Chapter 10

Use of Relative Entropy Statistics in Contingency Tables

Atıf Evren[1]

Erhan Ustaoğlu[2]

Elif Tuna[3]

Büşra Şahin[4]

Abstract

There are various information theoretic divergence measures used for determining associations between nominal variables. Among them, Shannon mutual information statistic is especially appealing, since its sampling properties are well-known. Although Shannon mutual information is more frequently used, Rényi and Tsallis mutual informations, as envelopes of various tools, provide much higher flexibility than Shannon mutual information. Indeed, Shannon mutual information is a special case of Kullback-Leibler divergence, Rényi, and Tsallis mutual informations. In this study, large sampling properties of Shannon, Rényi, Tsallis mutual information statistics are considered as well as Pearson, Tschuprow, Sakoda, Cramér, Hellinger, and Bhattacharyya measures. In simulations, the normality of most of the statistics, and the higher positive correlation coefficients between all these tools are observed. Their sampling variabilities are compared. Then by using Rényi and Tsallis mutual information statistics, correlation coefficients are estimated for 8 different scenarios, and 3 bivariate normal distributions.

1 Assoc. Prof. Dr.,Yildiz Technical University, Faculty of Sciences and Literature, Department of Statistics, Istanbul; aevren@yildiz.edu.tr, ORCID: 0000-0003-4094-7664

2 Assoc. Prof. Dr., Marmara University, Faculty of Management, Department of Management of Information Sciences, Istanbul; erhan.ustaoglu@marmara.edu.tr, ORCID: 0000-0002-9077-4370

3 Assoc. Prof. Dr., Yildiz Technical University, Faculty of Sciences and Literature , Department of Statistics; Davutpasa, Istanbul; eozturk@yildiz.edu.tr, ORCID: 0000-0001-8572-3109

4 Assist. Prof. Dr., Halic University, Faculty of Engineering, Department of Software Engineering, Istanbul; busrasahin@halic.edu.tr, ORCID: 0000-0003-3995-9238

https://doi.org/10.58830/ozgur.pub518.c2132

1. Introduction

Distance functions in statistics are applied frequently. Some basic statistics like variance, and standard deviation are simply some functions of Euclidean and Minkowski distances. Kolmogorov-Smirnov statistic can be derived from Chebyshev distance, etc. Some tests of independence consider the distance (or divergence) between a joint probability distribution, and the product of two marginal probability functions. Thus, distance or divergence measures are used in hypothesis testing, as well as clustering some multivariate data. In Bayesian methodology, to evaluate sample information, comparisons between conjugate distributions can be made by some divergence measures. Some procedures of the sequential analysis of Wald (Wald, 2004) are based on the divergence between two Bernoulli distributions.

Analysis of dependencies may also be realized by entropy and relative entropy formalism. Mathematical foundations of entropy and relative entropy can be found in Shannon (Shannon and Weaver, 1963) and Khinchin (Khinchin, 1957). Preliminaries can be found also in Rényi (Renyi, 2007), Ash (Robert, 1990), and Cover&Thomas (Cover and Thomas, 2006). Following Shannon's contribution to information theory, studies by Kullback (Kullback, 1997), Gokhale (Gokhale and Kullback, 1978), and Pardo (Pardo, 2006), rather focused on inferential issues of statistics within the framework of entropy.

Kulback considered entropy and mutual information (a special case of relative entropy) as the two key concepts of his system. Pardo discussed various divergence measures (including relative entropy) in goodness of fit testing, loglinear models and contingency tables. Singh (Singh, 1998) and Samilov (Şamilov, 2015) contributed in entropy-based parameter-estimation, and optimizing entropy. From econometrics (Ullah, 1996) to information-theoretic learning (Principe, 2010); the applications of entropy and divergence measures vary on a wide range.

In this study, Rényi and Tsallis mutual informations are considered in the context of two-way contingency tables. In case of nominal bivariate distributions, they can serve as an association measure between variables. In case of continuity of the variables, they may be used in estimating correlations after grouping bivariate continuous observations by the help of a contingency table as well.

2. Independence Tests for Contingency Tables

For contingency tables, the test of independence is based on the chi-square statistic

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij} - n.p_{ij}\right)^2}{n.p_{ij}} \tag{1}$$

where p_{ij} is the hypothesized probability for the ith column and jth row of the contingency table, having I columns, and J rows and n_{ij} is the observed frequency. If the variables are independent, $p_{ij} = p_{i+}p_{+j}$ for all i, j for which $p_{i+} = \sum_{j=1}^{J} p_{ij}$, $p_{+j} = \sum_{i=1}^{I} p_{ij}$. Based on n independent observations, substitution of the estimate $\hat{p}_{ij} = \hat{p}_{i+}\hat{p}_{+j} = \frac{n_{i+}}{n}\frac{n_{+j}}{n}$ into (1) yields

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij} - n_{i+}.n_{+j} / n\right)^2}{n_{i+}.n_{+j} / n} \tag{2}$$

If the variables are independent, this quantity is chi-square variable with $(I-1)(J-1)$ degrees of freedom. For an $I\,x\,J$ contingency table, let $q = Min(I\ ,J)$. It can be shown that $0 \le \chi^2 \le n(q-1)$ (Liebetrau, 1983). (2) is equivalent to

$$\varphi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij} - p_{i+}.p_{+j}\right)^2}{p_{i+}.p_{+j}} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{p_{ij}^2}{p_{i+}p_{+j}} - 1 \tag{3}$$

Note $0 \le \varphi^2 \le (q-1)$ with $\varphi^2 = 0$; in case of independence; and $\varphi^2 = q-1$; in case of perfect association. Alternatively, Pearson proposed

$$P = \left(\frac{\varphi^2}{1+\varphi^2}\right)^{1/2} \tag{4}$$

This measure is also called Pearson contingency coefficient. The range of P depends on the

dimensions of the contingency table. Sakoda (Sakoda, 1976) suggests

$$p^* = \frac{P}{p_{max}} = \left(\frac{q.\varphi^2}{(q-1).(1+\varphi^2)}\right)^{1/2} \tag{5}$$

Tschuprow considered the following

$$t=\left(\frac{\varphi^2}{\sqrt{(I-1).(J-1)}}\right)^{1/2} \tag{6}$$

Finally Cramér statistic is

$$v=\left(\frac{\varphi^2}{q-1}\right)^{1/2} \tag{7}$$

which is 1; in case of perfect association; and 0; in case of dependence; for all values of J and I.

3. Entropy as a Measure of Variability

The entropy of a statistical experiment is a measure of uncertainty. Various entropy formulations of probability distributions and various entropy measures are summarized well by Mihalowicz et al. (Mihalowicz et al.,2014) and Esteban&Morales (Estaban and Morales, 1995).

If the discrete random variable X takes values $x_1, x_2, \ldots, x_K$ with respective probabilities $p_1, p_2, \ldots, p_K$ on sample space S, then Shannon entropy is defined as

$$H_s=-\sum_{i=1}^{K} p_i.lnp_i \tag{8}$$

Rényi entropy is a generalization of Shannon entropy. It is defined as

$$H_{Ré}(\alpha)=\frac{ln\sum_{i=1}^{K} p_i^{\alpha}}{1-\alpha} \quad for \quad \alpha>0 \ and \ \alpha\neq 1 \tag{9}$$

As $\alpha \to 1$, Renyi entropy approaches to Shannon entropy. Another generalized form of

Shannon entropy is due to Tsallis. It is defined as

$$H_{Tsa}(\alpha)=\frac{1-\sum_{i=1}^{K} p_i^{\alpha}}{\alpha-1}, \ for \quad \alpha>0 \ and \ \alpha\neq 1 \tag{10}$$

Tsallis entropy approaches to Shannon entropy as $\alpha \to 1$. Asymptotic properties of several entropy measures are given by Zhang (Zhang, 2013).

4. Bivariate Distributions and Distance Measures

Suppose the pair of discrete random variables X and Y assume $x_1, x_2, \dots, x_n$ *and* $y_1, y_2, \dots, y_m$. If the joint probability function is denoted by $P_{X,Y}(x,y)$, then Shannon entropy is

$$H_S(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{X,Y}(x_i,y_j).logP_{X,Y}(x_i,y_j) \quad (11)$$

iii.1. **Renyi Divergence and Kullback-Leibler Information**

Rényi order-α divergence of q from p is defined as

$$D_{Ré}(p\,||\,q) = \frac{1}{\alpha-1} log \sum \frac{p^{\alpha}}{q^{\alpha-1}} \quad (12)$$

Here p and q are two discrete probability distribution functions. When $\alpha \to 1$

$$\lim_{\alpha\to1} \frac{1}{\alpha-1} log \sum \frac{p^{\alpha}}{q^{\alpha-1}} = \frac{0}{0}, \quad \lim_{\alpha\to1} D_{Ré}(p\,||\,q) = \lim_{\alpha\to1} \frac{\frac{d}{d\alpha}\left(log \sum \frac{p^{\alpha}}{q^{\alpha-1}}\right)}{\frac{d}{d\alpha}(\alpha-1)} = \sum p\, log\left(\frac{p}{q}\right) \quad (13)$$

The quantity, $\sum p\, log\left(\frac{p}{q}\right)$, is called relative entropy. It is the divergence between two probability distributions, p and q. Tsallis order-α divergence of q from p is defined as

$$D_{Tsa}(p\,||\,q) = \frac{1-\sum \frac{p^{\alpha}}{q^{\alpha-1}}}{1-\alpha} \quad (14)$$

iii.2. **Bhattacharyya and Hellinger Distances**

Alternatively, Bhattacharyya coefficient (B.C.), and Bhattacharyya distance between two distributions, p and q are defined as (Upton and Cook, 2006)

$$B.C. = \sum \sqrt{p.q} \quad (15)$$

$$D_{Bha}(p\,||\,q) = Arccos(B.C.) \quad (16)$$

Hellinger distance can also be calculated directly by Bhattacharyya coefficient;

$$D_{Hel}\left(p \,||\, q\right) = \sqrt{2\left(1 - B.C.\right)} \tag{17}$$

4.3. Shannon Mutual Information as a Measure of Independence

Shannon mutual information (or Kullback-Leibler divergence) is defined as

$$I(X,Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} P\left(X = x_i, Y = y_j\right) log \frac{P\left(X = x_i, Y = y_j\right)}{P\left(X = x_i\right).P(Y = y_j)} \tag{18}$$

If X and Y are independent, it is zero. For bivariate normal distribution, it is

$$I\left(X;Y\right) = -\frac{\ln\left(1-\rho^2\right)}{2} \tag{19}$$

Thus, the correlation coefficient can easily be estimated by

$$\hat{\rho} = \sqrt{1 - \exp(-2.\widehat{I\left(X;Y\right)})} \tag{20}$$

4.4. Bivariate Version of Renyi and Tsallis Divergences

Rényi order-α divergence of $P_{X,Y}\left(x,y\right)$ from $P_X\left(x\right)P_Y\left(y\right)$ is defined analogously,

$$D_{Ré}\left(P_{X,Y}\left(x,y\right) || P_X\left(x\right).P_Y\left(y\right)\right) = \frac{1}{\alpha - 1} log \sum\sum \frac{P_{X,Y}\left(x,y\right)^{\alpha}}{\left(P_X\left(x\right).P_Y\left(y\right)\right)^{\alpha-1}} \tag{21}$$

It is also called as Rényi mutual information. Tsallis order-α divergence is

$$D_{Tsa}\left(P_{X,Y}\left(x,y\right) || P_X\left(x\right).P_Y\left(y\right)\right) = \frac{1 - \sum\sum \frac{P_{X,Y}\left(x,y\right)^{\alpha}}{\left(P_X\left(x\right).P_Y\left(y\right)\right)^{\alpha-1}}}{1-\alpha} \tag{22}$$

In case of independence, Rényi and Tsallis mutual informations are zero as Shannon mutual information. As $\alpha \to 1$, Rényi and Tsallis mutual informations approach to Shannon mutual information.

4.5. **Bhattacharyya and Hellinger Measures for Testing independence**

Bivariate versions of Bhattacharyya and Hellinger distances are straightforward:

$$D_{Bha}\left(P_{X,Y}(x,y)\,||\,P_X(x).P_Y(y)\right)=Arccos\left(\sum\sum\sqrt{\frac{P_X(x).P_Y(y)}{P_{X,Y}(x,y)}}P_{X,Y}(x,y)\right) \quad (23)$$

$$D_{Hel}\left(P_{X,Y}(x,y)\,||\,P_X(x).P_Y(y)\right)=\sqrt{2\left(1-\sum\sum\sqrt{\frac{P_X(x).P_Y(y)}{P_{X,Y}(x,y)}}P_{X,Y}(x,y)\right)} \quad (24)$$

5. Experimental Results

In simulations, 8 discrete bivariate distributions are considered. 5 of them are 2x2, the rest are 3x4 contingency tables. The variables are assumed to be continuous for making comparisons between information-theoretic independence measures and classic correlation measures. The eight bivariate distributions presented are given below:

Table 1.1. 2x2 tables used in simulations having correlations 0.033, 0.88 and -0.88.

Distribution1			**Distribution2**			**Distribution3**		
X/Y	1	2	X/Y	1	2	X/Y	1	2
1	0.5	0,4	1	0.5	0.03	1	0.03	0.5
2	0.05	0.05	2	0.03	0.44	2	0.44	0.03
Correlation	0.033		Correlation	0.88		Correlation	-0.88	

Table 1.2 2x2 tables used in simulations having correlations 0.49 and -0.49.

Distribution4			**Distribution5**		
X/Y	1	2	X/Y	1	2
1	0.45	0.15	1	0.1	0.3
2	0.1	0.3	2	0.45	0.15
Correlation	0.49		Correlation	-0.49	

Table 1.3 3x4 tables used in simulations having correlations -0.86 and 0.012.

Distribution6					**Distribution7**				
X/Y	1	2	3	4	X/Y	1	2	3	4
1	0.01	0.01	0.01	0.27	1	0.1	0.1	0.05	0.05
2	0.01	0.01	0.37	0.01	2	0.1	0.1	0.15	0.05
3	0.27	0.01	0.01	0.01	3	0.12	0.04	0.1	0.04
Correlation	-0.86				Correlation	0.012			

Table 1.4 3x4 table used in simulations with correlation 0.86.

Distribution8				
X/Y	1	2	3	4
1	0.27	0.01	0.01	0.01
2	0.01	0.37	0.01	0.01
3	0.01	0.01	0.01	0.27
Correlation	0.86			

1000 independent observations are picked from each distribution. This procedure is repeated 1000 times. Then, the same experiments are repeated with 2000 independent observations, 2000 times. Simulations are realized by the help of Microsoft Excel.

Large sampling properties of 19 association measures studied are given below.

Table 2. The goodness of fit statistics used in determining associations.

Measure	**Explanation**	**Eqn. No.**	**Measure**	**Explanation**	**Eqn. No.**
I(X;Y)	Mut.Info (Shannon)	18.	R2	Mut. Info (Rényi) α=2	21.
Chi-square	Chi-square	3.	T2	Mut. Info (Tsallis) α=2	22.
Hellinger	Hellinger Distance	24.	R3	Mut. Info (Rényi) α=3	21.
Bhattacharyya	Bhattacharyya Distance	23.	T3	Mut. Info (Tsallis) α=3	22.

R0.25	Mut.Info (Rényi) α=0.25	21.	Cor. meas.	Correlation measure	20.
T0.25	Mut. Info (Tsallis) α=0.25	22.	Pears. cc.	Pearson cont. coef.	4.
R0.5	Mutual Info (Rényi) α=0.5	21.	Sakoda	Sakoda coef.	5.
T0.5	Mut. Info (Tsallis) α=0.5	23.	Tschupr.	Tschuprow coef.	6.
R1.5	Mut.Info (Rényi) α=1.5	21.	Cramér	Cramér coef.	7.
T1.5	Mut. Info (Tsallis) α=1.5	22.			

5.1. Normality results

The normality of each statistic has been investigated. The seven normality tests used are Shapiro Wilk W, Anderson Darling, Martinez Iglewicz, Kolmogorov-Smirnov, D'Agostino Skewness, D'Agostino Kurtosis, D'Agostino Omnibus tests. The percentage of normality observed by simulations, for each statistic is given in Table 3. In general, the tendency to distribute normal is obvious with few exceptions, namely, Tsallis and Rényi mutual informations with $\alpha = 0.25$ and $\alpha = 0.5$0 values.

Table 3. The tendency to normality of each statistics

Measure	% of Normality in Runs	Measure	% of Normality in Runs
I(X;Y)	75	T1.5	75
Chi-square	75	R2	75
Hellinger	87.5	T2	75
Bhattacharyya	87.5	R3	81.25
R0.25	0	T3	75
T0.25	12.5	Cor. meas.	81.25
R0.5	18.75	Pears. cc.	62.5
T0.5	31.75	Sakoda	75
R0.75	56.25	Tschuprow	81.25
T0.75	62.5	Cramér	81.25
R1.5	75		

5.2. Sampling variabilities and average correlations

Minimum standard deviations, and maximum average correlations (with the other association measures) are presented in Table 4. Pearson contingency coefficients, and T0.5 have generally the lowest variability. Mutual information, chi-square statistic, and T2 have maximum average correlations with others.

Table 4. Minimum standard deviations, maximum average correlations

Distribution	Population Correlation	Min. Std. Deviation	Measure	Maximum Ave. Corr.	Measure
1(1000x1000)	0.033	0.0005	T0.25	0.9134	I(X;Y)
1(2000x2000)	0.033	0.0003	T0.25	0.8914	I(X;Y)
2(1000x1000)	0.88	0.0063	Pears.cc	0.9963	Chi-sq(=T2).
2(2000x2000)	0.88	0.0045	Pears. cc.	0.9972	Chi-sq(=T2).
3(1000x1000)	-0,88	0.0063	Pears.cc	0.9974	Chi-sq(=T2).
3(2000x2000)	-0,88	0.0046	Pears. cc.	0.9974	Chi-sq(=T2).
4(1000x1000)	0.49	0.0045	T0.25	0.9981	R2
4(2000x2000)	0.49	0.0046	Pears. cc.	0.9974	Chi-sq(=T2).
5(1000x1000)	-0.49	0.0044	T0.25	0.998	I(X;Y)
5(2000x2000)	-0.49	0.003	T0.25	0.998	I(X;Y)
6(1000x1000)	-0.86	0.0019	Cramér	0.9826	I(X;Y)
6(2000x2000)	-0.86	0.0013	Pears. cc.	0.9856	I(X;Y)
7(1000x1000)	0.012	0.0025	T0.25	0.9916	R1.5
7(2000x2000)	0.012	0.0016	T0.25	0.9901	I(X;Y)
8(1000x1000)	0.86	0.0019	R0.25	0.9857	Chi-sq(=T2).
8(2000x2000)	0.86	0.0014	Cramér	0.9873	Chi-sq(=T2).

5.3. Estimating correlations by mutual information statistics

Correlation coefficients are estimated by plugging Shannon, Rényi and Tsallis mutual information quantities in (20). The best estimates for correlation coefficient, and related statistics are given in Table 5. All mutual information statistics are nonnegative; therefore, they only give information about the magnitudes of correlations. The distributions represented by 8 contingency tables are not bivariate normal, but equation (20) is still used intentionally. As a general tendency, for lower population correlations, Tsallis and Rényi estimates with low α parameters are observed to be successful. Similarly, for higher population correlations, selecting higher α parameters may work. For moderate correlations, Shannon mutual information performs well.

Table 5. Correlation estimates for 8 bivariate tables; normality assumption is violated.

Distribution	Population Correlation	Estimated Magnitude of Cor.	Best Measure
1(1000x1000)	0.033	0.036	T0.5
1(2000x2000)	0.033	0.031	R0.5
2(1000x1000)	0.88	0.928	R3
2(2000x2000	0.88	0.928	R3
3(1000x1000)	-0,88	0.887	R2
3(2000x2000)	-0,88	0.887	T2
4(1000X1000)	0.49	0.473	I(X;Y)
4(2000x2000)	0.49	0.471	I(X;Y)
5(1000x1000)	-0,49	0.472	I(X;Y)
5(2000x2000)	-0,49	0.473	I(X;Y)
6(1000x1000)	-0,86	0.864	R0.75
6(2000x2000)	-0,86	0.863	R0.75
7(1000x1000)	0.012	0.144	T0.25
7(2000x2000)	0.012	0.14	T0.25
8(1000x1000)	0.86	0.864	R0.75
8(2000x2000)	0.86	0.864	R0.75

5.4. Correlation estimates for bivariate normal distribution

Correlation estimates by various mutual information statistics, in case of bivariate normality, are made by simulating 1000 pairs. Simulations are realized by Microsoft Excel. Results are in Table 6.

Table 6. Correlation estimates for bivariate normal

Population Correlation	Absolute Value of Best Correlation Estimate (9 groups)	Absolute Value of Best Correlation Estimate (16 groups)
0	0.047(R0.25)	0.054(T0.25)
0.5	0.48(T2)	0.499(T2)
-0,5	0.521(R2)	0.501(R1.5)
0,75	0.754(T2)	0.763(R2)
-0,75	0.741(R2)	0.771(R2)

The number of groups used in discretizing normal data seems to be effective on mutual information statistics. In order to see this tendency, 9 and 16 groups are used alternatively. If a mutual information statistic underestimates correlation, two strategies may work: **i)** increasing the magnitude of α, or **ii)** increasing the number of categories of each variable in discretizing bivariate normal.

6. Discussion

Zografos (Zografos, 1993) gives asymptotic distributions of φ divergences whose special cases are Kullback -Leibler divergence, Renyi order-α divergence, etc. He has shown that the asymptotic distribution of this statistic is either normal, or a linear form in chi-square variables. Agresti (Agresti, 2002) mentions asymptotic normality property of functions of counts of a multinomial distribution. But, selecting α seems to be decisive in observing normality of mutual information statistics. Smaller α selections like 0.25 or 0.5 probably prevent normality. But these α values yielded the measures with least variations, irrespective of the scales of measurement of the variables. Behind this, some mutual information statistics, having α values near 2, showed the maximum average correlation with other independence statistics (They inform well about the dependencies of variables!). Finally, if the variables are continuous, or bivariate normal, Rényi and Tsallis mutual informations succeeded in estimating correlations with proper selections of α. Smaller α will probably yield smaller correlation estimates, whereas higher α will yield higher correlation estimates.

References

A.I. KHINCHIN, Mathematical Foundations of Information Theory, Dover Publications, Inc., New York, USA, 1957.

Abraham, WALD, Sequential Analysis, Dover Phoenix Editions, Dover Publications, Inc., Mineola, New York, USA, 2004.

Alaattin ŞAMİLOV, Entropi, İnformasyon ve Entropi Optimizasyon, Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic. Ltd. Şti., Ankara, Turkey, 2015.

Alan AGRESTI, Categorical Data Analysis, Second Edition, Wiley-Interscience, New Jersey, USA,2002.

Albert M. LIEBETRAU, Measures of Association, Series: Quantitative Applications in the Social Sciences, A Sage University Paper, 32, Sage Publications, The International Professional Publishers, USA, 1983.

Alfréd RENYI, Probability Theory, Dover Publications, Inc., Mineola, New York, USA, 2007.

Aman ULLAH, Entropy, Divergence and Distance Measures with Econometric Applications, Journal of Statistical Planning and Inference, 49(1996), 137-162.

Claude E. SHANNON, Warren WEAVER, The Mathematical Theory of Communication, University of Illinois Press, Urbana and Chicago, USA, 1963.

D.V. GOKHALE, Solomon KULLBACK, The Information in Contingency Tables, Marcel Dekker, Inc., New York, USA, 1978.

Graham UPTON, Ian COOK, A Dictionary of Statistics, Second edition, Oxford University Press, New York, USA, 2006.

José C PRINCIPE, Information Theoretic Learning, Rényi's Entropy and Kernel Perspectives, Springer, USA, 2010.

Joseph Victor MIHALOWICZ, Jonathan.V.,NICHOLS, Frank BUCHOLTZ, Handbook of Differential Entropy, CRC Press, Taylor&Francis Group, USA, 2014.

K ZOGRAFOS, Asymptotic Properties of Divergence Statistic and Its Applications in Contingency Tables, Int. J. Math. & Stat. Sci.,Vol. 2, No. 1, June 1993, 5-22.

Leandro PARDO, Statistical Inference Based on Divergence Measures, Chapman&Hall/CRC,Taylor&Francis Group, New York, USA, 2006.

M.D. ESTEBAN, D. MORALES, A summary on entropy statistics, Kybernetika, Vol.1(1995), No.4, 337-346.

Robert B. ASH, Information Theory, Dover Publications, Inc., New York, USA, 1990.

Sakoda, J. (1976). Measures of association for multivariate contingency tables. In Social Statistics Section, Proceedings of the American Statistical Association (pp. 777-780)

Solomon KULLBACK, Information Theory and Statistics, Dover Publications, Inc., Mineola, New York, USA, 1997.

Thomas M. COVER, Joy A. THOMAS, Elements of Information Theory, Second Edition, Wiley-Interscience, USA, 2006.

Vijay, P. SINGH, Entropy-Based Parameter Estimation in Hydrology, Kluwer Academic Publishers, Dordrecht, Netherlands, 1998.

Xing ZHANG, Asymptotic Normality of Entropy Estimators, The University of North Carolina at Charlotte, http://math.uncc.edu/sites/math.uncc.edu/files/2013_03.pdf