that even though the Hartley and min entropies for both distributions $\boldsymbol{\zeta_{16}}$, and $\boldsymbol{\eta_{16}}$ are a match, it is inevitable that there exists a value (or possibly set of values) α_{max} from inside the interval $\alpha_{max} \in [0+\varepsilon, \infty)$ or $(\{\alpha_{max}\} \subset [0+\varepsilon, \infty))$ where the distance between the entropy lines would reach maximum.

6 Tsallis Entropy and Tsallis Entropy α-Spectrum

Shannon's initial works in the field of information theory [8, 9, 10] lead to the development of entropy properties formulated in the Shannon-Khinchin axioms [12]. When presented to the scientific community, the model of Renyi entropy complied with the Shannon-Khinchin axioms formulated in [12] and was readily accepted due to that compliance.

Further work on various entropy and divergence models in 1960s and 1970s led to development of the generalized entropy models described in the relatively obscure scientific magazines of the Eastern Bloc and India. The two models were: Havrda-Charvat-Daroczy one parameter entropies [13, 14], and Sharma-Mittal two-parameter entropies [15].

The criticism of Renyi entropies for the insufficient stability properties, considered unacceptable for real world processes modeled in the experimental physics, by Lesche [16] in the early 1980s lead to rediscovery of one of the entropy generalizations (Havrda-Charvat-Daroczy entropy) by Constantino Tsallis, who in the late 1980s published the paper [17] rediscovering and refreshing some of the ideas presented earlier by Havrda, Charvat, and Daroczy [13, 14]. The Tsallis entropy (sometimes supplemented with the names of Havrda, Chrvat, and Daroczy, depending historical awareness) is defined as follows:

$$H_\alpha^T(\boldsymbol{\xi}) \stackrel{\text{def}}{=} \frac{1}{\alpha-1}\left(1-\sum_{i=1}^{n} p(\xi_i)^\alpha\right); \qquad (\mathbf{e6.1})$$

Despite the absence of the logarithm in the expression $(\mathbf{e6.1})$, the Tsallis entropy $H_\alpha^T(X)$ converges to Shannon entropy at $\alpha \to 1$ as shown in [17] and is Lesche stable. Like the other entropy functions, Tsallis entropy maps an n-dimensional simplex Δ_n into a real number for a given value of α and on all distributions defined in the n-dimensional simplex Δ_n the discrete uniform distribution U_n maximizes the entropy value for all values of $\alpha \in [0, \infty)$. These properties of the Tsallis entropy are not at all that surprising if one looks at the Taylor series decomposition of the function $\ln(x)$.

$$\ln(x) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i}(x-1)^i = (x-1) + \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i}(x-1)^i \qquad (\mathbf{e6.2})$$

$$-\ln(x) = \sum_{i=1}^{\infty} \frac{(-1)^{i}}{i}(x-1)^i = (1-x) + \sum_{i=2}^{\infty} \frac{(-1)^{i}}{i}(x-1)^i \qquad (\mathbf{e6.3})$$

Consider Renyi entropy expressed as $(\mathbf{e3.1})$ with logarithm base $\beta = e$, substitute log-function with the series representation $(\mathbf{e6.2})$ yields the following Renyi entropy expression:

$$H_\alpha(\boldsymbol{\xi}) \triangleq \frac{1}{1-\alpha}\ln\left[\sum_{i=1}^{n} p(\xi_i)^\alpha\right] = \frac{1}{1-\alpha}\sum_{j=1}^{\infty}\left(\frac{(-1)^{j+1}}{j}\left(\left[\sum_{i=1}^{n} p(\xi_i)^\alpha\right]-1\right)^i\right) \qquad (\mathbf{e6.4})$$

For the values $\alpha \to 0$ the argument of the logarithm $[\sum_{i=1}^{n} p(\xi_i)^{\alpha}]$ tends to zero, which reduces the residual tail weight of the Taylor series. By regrouping the summed elements, and substituting $p(\xi_i) \triangleq p_i$ in expression $(\mathbf{e6.4})$ to save space the relation of Tsallis and Renyi entropies can be expressed as follows:

$$H_{\alpha}(\boldsymbol{\xi}) = \frac{1}{\alpha - 1}\left(1 - \left[\sum_{i=1}^{n} {p_i}^{\alpha}\right]\right) + \frac{1}{\alpha - 1}\left(\sum_{j=2}^{\infty} \frac{(-1)^j}{j}\left(\sum_{i=1}^{n} {p_i}^{\alpha} - 1\right)^i\right) \qquad (\mathbf{e6.5})$$

The first summand in $(\mathbf{e6.5})$ is Tsallis entropy, hence $(\mathbf{e6.5})$ can be rewritten as follows:

$$H_{\alpha}(\boldsymbol{\xi}) = H_{\alpha}^{T}(\boldsymbol{\xi}) + \frac{1}{\alpha - 1}\left(\sum_{j=2}^{\infty} \frac{(-1)^j}{j}\left(\sum_{i=1}^{n} {p_i}^{\alpha} - 1\right)^i\right) \qquad (\mathbf{e6.6})$$

The interpretation of Tsallis entropy as a first degree approximation of Renyi entropy $(\mathbf{e6.6})$ by Taylor series, and the fact that series $(\mathbf{e6.2})$ converges on x values defined by $|x - 1| \leq 1$ (which translates to convergence in $x \in [0; 2]$) allows us to think of the Tsallis entropy as less non-linear (no log) version of the Renyi entropy, or as Renyi entropy without the Taylor series tail.

The finite discrete uniform distribution U_n maximizes the entropy value for all parameter values $\alpha \in [0, \infty)$, which means that for any random value produced distribution $\boldsymbol{\zeta}$ with probabilities from the n-dimensional simplex $(\overline{\boldsymbol{p}}(\boldsymbol{\zeta}) \in \Delta_n)$ the uniform distribution $H_{\alpha}^{T}(\boldsymbol{\xi} \sim U_n) \geqslant H_{\alpha}^{T}(\boldsymbol{\zeta})$. It is also easy to deduce from the expression $(\mathbf{e6.1})$ that the limit of Tsallis entropy for any distribution with *n* non-zero probabilities is $\lim_{\alpha \to 0} T_{\alpha}(U_n) = (n - 1)$. Combining this limit with the maximization inequality $H_{\alpha}^{T}(\boldsymbol{\xi} \sim U_n) \geqslant H_{\alpha}^{T}(\boldsymbol{\zeta})$ shows that at $\alpha \to 0$ the Tsallis entropy for any finite discrete distribution of cardinality not more than n will stay at or below $(n - 1)$.

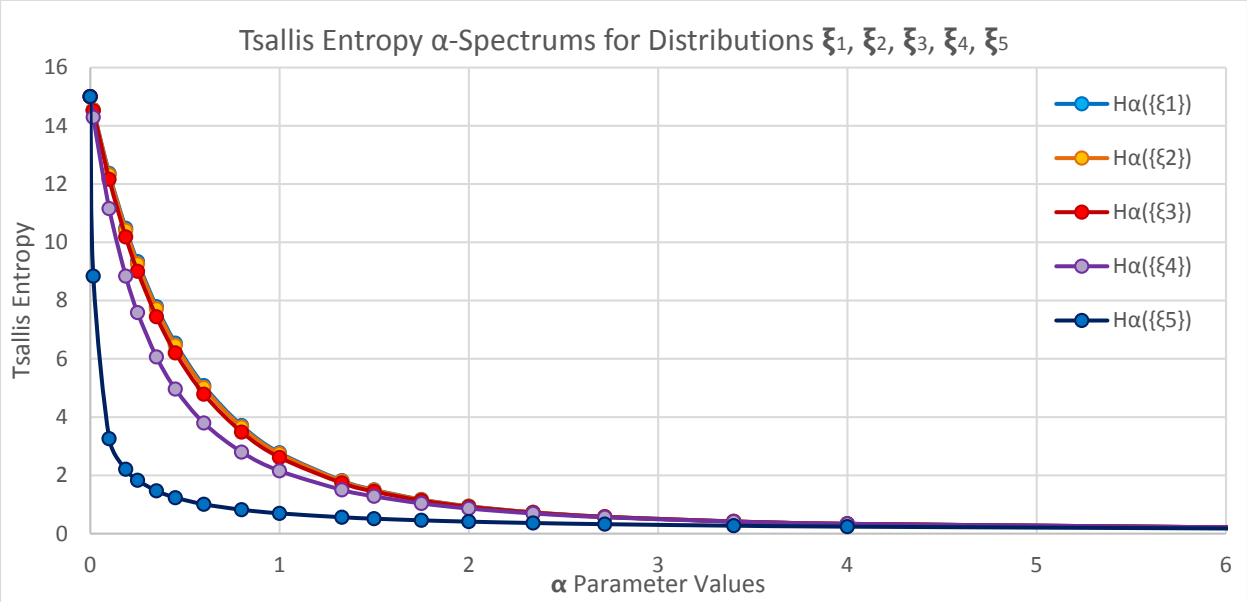
At another side of the α-spectrum, when $\alpha \to \infty$: the sum $\sum_{j=1}^{n} p(\xi_j)^{\alpha}$ tends to zero, the part of the Tsallis entropy $(\mathbf{e6.1})$ in the brackets tends to 1, and the $\left(\frac{1}{\alpha - 1}\right)$ tends to zero, which makes Tsallis entropy for any finite discrete distribution tend to zero at infinity $\lim_{\alpha \to \infty} H_{\alpha}^{T}(\boldsymbol{\zeta}) = 0$. These two properties of Tsallis entropy give a hint that there may be a range of α values where the plain differences in Tsallis entropy uniform and examined α-spectrums could reach the maximum value. This observation is important if one were to directly apply the expression $(\mathbf{e5.3})$ relating Renyi divergences and Renyi entropies to Tsallis entropies directly and obtain "Renyi-like" divergence of Tsallis entropies.

6.1 Note About Tsallis Entropy *α*-Spectrums for Example 4.1 Distributions

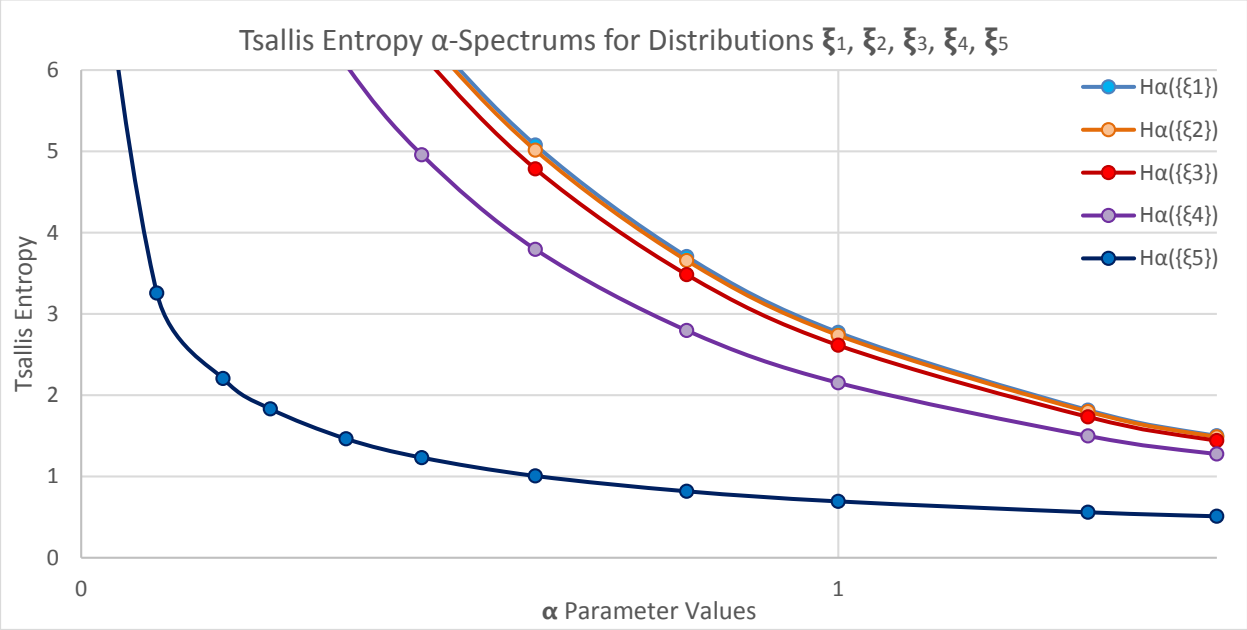
Examining the distributions from Example 4.1, it can be seen that the lines of Tsallis entropy α-spectrums appear too close to each other making it difficult to visually separate them, which is an unfortunate side-effect of the Lesche stability so coveted in physics. These 3 distributions from Example 4.1 will be further examined in the subsequent sections.

6.2 Tsallis Entropy *α*-Spectrums of the Example 4.2 Distributions

Consider the same synthetic finite discrete distributions described in the Example 4.2: $\boldsymbol{\xi_1 \sim U_{16}}$; $\boldsymbol{\xi_2 \sim U_{16} \pm 30\%}$; $\boldsymbol{\xi_3 \sim U_{16} \pm 60\%}$; $\boldsymbol{\xi_4 \sim N(16, 5)}$; $\boldsymbol{\xi_5 \sim N(16, 1)}$. For the given distributions Tsallis entropy α-spectrum graphs shape up as follows:



This graph displays already described properties of Tsallis entropies at the $\alpha \to 0$ and $\alpha \to \infty$. Due to Lesche stability, the uniform distribution $\boldsymbol{\xi_1}$ and modified uniform distributions $\boldsymbol{\xi_2}$ and $\boldsymbol{\xi_3}$ are located exceedingly close to each other, unlike in the Renyi entropy graph. To separate the graphs visually, the following rescaled fragment of the same chart is included below:



7 Differential "Renyi-like" Tsallis Entropy α-Spectrums Divergence

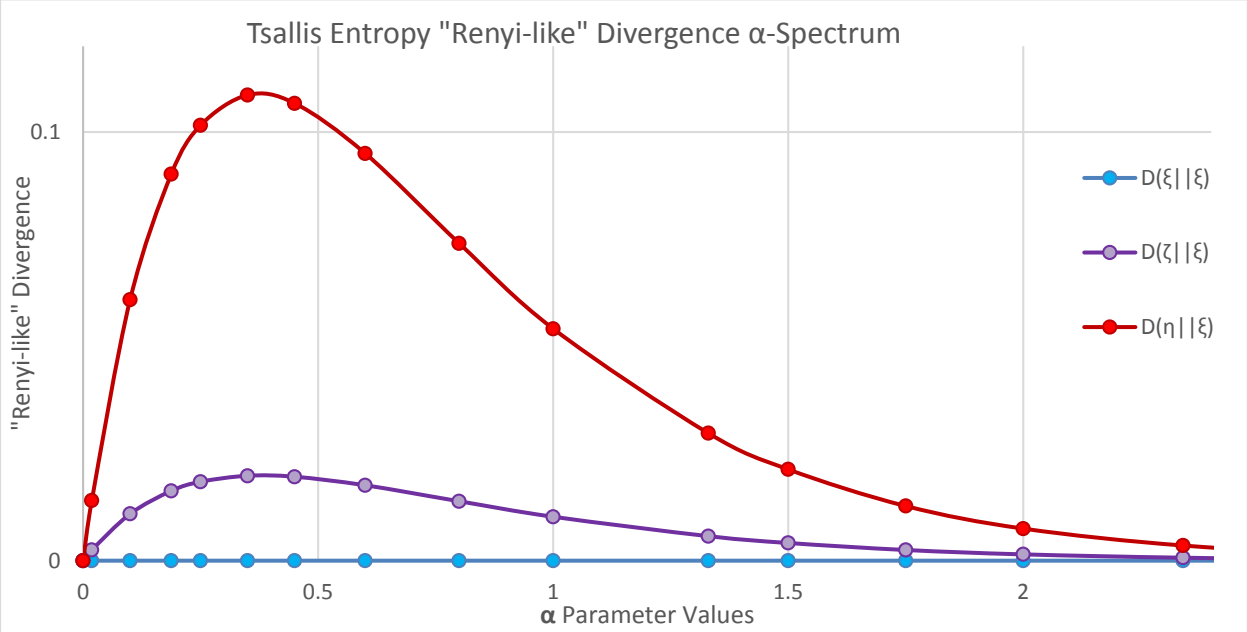
As previously explained, the expression (**e5.3**) relates the Renyi divergence of an arbitrary distribution to the uniform one and expresses the divergence in the terms of the difference in the corresponding Renyi entropies. Strictly speaking, the sequence of identities that lead to the expression (**e5.3**) would not yield the resulting expression similar to (**e5.3**) when applied to Tsallis entropy. On the other hand, the elegance and simplicity of the expression makes it quite universal. So, directly applying (**e5.3**) "in principle" to Tsallis entropy α-spectrums would yields a graph of "Renyi-like" divergence of Tsallis entropies. This quantity, computed for multiple values of α would yield "Renyi-like" divergence of Tsallis entropy α-spectrums and can be defined as follows.

$$D_{\alpha}^{Rl}(\boldsymbol{\eta}{\sim}P(\eta) \parallel \boldsymbol{\xi}{\sim}U_n) = H_{\alpha}^{T}(\boldsymbol{\xi}) - H_{\alpha}^{T}(\boldsymbol{\eta}) \qquad (\mathbf{e7.1})$$

7.1 Differential "Renyi-type" Divergence Applied to Tsallis Entropy *α*-Spectrums of the Example 4.1 Distributions

Plot Tsallis entropy "Renyi-like" divergences for these distributions of the cardinality 16:

- $\boldsymbol{\xi_{16}}$: uniform $\bar{p}(\boldsymbol{\xi_{16}}){\sim}U_{16}{=}\left\{\frac{1}{16};\frac{1}{16};\ldots;\frac{1}{16}\right\}$
- $\boldsymbol{\zeta_{16}}$: $\bar{p}(\boldsymbol{\zeta_{16}}) = \left\{\frac{1}{10};\frac{3}{50};\ldots;\frac{3}{50}\right\}$,
- $\boldsymbol{\eta_{16}}$: $\bar{p}(\boldsymbol{\eta_{16}}) = \left\{\frac{1}{10};\frac{1}{10};\frac{1}{10};\frac{1}{10};\frac{1}{20}\ldots\frac{1}{20}\right\}$



The difference between the distributions $\boldsymbol{\zeta_{16}}$ and $\boldsymbol{\eta_{16}}$ in this graph, compared to the graph of Renyi divergences plotted in 5.1, looks more obvious as it is peaking in perceptually confined space of α-values. As well as in example 5.1 the distribution $\boldsymbol{\eta_{16}}$ diverges from uniform distribution much further than $\boldsymbol{\zeta_{16}}$. Tsallis entropy and "Renyi-like" divergences seem to keep

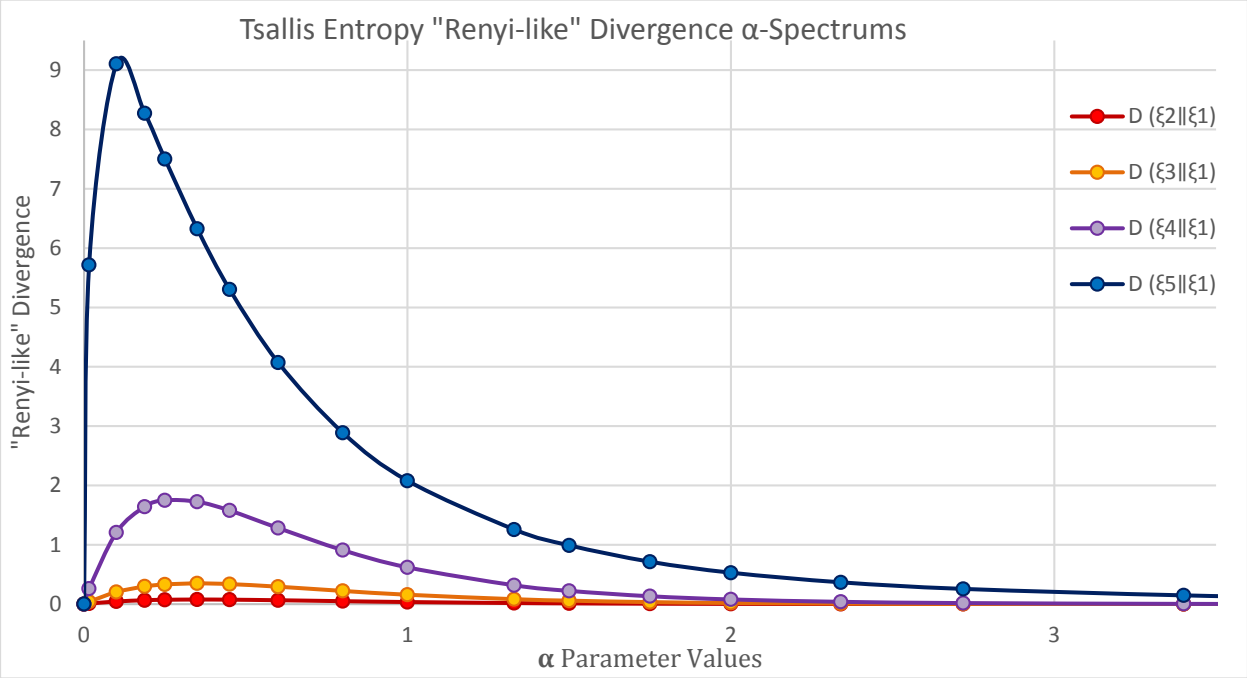
the "uniformity" quality of the distributions preserved, while making it easier to distinguish "more uniform" distributions from "less uniform".

7.2 Differential "Renyi-type" Divergence Applied to Tsallis Entropy *α*-Spectrums of the Example 4.2 Distributions

Consider the same synthetic finite discrete distributions described in 4.2:

$$\boldsymbol{\xi_1 \sim U_{16};\ \xi_2 \sim U_{16} \pm 30\%;\ \xi_3 \sim U_{16} \pm 60\%;\ \xi_4 \sim N(16,5);\ \ \xi_5 \sim N(16,1)}$$

The "Renyi-type" divergence from uniform computed for Tsallis entropy α-spectrums according to $(\mathbf{e7.1})$ results in the following chart:



As it was the case with example 7.1, the most notable property of the "Renyi-type" divergence quantity is that the computed values, specifying how different the entropy α-spectrum lines of the distribution are from the uniform distribution are concentrated and reach maximum in the relatively small range of the lower α-values of the spectrum. This particular property might be attractive from the statistical and computational points of view (subject to additional research and analysis) as at these smaller α-values of the α-spectrum all the probability values of the given probability distribution contribute to the computation of the resulting quantity, which contrasts the expression $(\mathbf{e4.3})$ for Renyi entropy at $\alpha \to \infty$ (min entropy) that utilizes only the largest probability value out of the whole distribution. The reason for the min entropy's popularity for uniformity evaluation is that the maximum probability event of the distribution is the "easiest to guess" event as well. The "easiest to guess" event corresponds to the worst case scenario, which allows to establish upper boundary for cryptographic applications, where it is frequently used. In security automation, artificial intelligence, and other entropy applications Tsallis entropy and "Renyi- like" divergence could be of more use due to such properties as relative ease of computing (no log) and the α-spectrum discriminating part being quite well confined.

8 Tsallis Divergence α-Spectrums

Similar to the Taylor series tail-cutting expression for Tsallis and Renyi entropies demonstrated in expressions $(\mathbf{e6.5}) - (\mathbf{e6.6})$ the Tsallis divergence α-spectrum can be derived from the Renyi divergence as follows:

$$D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\zeta}) \stackrel{\text{def}}{=} \frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n} p(\xi_i)^\alpha \, p(\zeta_i)^{1-\alpha}\right] - 1\right) \qquad (\mathbf{e8.1})$$

When applied to compute the Tsallis divergence for an arbitrary and uniform distributions the analytical outcome does not look as satisfyingly simple as expression $(\mathbf{e5.3})$ does for the analytically derived relation of the Renyi entropies and the Renyi divergence.

Let's assume that a discrete uniform random variable $\boldsymbol{\zeta}$ of n outcomes is defined as $\boldsymbol{\zeta}{\sim}U_n = \left\{\frac{1}{n}, \frac{1}{n}, \frac{1}{n} \dots, \frac{1}{n}\right\}$. If we describe an arbitrary distribution of random variable $\boldsymbol{\xi}$ as $\boldsymbol{\xi}{\sim}\bar{p}(\boldsymbol{\xi}) = \{p(\xi_1) = p_1, p(\xi_2) = p_2, \dots, p(\xi_n) = p_n\}$ the expression $(\mathbf{e8.1})$ would yield the following sequence of identities:

$$D_\alpha^T(\boldsymbol{\xi}{\sim}P \parallel \boldsymbol{\zeta}{\sim}U_n) = \frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n} {p_i}^\alpha \left(\frac{1}{n}\right)^{1-\alpha}\right] - 1\right) = \frac{1}{\alpha-1}\left(\left(\frac{1}{n^{(1-\alpha)}}\right)\sum_{i=1}^{n} {p_i}^\alpha - 1\right)$$

$$D_\alpha^T(\boldsymbol{\xi}{\sim}P \parallel \boldsymbol{\zeta}{\sim}U_n) = \frac{1}{\alpha-1}\left(\frac{\sum_{i=1}^{n} {p_i}^\alpha - n^{(1-\alpha)}}{n^{(1-\alpha)}}\right) = \left(\frac{n^{(1-\alpha)} - \sum_{i=1}^{n} {p_i}^\alpha}{(1-\alpha)n^{(1-\alpha)}}\right) \qquad (\mathbf{e8.2})$$

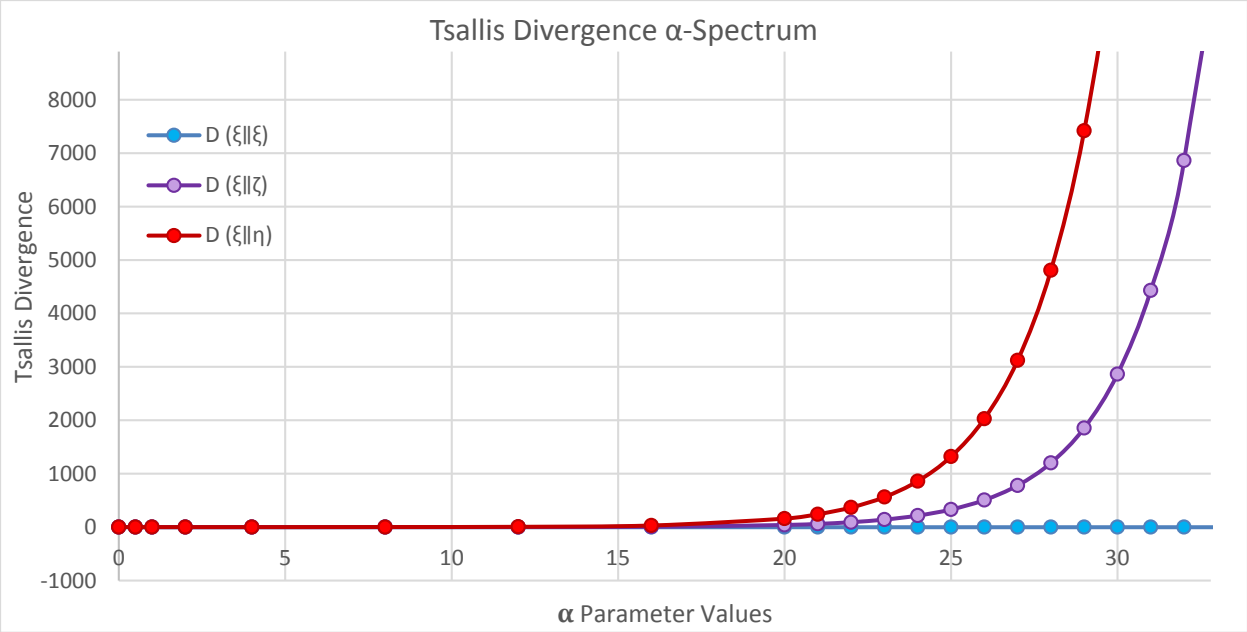
Unfortunately, the $(\mathbf{e8.2})$ does not come close to the simplicity of $(\mathbf{e6.6})$**.**

8.1 Tsallis Divergence *α*-Spectrums of the Example 4.1 Distributions

Examine the Tsallis divergences for the finite discrete distributions of the cardinality 16 from the uniform distribution $\boldsymbol{\xi_{16}}$:

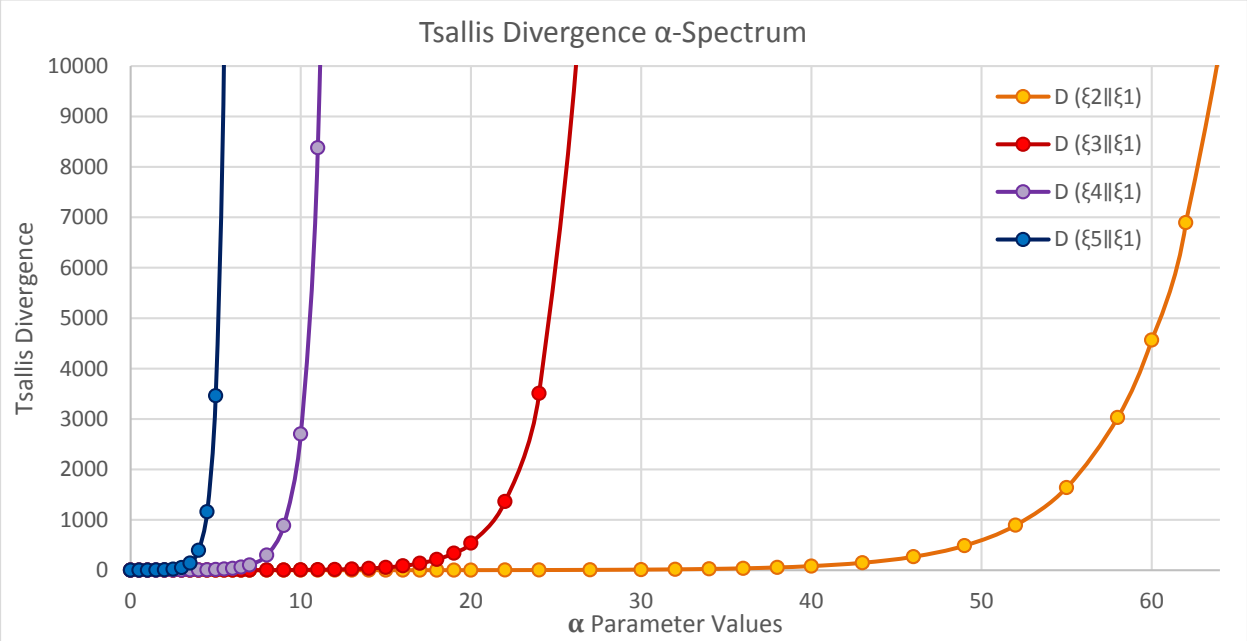
- $\boldsymbol{\xi_{16}}$: uniform $\bar{p}(\boldsymbol{\xi_{16}}){\sim}U_{16} = \left\{\frac{1}{16}; \frac{1}{16}; \dots; \frac{1}{16}\right\}$
- $\boldsymbol{\zeta_{16}}$: $\bar{p}(\boldsymbol{\zeta_{16}}){\sim}\left\{\frac{1}{10}; \frac{3}{50}; \dots; \frac{3}{50}\right\}$,
- $\boldsymbol{\eta_{16}}$: $\bar{p}(\boldsymbol{\eta_{16}}){\sim}\left\{\frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{10}; \frac{1}{20} \dots \frac{1}{20}\right\}$

The Tsallis divergence property noted while charting the graph is the high sensitivity to the order of the probability values in the distribution representation and consequentially the actual order of summation. In order to successfully plot the graphs below it is useful to notice, that all three distributions are sorted in ascending order. Until the sorting of the summands code was added the structure of the graph for obviously non-uniform distribution $\boldsymbol{\zeta_{16}}$ was almost indistinguishable from the graph of uniform distribution $\boldsymbol{\xi_{16}}$, which for the distributions of the higher cardinality may present quite a challenge, especially for the nearly uniform distributions or samples. As a result, the following chart was plotted:



8.2 Tsallis Divergence *α*-Spectrums of the Example 4.2 Distributions

Consider the same synthetic finite discrete distributions described in 4.2 and used in 6.2, and 7.2: $\boldsymbol{\xi_1 \sim U_{16}}$; $\boldsymbol{\xi_2 \sim U_{16} \pm 30\%}$; $\boldsymbol{\xi_3 \sim U_{16} \pm 60\%}$; $\boldsymbol{\xi_4 \sim N(16,5)}$; $\boldsymbol{\xi_5 \sim N(16,1)}$. Tsallis divergence from uniform distribution α-spectrums computed as $(\mathbf{e8.2})$ yields the following chart:



If Tsallis divergence α-spectra are used to assess the distribution's uniformity, the proximity of the sloping part of the graph to the point $\alpha = 0$ and the tangential angle (computed as either derivative or divided differences value) of the Tsallis divergence at the consequential values of α can be utilized as a practical indicator of the examined distribution's uniformity.

As in the case of the Renyi entropy, Tsallis divergence α-spectrum computed for the same distribution is zero, which can be easily deduced from the following identities:

$$D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\xi}) = -\frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n} p(\boldsymbol{\xi_i})^\alpha\, p(\boldsymbol{\xi_i})^{1-\alpha}\right]-1\right) = \frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n} p(\boldsymbol{\xi_i})^1\right]-1\right)$$

For any distribution element on the n-dimensional simplex Δ_n, considering the very definition of the simplex, the following must hold: $\sum_{i=1}^{n} p(\boldsymbol{\xi_i}) = 1$, hence

$$D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\xi}) = -\frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n} p(\boldsymbol{\xi_i})^1\right]-1\right) = \frac{1}{\alpha-1}(1-1) = \frac{0}{\alpha-1} \equiv 0 \text{ for } \forall\alpha \qquad (\mathbf{e8.3})$$

The structure of the expression $(\mathbf{e8.1})$ can explain the computational instability of Tsallis divergence α-spectrum at the larger values of α computed directly. At the values $\alpha \to \infty$ the corresponding $p(\zeta_i)^{1-\alpha}$ tends to infinity exponentially, while $p(\xi_i)^\alpha$ converges to zero. The computer math for the floating point multiplication leads to rounding errors at multiplication operations and addition. In some cases, it may be better to use the expression $(\mathbf{e8.1})$ in the following form:

$$D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\zeta}) \stackrel{\text{def}}{=} \frac{1}{\alpha-1}\left(\left[\sum_{i=1}^{n}\left(\frac{p(\xi_i)}{p(\zeta_i)}\right)^\alpha p(\zeta_i)\right]-1\right) \qquad (\mathbf{e8.4})$$

In other cases, sorting the summands helps. Unfortunately, for some distributions the rounding errors can mount to undesirable side effects. Fortunately, for the purposes of identifying uniformity, the expression $(\mathbf{e8.2})$ allows to perform two sorts in the course of computing the value of $D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\zeta})$. The expression $(\mathbf{e8.4})$ can be adopted for uniformly distributed $\boldsymbol{\zeta}$ as follows:

$$D_\alpha^T(\boldsymbol{\xi} \parallel \boldsymbol{\zeta}{\sim}U_n) = \frac{1}{\alpha-1}\left(\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(np(\xi_i)\right)^\alpha - 1\right) = \frac{1}{n(\alpha-1)}\left(\left[\sum_{i=1}^{n}\left(np(\xi_i)\right)^\alpha\right]-n\right) \qquad (\mathbf{e8.5})$$

The expressions $(\mathbf{e8.5})$ and $(\mathbf{e8.2})$ allow to sort the elements of the distribution in question and then the can be sorted as well before the execution of the addition. For computing Tsallis divergence from uniform distribution $\boldsymbol{\zeta}$ for the nearly uniform distribution $\boldsymbol{\xi}$ the expression $(\mathbf{e8.5})$ is giving the least rounding errors due to $np(\xi_i) \approx 1$, which naturally reduces the order difference between the summands in the expression.

9 Evaluation with Entropy and Divergence α-Spectrums

When analyzing entropy/randomness models or sources the traditional questions a researcher usually looks to answer are:

1. What is the maximum obtainable entropy of the model examined?

2. What is the average entropy of the model examined?
3. What is the worst case entropy of the model examined?

At the first glance it would seem that these 3 questions can easily be answered with the values of Hartley, Shannon and min entropies correspondingly.

The **1st** question can be answered by the property of either the Renyi or Tsallis entropies as the value of both the Renyi and Tsallis entropies at $\alpha \to 0$ essentially counts the number of distribution events with non-zero probability: as a logarithm of the count in the case of Renyi entropy or as the count reduced by one in the case of Tsallis entropy. Considering that the computational complexity of the Tsallis entropy is one logarithm function less computationally complex, the Tsallis entropy at $\alpha \to 0$: $H^T_{\alpha\to 0}(\boldsymbol{\xi})$ seems to be slightly more practical computationally. The Tsallis entropy $H^T_{\alpha\to 0}(\boldsymbol{\xi})$ is also easier to interpret as $(H^T_{\alpha\to 0}(\boldsymbol{\xi}) + 1)$ is an exact count of distribution events of non-zero probability. Though in the case of very large cardinality distributions the same argument can make use of Renyi entropy $H_{\alpha\to 0}(\boldsymbol{\xi})$ (equivalent to the Hartley entropy) more practical.

The **2nd** question can be answered by Shannon entropy value, which is theoretically the same asymptotical value of Tsallis and Renyi entropies at $\alpha \to 1$: $H^T_{\alpha\to 1+}(\boldsymbol{\xi})$, $H^T_{\alpha\to 1-}(\boldsymbol{\xi})$, $H_{\alpha\to 1+}(\boldsymbol{\xi})$, and $H_{\alpha\to 1-}(\boldsymbol{\xi})$ all theoretically should be converging to Shannon entropy. Though because of computationally complicated asymptotical behavior of both Tsallis and Renyi entropies at $\alpha \to 1$ the direct Shannon entropy calculation by using the expression $(\mathbf{e4.1})$ or the statistical methods from the NIST special publication 800-22 [1] are the preferred method of answering the question about the model's average entropy.

The answer to the **3rd** question seems straightforward and is given by the min entropy value $H_{\alpha\to\infty}(\boldsymbol{\xi})$, which is usually computed by using one of the formulae in the chain of identities of the expression $(\mathbf{e4.3})$**.** Though the question and the answer do not seem complicated, the simple distributions $\boldsymbol{\zeta_{16}}$, and $\boldsymbol{\eta_{16}}$ from the example 4.1 demonstrate that there are non-identical finite discrete probability distributions that may have same min-entropy and Hartley entropy. The distributions from the example 4.1 demonstrated that despite matching Hartley and min entropies the Renyi entropy α-spectrums can be different for all α, but $\alpha = 0$ and $\alpha \to \infty$. So, the general randomness of the distributions $\boldsymbol{\zeta_{16}}$, and $\boldsymbol{\eta_{16}}$ in all cases but the worst is different.

Unfortunately, there may be are multiple possible answers to the question "How **uniform** is the distribution produced by the model examined?". As the answer would depend on the information available and obtainable in the particular application. Considering the type, cardinality and computational resources available, the uniformity analysis can be performed in multiple ways, some of which were illustrated above.

Distribution uniformity evaluation can be performed by examining Renyi entropy α-spectrum explained and illustrated in section 4, which allows us to estimate uniformity of the distribution quite well. In the cases when the separation between the α-spectra is not sufficient, the min entropy can be of assistance. Renyi entropy α-spectrum analysis is effectively equivalent to the use of Renyi divergence described in the section 5 of the document. In the cases when Renyi divergence α-spectrum analysis is used the use of the expression $(\mathbf{e5.3})$ should be used, because

a direct use of expression $(\mathbf{e5}.\mathbf{1})$ can lead to the computational artifacts caused by the rounding and loss of precision during computation.

The Tsallis entropy at $\alpha \to 0$ (section 6) is an excellent candidate to replace Hartley entropy for estimating the count of non-zero probability events in the examined distribution. Though the Lesche stability makes the Tsallis entropy α-spectra less useful for distribution's uniformity identification or testing as explained in 6.1. On the other hand, the "Renyi-like" divergence of Tsallis entropy α-spectra, presented and illustrated in section 7, may be useful for uniformity analysis, security automation, and some artificial intelligence applications due to the divergence quantity spectrum being consistently concentrated in the confined range of the parameter α lower values.

Caution should be exercised when the Tsallis divergence α-spectrum (section 8) is used for identifying uniformity of the distributions. While allowing to evaluate the degree of uniformity present in the given distribution, the inherent computational instability of the Tsallis divergence given by expressions $(\mathbf{e8}.\mathbf{1}) - (\mathbf{e8}.\mathbf{2})$ demands to pay more attention to possible computational artifacts. For example, Tsallis divergence of two identical distributions is analytically zero as demonstrated by $(\mathbf{e8}.\mathbf{3})$. Never the less, when computed using either expression $(\mathbf{e8}.\mathbf{1})$ or $(\mathbf{e8}.\mathbf{2})$ instead of $(\mathbf{e8}.\mathbf{3})$ the computed values were diverging from the expected value of 0 in the conducted experiments. Hence, the Tsallis divergence α-spectrum should be used directly for identifying uniformity of the distributions very carefully. Though it is possible that there are analytical transformations capable of making Tsallis divergence computationally stable.

10 Generalized Entropy and Divergence Models

The main focus of this report is expansion of the traditional toolset of the Shannon [1] and min entropies [2, 3, 4] for randomness and uniformity analysis with more delicate tools allowing more detailed distribution analysis. The models based on Renyi and Tsallis entropies reviewed above are not in any way unique and further research into usability of the more general models of entropy may yield better methods for randomness and uniformity identification. A few more general entropy models are listed below.

Sharma-Mittal entropy is known [15] to generalize in one analytical expression Renyi, Tsallis, and a few other entropy models with two parameters, unlike one as in the cases of Renyi and Tsallis entropies. Sharma-Mittal entropy is usually defined as follows:

$$H_{\alpha,\beta}^{SM}(\boldsymbol{\zeta}) \stackrel{\text{def}}{=} \frac{1}{2^{(1-\beta)}-1}\left(\left[\sum_{i=1}^{n} {p_i}^{\alpha}\right]^{\frac{(\beta-1)}{(\alpha-1)}} - 1\right)$$

Sharma-Mittal entropy is actively researched in the field of information and coding theory [18, 19]. Sharma-Mittal entropy derived divergences, cross-entropies and relative entropies are also interesting topic of research.

Other generalized entropy models can be built by using the idea of Kolmogorov-Nagumo quasilinear functional average and an invertible function φ and function's inverse φ^{-1}instead of

the arithmetic GWM presented in section 3.1. The Kolmogorov-Nagumo quasilinear averages can replace the generalized averages yielding the entropy sometimes referred to as Tsallis quasilinear entropy [6, 20]:

$$H_{\alpha,\varphi}^{TQL}(\boldsymbol{\zeta}) \stackrel{\text{def}}{=} \log_\alpha \varphi^{-1}\left(\sum_{i=1}^{n} p_i \varphi({p_i}^{-1})\right) = \log_\alpha \varphi^{-1}\left(\sum_{i=1}^{n} p_i \varphi\left(\frac{1}{p_i}\right)\right)$$

Extensive analysis of entropies' statistical properties was performed in [21]. The analysis, challenging, and completeness examination of Shannon-Khinchin axioms is also an active research area [22]. Multiple models of distributions divergence, relative entropies, and cross-entropies are also intensively researched and may be readily applied to the problems of randomness and uniformity identification.

Appendix A—References

[1] National Institute of Standards and Technology Special Publication (SP), "800-22, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," NIST, Gaithersburg, MD, 2010.

[2] National Institute of Standards and Technology Special Publication (SP), "800-90A, Recommendation for Random Number Generation Using Deterministic Random Bit Generators," NIST, Gaithersburg, 2012.

[3] National Institute of Standards and Technology Special Publication (SP), "800-90B, Second Draft," NIST, Gaithersburg, 2016.

[4] National Institute of Standards and Technology Special Publication (SP), "800-90C, Draft," NIST, Gaithersburg, MD, 2012.

[5] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2 ed., Hoboken, NJ: John Wiley & Sons, 2006.

[6] J. C. Principe, Information theoretic learning : Renyi's entropy and kernel perspectives, New York, NY: Springer, 2010.

[7] A. Rényi, "On Measures of Entropy and Information," in *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960,* Berkeley, CA, 1961.

[8] C. E. Shannon, " A Mathematical Theory of Communication," *Bell System Technical Journal,* vol. 27, no. 4, pp. 379-423, July 1948.

[9] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* vol. 27, no. 4, pp. 623-666, October 1948.

[10] C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, Urbana, IL: University of Illinois, 1949.

[11] T. van Erven and P. Harremoes, "Rényi Divergence and Kullback-Leibler Divergence," *IEEE Transactions on Information Theory,* vol. 60, no. 7, pp. 3797 - 3820, 2014.

[12] A. Y. Khinchin, Mathematical Foundations of Information Theory, New York, NY: Dover Publications, 1957.

[13] J. Havrda and F. Charvat, "Quantification Method of Classification Processes. The Concept of Structural Entropy," *Kybernetika,* pp. 30-35, 1967.

[14] Z. Daroczy, "Generalized Information Functions," *Information and Control,* pp. 36-51, 1970.

[15] B. Sharma and D. Mittal, "New non-Additive Measures of Entropy for Discrete Probability Distributions," *Journal of Mathematical Sciences,* vol. 10, no. 1, pp. 28-40, 1975.

[16] B. Lesche, "Instabilities of Rényi entropies," *Journal of Statistical Physics,* vol. 27, no. 2, pp. 419-422, February 1982.

[17] C. Tsallis, "Possible Generalization of Boltzmann-Gibbs Statistic," *Journal of Statistical Physics,* vol. 52, no. 1, pp. 479-487, 1988.

[18] S. Kumar and A. Choudhary, "A Coding Theorem for the Information Measure of Order α and of Type β," *Asian Journal of Mathematics and Statistics,* 2011.

[19] S. Kumar and A. Choudhary, "Shrma-Mittal Entropy and Coding Theorem," *Tamsui Oxford Journal of Information and Mathematical Sciences,* vol. 29, no. 1, pp. 19-27, 2013.

[20] M. Masi, "A step beyond Tsallis and Rényi entropies," *Physics Letters A,* vol. 338, no. 3-5, pp. 217-224, 2005.

[21] M. D. Esteban and D. Morales, "A summary on entropy statistics," *Kybernetika,* vol. 31, no. 4, pp. 337-346, 1995.

[22] V. M. Ilić and M. S. Stanković, "Generalized Shannon–Khinchin axioms and uniqueness," *Physica A: Statistical Mechanics and its Applications,* vol. 411, pp. 138-145, 1 October 2014.