Chilean Journal of Statistics
Vol. 9, No. 2, September 2018, 55-64

Estimation of the probability function under special moments conditions using the maximum Shannon and Tsallis entropies

Zohre Nikooravesh*

Department of Basic Sciences, Birjand University of Technology, Birjand, Iran

(Received: July, 2018 · Accepted in final form: November, 2018)

Abstract

In this paper, the problem of the maximum of Shannon and Tsallis entropies is investigated by considering the inverse of the distribution function that is called quantile function. The proposed method is a further generalization of a direct method for quantile estimation, which used the integral-order probability weighted moments of in place of the product moments commonly used in the maximum entropy principle. This paper presents an advanced method combining simulation and optimization to determine the fractional probability weighted moments and the Lagrange multipliers associated with the quantile function. The numerical example presented, illustrates that the accuracy of the proposed fractional probability weighted moments based on the quantile function, is very high.

Keywords: Maximum entropy principle · Shannon entropy · Tsallis entropy · Weighted moments.

Mathematics Subject Classification: 94A17.

1. Introduction

The Principle of Maximum Entropy is based on the premise that when estimating the probability distribution, one should select a distribution which leaves the largest remaining uncertainty (i.e., the maximum entropy) consistent with the constraints. This way avoids the introduction of any additional assumptions or biases into the calculations. For the first time, Jaynes (1957) obtained a probabilistic distribution with a minimal elliptic under certain constraints using Shannon's maximum entropy. Many people, such as Zografos (2008), have studied the features, applications and generalization of the maximum entropy principle. This method has been used in many different branches of sciences that a comprehensive review of it can be found in Cover and Thomas (2006) and Pardo (2006).

Suppose that X is a random variable with an unknown distribution function $F(x)$. If $F(x)$ satisfy some special conditions, then we can find the form of a distribution function, from sampling characteristics. But these estimations are not possible or, if possible, are not reliable.

Tsallis (1988) proposed the generalization of the entropy by postulating a non-extensive entropy, (i.e., Tsallis entropy), which covers Shannon entropy in particular cases. This

*Corresponding author. Email: nikooravesh@birjandut.ac.ir

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

measure is non-logarithmic. Vila et al. (2011) investigated the application of three different Tsallis-based generalizations of mutual information to analyze the similarity between scanned documents. Another paper by Castelló et al. (2011) presented a study and a comparison of the use of different information-theoretic measures for polygonal mesh simplification by applying generalized measures from Information Theory such as Havrda-Charvát-Tsallis entropy and mutual information.

Tsallis entropy plays an essential role in non-extensive statistics, which is often called Tsallis statistics, so that many important results have been published from the various points of view by Tsallis (2009). As a matter of course, the Tsallis entropy and its related topics are mainly studied in the field of statistical physics. However, the concept of entropy is important not only in thermo-dynamical physics and statistical physics but also in information theory and analytical mathematics such as operator theory and probability theory.

In this paper, we deal with the estimation of the inverse distribution function, i.e. $x(F)$, which is called quantile function (QF). If there is a specific type of moments that we will define later, then we can maximize the entropy of the unknown function $x(F)$ under moments conditions and obtain the shape, and the parameters of $x(F)$. The proposed method is a further generalization of a direct method for quantile estimation, which used the probability weighted moments (PWMs) of integral orders in place of product moments commonly used with maximum entropy principle. We present an advanced method combining simulation and optimization to determine the fractional probability weighted moments and the Lagrange multipliers associated with the quantile function. The rest of this paper is dedicated to obtain estimations of the quantile function via the maximum entropy method by using weighted moments.

This paper is organized as follows: Section 2 defines probability weighted moments and introduces one approximation. Section 3 and Section 4 study the Shannon maximum entropy and Tsallis maximum entropy, respectively. Finaly Section 5 studies the accuracy of the obtained approximation for QF, with an example based on generalized Pareto Distribution (GPD).

2. Probability weighted moments

For a random variable X, the probability weighted moments (PWM), is defined as,

$$M_{r,s,t} = E[X^r F^s (1-F)^t] = \int_0^1 [x(F)]^r F^s (1-F)^t \mathrm{d}F. \tag{1}$$

Usually r, s, t are integers and $M_{r,s,t}$ is called integral-order probability weighted moment (IPWM). If at least one of these three values is a real (non-integer) and positive, $M_{r,s,t}$ is called the fractional probability weighted moment (FPWM). Two special cases below are simple and often used.

$$\begin{aligned} \alpha_t &= M_{1,0,t} = \int_0^1 x(F)(1-F)^t \mathrm{d}F, \\ \beta_s &= M_{1,s,0} = \int_0^1 x(F) F^s \mathrm{d}F. \end{aligned} \tag{2}$$

For an ordered sample, $x_1 \leq x_2 \leq \cdots \leq x_n$ (n is the number of sample data), α_t and β_s can be approximated from the following formula (Deng and Pandey, 2008).

$$a_t = \frac{1}{n} \sum_{i=1}^{n} \frac{\binom{n-i}{t} x_i}{\binom{n-1}{t}},$$
$$b_s = \frac{1}{n} \sum_{i=1}^{n} \frac{\binom{i-1}{s} x_i}{\binom{n-1}{s}}, \tag{3}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, for $0 \leq k \leq n$.

3. Shannon maximum entropy

We remind that for a random variable X with density f, Shannon entropy is defined by Shannon (1948) as follows,

$$H[f(x)] = -\int_{-\infty}^{+\infty} f(x) \ln(f(x)) \mathrm{d}x. \tag{4}$$

Let $x(F)$ represent the true but unknown quantile distribution of a random variable X, and $x_M(F)$ be the Mth order estimated model. The entropy of a QF is defined as,

$$H[x(F)] = -\int_{0}^{1} x(F) \ln(x(F)) \mathrm{d}F, \tag{5}$$

and the available information is presented in terms of sample FPWMs,

$$\int_{0}^{1} x_M(F) F^{\delta_i} \mathrm{d}F = b_i, \qquad i = 0, 1, \ldots, M; \delta_0 = 0. \tag{6}$$

The sample FPWMs were considered to be accurate up to Mth order, which means,

$$\int_{0}^{1} x_M(F) F^{\delta_i} \mathrm{d}F = \int_{0}^{1} x(F) F^{\delta_i} \mathrm{d}F = b_i, \qquad i = 0, 1, \ldots, M; \delta_0 = 0, \tag{7}$$

where δ_i for $i = 1, 2, \ldots, M$ are real positive or fractional numbers, b_i is a sample estimate of the population FPWM, for $i = 0, 1, \ldots, M$, and M is the highest order of FPWM considered in the analysis. Note that $x(F)$ in Equation (5) is not normalized, rather normalizing condition is included as an external constraint in Equation (6) corresponding to $\delta_0 = 0$.

In this way, the problem changes to the following optimization problem,

$$\begin{aligned}
&\max \int_0^1 x_M(F)\ln(x_M(F))\mathrm{d}F \\
&\text{s. t.} \\
&\qquad \int_0^1 x_M(F)\mathrm{d}F = b_0 \\
&\qquad \int_0^1 x_M(F)F^{\delta_i}\mathrm{d}F = b_i, \quad i = 1, 2, \ldots, M.
\end{aligned} \tag{8}$$

The optimization problem (8) satisfies the Lagrange multiplier conditions and it can be written as,

$$\begin{aligned}
\max &- \int_0^1 x_M(F)\ln(x_M(F))\mathrm{d}F - (\lambda_0 - 1)\left(\int_0^1 x_M(F)\mathrm{d}F - b_0\right) \\
&- \sum_{i=1}^{M} \lambda_i \left(\int_0^1 x_M(F)F^{\delta_i}\mathrm{d}F - b_i\right).
\end{aligned} \tag{9}$$

By differentiating Equation (9) with respect to $x_M(F)$ and by setting it equal zero, we have,

$$\begin{aligned}
&\frac{\partial}{\partial x_M(F)}\Bigg[- \int_0^1 x_M(F)\ln(x_M(F))\mathrm{d}F - (\lambda_0 - 1)\left(\int_0^1 x_M(F)\mathrm{d}F - b_0\right) \\
&\qquad - \cdots - \lambda_M \left(\int_0^1 x_M(F)F^{\delta_M}\mathrm{d}F - b_M\right)\Bigg] \\
&= -\ln(x_M(F)) + 1 - (\lambda_0 - 1) - \cdots - \lambda_M F^{\delta_M} \\
&= -\ln(x_M(F)) - \sum_{i=0}^{M} \lambda_i F^{\delta_i} \\
&= 0.
\end{aligned} \tag{10}$$

The following statement can be concluded,

$$\ln(x_M(F)) = -\sum_{i=0}^{M} \lambda_i F^{\delta_i} \quad \Rightarrow \quad x_M(F) = \exp\Big(-\sum_{i=0}^{M} \lambda_i F^{\delta_i}\Big). \tag{11}$$

Now by differentiating Equation (9) with respect to the other variables, we have,

$$\begin{aligned}
&\frac{\partial}{\partial \lambda_i}\Big[-\int_0^1 x_M(F)\ln(x_M(F))\mathrm{d}F-(\lambda_0-1)\Big(\int_0^1 x_M(F)\mathrm{d}F-b_0\Big)\\
&\qquad -\cdots-\lambda_M\Big(\int_0^1 x_M(F)F^{\delta_M}\mathrm{d}F-b_M\Big)\Big]\\
&=-\int_0^1 x_M(F)F^{\delta_i}\mathrm{d}F+b_i\\
&=0, \qquad (12)
\end{aligned}$$

for $i=0,1,\ldots,M$ one can be conclude that,

$$\int_0^1 x_M(F)F^{\delta_i}\mathrm{d}F=b_i, \quad i=0,1,\ldots,M. \tag{13}$$

By substituting, $x_M(F)$ we obtain,

$$\int_0^1 \exp\Big(-\sum_{i=0}^{M}\lambda_i F^{\delta_j}\Big)F^{\delta_i}\mathrm{d}F=b_i, \quad j=0,1,\ldots,M. \tag{14}$$

To find the unknown values λ_i, we must solve the non-linear system with $M+1$ unknown values and $M+1$ equations in the form of Equation (14). But the uncertainty of δ_i makes the problem more complicated.

To solve this problem, one can use a Monte Carlo simulation based optimization procedure. To begin with, fractionals δ_i, $i=1,\ldots,M$ are simulated uniformly from a selected interval $(0,h)$. Because, when we do not have any conditions, uniform distribution has the maximum entropy. Then, the Lagrangian multipliers λ_i, $i=0,\ldots,M$ are obtained by solving a set of nonlinear equations using the standard Newton-Raphson's method.

δ_i and λ_i obtained by this method satisfy all the constraints of the original problem. First by repeat the above process, we can obtain multiple δ_i and λ_i, then by selecting from among all obtained values of δ_i and λ_i we can have much more entropy than have other values. After that, we rewrite $x_M(F)=\exp(-\sum_{i=0}^{M}\lambda_i F^{\delta_i})$.

A measure of discrepancy between $x_M(F)$, and the exact QF, $x(F)$, can be given by the Kullback Leibler measure as (Deng and Pandey, 2008)

$$\mathrm{KL}(x,x_M)=\int_0^1 x(F)\ln\Big(\frac{x(F)}{x_M(F)}\Big)\mathrm{d}F=-H(x)-\int_0^1 x(F)\ln(x_M(F))\mathrm{d}F. \tag{15}$$

This measure signifies the difference between a true and an estimated probability distribution.

4. Tsallis Maximum Entropy

In this section, we describe the process of obtaining the quantile function by maximizing Tsallis entropy. To do this, first we define the Tsallis entropy, especially the maximum entropy problem, and then obtain an answer using mathematical software for an example.

Tsallis entropy for the random variable X is defined as follows,

$$H[f(x)] = -\int_{-\infty}^{+\infty} f^q(x)\ln_q(f(x))\mathrm{d}x, \tag{16}$$

where $\ln_q(x) = (x^{1-q}-1)/(1-q)$. Similar Equation (9), one can put,

$$\begin{aligned}\max -\int_0^1 x_M^q(F)\ln_q(x_M(F))\mathrm{d}F - (\lambda_0-1)\left(\int_0^1 x_M(F)\mathrm{d}F - b_0\right)\\ -\sum_{i=1}^{M}\lambda_i\left(\int_0^1 x_M(F)F^{\delta_i}\mathrm{d}F - b_i\right).\end{aligned} \tag{17}$$

Now we derive from Equation (17) with respect to $x_M(F)$ then we set it equal to zero. For this purpose we have,

$$\begin{aligned}&\frac{\partial}{\partial x_M(F)}\left[-\int_0^1 x_M^q(F)\ln_q(x_M(F))\mathrm{d}F - (\lambda_0-1)\left(\int_0^1 x_M(F)\mathrm{d}F - b_0\right)\right.\\ &\qquad\left. - \cdots - \lambda_M\left(\int_0^1 x_M(F)F^{\delta_M}\mathrm{d}F - b_M\right)\right]\\ &= -qx_M^{q-1}(F)\frac{x_M^{1-q}(F)-1}{1-q} + x_M^q(F)\frac{(1-q)x_M^{-q}}{1-q} - (\lambda_0-1) - \cdots - \lambda_M F^{\delta_M}\\ &= -\left[\frac{q}{1-q}x_M^{q-1}(F)+1\right] - (\lambda_0-1) - \lambda_1F^{\delta_1} - \cdots - \lambda_M F^{\delta_M}\\ &= -\left[\frac{q}{1-q}x_M^{q-1}(F)\right] - \sum_{i=0}^{M}\lambda_i F^{\delta_i} = 0.\end{aligned} \tag{18}$$

The following statement can be concluded,

$$x_M(F) = \left(1 + \frac{1-q}{q}\sum_{i=0}^{M}\lambda_i F^{\delta_i}\right)^{\frac{1}{q-1}}. \tag{19}$$

In the Figure 1, the estimate of the quantile function of the Equation (19) for $M = 2$ with parameters $b_1 = 0.50$, $b_2 = 0.75$ and $\delta_1 = 0.25$, $\delta_2 = 1.00$ has been shown. In this case λ_is are obtained for $M = 2$ in Table 1.

Table 1. Calculate of λ_i for $i = 0, 1, 2$

i	0	1	2
λ_i	0.6887	-1.6035	1.0613

5. Numerical examples on based Pareto distribution

This section evaluates the accuracy of the proposed estimate of quantile function by considering an example involving generalized Pareto distribution (GPD). This example is

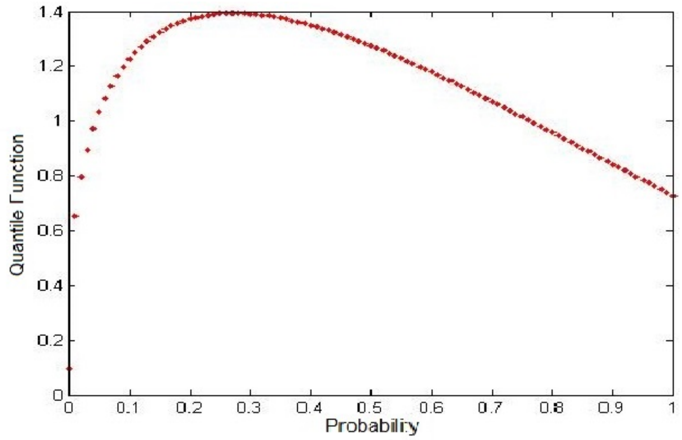

Figure 1. The approximation of QF for $M = 2$ with parameters $b_1 = 0.50$, $b_2 = 0.75$, $\delta_1 = 0.25$ and $\delta_2 = 1.00$.

intended to show that how well an estimated QF, $x_M(F)$, can approximate its exact counterparts $x(F)$. The cumulative distribution function of generalized Pareto Distribution is defined as,

$$F(x) = \begin{cases} 1 - [1 - \frac{cx}{d}]^{\frac{1}{c}}, & c \neq 0 \\ 1 - \exp(-x/d), & c = 0. \end{cases} \tag{20}$$

So we have the inverse of this function as,

$$x(F) = \begin{cases} \frac{d}{c}[1 - (1 - F)^c], & c \neq 0 \\ -d[\ln(1 - F)], & c = 0, \end{cases} \tag{21}$$

where c and $d \neq 0$ are constants. The range of x is $0 \leq x < \infty$ for $c \leq 0$ and $0 \leq x \leq c/d$ for $c > 0$. In Figure 2 and Figure 3 the diagrams of probability function F and quantile functions for $d = 1$, $c = -0.2$ are drawn respectively.

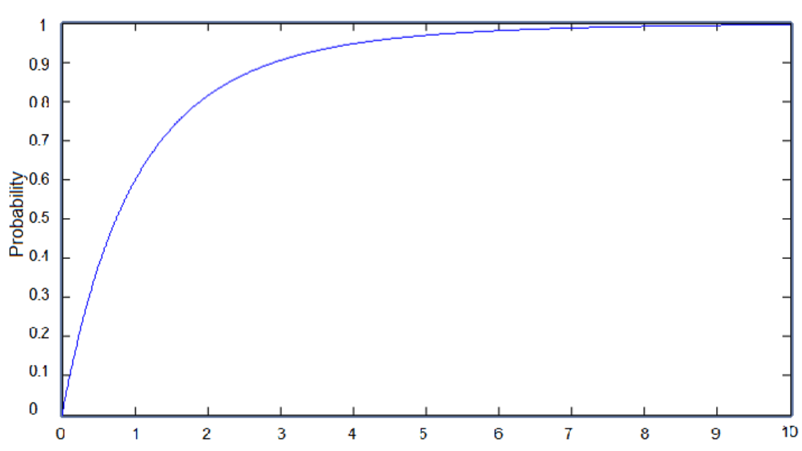

Figure 2. The probability function of GPD with $d = 1$, $c = -0.2$

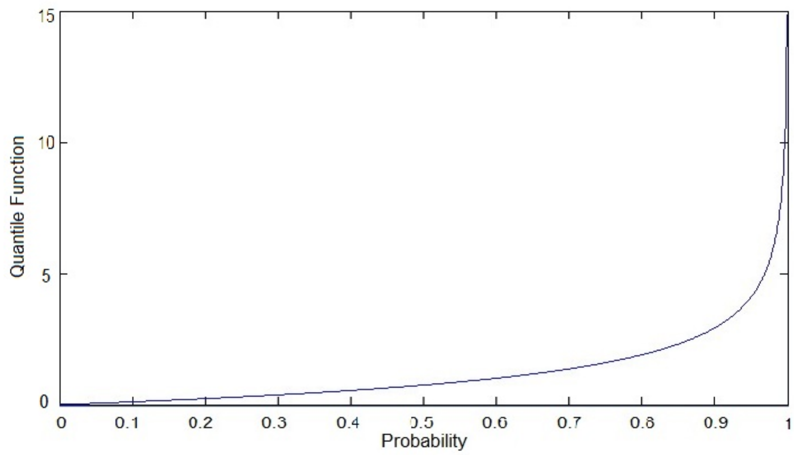

Figure 3. The quantile function of GPD with $d = 1$, $c = -0.2$

The $\alpha_t = M_{1,0,t}$ for the function GPD for $c \neq 0$ is obtained as follows,

$$\begin{aligned}\alpha_t = M_{1,0,t} &= \int_0^1 x(F)(1-F)^t \mathrm{d}F \\ &= \int_0^1 \frac{d}{c}[1-(1-F)^c](1-F)^t \mathrm{d}F \\ &= \frac{d}{c}\int_0^1 [1-G^c]G^t \mathrm{d}G \\ &= \frac{d}{(t+1)(c+t+1)}, \end{aligned} \tag{22}$$

and for $c = 0$, we get,

$$\alpha_t = M_{1,0,t} = \int_0^1 -d[\ln(1-F)](1-F)^t \mathrm{d}F \tag{23}$$

$$= -d\int_0^1 [\ln(G)]G^t \mathrm{d}G \tag{24}$$

$$= \frac{d}{(t+1)^2}, \tag{25}$$

where the variable $G = 1 - F$ is used.

For the amount of t, obtained results are given in Table 2.

Table 2. The calculation of α_t for $t = 1, 2, \cdots, 10$

t	1	2	3	4	5
α_t	0.2778	0.1190	0.0658	0.0417	0.0287
t	6	7	8	9	10
α_t	0.0210	0.0160	0.0126	0.0102	0.0084

Although for $\alpha_t = M_{1,0,t}$ could obtain explicit formulas In general, $M_{r,s,t}$ can't be written explicitly, because it leads to the solution of a complicated integral. But it should solve the corresponding integral with numerical methods. As an example in Table 3, the results of these calculations are given for the quantile function GPD with $c = -0.2$ and $d = 1$.

Table 3. The calculation of $M_{r,s,t}$

r	s	t	$M_{r,s,t}$
2	5	6	0.0001
1.20	1.75	2.50	0.0201

In Table 3, values r, s and t are randomly selected, and Simpson method have been used for solving integral with MATLAB software. For the quantile function, the optimal values of λ_i and δ_i have been obtained for $m = 1$ in Table 4.

Table 4. The calculation of λ_i and δ_i for $M = 1$

δ_0	δ_1	λ_0	λ_1
0	0.0001	-1894.4940	1893.5210

Figure 4 shows that the approximation accuracy of these parameters are remarkable. So the $F(x)$ with more approximation accuracy is as follows,

$$x_M(F) = -\ln(1-F)e^{-(1+\lambda_1)-\lambda_2(1-F)^{\delta_1}-\cdots-\lambda_M(1-F)^{\delta_M}}. \tag{26}$$

So,

$$x_1(F) = -\ln(1-F)e^{-(1-1894.4940)-1893.5210(1-F)^{0.0001}}. \tag{27}$$

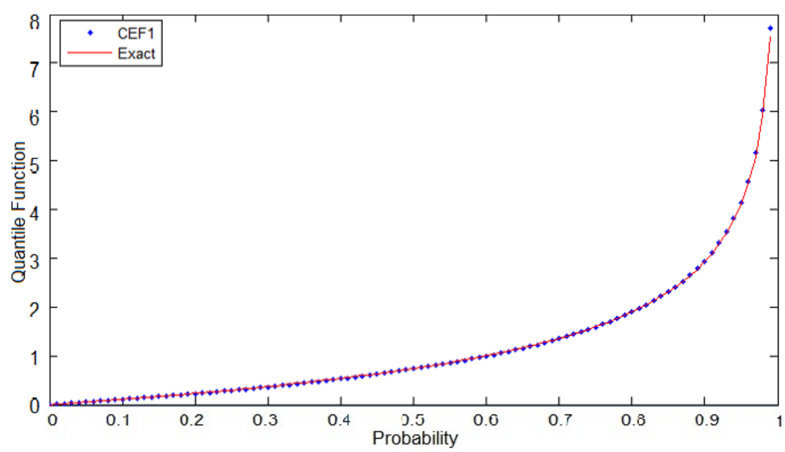

Figure 4. The approximation of QF with parameters λ_i and δ_i

Using equation (15) to calculate the error in Figure 4, we have

$$\mathrm{KL}(x, x_1) = \int_0^1 x(F)\ln\frac{x(F)}{x_1(F)}\mathrm{d}F \tag{28}$$

$$= -H(x) - \int_0^1 x(F)\ln(x_1(F))\mathrm{d}F \tag{29}$$

$$= 0.0017, \tag{30}$$

where this shows the high accuracy of the mentioned method.

For Shannon entropy, Deng and Pandey (2008) showed the proposed method exhibits significantly higher estimation accuracy. The statistical error, bias and RMSE, associated with FPWM method is smaller than that obtained from the use of IPWM method. From a practical example, conclusion can be made that FPWM based maximum entropy QF combines advantage of empirical QF (normal distribution) and IPWM based QF.

Conclusions

The paper presents a method to estimate the maximum of a quantile function from a small sample of data using Shannon's and Tsalli's maximum entropy methods. Our proposal is a further generalization of a direct method for quantile estimation, which used the probability weighted moments of integral orders in place of product moments commonly used with Shannon and Tsallis maximum entropy principles. A general estimation method was proposed in which Monte Carlo simulations and optimization algorithms were combined to estimate the fractional probability weighted moments and the Lagrange multipliers that would lead to the best-fit quantile function. We explored the relatively high accuracy of this method providing an example. A furue line of research is the study of the accuracy of this method for Tsallis entropy, analytically.

References

Castelló, P., González, C., Chover, M., Sbert M., and Feixas, M. (2011). Tsallis entropy for geometry simplification. Entropy 13, 1805-1828.

Cover, T.M., and Thomas, J.A. (2006). Elements of Information Theory. Wiley, New York.

Deng, J., and Pandey, M.D. (2008). Estimation of the maximum entropy quantile function using fractional probability weighted moments. Structural Safety 30, 307-319.

Jaynes, E.T. (1957). Information theory and statistical mechanics. Physical Review 106, 620-630.

Pardo, L. (2006). Statistical Inference Based on Divergence Measures. Chapman & Hall, London.

Shannon, C. (1948). A mathematical theory of communication. Bell Systems Technical Journal 27, 379-423.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics 52, 479-487.

Tsallis, C. (2009). Introduction to Nonextensive Statistical Mechanics. Springer, New York.

Vila, M., Bardera A., and Feixas, M. (2011). Tsallis mutual information for document classification. Entropy 13, 1694-1707.

Zografos, K. (2008). On some entropy and divergence type measures of variability and dependence for mixed continuous and discrete variables. Journal of Statistical Planning and Inference 138, 3899-3914.