MAXENTROPIC RECONSTRUCTION OF SOME PROBABILITY DISTRIBUTIONS WITH LINEAR INEQUALITY CONSTRAINTS

Vasile Preda,
Faculty of Mathematics and Informatics, University of Bucharest,
14 Academiei Street, Bucharest 010014, Romania
(preda@math.math.unibuc.ro)
Costel Bălcău,
Faculty of Mathematics and Informatics, University of Piteşti,
1 Târgu din Vale Street, Piteşti 110040, Romania
(cbalcau@linux.math.upit.ro)

Abstract

The maxentropic reconstruction is a technique for finding an unknown probability distribution from some known information. In this paper we obtain the maxentropic reconstruction of some probability distributions from the knowledge of a prior distribution and of some lower and upper bounds for the mean values of some random variables. For this we use the Csiszár's I-projection theorems and the geometric programming method. If some average values of the prior distribution are computed, we obtain a refined form of our solution. Finally, we give several examples for this approach.

Key words: relative entropy, I-projection, maxentropic reconstruction technique, geometric programming method, weak/strong duality theorems.

AMS subject classifications: 94A17, 90C25, 90C46.

1 INTRODUCTION

The purpose of a maxentropic reconstruction is to estimate a positive function from incomplete data. This technique has several useful applications in a variety of fields such as astronomy [18], traffic distribution [26], X-ray crystallography [25], quantum mechanics [10], noisy data [8], tomography [13] and applications of Markov chains [14, 23].

The problem of maxentropic reconstruction of a probability distribution with constraints expressed by mean values has played a central role in statistics. Let us consider a random variable X. We are given only some partial information about the probability distribution of X. Based to the *maximum entropy principle*, introduced by Jaynes [17] and Kullback [19] we should choose the probability distribution that is consistent with the given information but maximizes the entropy or minimizes the relative entropy. It is well-known that by using this method some main probability distributions have been reobtained [17, 19, 16, 9, 21, 22, 5, 15].

Borwein, Lewis, Limber and Noll [3, 2] present a variety of new entropy measures, a duality theory and numerical results for maxentropic reconstruction of an unknown density function on the basis of certain measurements. Gzyl [11, 12] uses the maxentropic reconstruction method for solving ill-posed linear inverse problems and for obtaining some families of probability distributions. More recently, Bnouhachem and Liu [1] use an alternating direction method for maxentropic reconstruction problems subject to simple constraints sets.

In this paper, we extend the maxentropic reconstruction technique to find an unknown continuous probability distribution from the knowledge of a prior distribution and of some bounds for the mean values of some random variables. In Section 2 we define this problem as an entropy optimization primal problem with linear inequality constraints. For solving this problem, in Section 3 we apply the *geometric programming method* [6, 7, 20] to derive a convex dual problem. By using the *Csiszár's I-projection theorems* [4] we can show some duality theorems. In Section 4 we derive a refined form of the dual problem when some average values of the prior distribution are computed. In Section 5 we present three particular cases, when the dual problem can be more simplified. Finally, we apply our approach to obtain the power distribution and the Pareto distribution.

2 PROBLEM STATEMENT

Let $(V, \mathcal{B}(V), m_L)$ be a measure space, where $V \subseteq \mathbb{R}$ is an interval, $\mathcal{B}(V)$ denotes the Borel sets of V and m_L denotes the Lebesque measure on the real line. Let us consider a prior probability distribution ν on $(V, \mathcal{B}(V))$ defined by the density function $q : V \to \mathbf{R}$, $q(x) \geq 0,\ \forall x \in V,\ \displaystyle\int_V q(x)dx = 1.$

Remark 1 *The prior distribution ν may be given on the basis of some experiences or it may be obtained from theoretical models.*

We denote by $\mathcal{M}(V, \nu)$ the set of all probability density functions p on $(V, \mathcal{B}(V), m_L)$ such that the probability distribution μ defined by p is absolutely continuous with respect to the prior distribution ν.

To formulate our problem, let f_i, $i \in \{1, \ldots, n\}$, be real continuous functions defined on V, and let c_i , $i \in \{1, \ldots, n\}$, be real constants. Our problem consists of finding continuous probability distributions μ on $(V, \mathcal{B}(V))$ with density function $p \in \mathcal{M}(V, \nu)$ that are consistent with the following inequality constraints of mean values type

$$\int_V p(x) f_i(x) dx \geq c_i,\ \forall i \in \{1, \ldots, n\},$$

where the functions f_i and the constants c_i, $i \in \{1, \ldots, n\}$, are given.

For solving this problem, we apply the maxentropic reconstruction technique. Thus we consider the following entropy optimization problem with linear inequality constraints

$$(P) : \left| \begin{array}{l} \displaystyle\min_{p \in \mathcal{M}(V,\nu)} H(p; q) = \int_V p(x) \ln \frac{p(x)}{q(x)} dx \quad \text{s.t.} \\ \displaystyle\int_V p(x) f_i(x) dx \geq c_i,\ \forall i \in \{1, \ldots, n\}. \end{array} \right.$$

We assume that the program (P) is consistent and that

$$\int_V q(x) \exp(s|x|) dx < \infty,\ \forall s \in \mathbb{R}. \tag{2.1}$$

Remark 2 *The objective function $H(p; q)$ of program (P) is the relative entropy of p with respect to q (the cross-entropy of q with respect to p; the Kullback-Leibler number). This function is a strictly convex function on p (see [15], for example) and hence it follows that the program (P) has an unique optimal solution.*

3 DUALITY

Using the geometric programming method (see [6, 7, 20], for example), now we are ready to define a convex dual problem for program (P) as follows

$$(D): \left| \begin{array}{l} \max\limits_{\lambda \in \Lambda} G(\lambda) = \sum\limits_{i=1}^{n} \lambda_i c_i - \ln Z(\lambda) \quad \text{s.t.} \\ \displaystyle\int_V q(x)[f_i(x) - c_i] \exp\left[\sum_{k=1}^{n} \lambda_k f_k(x)\right] dx \geq 0, \ \forall i \in \{1, \ldots, n\}, \\ \lambda_i \geq 0, \ \forall i \in \{1, \ldots, n\}, \end{array} \right.$$

where $Z(\lambda) = \displaystyle\int_V q(x) \exp\left[\sum_{k=1}^{n} \lambda_k f_k(x)\right] dx$ and $\Lambda = \{\lambda \in \mathbb{R}^n \mid Z(\lambda) < \infty\}$.

Remark 3 *The Hessian of the dual objective function G is negative definite (see [25]), and then G is a strictly concave function. According to the Fenchel duality (see [24]), if the dual program (D) has an interior feasible solution then the dual program (D) has also an unique optimal solution. The assumption (2.1) ensures the differentiability of G.*

Now we provide the following duality theorems.

Theorem 1 (weak duality) *If p is a feasible solution of program (P) and λ is a feasible solution of program (D), then*

$$H(p; q) \geq G(\lambda). \tag{3.1}$$

Moreover, the equality holds only if

$$p(x) = \frac{q(x)}{Z(\lambda)} \exp\left[\sum_{k=1}^{n} \lambda_k f_k(x)\right], \ \forall x \in V. \tag{3.2}$$

Proof: By the definitions of programs (P) and (D) and using the Jensen's inequality for the concave function $\ln x$ we have

$$\sum_{i=1}^{n} \lambda_i c_i - H(p; q) \leq \ln \int_V p(x) \frac{q(x)}{p(x)} \exp\left[\sum_{i=1}^{n} \lambda_i f_i(x)\right] dx = \ln Z(\lambda).$$

The inequality (3.1) is proved. Applying the equality part of Jensen's inequality we obtain that the inequality (3.1) becomes an equality only if

$$\frac{q(x)}{p(x)} \exp\left[\sum_{i=1}^{n} \lambda_i f_i(x)\right] = C, \ \forall x \in V,$$

C being a constant with respect to x. By the fact that $\int_V p(x)dx = 1$ we derive that $C = Z(\lambda)$, and hence the equality (3.2) is also proved. ∎

Theorem 2 (strong duality) *If λ^* is an optimal solution of dual program (D), then the function p^* defined by*

$$p^*(x) = \frac{q(x)}{Z(\lambda^*)} \exp\left[\sum_{k=1}^{n} \lambda_k^* f_k(x)\right], \ \forall x \in V \tag{3.3}$$

is the optimal solution of program (P) and $\quad H(p^*; q) = G(\lambda^*)$.

Proof: Obviously, $\int_V p^*(x)dx = 1$, and for all $i \in \{1, \ldots, n\}$ we have

$$\int_V p^*(x) f_i(x) dx \geq \frac{1}{Z(\lambda^*)} \int_V q(x) c_i \exp\left[\sum_{k=1}^{n} \lambda_k^* f_k(x)\right] dx = c_i.$$

Therefore p^* is a feasible solution of program (P). We shall show, by contradiction, that for all $i \in \{1, \ldots, n\}$ the following equality holds

$$\lambda_i^* \int_V q(x)[f_i(x) - c_i] \exp\left[\sum_{k=1}^{n} \lambda_k^* f_k(x)\right] dx = 0. \tag{3.4}$$

Indeed, if $\lambda_i^* > 0$ and $\int_V q(x)[f_i(x) - c_i] \exp\left[\sum_{k=1}^{n} \lambda_k^* f_k(x)\right] dx > 0$, then $\partial_i G(\lambda^*) = 0$ and it follows that $\int_V q(x)[f_i(x) - c_i] \exp\left[\sum_{k=1}^{n} \lambda_k^* f_k(x)\right] dx = 0$, contradiction. Thus (3.4) holds. From (3.3) and (3.4) we obtain that

$$H(p^*; q) = \sum_{i=1}^{n} \lambda_i^* \int_V p^*(x) f_i(x) dx - \ln Z(\lambda^*) \int_V p^*(x) dx = G(\lambda^*).$$

On the other hand, if p is an arbitrary feasible solution of program (P), in the same manner as above we obtain that

$$\int_V p(x) \ln \frac{p^*(x)}{q(x)} dx = \sum_{i=1}^{n} \lambda_i^* \int_V p(x) f_i(x) dx - \ln Z(\lambda^*) \int_V p(x) dx \geq G(\lambda^*).$$

It follows that $H(p^*; q) \leq \int_V p(x) \ln \frac{p^*(x)}{q(x)} dx$, for every feasible solution p of program (P). Therefore, by Csiszár characterization of minimum relative entropy [4], we conclude that p^* is the optimal solution of program (P). ■

Remark 4 *Using the equivalence $\int_V p(x) f_i(x) dx = c_i$ if and only if $\int_V p(x) f_i(x) dx \geq c_i$ and $\int_V p(x)(-f_i(x)) dx \geq -c_i$, we regain the duality results concerning minimum relative entropy with linear equality constraints (see Ihara [15]). In this particular case, we can remove all the constraints of the dual program (D), these constraints being satisfied by the dual optimal solution since $\partial_i G(\lambda^*) = 0$, $\forall i \in \{1, \ldots, n\}$. Therefore, in this case the dual program (D) will be an unconstrained convex maximization problem.*

4 A REFINED FORM OF THE DUAL PROGRAM

If the average values $\int_V q(x)f_i(x)dx$, $i \in \{1, \ldots, n\}$ of the prior probability distribution are known, then we can obtain a simplified form of the dual program (D). Denote

$$I_1 = \left\{ i \in \{1, \ldots, n\} \mid \int_V q(x)f_i(x)dx \geq c_i \right\},\ I_0 = \{1, \ldots, n\} \setminus I_1.$$

The dual program (D) has now the following simplified form

$$(D_0) : \left| \begin{array}{l} \max\limits_{\lambda \in \Lambda_0} G_0(\lambda) = \sum\limits_{i \in I_0} \lambda_i c_i - \ln Z_0(\lambda) \quad \text{s.t.} \\ \int_V q(x)[f_i(x) - c_i] \exp\left[\sum\limits_{k \in I_0} \lambda_k f_k(x) \right] dx \geq 0,\ \forall i \in I_1, \\ \lambda_i \geq 0,\ \forall i \in I_0, \end{array} \right.$$

where $Z_0(\lambda) = \int_V q(x) \exp\left[\sum\limits_{k \in I_0} \lambda_k f_k(x) \right] dx$ and $\Lambda_0 = \left\{ \lambda \in \mathbb{R}^{|I_0|} \mid Z_0(\lambda) < \infty \right\}$.

By changing (D) into (D_0) we reduce both the number of variables and the number of constraints of dual program. Similarly to Theorem 1 it can prove the following duality theorem.

Theorem 3 (weak duality) *If p is a feasible solution of program (P) and λ is a feasible solution of program (D_0), then* $H(p; q) \geq G_0(\lambda)$.

The equality holds only if $p(x) = \dfrac{q(x)}{Z_0(\lambda)} \exp\left[\sum\limits_{k \in I_0} \lambda_k f_k(x) \right],\ \forall x \in V.$

Theorem 4 (strong duality) *If λ^* is an optimal interior solution of dual program (D_0), then the function p^* defined by*

$$p^*(x) = \frac{q(x)}{Z_0(\lambda^*)} \exp\left[\sum_{k \in I_0} \lambda^*_k f_k(x) \right],\ \forall x \in V$$

is the optimal solution of program (P) and $H(p^*; q) = G_0(\lambda^*)$.

Proof: Obviously, $\int_V p^*(x)dx = 1$. For every $i \in I_0$, λ^* is an interior dual feasible solution with respect to the component λ_i, and hence $\partial_i G_0(\lambda^*) = 0$. It follows that $\int_V p^*(x)f_i(x)dx = c_i,\ \forall i \in I_0$.

For every $i \in I_1$, using the definition of program (D_0) we obtain that

$$\int_V p^*(x)f_i(x)dx \geq \frac{1}{Z_0(\lambda^*)} \int_V q(x)c_i \exp\left[\sum_{k \in I_0} \lambda^*_k f_k(x) \right] dx = c_i.$$

Thus p^* is a feasible solution of program (P). In the same way as in the proof of Theorem 2 we derive that $H(p^*; q) = G_0(\lambda^*) \leq \int_V p(x) \ln \dfrac{p^*(x)}{q(x)} dx$, for every feasible solution p of program (P) and we conclude that p^* is the optimal solution of program (P). ■

5 SOME SPECIAL CASES AND EXAMPLES

Case 1 If $I_1 = \{1, \ldots, n\}$, then it follows that the prior probability density function q is a feasible solution for primal program (P). Using the following well-known property of relative entropy (see [15], for example)

$$H(p;q) \geq 0 = H(q;q),\ \forall p \in \mathcal{M}(V, \nu),$$

we derive that the prior probability density function q is the optimal solution of program (P). In this case the optimal solution of program (D) is $\lambda^* = 0$.

Case 2 If $I_1 = \emptyset$, then the dual program (D_0) has the following form

$$(D_0):\ \left|\ \begin{array}{l} \max\limits_{\lambda \in \Lambda} G(\lambda) = \sum\limits_{i=1}^{n} \lambda_i c_i - \ln Z(\lambda) \quad \text{s.t.} \\ \lambda_i \geq 0,\ \forall i \in \{1, \ldots, n\}. \end{array}\right.$$

Case 3 If $I_1 = \{1, \ldots, n\} \setminus \{k\}$, where $k \in \{1, \ldots, n\}$, then the dual program (D_0) has the following form

$$(D_0):\ \left|\ \begin{array}{l} \max\limits_{\lambda \in \Lambda_0} G_0(\lambda) = \lambda c_k - \ln \int_V q(x) \exp[\lambda f_k(x)] dx \quad \text{s.t.} \\ \int_V q(x)[f_i(x) - c_i] \exp[\lambda f_k(x)] dx \geq 0,\ \forall i \in \{1, \ldots, n\} \setminus \{k\}, \end{array}\right.$$

where $\Lambda_0 = \left\{ \lambda \in \mathbb{R} \mid \int_V q(x) \exp[\lambda f_k(x)] dx < \infty \right\}$.

In this case we can remove the nonnegativity restriction $\lambda \geq 0$ of the dual program, this constraint being satisfied by the dual optimal solution.

Example 1 Let $V = [0, \theta]$ and consider as a prior distribution the *power distribution* ν of parameters α and θ, with the density function

$$q(x) = \alpha \theta^{-\alpha} x^{\alpha - 1},\ \forall x \in [0, \theta],$$

α and θ being positive real numbers. We remark that $E_\nu[\ln X] = -\frac{1}{\alpha} + \ln \theta$.

Consider the maxentropic reconstruction of probability distributions μ on $(V, \mathcal{B}(V))$ with density function $p \in \mathcal{M}(V, \nu)$ that are consistent with the following inequality constraints of mean values type

$$-\frac{1}{a} + \ln \theta \leq \int_0^\theta p(x) \ln x dx \leq -\frac{1}{b} + \ln \theta,$$

where the real constants a, b are given, $0 < a \leq b$. Now, the primal optimization problem (P) has the following form

$$(P_1):\ \left|\ \begin{array}{l} \min\limits_{p \in \mathcal{M}(V,\nu)} H(p;q) = \int_0^\theta p(x) \ln \frac{p(x)}{q(x)} dx \quad \text{s.t.} \\ \int_0^\theta p(x) \ln x dx \geq -\frac{1}{a} + \ln \theta,\ \int_0^\theta p(x)(-\ln x) dx \geq \frac{1}{b} - \ln \theta. \end{array}\right.$$

According to the above results, we have three situations.

Case i) If $\alpha \in [a,b]$, then $-\frac{1}{a} + \ln\theta \leq \int_0^\theta q(x)\ln x dx \leq -\frac{1}{b} + \ln\theta$, and hence the optimal solution of problem (P_1) is $p^* = q$.

Case ii) If $\alpha < a$, then $\int_0^\theta q(x)\ln x dx < -\frac{1}{a} + \ln\theta$ and $\int_0^\theta q(x)(-\ln x)dx \geq \frac{1}{b} - \ln\theta$. Therefore we obtain the following dual problem

$$(D_1): \max_{\lambda\in\Lambda_1} G_1(\lambda) = -\frac{\lambda}{a} + \ln\frac{\alpha+\lambda}{\alpha},$$

where $\Lambda_1 = (-\alpha, \infty)$. Obviously, this dual has the optimal solution $\lambda^* = a - \alpha$, and hence the primal problem (P_1) has the optimal solution

$$p^*(x) = \frac{q(x)\exp(\lambda^*\ln x)}{\displaystyle\int_0^\theta q(t)\exp(\lambda^*\ln t)dt} = a\theta^{-a}x^{a-1},\ \forall x \in [0,\theta],$$

i.e. the density of the power distribution of parameters a and θ.

Case iii) If $\alpha > b$, then in the same manner as in the previous case we obtain the dual problem $(\widetilde{D}_1): \max_{\lambda\in(-\infty,\alpha)} \widetilde{G}_1(\lambda) = \frac{\lambda}{b} + \ln\frac{\alpha-\lambda}{\alpha}$, with the optimal solution $\lambda^* = \alpha - b$, and the problem (P_1) has as optimal solution the density of the power distribution of parameters b and θ.

Example 2 Let $V = [\theta, \infty)$ and consider as a prior distribution the *Pareto distribution* ν of parameters (α, θ), with the density function

$$q(x) = \alpha\theta^\alpha x^{-\alpha-1},\ \forall x \in [\theta, \infty),$$

α and θ being positive real numbers. We remark that $E_\nu[\ln X] = \frac{1}{\alpha} + \ln\theta$.

Consider the maxentropic reconstruction of probability distributions μ on $(V, \mathcal{B}(V))$ with density function $p \in \mathcal{M}(V,\nu)$ that are consistent with the following inequality constraints of mean values type

$$\frac{1}{a} + \ln\theta \leq \int_\theta^\infty p(x)\ln x dx \leq \frac{1}{b} + \ln\theta,$$

where the real constants a, b are given, $0 < b \leq a$. Similarly to Example 1 we can obtain that the maxentropic distribution μ^* is also a Pareto distribution of parameters (α, θ), (a, θ), and (b, θ), accordingly as $\alpha \in [b, a]$, $\alpha > a$, and $\alpha < b$, respectively.

Remark 5 *For equality constraints, the power and the Pareto distributions can be directly obtained by using the maximum entropy principle (see [22]).*

BIBLIOGRAPHY

[1] BNOUHACHEM, A.; LIU, Z.B. (2004), *Alternating direction method for maximum entropy subject to simple constraints sets*, JOTA, 121, 259-277.

[2] BORWEIN, J.M.; LEWIS, A.S.; LIMBER, M.N.; NOLL, D. (1995), *Maximum entropy reconstruction using derivate information, Part 2: Computational results*, Numerische Mathematik, 69, 243-256.

[3] BORWEIN, J.M.; LEWIS, A.S.; NOLL, D. (1996), *Maximum entropy reconstruction using derivate information, Part 1: Fisher information and convex duality*, Mathematics of Operations Research, 21, 442-468.

[4] CSISZÁR, I. (1975), *I-divergence geometry of probability distributions and minimization problems*, Ann. Probability, 3, 146-158.

[5] DUMITRESCU, M. (1986), *The application of the principle of minimum cross-entropy to the characterization of the exponential-type probability distributions*, Ann. Inst. Stat. Math., 38, 451-457.

[6] ERLANDER, S. (1981), *Entropy in linear programs*, Mathematical Programming, 21, 137-151.

[7] FANG, S.-C.; RAJASEKERA, J.R; TSAO, H.-S.J. (1997), *Entropy Optimization and Mathematical Programming*, Kluwer Academic, Boston.

[8] GOLAN, A.; GZYL, H. (2002), *A generalized maxentropic inversion procedure for noisy data*, Appl. Math. Comput., 127, 249-260.

[9] GUIAŞU, S. (1990), *A classification of the main probability distributions by minimizing the weighted logarithmic measure of deviation*, Ann. Inst. Statist. Math., 42, 269-279.

[10] GUIAŞU, S. (2001), *Quantum Mechanics*, Nova Science Publishers, New York.

[11] GZYL, H. (2000), *Maxentropic reconstruction of some probability distributions*, Studies in Appl. Math., 105, 235-243.

[12] GZYL, H. (2002), *Ill-posed linear inverse problems and maximum entropy in the mean*, Acta Cientifica Venezolana, 53, 74-93.

[13] GZYL, H. (2002), *Tomographic reconstruction by maximum entropy in the mean: unconstrained reconstructions*, Appl. Math. Comput., 129, 157-169.

[14] GZYL, H.; VELÁSQUEZ, Y. (2002), *Reconstruction of transition probabilities by maximum entropy in the mean*, Bayes. Infer. and M.E.M. in Sci. and Eng., Am. Inst. of Phys., 617, 192-203.

[15] IHARA, S. (1993), *Information Theory for Continuous Systems*, World Scientific, Singapore.

[16] INGARDEN, R.S.; KOSSAKOWSKI, A. (1971), *Poisson probability distribution and information thermodynamics*, Bull. Acad. Polon. Sci. Ser. Math., 19, 83-86.

[17] JAYNES, E.T. (1957), *Information theory and statistical mechanics*, Phys. Rev., 106, 620-630.

[18] JAYNES, E.T. (1982), *On the rationale of maximum entropy methods*, IEE Proceedings, 70, 939-952.

[19] KULLBACK, S. (1959), *Information Theory and Statistics*, Wiley, New York.

[20] PETERSON, E.L. (1976), *Geometric programming*, SIAM Review, 19, 1-45.

[21] PREDA, V. (1982), *The Student distribution and the principle of maximum entropy*, Ann. Inst. Statist. Math., 34, 335-338.

[22] PREDA, V. (1984), *Informational characterizing of the Pareto and power distributions*, Bull. Math. Soc. Sci. Math. Repub. Soc. Roum., 28, 343-346.

[23] PREDA, V.; BĂLCĂU, C. (2003), *On maxentropic reconstruction of countable Markov chains and matrix scaling problems*, Studies in Appl. Math., 111, 85-100.

[24] ROCKAFELLAR, R.T. (1970), *Convex Analysis*, Princeton Univ. Press., New Jersey.

[25] WU, Z.; PHILLIPS, G.; TAPIA, R.; ZHANG, Y. (2001), *A fast Newton algorithm for entropy maximization in phase determination*, SIAM Review, 43, 623-642.

[26] YANG, H.; WONG, S.C. (1999), *The most likely equilibrium traffic queuing pattern in a capacity constrained network*, Annals of Operations Research, 87, 73-85.