On Asymptotic Normality of Entropy Measures

Atıf Evren
Yildiz Technical University, Faculty of Sciences and Literature
Department of Statistics, Davutpasa, Esenler, 34210, Istanbul
Turkey

Erhan Ustaoğlu
Marmara University, Faculty of Administrative Sciences
Department of Management Information Sciences, Bahçelievler, 34180 Istanbul
Turkey

Abstract

Since distributions of qualitative variables can be represented by multinomial distributions, the role of multinomial distribution in entropy considerations is essential in statistics. Moreover for larger sample sizes multinomial distributions can be approximated well by multivariate normal distributions. The measures of qualitative variations depend on either class frequencies or some functional forms of class frequencies. Therefore the connection between qualitative variation statistics and normality seems straightforward for larger sample sizes. Asymptotic distributions of Shannon, Rényi and Tsallis entropies make some hypothesis testing and inferential techniques applicable to qualitative variations because some entropy measures are also frequently used in qualitative variation calculations. In this study, first we will give few examples of such applications by three entropy measures. Then we make a comparison between the performances of these three entropy measures. Finally, the degree of uncertainty, which is a significant factor that affects the speed of convergence to normality, is emphasized.

Keywords: Asymptotic normality, Rényi entropy, Shannon entropy, Tsallis entropy

Introduction

For Boltzmann, the entropy of a physical system is the measure of disorder. The entropy of statistical mechanics is Boltzmann's constant times the natural logarithm of the number of possible states. The entropy of a statistical experiment, on the other hand, can be evaluated as a measure of uncertainty before a statistical experiment takes place. So in a statistical sense, entropy and the amount of information are two closely related concepts. Since, uncertainty is not present after experimentation; entropy can be viewed as the amount of information that can be gathered through sampling.

For some introductory concepts and applications of statistical entropy, one can refer to Renyi (2007-a, 2007b), Pierce (1980), Khinchin (1957), Ash (1990), Cover and Thomas (2006) and Reza (1994). For more advanced topics in entropy (and especially for statistical applications of entropy concepts) Pardo (2006), and Esteban & Morales(1995) should be highlighted. Finally as a comprehensive study on different entropy measures, Ullah A. (1996) should be emphasized.

Some of the frequently used entropy measures are Shannon, Rényi and Tsallis entropies. Among them maybe the most popular one is Shannon entropy. Rényi entropy has gained popularity, recently, especially for the purposes of information sciences. The role that Rényi entropy played in machine learning is crucial (Principe, 2010). Another popular entropy measure, which is based on a different parameterization technique, is Tsallis entropy. Gini Concentration Index is a special case of Tsallis entropy as well as Shannon entropy is the limit of Rényi entropy as α approaches to unity. Therefore Rényi and Tsallis entropies serve as envelopes to bring researchers more flexibility for further analysis.

Qualitative Variation, Entropy and Multinomial Distributions

Especially, when the random variable is qualitative, it is impossible to calculate the mean, variance and standard deviation. In such cases, measures based on frequencies of each category are to be used to measure qualitative variation. Among several qualitative variation measures, entropy measures have gained familiarity, recently. Finally, it should be noted that a suitable form of multinomial distribution could model a qualitative distribution.

The Relation between Entropy Measures and Multivariate Normality

Increasing the sample size of the binomial distribution with parameter θ tends to a normal distribution with mean $n\theta$ and variance$n\theta\,(1-\theta)$. A similar result holds for the multinomial distribution. If the probability that a random observation comes from the ith class is π_i (i=1, 2, k), then the observed frequencies f_i will tend to a multivariate normal distribution with means $n\pi_i$ and the as sample size n increases indefinitely (Agresti, 2002). The variance-covariance matrix V will be given as below;

$$V = n\begin{bmatrix} \pi_1(1-\pi_{1)} & -\pi_1\pi_2 & \cdots & -\pi_1\pi_k \\ -\pi_1\pi_2 & \pi_2(1-\pi_{2)} & \cdots & -\pi_2\pi_k \\ \cdots & \cdots & \cdots & \cdots \\ -\pi_1\pi_k & \cdots & \cdots & \pi_k(1-\pi_{k)} \end{bmatrix} \qquad (1)$$

Delta Method for Function of Random Variable and Asymptotic Normality

Let T_n denote a statistic based on a sample size n. For large sample sizes, suppose that T_n is approximately normally distributed about θ with approximate standard error ${}^{\sigma}/_{\sqrt{n}}$. More precisely, as $n \to \infty$, the cdf of $\sqrt{n}(T_n-\theta)$ converges to a $N(0, \sigma^2)$ cdf. Then

$$\sqrt{n}(T_n-\theta) \xrightarrow{d} N(0, \sigma^2) \; (2)$$

For a function g which is at least twice differentiable, the limiting distribution of g (T_n) can be derived as

$$\sqrt{n}(g(T_n)-g(\theta)) \xrightarrow{d} N(0, \sigma^2\left[g'(\theta)\right]^2) \qquad (3)$$

This, which is called as “delta method”, forms the necessary link between entropy estimators and asymptotic normality.

Shannon Entropy

Let the discrete random variable X takes on the value$x_1, x_2, \dots, x_K$with respective probabilities$p_1, p_2, \dots, p_K$. Shannon entropy is defined as

$$H_s = -\sum_{i=1}^{K} p_i log p_i \qquad (4)$$

The unit of entropy is bit if the base of the logarithm is taken to be 2 and nat if the base is Nauperian (Garcia, 1994). In case of maximum heterogeneity (i.e., the each probability is equal to each other)$H_s = logK$.The upper limit of Shannon entropy depends on the number of categories; K. Let $\widehat{H}$ is the estimator of Shannon entropy. It is calculated as

$$\widehat{H} = -\sum_{i=1}^{K} \hat{p}_i log \hat{p}_i \; (5)$$

Here $\hat{p}_i$ probabilities are estimated by maximum likelihood method. Although this estimator is biased, increasing the sample size can reduce the amount of bias. The variance of Shannon entropy is found as (Zhang Xing, 2013).

$$Var(\widehat{H}) = \frac{1}{n}(\sum_{i=1}^{K} p_i ln^2 p_i - H^2) + \frac{K-1}{2n^2} + O(n^{-3})) \; (6)$$

Note that H can be normalized dividing it by “log K”. In this case the variance of this normalized version can also be obtained easily because of the linear relationship between Shannon entropy and its normalized version. Since log K is greater than one, one can use standardized versions to decrease variability.

Rényi Entropy

Rényi entropy is defined as

$$H_R = \frac{log \sum_{i=1}^{K} p_i^{\alpha}}{1-\alpha} \quad for \;\; \alpha > 0 \; and \; \alpha \neq 1 \qquad (7)$$

Rényi entropy is also called as α type of entropy (Ullah, A., 1996). As the parameter αapproaches unity, Rényi entropy approaches to Shannon entropy. Thus Shannon entropy is a special case of Rényi entropy. Pielou suggests using Rényi entropy with $\alpha = 2$ as a diversity index (Fattorini, 2003). The variance of Rényi entropy is given as follows (Pardo, 2006);

$$Var(\widehat{H}_R) = \frac{1}{n}\left[\left(\frac{\alpha}{\alpha-1}\right)^2 \left(\sum_{i=1}^{K} p_i^{\alpha}\right)^{-2} \left(\sum_{i=1}^{K} p_i^{2\alpha-1} - \left(\sum_{i=1}^{K} p_i^{\alpha}\right)^2\right)\right] \quad (8)$$

Note that normalized versions of Rényi entropies can be formulated directly as discussed in (1.3). Like Shannon entropy, the value of Rényi entropy is log K in case of maximum entropy.

Tsallis Entropy

Tsallis (or Havrda-Charvat) entropy is known as

$$H_T = \frac{1-\sum_{i=1}^{K} p_i^{\alpha}}{\alpha-1}, \ for \ \ \alpha > 0 \ and \ \alpha \neq 1 \quad (9)$$

For $\alpha = 2$ Tsallis entropy is identical to Gini Concentration Index. The variance of this entropy estimator is (Pardo, 2006);

$$Var(\widehat{H}_T) = \frac{1}{n}\left[\left(\frac{\alpha}{\alpha-1}\right)^2 \left(\sum_{i=1}^{K} p_i^{2\alpha-1} - \left(\sum_{i=1}^{K} p_i^{\alpha}\right)^2\right)\right] \quad (10)$$

The maximum value of Tsallis entropy can be calculated as $\frac{1-K^{1-\alpha}}{\alpha-1}$ by setting $p_i = \frac{1}{K}$ in equation (9). It should also be noted that parameterization (i.e., choosing alpha value arbitrarily) does have an effect on maximum entropy as well as the number of categories. Unlike Shannon and Rényi entropies this quantity may be greater than or less than unity. So when one considers Tsallis entropy, using normalized version to decrease variability may not work all the time.

Asymptotic Sampling Distributions of Entropy Measures

In literature there are other entropy measures than Shannon, Rényi and Tsallis entropies. For asymptotic properties of entropy estimators one can refer to Pardo (2006) and Esteban & Morales (1995). To summarize let $\widehat{\boldsymbol{H}}$ be any entropy estimator whose expected value is $H = E(\widehat{H})$. If the number of categories K is finite, and the sample size is sufficiently large, the statistic $\frac{\widehat{H}-H}{\sqrt{Var\,(\widehat{H})}}$ fits standard normal distribution. It should be noted that entropy estimators are biased but the amount of bias is low especially when sample sizes are large enough. Therefore an approximate $100(1-\alpha)\%$ confidence interval for any entropy measure can be obtained as

$$\widehat{H} \pm Z_{\alpha/2}\sqrt{Var(\widehat{H})} \quad (11)$$

Here $Z_{\alpha/2}$ value is the abscissa of standard normal variable corresponding to a right-tail probability of $\alpha/2$. For some hypothesis testing examples on some entropy measures, one might refer to Magurran (1988) and Agresti (1978). For testing the equality of entropies of two populations by all three measures (Shannon, Rényi and Tsallis entropies), we consider two hypothetical frequency distributions on marital status of some people as shown in Table I. The summarizing statistics on three entropy measures; Shannon entropy; Rényi entropy (α=2), Tsallis entropy (α=2) are given in Table II. Since all three measures distribute normally asymptotically, and the probability distributions of each community on marital status are assumed to be independent, the following hypothesis testing procedure will be repeated for three entropy measures:

H_0: Entropies are equal
H_A: Entropies are not equal
Test statistic T is calculated as

$$T = \frac{Estimated\ entropy\ of\ A - Estimated\ entropy\ of\ B}{\sqrt{Variance\ of\ entropy\ of\ A + Variance\ of\ entropy\ of\ B}} \quad (12)$$

Assume α=0.05
Decision rule:
Accept H_0 if $-1.96 \leq T \leq 1.96$
Reject H_0 otherwise.

By any of these three measures, we can conclude that entropies are not equal as shown in Table III.

Simulation Study

52 different simulations on 6 different categorical distributions are run to analyze asymptotic behavior of 9 different entropy measures. Simulations are realized by macros on Microsoft Excel. Entropy measures that are under consideration are Shannon, Rényi for $\alpha = 0.5, 0.99, 1.5, 2$ and Tsallis for $\alpha = 0.5, 0.99, 1.5, 2$. The six probability distributions used in simulations are given in Table IV, Table V and Table VI.

Results

Simulation results on normality of entropy estimators for 52 trials are given in Table VII**.**

1. All entropy measures underestimate population entropies. Yet the bias is small.
2. All entropy measures are very highly correlated to each other. To give a better impression on this, the correlation matrix of entropy estimators for 34th model is given in Table VIII, since all models have this property more or less.
3. The coefficients of variation for different (normalized) entropy measures vary between 0.01% and 14.7%. Yet, it is generally the case that coefficients of variation scores are frequently very low. The minimum, average and maximum values for coefficients of variation (for normalized versions) are given in Table IX. It can also be verified that minimum range scores are generally obtained for Shannon and Ré(0.5) entropies. This situation is summarized by Figure I.
4. As a general tendency, distributions of entropy estimators tend to normality, as sample sizes and number of runs increase indefinitely. One exception is the case of maximum entropy (i.e., uniform distributions) where the rate of convergence to normality is low. This phenomenon should be underlined. For other instances, normality can be reached even for smaller sample sizes (250 or higher). As an example, the frequency distributions of all entropy measures for 34th model are given in Table X.

The non-normality of Ré(0.99) and Tsa(0.99) entropies are probably due to the fact that these two entropies are undefined at $\alpha = 1$. Other entropy measures seem fit well some forms of normal distribution. Note that normality tests of various entropy measures are realized by NCSS (2004). For illustrative purposes, the tendency of entropy measures to normality for 34th model can be checked by Figure II. Although it is impossible to summarize all entropy statistics, it can still be said that all entropy measures studied by 34th simulation have the normality property.

Conclusion

Asymptotic normality of nine entropy measures for some hypothetical multinomial populations is studied. Normalized versions of various entropy measures are considered to be able to make sound comparisons since different alpha values correspond to totally different transformations on probabilities.

The first result is that all entropy measures whether they are normalized or not, are highly and positively correlated to each other. This is important because one can arbitrarily select any of these entropy measures for further analysis. Yet we recommend considering normalized versions. Because in such cases, all normalized entropy scores fall between zero and one, a property which facilitates statistical tests on equality of means. Besides when considering Shannon and Rényi entropies, using normalized versions is beneficial because of lower variances. Secondly, the normality of all nine entropy measures seems apparent as sample sizes and number of runs increase indefinitely.

The third result is that in case of maximum entropy, the rate of convergence to normality is low. This is a factor that affects the validity of interval estimates and hypothesis-tests based on asymptotic normality. In such cases some probability inequalities like Chebyshev's inequality may be useful due to the fact that the asymptotic variances of these entropy measures are relatively low.

References

Agresti,A., Agresti,B.F. (1978). Statistical Analysis of Qualitative Variation**,** Sociological Methodology, Vol.9, 204-237.

Agresti, A. (2002). Categorical Data Analysis, Wiley Interscience (Second Edition).

Ash, R.B.(1990). Information Theory, Dover Publications, NY.

Cover, T.M.,Thomas, J.A.(2006). Elements of Information Theory, WileyInterscience (Second Edition), Hoboken, New Jersey.

Esteban, M.D., Morales, D.(1995). A summary on entropy statistics, Kybernetika, Vol. 31, No.4, 337-346.
Fattorini,L. (2003). Statistical Analysis of Ecological Diversity, Environmetrics, Unesco EOLSS Sample Chapters.
Garcia,A.L. (1994). Probability and Random Processes for Electrical Engineering, Addison-Wesley Longman (Second Edition), 162-172.
Khinchin,A.I (1957). Mathematical Foundations of Information Theory, Dover Publications, NY, 7.
Magurran, A.E. (1988). Ecological Diversity and Its Measurement, Princeton University Press, New Jersey
Pardo,L. (2006). Statistical Inference Measures Based on Divergence Measures, Chapman & Hall/CRC, 99.
Principe,J.C. (2010). Information Theoretic Learning Rényi’s entropy and Kernel Perspectives, Information Science and Statistics, Springer.
Pierce,J.R. (1980). An Introduction to Information Theory Symbols, Signals and Noise, Dover Publications, Second, revised edition, NY, 21-23, 202.
Rényi, A. (2007-a). Probability Theory, Dover Publications, NY, 554.
Rényi, A. (2007-b). Foundations of Probability, Dover Publications, NY, 23.
Reza,F.M. (1994). An Introduction to Information Theory, Dover Publications, NY.
Ullah,A. (1996). Entropy, Divergence and Distance Measures with Econometric Applications, Journal of Statistical Planning and Inference49, 137-62, Elsevier.
Zhang,Xing (2013). Asymptotic Normality of Entropy Estimators, proposal submitted to the faculty of The University of North Carolina at Charlotte in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics, Charlotte.

Tables

Table I: Two hypothetical distributions on marital status of some people

Group	Single	Married	Separated	Divorced	Widowed	Total
A	340	232	201	105	122	1000
P(A)	0.34	0.232	0.201	0.105	0.122	1
B	320	196	188	144	152	1000
P(B)	0.32	0.196	0.188	0.144	0.152	1

Table II: Summarizing statistics of three entropy measures on two frequency distributions

Summarizing Statistics	Group A	Group B
Shannon's Entropy	2.195133	2.255867
Rényi'sEntropy(α=2)	2.084768	2.184425
Tsallis' Entropy(α=2)	0.764266	0.78
Estimated Variance of Shannon's Entropy	0.000347	0.000201
Estimated Std. Deviation of Shannon's Entropy	0.018617	0.014163
Estimated Variance of Rényi's Entropy	0.000527	0.000417
Estimated Std. Deviation ofRényi's Entropy	0.022946	0.020409
Estimated Variance of Tsallis' Entropy	2.93E-05	2.02E-05
Estimated Std. Deviation of Tsallis' Entropy	0.005409	0.00449

Table III: Test Statistics for three entropy measures

Entropy Measure	Test statistics (T)
Shannon entropy	-2.596
Rényi entropy	-3.245
Tsallis entropy	-2.238

Table IV: First and second probability distributions used for random number generation

Category	Probability	Category	Probability
1	0.25	1	0.125
2	0.25	2	0.125
3	0.25	3	0.25
4	0.25	4	0.5

Distribution no.	1	Distribution no.	2

Table V: Third and fourth probability distributions used for random number generation

Category	Probability	Category	Probability
1	0.167	1	0.05
2	0.167	2	0.05
3	0.167	3	0.1
4	0.167	4	0.1
5	0.167	5	0.2
6	0.167	6	0.5
Distribution no.	3	Distribution no.	4

Table VI: Fifth and sixth probability distributions used for random number generation

Category	Probability	Category	Probability
1	0.1	1	0.05
2	0.1	2	0.05
3	0.1	3	0.05
4	0.1	4	0.05
5	0.1	5	0.05
6	0.1	6	0.05
7	0.1	7	0.1
8	0.1	8	0.1
9	0.1	9	0.1
10	0.1	10	0.4
Distribution no.	5	Distribution no.	6

Table VII: Simulation results on normality of entropy estimators for 52 trials.

Simulation	Distribution No	Explanation	Sample Size	Number of runs	Normality ?
1	1	Max. Ent.Dist.	50	250	No.
2	1	Max. Ent.Dist.	50	500	No.
3	1	Max. Ent.Dist.	500	250	No.
4	1	Max. Ent.Dist.	500	505	No.
5	1	Max. Ent.Dist.	1000	502	No
6	1	Max. Ent.Dist.	1000	1000	No.
7	1	Max. Ent.Dist.	2000	500	No
8	1	Max. Ent.Dist.	2000	1000	No.
9	1	Max. Ent.Dist.	4000	500	No.
10	1	Max. Ent.Dist.	4000	1000	No.
11	2		50	250	No
12	2		50	500	No.
13	2		500	250	Mostly yes.
14	2		500	505	Mostly yes
15	2		1000	502	Mostly yes.
16	2		1000	1000	Mostly yes.
17	2		1000	2000	Mostly yes.
18	2		2000	1000	Mostly yes.
19	2		2000	2000	Mostly yes.
20	2		2500	351	Yes.
21	2		2500	553	Yes.
22	3	Max. Ent.Dist.	50	618	No.
23	3	Max. Ent.Dist.	250	290	No.
24	3	Max. Ent.Dist.	500	501	No.
25	3	Max. Ent.Dist.	2500	250	No.
26	3	Max. Ent.Dist.	2500	500	No.
27	3	Max. Ent.Dist.	1000	1000	No.
28	4		50	273	Partly no.
29	4		50	618	Partly no.
30	4		50	1155	No,.
31	4		250	290	No
32	4		500	250	No.
33	4		500	500	No.
34	4		1000	1000	Mostly yes.
35	4		2500	250	Partly yes.
36	4		2500	500	Yes.

37	5	Max. Ent.Dist.	250	250	No.
38	5	Max. Ent.Dist.	250	500	No.
39	5	Max. Ent.Dist.	250	1500	No.
40	5	Max. Ent.Dist.	500	250	No.
41	5	Max. Ent.Dist.	500	500	No.
42	5	Max. Ent.Dist.	500	1000	No.
43	5	Max. Ent.Dist.	1000	1000	No.
44	5	Max. Ent.Dist.	2500	250	No.
45	6		250	250	Mostly yes
46	6		250	500	Partly yes.
47	6		250	1500	No.
48	6		1000	1000	Mostly yes.
49	6		1000	4000	Yes.
50	6		2972	2000	Yes
51	6		4000	500	Mostly yes.
52	6		4000	1000	Mostly yes.

Table VIII: Correlation matrix of entropy estimators for model 34

	Shannon	Ré(0.5)	Ré(0.99)	Ré(1.5)	Ré(2)	Ts(0.5)	Ts(0.99)	Ts(1.5)	Ts(2)
Shannon	1	0.98	0.84	0.99	0.97	0.98	0.84	0.99	0.97
Ré(0.5)	0.98	1	0.83	0.95	0.91	0.99	0.83	0.95	0.91
Ré(0.99)	0.84	0.83	1	0.83	0.82	0.83	1	0.83	0.82
Ré(1.5)	0.99	0.95	0.83	1	0.99	0.95	0.83	0.99	0.99
Ré(2)	0.97	0.91	0.82	0.99	1	0.91	0.82	0.99	0.99
Ts(0.5)	0.98	0.99	0.83	0.95	0.91	1	0.83	0.95	0.91
Ts(0.99)	0.84	0.83	1	0.83	0.82	0.83	1	0.83	0.82
Ts(1.5)	0.99	0.95	0.83	0.99	0.99	0.95	0.83	1	0.99
Ts(2)	0.97	0.91	0.82	0.99	0.99	0.91	0.82	0.99	1

Table IX: Coefficients of Variation for various normalized entropies

Coefficient of Variation				
Index	Minimum	Mean	Maximum	Range
SH	0.000242	0.010516	0.03676876	0.036527
Ré(0.5)	0.0001354	0.0101	0.05907258	0.058937
Ré(0.99)	0.0048117	0.025334	0.13464958	0.129838
Ré(1.5)	0.0003991	0.026556	0.12136609	0.120967
Ré(2)	0.0004916	0.03449	0.14731634	0.146825
Tsa(0.5)	0.0001877	0.014241	0.08300134	0.082814
Tsa(0.99)	0.0049494	0.025838	0.13448719	0.129538
Tsa(1.5)	0.0002767	0.019177	0.08900824	0.088732
Tsa(2)	0.0001928	0.018698	0.09058598	0.090393

Table X: Normality tests of Shannon entropies calculated in 34

Tests	Shannon	Ré(0.5)	Ré(0.99)	Ré(1.5)	Ré(2)	Ts(0.5)	Ts(0.99)	Ts(1.5)	Ts(2)
Shapiro-Wilk W	normal	normal	no	normal	normal	normal	no	normal	normal
Anderson-Darling	normal	normal	no	normal	normal	normal	no	normal	normal
Martinez-Iglewicz	normal	normal	no	normal	normal	normal	no	normal	normal
Kolmogorov-Smirnov	normal	normal	no	normal	normal	normal	no	normal	normal
D'AgostinoSkewness	normal	no	no	normal	normal	normal	no	normal	normal
D'Agostino Kurtosis	normal	normal	no	normal	normal	normal	no	normal	normal
D'Agostino Omnibus	normal	no	no	normal	normal	normal	no	normal	normal

Figures

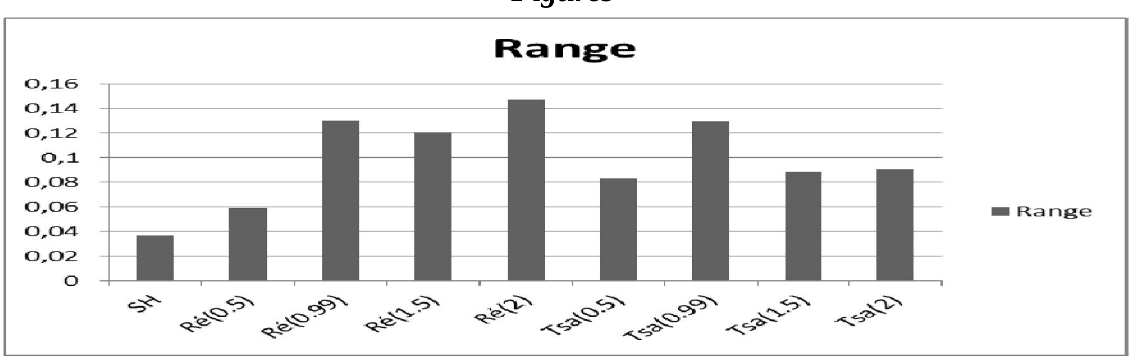

Figure I: Ranges of coefficients of determination calculated for various entropy measures

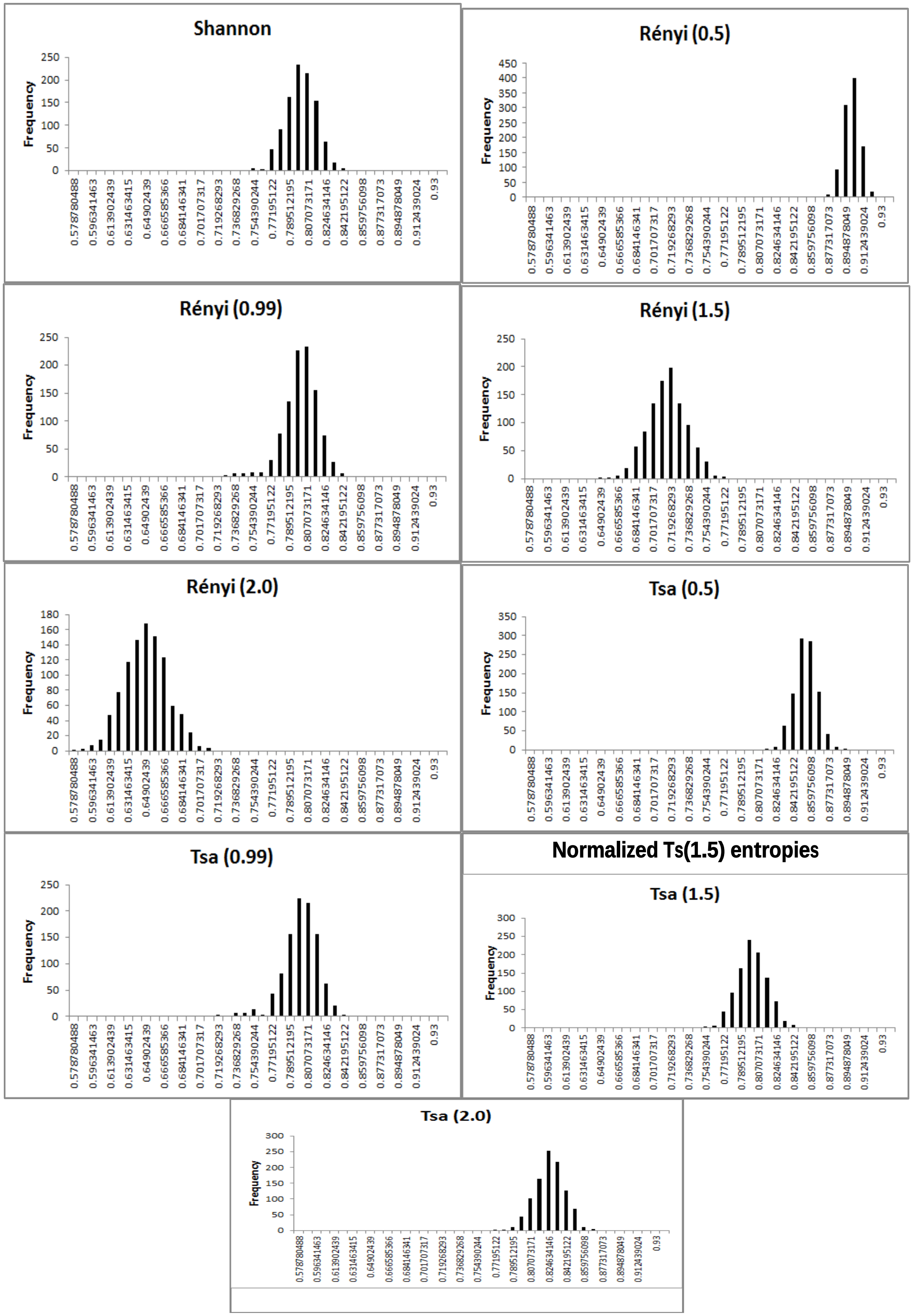

Figure II: Frequency distributions of entropies calculated for 34th model