The generalised maximum $\alpha$ entropy principle

Manije Sanei Tabass  
Department of Statistics,  
School of Mathematical Sciences,  
University of Sistan and Baluchestan,  
Zahedan, Iran  
Email: manijesanei@gmail.com

G.R. Mohtashami Borzadaran*  
Department of Statistics,  
Faculty of Mathematical Sciences,  
Ferdowsi University of Mashhad,  
Mashhad, Iran  
Email: grmohtashami@um.ac.ir  
*Corresponding author

Abstract: Generalisations of maximum entropy principle (MEP) and minimum discrimination information principle (MDIP) are described by Kapur and Kesavan (1989). In this paper, we used generalised entropies and replaced Shannon entropy with Tsallis entropy when $\alpha = 2$. The generalisation has been achieved by the entropy maximisation postulate and examining its consequences. The inverse principles which are inherent in the maximum $\alpha$ entropy and minimum discrimination $\alpha$ entropy are made in the new methodology.

Keywords: Shannon entropy; Tsallis entropy; Tsallis divergence; generalised maximum entropy principle; GMEP; maximum entropy principle; MEP; minimum discrimination.

Reference to this paper should be made as follows: Tabass, M.S. and Borzadaran, G.R.M. (2018) ‘The generalised maximum $\alpha$ entropy principle’, Int. J. Mathematics in Operational Research, Vol. 12, No. 1, pp.129–137.

Biographical notes: Manije Sanei Tabass received his BS in Statistics from Birjand University in Iran and MSc in Statistics from Ferdowsi University of Mashhad, Mashhad, Iran. She received her PhD in Statistics from Ferdowsi University of Mashhad, Mashhad, Iran. She is currently a Professor Assistant in Statistics in the Department of Statistics, School of Mathematical Sciences, University of Sistan and Baluchestan. Zahedan, Iran. Her research interests include information theory and generalised maximum entropy.

G.R. Mohtashami Borzadaran received his BS in Statistics from the Faculty of Informatics and Statistics in Iran and MSc in Statistics from Shahid Beheshti University in Iran and PhD in Statistical Inference from Sheffield University, Sheffield, England (UK) in 1997. He is currently a Professor in Statistics in the Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Iran. His research is focused on information theory and reliability.
This paper is a revised and expanded version of a paper entitled ‘The generalized maximum $\alpha$ entropy principle’ presented at 9th Seminar of Probability and Random Processes, University of Sistan and Baluchestan, Zahedan, Iran, 11–12 September 2013.

1 Introduction

Shannon entropy had been introduced in 1948 and after that, generalised entropies have been obtained. These entropies which are order of $\alpha$ cover Shannon entropy in particular cases. Tsallis proposed the generalisation of the entropy by postulating a non-extensive entropy in 1988. This measure is non-logarithmic and is obtained through the joint generalisation of the averaging procedures and the concept of information gain. Tsallis entropy has haunted physicists, chemists, mathematicians, engineers, economists and others since long! (see for example: Tirnakli and Torres, 2000; Telesca, 2011; Vallianatos et al., 2014; Kleiber, 2005, ....).

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum entropy estimate. The maximum entropy principle (MEP) is characterising some unknown events in a statistical model and we should always choose the one that has maximum entropy. The MEP and the associated generalised maximum entropy principle (GMEP) have been enunciated in Kesavan and Kapur (1990). Here, replacing Shannon entropy with Tsallis entropy of order $\alpha$ (for $\alpha = 2$) we introduce generalised maximum $\alpha$ entropy principle. Tsallis entropy has applications in a broad range of Science. Particularly for $\alpha = 2$ Tsallis entropy has applications in Finance.

We use Tsallis entropy instead of Shannon, so we use a different measure that can be used in the statistical estimation processes. On the other hand, in many of problems, Tsallis entropy (particularly for $\alpha = 2$) has more applications than a logarithmic measure (as Shannon entropy). It is easier working with a linear transformation (as Tsallis) of a logarithmic. The generalised maximum $\alpha$ entropy principle which is the subject generalisation of the matter of this paper, is a $M\alpha$EP (maximum $\alpha$ entropy principle). Furthermore, $M\alpha$EP provides the requisite background for the formulation of this New The $M\alpha$EP provides a methodology for principle distribution, which identifying the probability have maximum $\alpha$ entropy based on given information. Indeed given the three probabilistic entities, namely, the $\alpha$ entropy measure, the set of moment constraints and the probability distribution. As stated earlier, the identification is principle of maximisation of the $\alpha$ based on the entropy measure subject to the given constraints.

The generalised maximum $\alpha$ entropy principle (GM$\alpha$EP) addresses itself to the determination of any one the three when the remaining two probabilistic entities are specified.

The GM$\alpha$EP then spells out deductive procedures for the determination of the unspecified entity when the other two are specified. In this relation, we state three principles, we wish to determine the probability distribution that maximises the entropy measure ($A$), given the probability distribution for $p$, and the entropy measure $\phi(\cdot)$, determine one or more probability constraints that yield the given probability distribution ($B$) and given the constraints $g_1(x_i), g_2(x_i),..., g_m(x_i)$ and the probability distribution $p_i$, ...
determine the best entropy measure that maximized subject to the given constraints (c).

The principle MEP implies the determination of the probability distribution proceeding from the Shannon entropy and a given set of constraints (Cover and Thomas, 2006). However, MDI principle implies the probability distribution proceeding the Shannon entropy and given set of constraints and furthermore a prior probability Distribution Q.

After Shannon (1948), generalisation forms of Entropy are introduced. One of them is Tsallis entropy, that implies the Shannon entropy when \( \alpha \to 1 \). We remind that this paper is written on the basis of the paper of Kesavan and Kapur (1989) with the exception that we use the Tsallis entropy instead of the Shannon entropy. Here, we generalise this formalisms (MEP and MDI), with replacing the Shannon entropy by the Tsallis entropy when \( \alpha = 2 \) and we obtain most unbiased probability distribution proceeding the Tsallis entropy and given set of constraints. A formalism in GMEP is presented.

The principles of maximum \( \alpha \) entropy and minimum discrimination information (MDI) is reviewed and an illustrative example is given.

2 The generalised maximum \( \alpha \) entropy principle

Given the three probabilistic entities, namely the entropy measure, the set of moment constraints and the probability distribution. The MEP provides a methodology for identifying probability distribution, based on a knowledge of the first two entities. In generalised MEP, the formalism renders it possible to determine any one of the probabilistic quantities when the other two are specified.

Replacing Shannon entropy with \( \alpha \) entropy, we present the principle of maximum \( \alpha \) entropy.

2.1 The maximum \( \alpha \) entropy version

In continuous, we state the principle of maximum \( \alpha \) entropy and give illustrative example.

For a probability distribution, \( P = (p_1, \ldots, p_n) \). The Shannon entropy (\( H(P) \)) and Tsallis entropy of order \( \alpha \) (\( S_\alpha(P) \)) are respectively:

\[
H(P) = -\sum_{i=1}^{n} p_i \log p_i,
\]

\[
S_\alpha(P) = \frac{1}{\alpha-1} \left[ \sum_{i=1}^{n} p_i^\alpha - 1 \right], \quad \alpha > 0, \alpha \neq 1
\]

Let \( \varphi(.) \) be a convex function, in general the measure of entropy is:

\[
H_\varphi(P) = -\sum_{i=1}^{n} \varphi(p_i),
\]

If \( g_r(x_i), r = 1, \ldots, m \) are functions that satisfying

\[
\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g_r(x_i) = a_r, \quad r = 1, 2, \ldots, m.
\]
Using the method of Lagrange multipliers, we maximise (3) subject to the \((m+1)\) constraints in (4). The first derivative of \(\varphi(p)\) get under expression:

\[
\sum_{i=1}^{n} \varphi(p_i) + \lambda_0 \left( \sum_{i=1}^{n} p_i - 1 \right) + \sum_{i} \lambda_i \left( \sum_{i=1}^{n} p_i g_i(x_i) - a_i \right) = 0
\]

\[
\Rightarrow \Phi'(p_i) = \lambda_0 + \sum_{i} \lambda_i g_i(x_i)
\]

2.1.1 The direct principle

Given the entropy measure \(\varphi(.)\) and the constraint mean values of \(g_1(x), \ldots, g_m(x)\), we wish to determine the probability distribution that maximises 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 \(p_i\).

2.1.2 The first inverse problem (determination of constraints)

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

2.1.3 The second inverse problem (determination of the entropy measure)

In this case, given the constraints \(g_1(x), \ldots, g_m(x)\) and the probability distribution \(p_i\), determine the most unbiased entropy measure that when maximised subject to the given constraints. Yields the given values in to (5) and get a different equation that can be solved for \(\varphi(.)\). Once \(\varphi(.)\) is known, we can determine the entropy measure

\[
H_\varphi(P) = -\sum_{i=1}^{n} \varphi(p_i)
\]

we now illustrate the \(\alpha\)EP version of generalised MEP on the basis of an example.

Example 1:

Let

\[
\varphi(p_i) = p_i(p_i - 1)
\]

and

\[
H_\varphi(P) = -\sum_{i=1}^{n} \varphi(p_i)
\]

We substitute \(\varphi(p_i)\), gives

\[
H_\varphi(P) = -\sum_{i=1}^{n} \varphi(p_i)
\]

\[
H_\varphi(P) = -\sum_{i=1}^{n} \varphi(p_i) = -\sum_{i=1}^{n} p_i(p_i - 1)
\]
The generalised maximum $\alpha$ entropy principle

in this case $H_\alpha(P)$ is Tsallis entropy when $\alpha = 2$.

For random variables $X_i$, $i = 1, 2, \ldots, n$, the constraints be:

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

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

$$\varphi'(p_i) = \lambda_0 + \lambda_1 x_i \Rightarrow 2p_i - 1 = \lambda_0 + \lambda_1 x_i$$

$$\Rightarrow p_i = \frac{\lambda_0 + \lambda_1 x_i + 1}{2}$$

$$\Rightarrow 1 = \frac{\lambda_0 + 1}{2} + \frac{\lambda_1}{2} \sum_{i=1}^{n} x_i$$

$$\Rightarrow \frac{\lambda_0 + 1}{2} = 1 - \frac{\lambda_1}{2} \sum_{i=1}^{n} x_i$$

$$\Rightarrow p_i = \frac{\lambda_1}{2} x_i + \frac{1}{n} - \frac{\lambda_1}{2n} \sum_{i=1}^{n} x_i$$

$$\Rightarrow p_i = \frac{\mu}{2} x_i + \frac{1}{n} - \frac{\mu}{2n} \sum_{i=1}^{n} x_i$$

Which $\mu = \lambda_1$ and The Lagrange multiplier $\mu$ can be found from

$$\hat{x} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\mu}{2} \sum_{i=1}^{n} x_i^2 - \frac{\mu}{2n} \left( \sum_{i=1}^{n} x_i \right)^2$$

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.

2.1.4 Determination of constraints (the first inverse problem)

If $a$, $b$, $c$ are constants and the constraints are determined from a knowledge of equations (6) and (9), Substituting in to equation (5), we get

$$\lambda_0 + \lambda_1 g_1(x) + \cdots + \lambda_m g_m(x) = 2p_i - 1$$

$$= 2 \left( \frac{\mu}{2} x_i - \frac{\mu}{2n} \sum_{j=1}^{n} x_j + \frac{1}{n} \right) - 1$$

$$= \mu x_i - \mu \bar{x} + \frac{2}{n} - 1,$$

$$\lambda_0 = -\frac{\mu}{n} \bar{x} + \frac{2}{n} - 1,$$

$$\lambda_1 = \mu, \quad g_1(x) = x_i$$

Hence, the constraints are:

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i x_i = \hat{x}$$
2.1.5 Determination of the entropy measure (second inverse problem)

The differential equation is:
\[ \varphi'(p_i) = \lambda_0 + \lambda_i x_i \]

and from (8)
\[ \varphi'(p_i) = a + b p_i \]
\[ \Rightarrow \varphi(p_i) = a p_i + \frac{b}{2} p_i^2 + c \]

\[ H_\varphi (P) = - \sum_{i=1}^{n} \varphi (p_i) = -a - \frac{b}{2} \sum_{i=1}^{n} p_i^2 - cn \]
\[ \Rightarrow b = 2 \]
\[ -a - cn = 1 \]
\[ \Rightarrow H_\varphi (p) = - \sum_{i=1}^{n} p_i^2 + 1 \]
\[ = 1 - \sum_{i=1}^{n} p_i \]
\[ = 1 - \sum_{i=1}^{n} p_i (p_i - 1). \]

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

3 Formalism of the MDI principle

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:
\[
D_{\alpha} (P \parallel Q) = \frac{1}{\alpha - 1} \sum_{i=1}^{n} \left[ \left( \frac{p_i}{q_i} \right)^{\alpha - 1} - 1 \right] p_i ,
\]
\( \alpha > 0, \alpha \neq 1. \)

This measure is always greater than zero and has a global minimum value of zero when the two distributions are identical. If \( Q \) is the uniform distribution
\[
\left( Q = U = \left( \frac{1}{n} , \frac{1}{n} , \ldots , \frac{1}{n} \right) \right)
\]
we have
\[
D_{\alpha} (P \parallel Q) = D_{\alpha} (P \parallel U)
\]
\[
= \frac{1}{\alpha - 1} n^{\alpha - 1} \sum_{i=1}^{n} \left( p_i^{\alpha - 1} - 1 \right) p_i 
\]
\[
= n^{\alpha - 1} \left[ S_\alpha (U) - S_\alpha (P) \right]
\]
where $S_d(U)$ is the Tsallis entropy associated with the uniform distribution.

Minimising $D_{\alpha_d}(P \parallel U)$ would entail maximising $S_d(P)$.

If $\varphi(.)$ be a convex function, Kulback-Liebler divergence measure is:

$$D(P \parallel Q) = \sum_{i=1}^{n} q_i \varphi \left( \frac{p_i}{q_i} \right),$$

be the measure of the directed divergence (Kullback and Liebler, 1951). The constraints be:

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

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

$$\sum_{i=1}^{n} q_i \varphi \left( \frac{p_i}{q_i} \right) + (\lambda_0 + 1) \left( \sum_{i=1}^{n} p_i - 1 \right) + \sum_{r=1}^{m} \lambda_r \left( \sum_{i=1}^{n} g_r p_r - a_r \right) = 0$$

$$\Rightarrow \varphi \left( \frac{p_i}{q_i} \right) = -\lambda_0 - 1 - \lambda_1 g_1(x_i) - \cdots - \lambda_m g_m(x_i) = -\frac{\sum_{i=1}^{n} \varphi \left( \frac{p_i}{q_i} \right)}{\sum_{r=1}^{m} g_r p_r - a_r}.$$

The Tsallis divergence measure when $a = 2$ is:

$$D_{\alpha_d}(P \parallel Q) = \frac{1}{2-1} \sum_{i=1}^{n} \left[ \frac{p_i}{q_i} - 1 \right] p_i$$

$$= \sum_{i=1}^{n} \left[ \frac{p_i}{q_i} - 1 \right] p_i$$

If

$$\varphi \left( p_i \right) = p_i (p_i - 1)$$

we have

$$\varphi \left( \frac{p_i}{q_i} \right) = \frac{p_i}{q_i} \left( \frac{p_i}{q_i} - 1 \right)$$

and so

$$D_{\alpha_d}(P \parallel Q) = \sum_{i=1}^{n} \left[ \frac{p_i}{q_i} - 1 \right] p_i$$

$$\varphi \left( \frac{p_i}{q_i} \right) = 2 \left( \frac{p_i}{q_i} - 1 \right)$$

$$= \lambda_0 + 1 \left( -\lambda_1 g_1(x_i) - \cdots - \lambda_m g_m(x_i) \right)$$

$$\Rightarrow$$

$$p_i = -\frac{\lambda_0 - \lambda_1 g_1(x_i) - \cdots - \lambda_m g_m(x_i)}{2} q_i.$$
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 $p_i^v, q_i^v$ and $\varphi(.)$ are known, (25) determines the constraint functions $g_1(.), g_2(.), ..., g_m(.)$.

c Determination of the divergence measure (second inverse problem)
If $p_i^v, q_i^v$ and $g(x_i)$ are known, (25) and (26) determine $\varphi(.)$ and as such determines the divergence measure $D_{\alpha}(P \parallel Q)$.

d Determination of a priori distributions (third inverse problem)
Finally, if $p_i^v, q_i^v$ and $\varphi(.)$ are known, (25) and (26) determine the $q_i^v$.

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 Conclusions
GMEP and MEP 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 non-additive 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 maximised.

References


