The Generalized Maximum α Entropy Principle

Sanei Tabass, M[1]* and Mohtashami Borzadaran, G. R[1]

[1] Department of Statistics, Ferdowsi University Of Mashhad, Mashhad, Iran.

Abstract. Generalizations of Maximum entropy principle(MEP) and minimum discrimination information principle (MDIP)are described by Kapur and Kesavan(1989). In this paper we used generalized entropies and replaced Shannon entropy with Tsallis entropy when $\alpha = 2$. The generalization has been achieved by the entropy maximization postulate and examining its consequences. The inverse principles which are inherent in the maximum α entropy and minimum discrimination α entropy are made in the new methodology. An example is given to illustrate the power and scope of the generalized maximum α entropy that follows from the entropy maximization postulate.

Keywords.Shannon entropy, Tsallis entropy, Tsallis divergence, Generalized Maximum Entropy, Minimum discrimination

1 Introduction

The generalized maximum α entropy principle which is the subject matter of this paper, is a generalization of the MαEP (maximum α entropy principle)in the sense that the latter forms an important constituent of it. Furthermore it is the MαEP that provides the requisite background for the formulation of this new principle. Given the three probabilistic entities, namely, the α entropy measure, the set of moment constraints and the probability distribution, the MαEP provides a methodology for identifying the most unbiased probability distribution, based on a knowledge of the first two entities. As stated earlier, the identification is based on the principle of maximization of the α entropy measure subject to the given constraints.
The GMαEP (generalized maximum α entropy principle) addresses itself to the determination of any one the three when the remaining two probabilistic entities are specified. The philosophical underpinning of the GMαEP rests on the α entropy maximization postulate, which states it is the maximum information theoretic entropy that is always the controlling quantity with respect to the states of the three mutually coupled probabilistic entities. The GMαEP then spells out deductive procedures for the determination of the unspecified entity when the other two are specified. The principle MEP (Maximum entropy principle) implies the determination of the most unbiased probability distribution proceeding from Shannon entropy and a given set of constraints and MDI principle implies the most unbiased probability distribution proceeding Shannon entropy and given set of constraints and furthermore a prior probability distribution Q. In this paper, we generalize this formalisms (MEP and MDI), with replacing Shannon entropy by Tsallis entropy when $\alpha = 2$ and we obtain most unbiased probability distribution proceeding Tsallis entropy and given set of constraints. A formalism in generalized maximum entropy principle is presented. The principles of Maximum

α entropy and Minimum discrimination information is reviewed and an illustrative example is given.

2 The generalized maximum α entropy principle

Generalized maximum entropy principle possible to determine any one of the probabilistic quantities when the other two are specified.

2.1 The maximum α entropy version

In continuous, we review the principle of maximum α entropy and give illustrative example. For a probability distribution, $P = (p_1, ..., p_n)$ The Shannon and Tsallis entropies are respectively :

$$H(P) = -\sum_i p_i log p_i \tag{1}$$

and Tsallis entropy is:

$$S_\alpha(P) = \frac{1}{\alpha - 1}[\sum_i P_i^\alpha - 1], \alpha > 0, \alpha \neq 1. \tag{2}$$

Let $\phi(.)$ be a convex function,

$$H(P) = -\sum_{i=1}^{n} \phi(p_i) \tag{3}$$

be the measure of entropy and the constraints be

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g_r(x_i) = a_r, r = 1, 2, ..., m. \tag{4}$$

Using the method of Lagrange multipliers, we maximize (3) subject to the $(m + 1)$ constraints in (4) and get an expression for the first derivative of $\phi(P)$ as,

$$-\sum \phi(p_i) + \lambda_0(\sum p_i - 1) + \sum \lambda_r(\sum p_i g_r(x_i) - a_r) \Rightarrow$$
$$\phi'(p_i) = \lambda_0 + \sum_{r=1}^{m} \lambda_r g_r(x) \tag{5}$$

A) The direct principle
Given the entropy measure $\phi(.)$ and the constraint mean values of $g_1(x), ..., g_m(x)$, we wish to determine the probability distribution that maximizes the entropy measure. Using (5) to substitute in to (4), we can solve for the $(m + 1)$ Lagrange multipliers which in turn yield the probabilities p_i.

B) The first inverse problem (determination of constraints)

Given the probability distribution for p_i and the entropy measure $\phi(.)$, determine one or more probability constraints that yield the given probability distribution when the entropy measure is maximized subject to these constraints. Since we know $\phi(p_i)$, we also know $\phi'(p_i)$, and hence the right hands of (5) can be determined. This will allow us to identify the values for $g_1(x_i), ..., g_m(x_i)$ by matching terms and thus a most unbiased set of constraints (4).

C) The second inverse problem (determination of the entropy measure)
In this case, given the constraints $g_1(x_i), g_2(x_i), ..., g_m(x_i)$ and the probability distribution p_i, determine the most unbiased entropy measure that when maximized subject to the given constraints. Yields the given values in to (5) and get a different equation that can be solved for $\phi(.)$. Once $\phi(.)$ is known, we can determine the entropy measure $H(P) = -\sum_{i=1}^{n} \phi(p_i)$. We now illustrate the MαEP version of generalized maximum α entropy principle on the basis of an example.

Example1: Let

$$\phi(p_i) = p_i(p_{i-1}) \tag{6}$$

and $H(P) = -\sum_i \phi(p_i)$. We substitute $\phi(p_i)$, gives $H(P) = -\sum_i \phi(p_i) = -\sum_i p_i(p_{i-1})$, in this case $H(P)$ is Tsallis entropy when $\alpha = 2$.
Also the constraints be

$$\sum_{i=1}^{n} p_i = 1,\ \hat{x} = \sum_{i=1}^{n} p_i x_i, \tag{7}$$

and p_i that obtain from this equation is taken as the probability distribution

$$\begin{aligned}
\phi'(p_i) &= \lambda_0 + \lambda_1 x_i \Rightarrow 2p_i - 1 = \lambda_0 + \lambda_1 x_i && (8)\\
\Rightarrow p_i &= \frac{\lambda_0 + \lambda_1 x_i + 1}{2}\\
\Rightarrow 1 &= \frac{\lambda_0 + 1}{2} n + \frac{\lambda_1}{2} \sum x_i \Rightarrow \frac{\lambda_0 + 1}{2} n = 1 - \frac{\lambda_1}{2} \sum x_i\\
\Rightarrow p_i &= \frac{\lambda_1}{2} x_i + \frac{1}{n} - \frac{\lambda_1}{2n} \sum x_i\\
\Rightarrow p_i &= \frac{\mu}{2} x_i + \frac{1}{n} - \frac{\mu}{2n} \sum x_i. && (9)
\end{aligned}$$

The Lagrange multiplier μ can found from

$$\hat{x} = \frac{\sum x_i}{n} + \frac{\mu}{2} \sum x_i^2 - \frac{\mu}{2n} (\sum x_i)^2 \tag{10}$$

On the basis of the formalism presented earlier, we wish to demonstrate solutions to the one direct and two inverse problems for this specific example.
D)Determination of constraints (The first inverse problem)
The constraints are determined from a knowledge of equations (6) and (9), Substituting in to equation (5), we get

$$\begin{aligned}\lambda_0 + \lambda_1 g_1(x) + ... + \lambda_m g_m(x) &= 2p_i - 1 \\ &= 2(\frac{\mu}{2}x_i - \frac{\mu}{2n}\sum x_i + \frac{1}{n}) - 1 \\ &= \mu x_i - \mu\bar{x} + \frac{2}{n} - 1 \Rightarrow\end{aligned} \tag{11}$$

$$\begin{cases}\lambda_0 = -\frac{\mu}{n}\bar{x} + \frac{2}{n} - 1, \\ \lambda_1 = \mu, g_1(x) = x_i\end{cases} \tag{12}$$

Hence, the constraints are:

$$\sum_i^n p_i = 1 \; and \; \sum_i^n p_i x_i = \hat{x} \tag{13}$$

E) Determination of the entropy measure (second inverse problem)
Consider the following differential equation:

$$\begin{aligned}\phi'(p_i) &= \lambda_0 + \lambda_1 x_i \; and \; from \; (8) \; = a + bp_i \\ \Rightarrow \phi(p_i) &= ap_i + \frac{b}{2}p_i^2 + c \\ H(P) &= -\sum_{i=1}^n \phi(p_i) = -a - \frac{b}{2}\sum_{i=1}^n p_i^2 - cn\end{aligned}$$

$$\Rightarrow \begin{cases} b = 2, \\ -a - cn = 1\end{cases} \Rightarrow H(P) = -\sum_{i=1}^n p_i^2 + 1 = 1 - \sum_{i=1}^n p_i^2 = -\sum_{i=1}^n p_i(p_i - 1) \tag{14}$$

We get the entropy $H(P) = -\sum_{i=1}^n p_i(p_i - 1)$ which is Tsallis entropy when $\alpha = 2$.

2.2 Generalized maximum α entropy, which takes in to account the generalized averaging procedures

Generalized entropies is based on the joint generalization of the concept of information gain and the averaging procedure. Tsallis (1988) also proposed the generalization of the entropy by postulating a non-extensive entropy, (i.e.Tsallis entropy), which covers Shannon entropy in particular case. This measure is non-logarithmic. Tsallis entropy (which is non additive), is one of the generalized entropy measures, and is based on the generalization of the information gain only, since it preserves the linear averaging procedure of the Shannon entropy.

This form of entropy is obtained through the joint generalization of the averaging procedures and the concept of information gain. Tsallis entropy is not extensive but generalizes the concept of the information gain and is obtained by the linear averaging procedure. As seen before Tsallis entropy preserves the same averaging procedure as Shannon entropy, i.e, linear averaging, but generalizes the concept of information by deforming the logarithmic function. Now, Tsallis entropy rewrites as follows:

$$S_\alpha(f(x)) =< ln_\alpha(\frac{1}{p}) >_{lin}$$

where $ln_\alpha(x)$ is α-logarithm given by

$$ln_\alpha(x) = \frac{x^{1-\alpha} - 1}{1-\alpha}, \tag{15}$$

It is evident that, the functional to be maximized must be of the form

$$\Omega_\alpha =< ln_\alpha(\frac{1}{p}) >_{lin} -\lambda_0 < . >_{lin} -\lambda_1 < .. >_{lin} \tag{16}$$

where λ_0, λ_1 are as usual Lagrange multipliers.
In order to determine the form of the function $f(x)$ to be used in the linear averaging procedure, Tsallis et.al. (2009) have found

$$f_{\alpha n}(x) = \frac{f^{1+n(\alpha-1)}(x)}{\int f^{1+n(\alpha-1)}(x)dx} \qquad n = 0, 1, 2, ... \tag{17}$$

A function $f(x)$ can alternatively be determined by its α-moments (as a result of rewriting the definition of the information gain through α- logarithm). They entitled relation (17), generalized escort distribution. The probability density that must be used in equation (16), is the generalized escort distribution is given by equation (17). If we want to obtain the linear average of a constant, we substitute $n = 0$ in equation (17). For any first moment, we substitute $n = 1$ and so on. There for, we write, for the linear average of 1 as

$$< 1 >_{lin}= \int dx f_{\alpha 0}(x) = \frac{\int f(x)dx}{\int f(x)dx} = 1 \tag{18}$$

Next, we consider the linear average of the x

$$< x >_{lin}= \int x f_{\alpha 1}(x)dx = \frac{x f^\alpha(x)dx}{\int f^\alpha(x)dx} \tag{19}$$

The maximization of the functional in equation (16) subject to constraints equations (18) and (19), yields the well-known stationary distribution

$$f(x) = \frac{[1 - \lambda_0(1-\alpha)\frac{x-u_\alpha}{\int f(x)^\alpha dx}]^{\frac{1}{1-\alpha}}}{\int [1 - \lambda_0(1-\alpha)\frac{x-u_\alpha}{\int f(x)^\alpha dx}]^{\frac{1}{1-\alpha}} dx} \tag{20}$$

3 Formalism of the Minimum discrimination information principle (MDI)

Here we briefly review the principle of Minimum discrimination information. In order to discriminate the probability distribution P from Q, the measure, Tsallis divergence is introduced as follow:

$$D_{S\alpha}(P||Q) = \frac{1}{\alpha-1}\sum_{i=1}^{n}[(\frac{f_i}{g_i})^{\alpha-1}-1]f_i, \alpha>0, \alpha\neq 1. \tag{21}$$

This measure always≥ 0, and has a global minimum value of zero when the two distributions are identical. If Q is the uniform distribution ($Q = U = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$) we have

$$\begin{aligned} D_{S\alpha}(P||Q) = D_{S\alpha}(P||U) &= \frac{1}{\alpha-1}n^{\alpha-1}\sum_{i=1}^{n}(p_i^{\alpha-1}-1)p_i \\ &= n^{\alpha-1}[S_\alpha(U) - S_\alpha(P)] \end{aligned}$$

where $S_\alpha(U)$ is the Tsallis entropy associated with the uniform distribution. Minimizing $D_{S\alpha}(P||U)$ would entail maximizing $S_\alpha(P)$.
If $\phi(.)$ be a convex function and let

$$D(P||Q) = \sum_{i=1}^{n} q_i\phi(\frac{p_i}{q_i}) \tag{22}$$

be the measure of directed divergence. The constraints be

$$\sum_{i=1}^{n} p_i = 1\ ,\ \sum_{i=1}^{n} p_i g_r(x_i) = a_r,\ r = 1, 2, ..., m. \tag{23}$$

Minimizing (22) subject to (23), we get

$$\begin{aligned} &\sum q_i\phi(\frac{p_i}{q_i}) + (\lambda_0+1)(\sum p_i - 1) + \sum \lambda_r(\sum g_r p_r - a_r) = 0 \\ &\Rightarrow \phi'(\frac{p_i}{q_i}) = -(\lambda_0+1) - \lambda_1 g_1(x_1) - ... - \lambda_m g_m(x_i) \end{aligned} \tag{24}$$

Tsallis divergence measure when $\alpha = 2$ is:

$$D_{S2}(P||Q) = \frac{1}{2-1}\sum_{i=1}^{n}[(\frac{p_i}{q_i})-1]p_i = \sum_{i=1}^{n}(\frac{p_i}{q_i}-1)p_i$$

If $\phi(p_i) = p_i(p_i-1)$ we have $\phi(\frac{p_i}{q_i}) = \frac{p_i}{q_i}(\frac{p_i}{q_i}-1)$ and so

$$D_{s2}(P||Q) = \sum(\frac{p_i}{q_i}-1)p_i$$

$$\phi'(\frac{p_i}{q_i}) = 2\frac{p_i}{q_i} - 1 = -(\lambda_0+1) - \lambda_1 g_1(x_i) - ... - \lambda_m g_m(x_i) \tag{25}$$

$$\Rightarrow p_i = \frac{-\lambda_0 - \lambda_1 g_1(x_i) - ... - \lambda_m g_m(x_i)}{2} q_i \tag{26}$$

A) The direct problem (Determination of probability distribution)
If q_i and $g_1(x_i), g_2(x_i), ..., g_m(x_i)$ are known, (25) determines the $p_1, p_2, ..., p_n$.

B) First inverse problem (Determination of the constraints)
If $p_i^{`}s$, $q_i^{`}s$ and $\phi(.)$ are known, (25) determines the constraint functions $g_1(.), g_2(.), ..., g_m(.)$.

C) Determination of the divergence measure (Second inverse problem)
If $p_i^{`}s$, $q_i^{`}s$ and $g_r(x_i)$ are known, (25) and (26) determine $\phi(.)$ and as such determines the divergence measure $D_{s\alpha}(P||Q)$.

D) Determination of a priori distributions (Third inverse problem)
Finally, if $p_i^{`}s$, $q_i^{`}s$ and $\phi(.)$ are known, (25) and (26) determine the $q_i^{`}s$. It has to be shown that if any three of the aforementioned are given, then the fourth is the most unbiased one. That is, the fourth is such that the observed probability distribution is a minimum discrimination information probability distribution(MDIPD).

4 Conclusion

Generalized maximum entropy principle and maximum entropy principle have found useful applications in a wide variety of problems. But there are other applications, as illustrated by example in this paper. we replaced Shannon entropy and Kullback divergence with Tsallis entropy and Tsallis divergence respectively. The first inverse problem is addressed to the determination of a set of unbiased constraints. The second inverse problem focuses on determining the most unbiased entropy measure when the other two probabilistic entities are given. The nonadditive Tsallis entropy preserves the linear averaging procedure in its definition. with considering this point, the constraints have shown with linear averaging procedure and Tsallis entropy maximized.

References

Bagsi, G, B. and Tirnakli, U. (2009). *On the way towards a generalized maximization procedure*. Physics Letters A, 373, 3230-3234.

Bercher, J. F. (2008). *Tsallis distribution as a standard maximum entropy solution with tail constraint*. Physics Letters A, 372, 5657-5659.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Second Edition, Wiley InterScience.

Csiszar, I. (1995). *Generalized cut-off rates and Renyi information measures*. IEEE Trans.Inf. Theory, 41, 26-34.

Harvda, J. and Charvat, F. (1967).*Quantification method of classification processes: Concept of structural $\alpha-$ entropy*. Kybernetika, 3, 30-35.

Kesavan, H. K. and Kapur, J. N. (1989). *The generalized maximum entropy principle*. IEEE Transactions on systems, Man, and Cybernetics, Vol 19, 5, 1042-1052.

Kesavan, H. K. and Kapur, J. N. (1990). *On the families of solutions to generalized maximum entropy and minimum cross-entropy problems*. International Journal of General Systems, vol 16, 3, 199-214.

Shannon, C. E.(1948). *A mathematical theory of communication* . Bell system Tech., 27, 379-423.

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

Tsallis, C. (2002). *Entropic non-extensivity: A possible measure of complexity*. Chaos, Solitons and Fractals, 13, 371-391.

Tsallis, C., Plastino, A. R. and Alvarez, R. F. (2009). *Escort mean values and the characterization of power-low-decaying probability densities* . Journal of Mathematical physics, 50, 043303.