Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey, May 27-29, 2006 (pp617-622)

An Approach to Generalize Entropy Optimization Principles and Related Aspects of Newton's Method

ALADDIN SHAMILOV
Anadolu University
Faculty of Science
Department of Statistics
26470, Eskisehir
TURKEY

Abstract: - In the present study, we have defined, so-called, MaxEnt and MinxEnt functionals on the set of corresponding moment vector functions via the MaxEnt and MinxEnt optimization measures. By virtue of these functionals we have obtained two new distributions – MinMaxEnt and MaxMinxEnt distributions. The approach to obtain mentioned distributions can be formulated as a generalization of MaxEnt and MinxEnt optimization principles. Moreover, we have proposed a proof of existence and uniqueness of solution of functional equation with respect to Langrange multipliers and a rule to choose an initial point for Newton's approximations.

Key-words; MinMaxEnt, MaxMinxEnt, Generalize entropy optimization principle, Newton's method.

1 Introduction

Several optimization principles are formulated and methods realizing these principles are suggested in [2]. Optimizations principles can be applied to different statistical problems [2; 7; 5; 8].

Two of the optimization principles are as follows:

Principle 1: Out all probability distributions satisfying given constraints, choose the distribution that is closest to the uniform distribution.

Principle 2: Out all probability distributions satisfying the given constraints, choose the distribution that is closest to the given a priori distribution.

In [6] a generalization of MaxEnt Principle is proposed. In this study, we have defined, so-called, MaxEnt and MinxEnt continuous functionals on the compact set of corresponding moment vector functions via the MaxEnt and MinxEnt optimization measures.

The moment vector function giving the least value to MaxEnt functional and the moment vector function giving the greatest value to MinxEnt functional generate two distributions. We call this distributions correspondingly MinMaxEnt and MaxMinxEnt distributions.

The approach to obtain MinMaxEnt and MaxMinxEnt distributions can be formulated as a generalization of MaxEnt and MinxEnt optimization principles.

The mentioned generalization of entropy optimization principles allows to select the distributions which are as well as possible closer to observed distribution in the sense of corresponding measures.

Moreover, we have considered some questions concerned with the application of Newton's method to obtain MinMaxEnt and MaxMinxEnt distributions. Problems of this kind are the establishment of existence and uniqueness of solution of functional equation araised in the present situation and the selection of the initial point for approximations.

2 Definition of MaxEnt Functional

The problem of maximizing MaxEnt measure

$$H = -\sum_{i=1}^{n} p_i \ln p_i \qquad (2.1)$$

subject to constraints

$$\sum_{i=1}^{n} p_i g_j(x_i) = \mu_j, \quad j = 0,1,2,\ldots, m \qquad (2.2)$$

where, $\mu_0 = 1,\ g_0(x) = 1,\quad p_i \ge 0\quad i = 1,2,\ldots, n;$ $m+1 < n,$
has solution

$$p_i = e^{-\sum_{j=0}^{m} \lambda_j g_j(x_i)}, \quad i = 1,2,\ldots n \qquad (2.3)$$

where λ_i $(i = 0,1,\ldots, m)$ are Langrange multipliers. Consequently,

$$H_{\max} = -\sum_{i=1}^{n} e^{-\sum_{j=0}^{m} \lambda_j g_j(x_i)} \left(-\sum_{j=0}^{m} \lambda_j g_j(x_i) \right) \qquad (2.4)$$
$$= \sum_{j=0}^{m} \lambda_j \mu_j .$$

If the distribution $p^{(0)} = \left(p_1^{(0)}, \ldots, p_n^{(0)}\right)$ is given, then one can obtain moment vector value $\mu = (1, \mu_1, \ldots, \mu_m)$ for the each moment vector function $g(x) = (1, g_1(x), \ldots, g_m(x))$ and H_{max} can be considered as a functional depends on moment vector function g(x). We call this functional MaxEnt functional. Consequently, we will use the notation U(p) and U(g) interchangeably to denote the maximum value of H corresponding to
$p = (p_1, \ldots, p_n)$
or
$g(x) = (1, g_1(x), \ldots, g_m(x))$.
It's proved as follows:

Theorem 1: If U(g) is the above defined MaxEnt functional derived by maximizing (2.1) subject to constraints (2.2), then U(g) is continuous on the set of continuous moment vector functions from $C_{[a,b]}$ and reaches its least and greatest values in the given compact set K, $K \subset C_{[a,b]}$

Remark: MaxEnt and MinxEnt functionals for continuous- variate distributions are defined simularly. Mentioned functionals also are continuous on the compact set of moment vector functions.

3 Discrete MinMaxEnt Distribution

Let U(g) be a functional defined in Section 2 which is derived by maximizing (2.1) subject to constraints (2.2) and K be the compact set of moment vector functions g(x). Then according to the Theorem 1, U(g) reaches its least and greatest values in this compact set.

Suppose that
$\min_{g \in K} U(g) = U\left(g^{(1)}\right);\ \max_{g \in K} U(g) = U\left(g^{(2)}\right).$

Consequently,

$$U\left(g^{(1)}\right) \le U\left(g^{(2)}\right). \qquad (3.1)$$

We call the distribution $p^{(1)} = \left(p_1^{(1)}, \ldots, p_n^{(1)}\right)$ corresponding to the moment vector function $g^{(1)}(x)$ as discrete MinMaxEnt distribution.

Theorem 2: If $p^{(0)} = \left(p_1^{(0)}, \ldots, p_n^{(0)}\right)$ is given distribution, $p^{(1)} = \left(p_1^{(1)}, \ldots, p_n^{(1)}\right)$ is the MinMaxEnt distribution and $p^{(2)} = \left(p_1^{(2)}, \ldots, p_n^{(2)}\right)$ is the distribution corresponding to moment vector function $g^{(2)}$(x) in (3.1), then inequalities

$$H\left(p^{(0)}\right) \le U\left(p^{(1)}\right) \le U\left(p^{(2)}\right),$$
or
$$H\left(p^{(0)}\right) \le U\left(g^{(1)}\right) \le U\left(g^{(2)}\right), \qquad (3.2)$$

where H is entropy function, U is MaxEnt functional defined beforehand, are satisfied.

Otherwise out all probability distributions corresponding to moment vector functions g(x) from compact set K, the MinMaxEnt distribution is the closest to the given distribution $p^{(0)} = \left(p_1^{(0)}, \ldots, p_n^{(0)}\right)$.

4 Continuous–Variate MinMaxEnt Distribution

Let $f_0(x)$ be given probability density function (p.d.f.). The problem of maximizing continuous-variate MaxEnt measure

$$H = -\int_a^b f(x)\ln f(x)dx\,, \tag{4.1}$$

where $g_0(x)=1,\quad \mu_0=1,$ has solution

$$f(x) = e^{-\sum_{j=0}^{m}\lambda_j g_j(x)}\,, \tag{4.2}$$

where $\lambda_i \;\; (i=0,1,\ldots,m)$ are Langrange multipliers.

Consequently,

$$H_{\max} = -\int_a^b e^{-\sum_{j=0}^{m}\lambda_j g_j(x)}\left(-\sum_{j=0}^{m}\lambda_j g_j(x)\right)dx = \sum_{j=0}^{m}\lambda_j\mu_j\,. \tag{4.3}$$

If the p.d.f. $f_0(x)$ is given, then one can obtain moment vector value $\mu=(1,\mu_1,\ldots,\mu_m)$ for each moment vector function $g(x)=(1,g_1(x),\ldots,g_m(x))$ and H_{max} can be considered as a functional dependent on moment vector functions g(x). We call this functional continuous-variate MaxEnt functional and use the notation U(f) and U(g) interchangeably to denote the maximum value of H corresponding to p.d.f. f(x) or moment vector function g(x). Functional U(g) is continuous on compact set K of moment vector functions g(x).

Suppose that,

$$\min_{g\in K} U(g)=U\left(g^{(1)}\right);\ \max_{g\in K} U(g)=U\left(g^{(1)}\right),$$

then

$$U\left(g^{(1)}\right)\le U\left(g^{(2)}\right). \tag{4.4}$$

We call the distribution $f_1(x)$ corresponding to moment vector function $g^{(1)}(x)$, MinMaxEnt continuous variate distribution.

Theorem 3: Let $f_0(x)$ be given p.d.f., $f^{(1)}(x)$ be the continuous variate MinMaxEnt distribution and $f^{(2)}(x)$ be the distribution corresponding to moment vector function $g^{(2)}(x)$ in (4.4), then the inequalities,

$$H\left(f^{(0)}\right)\le U\left(f^{(1)}\right)\le U\left(f^{(2)}\right),$$

or

$$H\left(f^{(0)}\right)\le U\left(g^{(1)}\right)\le U\left(g^{(2)}\right), \tag{4.5}$$

where H is entropy function, U is MaxEnt functional defined beforehand, are satisfied.

Otherwise, out all probability distributions corresponding to moment vector functions g(x) from compact set K the continuous-variate MinMaxEnt distribution is the closest to the given p.d.f. $f_0(x)$.

5 Discrete MaxMinxEnt Distribution

Let $p^{(0)}=\left(p_1^{(0)},\ldots,p_n^{(0)}\right)$ be given probability distribution. The problem of minimizing discrete MinxEnt probability measure

$$D(p:q)=\sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i} \tag{5.1}$$

subject to constraints (2.2) has solution

$$p_i = q_i e^{-\sum_{j=0}^{m}\lambda_j g_j(x_i)}\,,\quad i=1,2,\ldots n \tag{5.2}$$

where $\lambda_1,\ldots,\lambda_m$ are Langrange multipliers.

Suppose that the probability distribution $p^{(0)}=\left(p_1^{(0)},\ldots,p_n^{(0)}\right)$ is given, then one can obtain moment vector value $\mu=(1,\mu_1,\ldots,\mu_m)$ for each moment vector function $g(x)=(1,g_1(x),\ldots,g_m(x))$ and D can be considered as a functional dependent on moment vector functions g(x).

We call this functional MinxEnt functional and use the notation V(f) and V(g) interchangeably to denote the minimum value of D corresponding to probability distribution p or moment vector function g(x). Functional V(g) is continuous on compact set K of moment vector functions g(x).

Let

$$\max_{g\in K} V(g)=V\left(g^{(1)}\right)\ \min_{g\in K} V(g)=V\left(g^{(2)}\right),$$

then

$$V\left(g^{(1)}\right)\ge V\left(g^{(2)}\right). \tag{5.3}$$

We call the distribution $p^{(1)}=\left(p_1^{(1)},\ldots,p_n^{(1)}\right)$ corresponding to moment vector function $g^{(1)}(x)$ discrete MaxMinxEnt distribution.

Theorem 4: Assume that, $p^{(0)} = \left(p_1^{(0)}, \ldots, p_n^{(0)}\right)$ is given probability distribution, $p^{(1)} = \left(p_1^{(1)}, \ldots, p_n^{(1)}\right)$ is the discrete MaxMinxEnt distribution and $p^{(2)} = \left(p_1^{(2)}, \ldots, p_n^{(2)}\right)$ is the distribution corresponding to moment vector function $g^{(2)}$ (x) in (5.3). Then inequalities,

$$D\left(p^{(0)} : q\right) \geq V\left(p^{(1)}\right) \geq V\left(p^{(2)}\right)$$

or

$$D\left(p^{(0)} : q\right) \geq V\left(g^{(1)}\right) \geq V\left(g^{(2)}\right) \qquad (5.4)$$

where D is Kullback-Leibler measure, V is MinxEnt functional defined beforehand, hold.

Otherwise, out all probability distributions corresponding to moment vector functions g(x) from compact set K the discrete MaxMinxEnt distribution is the closest to the given probability distribution $p^{(0)} = \left(p_1^{(0)}, \ldots, p_n^{(0)}\right)$.

6 Continuous-Variate MaxMinxEnt Distribution,

Let $f_0(x)$ be given p.d.f., $\int_a^b f_0(x)dx = 1$. The problem of minimizing continuous-variate MinxEnt measure

$$D(f : q) = \int_a^b f(x) \ln \frac{f(x)}{q(x)} dx \qquad (6.1)$$

subject to constraints

$$\int_a^b f(x) g_j(x) dx = \mu_j, \quad j = 0,1,2,\ldots, m \qquad (6.2)$$

where $g_0(x) = 1$, $\mu_0 = 1$ has solution

$$f(x) = q(x) e^{-\sum_{j=0}^{m} \lambda_j g_j(x)},$$

where $\lambda_1, \ldots, \lambda_m$ are Langrange multipliers.

Therefore,

$$D_{\min} = \int_a^b q(x) e^{-\sum_{j=0}^{m} \lambda_j g_j(x)} \left(-\sum_{j=0}^{m} \lambda_j g_j(x)\right) dx \qquad (6.3)$$

$$= -\sum_{j=0}^{m} \lambda_j \mu_j.$$

If the p.d.f. $f_0(x)$ is given, then one can obtain moment vector value $\mu = (1, \mu_1, \ldots, \mu_m)$ for each moment vector function $g(x) = (1, g_1(x), \ldots, g_m(x))$ and $D_{\min}$ can be considered as a functional dependent on moment vector functions g(x). We call this functional continuous variate MinxEnt functional and use the notation V(f) and V(g) interchangeably to denote the minimum value of D corresponding to p.d.f. f(x) or moment vector function g(x). Functional V(g) is continuous on compact set $K \subset C[a,b]$ of moment vector-functions g(x).

Let $\max_{g \in K} V(g) = V\left(g^{(1)}\right)$ $\min_{g \in K} V(g) = V\left(g^{(2)}\right)$,

then

$$V\left(g^{(1)}\right) \geq V\left(g^{(2)}\right). \qquad (6.4)$$

We call the distribution $f_1(x)$ corresponding to moment vector function $g^{(1)}(x)$ continuous MaxMinxEnt distribution.

Theorem 5: Let $f_0(x)$ be given p.d.f., $f^{(1)}(x)$ be the continuous variate MaxMinxEnt distribution and $f^{(2)}(x)$ be the distribution corresponding to moment vector function $g^{(2)}(x)$ in (6.4). Then the inequalities

$$D\left(f^{(0)} : q\right) \geq V\left(f^{(1)}\right) \geq V\left(f^{(2)}\right)$$

or

$$D\left(f^{(0)} : q\right) \geq V\left(g^{(1)}\right) \geq V\left(g^{(2)}\right),$$

where D is Kullback-Leibler measure, V is MinxEnt functional defined beforehand, are satisfied.

Otherwise, out all probability distributions corresponding to moment vector functions g(x) from compact set K, $K \in C[a,b]$ the continuous variate MaxMinxEnt distribution is the closest to the given probability distribution $f_0(x)$.

7 On choosing an initial point for Newton's approximations

The problem obtaining MinMaxEnt and MaxMinxEnt probability distributions is concerned with evaluation of Langrange multipliers encountered in formulas of MaxEnt and MinxEnt probability distributions.

Mentioned problem can be solved, for example, by using Newton's method. Therefore it is important to consider main questions of Newton's method. These questions are:

- The existence and uniqueness of solution of functional equation arised in determination of MinMaxEnt and MaxMinxEnt distributions

- The selection of the initial point for Newton's approximations.

Note that, the established result is useful not only in evaluation of Lagrange multipliers but also in the other applications of Newton's method and is an alternative to [7].

We shall consider mentioned questions for discrete MinMaxEnt distribution. Note that the developed outline is also suitable for the other distributions.

Langrange multipliers $\lambda_1, \ldots, \lambda_m$ are determined as solution of system of equations:

$$\left.\begin{array}{l} \sum_{i=1}^{n} e^{-\sum_{j=0}^{m} \lambda_j g_j(x_i)} = 1, \quad g_0(x) = 1 \\ f(\lambda) = \eta \end{array}\right\} \quad (7.1)$$

where $f = (f_1, f_2, \ldots, f_m)'$,

$\eta = (\eta_1, \eta_2, \ldots, \eta_m)'$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$,

$$f_j(\lambda) = \sum_{i=1}^{n} g_j(x_i) e^{-\sum_{j=0}^{m} \lambda_j g_j(x)}, \quad j = 1, 2, \ldots, m.$$

Remind that (7.1) is received by inserting (2.2) into (2.3). From first equation (7.1) follows:

$$\lambda_0 = \ln \sum_{i=1}^{n} e^{-\sum_{j=1}^{m} \lambda_j g_j(x_i)}. \quad (7.2)$$

Consequently, in order to obtain $\lambda_1, \ldots, \lambda_m$ it is sufficient to solve equation

$$f(\lambda) = \eta. \quad (7.3)$$

Let

$$B = \left\{ y \in E^m : |y_j| \le a_j \right\}, \quad a_j = \sum_{i=1}^{n} |g_j(x_i)|,$$

then $f : E^m \to B$. $f'(\lambda)$ is variance-covariance matrix. Therefore if random variables $g_1(x), \ldots, g_m(x)$ are linearly independent, then $f'(\lambda)$ is positive defined matrix and its all eigenvalues are positive [4; 1]. Let χ_0 be the least eigenvalue of $f'(\lambda)$, then the inequality

$$\left\| [f'(\lambda)]^{-1} \right\| \le \frac{1}{\chi_0} \quad (7.4)$$

holds.

We have proved that $f''(\lambda)$ is bounded in norm for all $\lambda \in E^m$. So in order to apply Newton's method to equation (7.3), the number $\left\| \eta - \eta^{(0)} \right\|$ must be sufficiently small, where $\eta^{(0)} = f(\lambda^{(0)})$. If the mentioned condition is fulfilled, then $\lambda^{(0)}$ can be taken as initial point for Newton's apprximations.

Now assume that this condition is not fulfilled. Since $f'(\lambda)$ is non-singular according to theorem on locally reversion [3; 9], one-to-one mapping exists between small neighbourhoods of $\lambda^{(0)}$ and $\eta^{(0)}$. In other words $f(\lambda)$ is locally one-to-one. Let $\eta^{(1)}$ be a point of segment $[\eta^{(0)}, \eta]$, $\left(x(t) \in [\eta^{(0)}, \eta], x(t) = (\eta - \eta^{(0)}) t + \eta^{(0)};\ 0 \le t \le 1\right)$, which belongs to small neighbourhood of $\eta^{(0)} : \eta^{(1)} = (\eta - \eta^{(0)}) \frac{\varepsilon}{\left\| \eta - \eta^{(0)} \right\|} + \eta^{(0)}, \quad \varepsilon > 0$.

Then $\lambda^{(0)}$ is acceptable as initial point of Newton's approximations and by using Newton's method it is possible to find that

$$\lambda^{(1)} = f^{-1}(\eta^{(1)}). \quad (7.5)$$

Proceeding the fixed process along segment $[\eta^{(0)}, \eta]$ it is available to find point $\eta^{(n)}$, $\eta^{(n)} \in [\eta^{(0)}, \eta]$ such that $\left\| \eta - \eta^{(n)} \right\| < \varepsilon$ and $\lambda^{(n)} = f^{-1}(\eta^{(n)})$. Therefore, $\lambda^{(n)}$ is acceptable as initial point of Newton's approximations for solution of equation (7.3).

The fact that the number n of intervals $[\eta^{i+1}, \eta^{i}]$ is finite can be established by Borel's Lemma.

Remark 1: The existence of solution of equation (7.3) can be proved only by applying the theorem on locally reversion. But the application of Newton's method allows to obtain approximately $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(n)}$ which among $\left[\eta^{(0)},\eta\right]$ lead to η. Since each convex function could have a unique maximum, the uniqueness of solution of equation (7.3) follows from the convexity of Entropy function H. So f is one-to-one on all of E^m.

Remark 2: Practically, it is convenietly to choose the initial point for Newton's approximations as $\lambda^{(0)}=(0,\ldots,0)$, then $\lambda_0=\ln n$.

8 Conclusion

The proved Theorems 2-5 for discrete distributions and continuous-variate distributions in the forms of MinMaxEnt and MaxMinxEnt distributions obtain the closest distributions to given in dependence on considering moment vector functions and optimization measures.

In the case of discrete distributions the application of known optimization principles require to satisfy inequality m+1<n. But the Theorem 2 and 4 shows that it is possible to use moment functions of number m, when m+1>n. This situation can be described in the following form.

Let the number m of given moment functions such that m+1>n. Then classical optimization principles immediately aren't applicable. But out all moment vector functions of the form $(1,g_1(x))$ according Theorems 2,4 it is possible to choose the moment vector function generating the distribution closest to the given probability distribution in the sense of corresponding optimization measure. Let mentioned moment vector function be $\left(1,g_1^{(0)}(x)\right)$. Furthermore, by testing all combinations of pair $\left(1,g_1^{(0)}(x)\right)$ and the other given moment vector functions out all moment vector functions $\left(1,g_1^{(0)}(x),g_2^{(0)}(x)\right)$ according to Theorems 2-4 one can choose moment vector function $\left(1,g_1^{(0)}(x),g_2^{(0)}(x)\right)$ generating the distribution which is the closest to the given distribution. Proceeding this process, it is possible to choose moment vector function $\left(1,g_1^{(0)}(x),\ldots,g_k^{(0)}(x)\right)$, where k+1<n that generates the distribution among all distributions generated by k+1 dimensional moment vector functions.

The process described in Section 7 enables to obtain an initial point $\lambda^{(n)}$ for Newton's approximations for solution of equation $f(\lambda)=\eta$ by starting from arbitrary point $\lambda^{(0)}$, $\lambda^{(0)}\in E^m$ and succesfully applying Newton's method to equations $f\left(\lambda^{(i)}\right)=\eta^{(i)}$, $i=0,1,\ldots,n$.

The established result concerned with Newton's method is useful not only for evaluation of Langrange multipliers but also in the other applications and is an alternative to [7].

References

[1] Cramer H., *Mathematical Methods of Statistics*, Princeton University Pres, 1966.

[2] Kapur, J. N. and Kesavan, H. K., *Entropy Optimization Principles with Applications,* Academic Press, Inc., USA, 1992.

[3] Munkres J. R., *Analysis on Manifolds*, Westview Press, 1991.

[4] Papoulis A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Inc., 1991.

[5] Romera E., Angulo J. C. and Dehesa J. S., Reconstruction of a Density From its Entropic Moments, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings*, Volume 617, 2002, pp. 449-457.

[6] Shamilov A. Mert Kantar Y., On a Distribution Minimizing Maximum Entropy, *Ordered Statistical Data: Approximations, Bounds and Characterizations*, Izmir University of Economics, 2005, pp. 56.

[7] Ximing W., Calculation of Maximum Entropy Densities with Application to Income Distribution, *Journal of Econometrics*, 115, 2003, pp.347-354.

[8] Grendar M., Grendar Jr. and M., Maximum Entropy Method with Non-Linear Moment Constraints: Challenges, *Bayesian inference and Maximum Entropy methods in Science and Engineering*, 2004, pp. 97-109.

[9] Andre I. K., *Advanced Calculus with Applications in Statistics*, A John Wiley&Sons, Inc. Publication, 2003