13th Iranian Statistics Conference
Shahid Bahonar University of Kerman
23-25 August 2016

# Generalized entropy maximization under a general constraints

Gholamreza Mohtashami Borzadaran*, Ali Khosravi Tanak

Department of Statistics, Faculty of Mathematics, Ferdowsi university of Mashhad.

**Abstract**

The maximum entropy principle is a flexible and powerful tool for approximation of distributions. hitherto a lot of work has been done in terms of Shannon entropy. In this paper, we first extend the maximum entropy approach from Shannon entropy to generalized entropy. Next, as a special case, we find the maximum quadratic entropy distribution to approximate the model of income distribution with a given mean and Gini index.

**Keywords:** Maximum entropy, Generalized entropy, Inequality measures, Quadratic entropy, Gini index.

**Mathematics Subject Classification (2010):** 94A17, 62F30, 91B82.

## 1 Introduction

As Verdú (1998) describes Information Theory in his review paper, Information Theory is a unifying theory with profound intersections with Probability, Statistics, Computer Science, and other fields. Information Theory continues to set the stage for the development of communications, data storage and processing, and other information technologies. Historically, Information Theory was born in 1948, when Shannon published his famous paper "A mathematical theory of communication." Motivated by the problem of efficiently transmitting information over a noisy communication channel, he introduced a revolutionary new probabilistic way of thinking about communication and simultaneously created the first truly mathematical theory of entropy. In his paper, the concept of entropy were proposed as a measure of uncertainty of a random variable.

The concept of information is closely linked with the concept of uncertainty or surprise. In other word, Uncertainty or surprise can be considered as different shades of information, entropy comes in handy as a measure thereof. Consider a random experiment with outcomes $x_1, x_2, ..., x_N$ with probabilities $p_1, p_2, ..., p_N$, respectively; one can say that these outcomes are the values that a discrete random variable $X$ takes on. The measure of uncertainty about the collection of events is called entropy. Thus, entropy can be interpreted as a measure of uncertainty about the event prior to the experimentation. The information associated with the outcome $\{X = x_i\}$ is denoted by

$$h(p_i) = -\log p_i. \tag{1}$$

This concept can be extended to a series of $N$ events occurring with probabilities $p_1, p_2, ..., p_N$, which then leads to the Shannon entropy as the expected value of this series as

$$H = -\sum_{i=1}^{N} p_i \log p_i, \tag{2}$$

*Speaker: grmohtashami@um.ac.ir

Most environmental processes are continuous in nature. The entropy concepts presented for discrete variables can be extended to continuous random variables. Let X be a random variable having a continuous cumulative density function (cdf) $F$ with probability density function (pdf) $f$, then the basic uncertainty measure for distribution $F$ is defined as

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx, \tag{3}$$

provided the integral exists. The Shannon entropy satisfies a number of desiderata, such as continuity, symmetry, additivity, expansibility, recursivity, and others (Shannon and Weaver (1949)).

There are a generalization of entropy which is proposed by Csiszár (1973). The generalized entropy named $\phi$-entropy, is defined as

$$H_\phi(f) = \int_{-\infty}^{\infty} \phi[f(x)] dx, \tag{4}$$

where $\phi$ is a concave function. Shannon entropy and quadratic entropy which is proposed by Burbea (1984) are most known special cases of $H_\phi(f)$. In Table 1, you see some informational entropies which can be represent by $H_\phi(f)$.

Table 1: Generalized entropies.

| Name of entropy | $\phi(x)$ |
|---|---|
| Shannon (1948) | $-x \log f(x)$ |
| Quadratic entropy, Burbea and Rao (1982) | $x - x^2$ |
| Harvda and Charvat (1967) | $\frac{x^s}{1-s} - x, \quad 0 < s, s \neq 1$ |
| Kapur (1972) | $\frac{x^s+(1-x)^s-1}{1-s}, \quad s \neq 1$ |
| Burbea (1984) | $\frac{x^s-(1+x)^s+1+(s-1)^{-1}(2^s-2)x}{s-2}, \quad s \neq 2$ |

In economics and the social sciences, inequality can be defined as the dispersion of the distribution of income or some other welfare indicator. There are various ways to measure inequality. The Lorenz curve developed by Lorenz (1905) is perhaps the most fundamental tool used to measure income inequality. Graphically, the Lorenz curve gives the proportion of total societal income accruing to the lowest earning proportion of income earners. Let $X$ denote a random variable with cumulative distribution function (cdf) $F$ supported in $(0, \infty)$ and mean $E(X) = \mu$. The Lorenz curve is defined as follows:

$$L_F(u) = \frac{1}{\mu} \int_0^u F^{-1}(x) dx \ , \quad 0 \leq u \leq 1, \tag{5}$$

where $F^{-1}(x) = \inf\{t : F(t) \geq x\}$. If $F$ is income distribution, then $L_F(u)$ denotes the fraction of total income which is in the hands of the $u$th fraction of population possessing the lowest income. This representation forms the basis of many common inequality measures; among them the Gini index which was proposed by Gini (1914) is a famous and well-known measure. The Gini index, $G(F)$, is defined as twice the area between the considered Lorenz curve and the line of perfect equality $L_F(u) = u$,

$$G(F) = 2\int_0^1 (u - L_F(u)) du = 1 - 2\int_0^1 L_F(u) du. \tag{6}$$

The Gini index takes values in the unit interval $[0, 1]$. A low Gini index indicates more equal income distribution, while a high Gini index indicates more unequal distribution. Various generalizations of the Gini index have already been suggested in the literature. Kakwani (1980) and Yitzhaki (1983) proposed a family of generalized Gini indices by introducing different weighting functions for the area under the Lorenz curve

$$G_\nu(F) = 1 - \int_0^1 \nu(\nu-1)(1-u)^{\nu-2} L_F(u) du, \ \nu > 1. \tag{7}$$

In the case of $\nu = 2$, we have the Gini index. When $\nu$ increases, higher weights are attached to small incomes.

Pietra (1932) proposed another inequality measure that is most useful and appropriate in the case of asymmetric and skewed probability distributions. The Pietra index, $P(F)$, is defined as the maximal vertical deviation between the Lorenz curve and the equality line,

$$P(F) = \max_{0\leq u\leq 1} \{u - L_F(u)\}. \tag{8}$$

See Kleiber and Kotz (2003) and the references therein for more details about inequality measures.

Approximation of distributions is a fundamental problem in statistical data analysis. The maximum entropy principle proposed by Jaynes (1957), gives us a general way of approximating a distribution. According to this principle, the best approach is to ensure that the approximation satisfies any constraints on the unknown distribution that we are aware of, and that subject to those constraints, the distribution should have maximum entropy. The problem of maximizing entropy subject to some constraint such as moments has been studied by many authors. For example, see Kagan et al. (1973) and Kapur (1989). Recently, some works have been done in the subject of entropy maximization under the constraints on the inequality measures. Eliazar and Sokolov (2010a) found the distribution that maximizes entropy subject to a given mean and Gini index. Also, Eliazar and Sokolov (2010b) obtained the distribution that maximizes entropy subject to a given mean and Pietra index. Khosravi et al. (2015) found the maximum entropy under the constraint on generalized Gini index. In this paper, we intend to develop their results in terms of generalized entropy.

The rest of the paper is as follows: In Section 2 we review some results on entropy maximization under some moment constraints and inequality measures constraints. Section 3 is devoted to our result in maximization of generalized entropy with a general constraint. In Section 5, we obtain maximum quadratic entropy distribution under the constraints on mean and Gini index.

# 2 Maximum entropy

As pointed out by Jaynes (1957), when an inference is made on the basis of incomplete information, it should be drawn from the probability distribution that maximizes the entropy subject to the constraints on the distribution. The resulting maximum entropy probability distribution corresponds to a distribution which is consistent with the given partial information, but has maximum uncertainty or entropy associated with it.

## 2.1 Maximum entropy under constraints on moments

Let $X$ be a continuous random variable with pdf $f(x)$ and let the following constraints hold,

$$\begin{cases} \int_{-\infty}^{\infty} f(x)dx = 1 \\ \int_{-\infty}^{\infty} h_i(x)f(x)dx = \theta_i, \; i = 1, 2, ..., m. \end{cases} \tag{9}$$

Kagan et al. (1973) showed that the pdf of maximum entropy distribution under the constraints (9) is given by

$$f(x) = A\exp[-\lambda_1 h_1(x) - \lambda_2 h_2(x) - ... - \lambda_m h_m(x)],$$

where $A$, $\lambda_1$, $\lambda_2$, ..., $\lambda_m$ are to be obtained by using the constraints (9). In the following, some well-known special cases are presented.

- **If the range of the random variable $X$ is $[0, 1]$:** Within the class of probability distributions supported on the unit interval, the entropy maximizer is the uniform distribution.

- **If the range of the random variable $X$ is $[0, \infty)$:** Within the class of probability distributions supported on non-negative real numbers, and possessing a given mean, the entropy maximizer is the exponential distribution.

- **If the range of the random variable $X$ is $(-\infty, \infty)$:**
  1. In the constraints (9), if $h_1(x) = (x-a)^2$, where $a$ is a fixed real number, then the maximum entropy distribution is normal with mean $a$ and variance $\theta_1$.

  2. In the constraints (9), if $h_1(x) = x$ and $h_2(x) = (x-\theta_1)^2$, that is when the mean and the variance of $X$ are prescribed to be $\theta_1$ and $\theta_2$ respectively, then the maximum entropy distribution is normal with mean $\theta_1$ and variance $\theta_2$.

  3. **Dispersion:** One the most basic approach to gauge statistical heterogeneity is the notion of dispersion: measuring the fluctuations of the probability distribution around its mean. The dispersion is given by the functional
  $$D(f) = \left(\int_{-\infty}^{\infty} |x-\mu|^p f(x)dx\right)^{\frac{1}{p}}, \ p \geq 1,$$
  The greater $D(f)$, the more scattered and heterogeneous the probability distribution and the smaller $D(f)$, the more concentrated and homogeneous the probability distribution. In the case $p = 2$, the dispersion equals the standard deviation, and the square dispersion equals the variance of distribution. In the constraints (9), if $h_1(x) = x$ and $h_2(x) = |x-\theta_1|^p$, that is when the mean and the dispersion of $X$ are prescribed to be $\theta_1$ and $\theta_2 = \sigma^p$, respectively, then the Subbotin's distribution with the following density function has maximum entropy,
  $$f(x) = \frac{\phi(p)}{\sigma} \exp\left(-\frac{1}{p}\left|\frac{x-\theta_1}{\sigma}\right|^p\right), -\infty < x < \infty, \tag{10}$$
  where $\phi(p) = \frac{p^{1-1/p}}{2\Gamma(1/p)}$ and $\Gamma(.)$ is the complete gamma function. In the case $p = 2$, (10) is the pdf of the normal distribution.

## 2.2 Maximum entropy under constraints on inequality measures

Since among distributions supported on non-negative real line, the experience has shown that the exponential distribution doesn't fit to income data well, looking for the maximum entropy distributions under other constraint has been continued. Holm (1993) derived a family of maximum entropy quantile functions under the constraints on the mean and Gini index. Ryu (2008) determined the functional form of the share function (as a pdf) by the maximum entropy principle under the constraint on the Bonferroni index. Yaghoobi et al. (2014) derived a family of maximum Tsallis entropy quantile functions under the constraints on the mean and Gini index. In this subsection, this question is answered: "What is the maximum entropy distribution subject to a given mean and a given measure of inequality?" The answer is provided when inequality measure is Gini index, Pietra index and generalized Gini index.

### 2.2.1 Gini index

Let $X$ be a non-negative random variable with pdf $f$ and cdf $F$. Here the problem of maximizing entropy subject to a given mean and a given Gini index is stated. In other words, the problem is maximizing the entropy of $X$ subject to the following constraints,

$$\begin{cases} \int_0^\infty f(x)dx = 1, \\ \int_0^\infty xf(x)dx = \mu, \\ G(F) = \gamma. \end{cases}$$

Eliazar and Sokolov (2010a) showed that under the aforementioned constrains, the survival function ($\bar{F} = 1 - F$) of the maximum entropy distribution is given by

$$\bar{F}(x) = \frac{1}{\sigma \exp(\rho x) + (1 - \sigma)}, \quad x \geq 0,$$

where $\sigma$ is positive real valued parameter depending on $\gamma = 1 + \frac{1}{\sigma - 1} - \frac{1}{\log \sigma}$ and $\rho = \frac{\log \sigma}{(\sigma - 1)\mu}$.

### 2.2.2 Pietra index

Eliazar and Sokolov (2010b) showed that within the class of pdfs possessing a given mean and a given Peitra index that is under the constraints

$$\begin{cases} \int_0^\infty f(x)dx = 1, \\ \int_0^\infty x f(x)dx = \mu, \\ P(F) = \eta, \end{cases}$$

the entropy maximizer is bi-exponential pdf

$$f(x) = \begin{cases} c_1 \exp(\alpha x) & \text{if } 0 < x < \mu, \\ c_2 \exp(-\beta x) & \text{if } \mu < x < \infty, \end{cases}$$

where $\alpha$ and $\beta$ are real exponents depending on $\mu$ and $\eta$; $c_1$ and $c_2$ are normalizing coefficients satisfying the relation $\log(c_2/c_1) = (\alpha + \beta)\mu$.

### 2.2.3 Generalized Gini index

Khosravi et al. (2015) extend the result of Eliazar and Sokolov (2010a) by replacing generalized Gini index with Gini index and showed under the constraints

$$\begin{cases} \int_0^\infty f(x)dx = 1, \\ \int_0^\infty x f(x)dx = \mu, \\ G_\nu(F) = \gamma, \end{cases} \tag{11}$$

the maximum entropy distribution has the following survival function,

$$\bar{F}(x) = \left( \frac{1}{c_1 \exp(c_2 x) + (1 - c_1)} \right)^{\frac{1}{\nu - 1}}, \quad x \geq 0, \tag{12}$$

where $c_1$ and $c_2$ depend on $\mu$ and $\gamma$, and are obtained from (11).

## 3 Maximum generalized entropy

In this section, the generalized entropy (4) is considered and we intend to find the maximum generalized entropy under the general constraints. For this purpose, we present the following theorem.

**Theorem 3.1.** *Necesssary and sufficient condition for the generalized entropy $H_\phi$ to have maximum for a given cdf F(x) with pdf f(x), under the convex (in f) constraints*

$$\int_{-\infty}^{\infty} G_i(F(x), f(x), x)dx = \theta_i \ , \ i = 1, 2, ..., m,$$

*is that F(x) satisfy the equation:*

$$\sum_{i=1}^{m} \lambda_i \left( \frac{\partial G_i}{\partial F} - \frac{d}{dx} \frac{\partial G_i}{\partial f} \right) + f' \phi''(f) = 0, \tag{13}$$

*where $\lambda_1, \lambda_2, ..., \lambda_m$ are Lagrange multipliers.*

*Proof.* Since $-H_\phi(f)$ is a convex functional, it is sufficient to $F(x)$ satisfy the Euler's equation for the corresponding Lagrangian, that is

$$\frac{\partial L}{\partial F} - \frac{d}{dx}\frac{\partial L}{\partial f} = 0, \tag{14}$$

where

$$L(F, f, x) = \sum_{i=1}^{m} \lambda_i G_i(F, f, x) - \phi(f).$$

From Euler's equation (14) we have

$$\sum_{i=1}^{m} \lambda_i \frac{\partial G_i}{\partial F} - \frac{d}{dx}\left(\sum_{i=1}^{m} \lambda_i \frac{\partial G_i}{\partial f} - \phi'(f)\right) = 0.$$

The above equation hold iff the equation (13) hold. □

Theorem 3.1 give us a general rule for finding the maximum generalized entropy. All afformentioned maximum entropy distributions can be obtained using the equation (13). For example, when we have Shannon entropy ($\phi(x) = -x\log(x)$) and $G_1(F, f, x) = f(x)$;

- If $G_i(F, f, x) = h_i(x)f(x), i = 2, ..., m$, then the results of section 2.1 is obtained.
- If $G_2(F, f, x) = xf(x)$ and $G_3(F, f, x) = [F(x) - 1]^2$ then the results of section 2.2.1 is obtained.
- If $G_2(F, f, x) = xf(x)$ and $G_3(F, f, x) = \max(0, x - \mu)f(x)$ then the results of section 2.2.2 is obtained.
- If $G_2(F, f, x) = xf(x)$ and $G_3(F, f, x) = [F(x) - 1]^\nu$ then the results of section 2.2.3 is obtained.

# 4 Maximum quadratic entropy under the constraint on Gini index

In this section, within the class of distributions supported on the positive real line, we intend to find the distribution that maximizes the quadratic entropy as a special case of generalized entropy under the constraints

$$\begin{cases} \int_0^\infty f(x)dx = 1, \\ \int_0^\infty xf(x)dx = \mu, \\ G(F) = \vartheta. \end{cases} \tag{15}$$

The result is stated in the next theorem.

**Theorem 4.1.** *The survival function of quadratic entropy maximizer distribution under the constraints (15) is*

$$\bar{F}(x) = \frac{e^{\alpha(2\beta - x)} - e^{\alpha x}}{e^{2\alpha\beta} - 1}, \quad 0 \le x \le \beta, \tag{16}$$

*where $\alpha$ and $\beta$ depend on $\mu$ and $\vartheta$, and are obtained from (15).*

*Proof.* From theorem 3.1, the maximum quadratic entropy must satisfy in equation (13), when $\phi(f) = f - f^2$. So we have the following equation:

$$-\lambda_2 - 2\lambda_3\bar{F}(x) - 2f'(x) = 0.$$

Since $\bar{F}(x)$ is a survival probability function, $\lambda_2$ must be equal zero. Thus, we arrive at the differential equation

$$\lambda_3\bar{F}(x) + f'(x) = 0.$$

The solution of this differential equation is

$$\bar{F}(x) = c_1 e^{\sqrt{\lambda_3}x} + c_2 e^{-\sqrt{\lambda_3}x}.$$

Again, since $\bar{F}(x)$ is a survival probability function, the support of the distribution should be limited. Thus, we arrive at the survival function in (16). □

# Conclusion

The principle of maximum entropy is a technique that can be used to estimate distribution of populations. In this paper, we studied a generalization of maximum Shannon entropy in terms of $\phi$-entropy. Using it the maximum quadratic entropy distribution as an approximation of income distribution was found.

# References

Burbea, J. (1984), The Bose-Einstein entropy of degree alpha and its Jensen difference, *Journal of Multivariate Analysis*, **25**, 225–240.

Burbea, J. and Rao, C. R. (1982), Entropy differential metric, distance and divergence measures in probability spaces: A unified approach, *Journal of Multivariate Analysis*, **12 (4)**, 575–596.

Csiszár, I. (1973), Generalized entropy and quantization problems, *Transactions of the 6th Praque Conference on Information Theory, Statistical Decision Functions, Random Processes (Academia, Prague).*

Eliazar, I. and Sokolov, M. (2010a), Maximization of statistical heterogeneity: from Shannon's entropy to Gini's index, *Physica A*, **389**, 3023–3038.

Eliazar, I. and Sokolov, M. (2010b), Measuring statistical heterogeneity: The Pietra index, *Physica A*, **389**, 117–125.

Gini, C. (1914), Sulla misura della concentrazione e della variabilita' dei caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **73**, 1203–1248.

Havrda, J. and Charvát, F. (1967), Quantification method of classification processes. Concept of structural $a$-entropy. *Kybernetika*, **3**, 30–35.

Holm, J. (1993), Maximum Entropy Lorenz Curves. *Journal of Econometrics*, **59**, 377–389.

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

Kagan, A. M., Linnik, Y. M. and Rao, C. R. (1973),*Characterization Problems in Mathematical Statistics*, John Wiley, New York.

Kakwani, N. (1980), On a class of poverty measures, *Econometrica*, **48**, 437–446.

Kapur, J. N. (1972), Measures of uncertainty, Mathematical programming and Physics. *Journal of the Indian Society of Agricultural Statistics*, **24**, 46–66.

Kapur, J. N. (1989),*Maximum Entropy Models in a Science and Engineering*, John Wiley, New York.

Khosravi Tanak, A., Mohtashami Borzadaran, G. R. and Ahmadi, J. (2015), Entropy maximization under the constraints on the generalized Gini index and its application in modeling income distributions. *Physica A: Statistical Mechanics and its Applications*, **438**, 657–666.

Kleiber, C. and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken, New Jersey.

Lorenz, M. O. (1905), Methods of measuring the concentration of wealth, *Quarterly Publications of the American Statistical Association*, **9**, 209–219.

Pietra, G. (1932), Nuovi contributi alla methodologia degli indici di variabilita e di concentrazione, *Atti del Reale Instituto Veneto di Scienze*, Lettere ed Arti, **91**, 989–1008.

Ryu, H. K. (2008), Maximum entropy estimation of income distributions from Bonferroni indices. In Modeling Income Distributions and Lorenz Curves, Springer New York, 193–210.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423.

Shannon, C. E. and Weaver, W. (1949), *The mathematical theory of information*, University of Illinois Press.

Verdú, S. (1998), Fifty years of Shannon theory, *IEEE Transactions On Information Theory*, **44**, 2057–2077.

Yaghoobi Avval Riabi, M., Mohtashami Borzadaran, G. R. and Yari, G. , (2014), Tsallis Maximum Entropy Lorenz Curves, *Journal of Statistical Research of Iran*, **11**, 41–56.

Yitzhaki, S. (1983), On an extension of the Gini inequality index, *International Economic Review*, **24**, 617–628.