JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 185, 596–604 (1994)

Moment-Matching and Best Entropy Estimation

PETER BORWEIN*

Department of Mathematics, Statistics and C.S., Dalhousie University, Halifax, N.S., Canada B3H 3J5

AND

ADRIAN S. LEWIS*,†

Department of Combinatorics and Optimization University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Submitted by Augustine O. Esogbue

Received April 15, 1991

Given the first n moments of an unknown function $\bar{x}$ on the unit interval, a common estimate of $\bar{x}$ is $\psi(\pi_n)$, where π_n is a polynomial of degree n taking values in a prescribed interval, ψ is a given monotone function, and π_n is chosen so that the moments of $\psi(\pi_n)$ equal those of $\bar{x}$. This moment-matching procedure is closely related to best entropy estimation of $\bar{x}$: two classical cases arise when ψ is the exponential function (corresponding to the Boltzmann–Shannon entropy) and the reciprocal function (corresponding to the Burg entropy). General conditions ensuring the existence and uniqueness of π_n are given using convex programming duality techniques, and it is shown that the estimate $\psi(\pi_n)$ converges uniformly to $\bar{x}$ providing $\bar{x}$ is sufficiently smooth.

1. INTRODUCTION

Suppose $\bar{x}$ is an unknown function on $[0, 1]$ taking values in some prescribed interval $[\alpha, \beta]$ (possibly infinite). We wish to estimate $\bar{x}$ on the basis of its first n moments, $\int_0^1 s^i x(s)\, ds$, for $i = 0, \ldots, n$. Such problems

* Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

† email: aslewis @ orion.uwaterloo.ca

0022-247X/94 $6.00

and their trigonometric and multidimensional analogues serve as a model for a wide variety of physical measurement problems. One extremely popular approach may be termed best entropy estimation: the estimate is selected to be that function which minimizes a certain measure of entropy, $\int_0^1 \phi(x(s))\,ds$, subject to the given moment constraints. The two classical choices have been the Boltzmann–Shannon and Burg entropies (see [13] and [14]), the choice between them being very controversial. For surveys, see [12] and [16] (containing in total almost 700 references). More recently, other entropy measures have been suggested, including the L_2 entropy [10], [11], and general families of entropies proposed in [20], [19], and by an elegant probabilistic discussion in [5].

One important criterion for comparing the various choices of entropy is the convergence of the estimates of $\bar{x}$ as the number of given moments grows (see for example [24]). Convergence in various senses of the Boltzmann–Shannon estimates is studied in [18], [9], [7], [3], and [1]. For more general entropies, convergence questions are discussed in [15] and [5], and a convergence result for the Burg entropy appears in [8] based on earlier work of [23].

The aim of this paper is to give general conditions which ensure uniform convergence of the best entropy estimates to the unknown function $\bar{x}$. The approach (see Section 2) is to apply a simple approximation-theoretic argument to the moment-matching procedure described in the abstract to deduce uniform convergence from interpolation properties of the estimates. This makes it clear, for example, that the uniform convergence of the Burg entropy estimates to a sufficiently smooth strictly positive function $\bar{x}$ is not "merely a fortunate accident," as claimed in the conclusions of [24].

The equivalence of the moment-matching procedure to best entropy estimation has been observed widely for special cases in the applied literature, the (loose) justification generally being via Lagrange multipliers attached to the moment constraints. However, the existence question for the polynomials π_n is quite delicate. Section 3 is devoted to a rigorous explanation based on convex programming duality.

2. Moment-Matching

Throughout this paper we shall assume $-\infty \le p < q \le +\infty$, $-\infty \le \alpha < \beta \le +\infty$, and $\psi\colon (p, q) \to (\alpha, \beta)$ is a continuously invertible, strictly increasing function satisfying

$$\liminf_{r \uparrow q} (q - r)\psi(r) > 0, \text{ if } q < +\infty, \text{ and} \tag{1}$$

$$\limsup_{r \downarrow p} (r - p)\psi(r) < 0, \text{ if } p > -\infty. \tag{2}$$

(In particular, $\beta = +\infty$ if $q < +\infty$, and $\alpha = -\infty$ if $p > -\infty$.) Two cases of particular interest are

$$p = -\infty,\ q = +\infty,\ \alpha = 0,\ \beta = +\infty, \text{ and } \psi(r) = e^r; \tag{3}$$

$$p = -\infty,\ q = 0,\ \alpha = 0,\ \beta = +\infty, \text{ and } \psi(r) = -1/r. \tag{4}$$

We consider a function $\bar{x}$ in $L_1[0, 1]$ satisfying

$$\alpha \le \bar{x}(s) \le \beta, \text{ almost everywhere, and} \tag{5}$$

$$\bar{x}(s) \in (\alpha, \beta), \text{ on a set of positive measure.} \tag{6}$$

We wish to match the moments of $\bar{x}$ by the image of a polynomial under ψ. The result below, which we prove in the following section, shows that under the above conditions this problem has a unique solution.

THEOREM 2.1 *There exists a unique polynomial π_n of degree n which satisfies*

$$p < \pi_n(s) < q, \text{ for all } s \text{ in } [0, 1], \text{ and} \tag{7}$$

$$\int_0^1 \psi(\pi_n(s)) s^i \, ds = \int_0^1 \bar{x}(s)\, s^i \, ds, \text{ for } i = 0, \ldots, n. \tag{8}$$

LEMMA 2.2 *Suppose $\bar{x}$ is piecewise continuous and π_n is the polynomial of Theorem* 2.1. *Then $\pi_n - \psi^{-1}(\bar{x})$ has at least $(n + 1)$ sign-changes, or is identically zero.*

Proof. (We interpret $\psi^{-1}(\alpha) := p$ and $\psi^{-1}(\beta) := q$.) If $\pi_n - \psi^{-1}(\bar{x})$ has at most n changes of sign, so does $\psi(\pi_n) - \bar{x}$, so we can choose a nonzero polynomial ξ of degree n, with the same sign. However, (8) implies

$$\int_0^1 (\psi(\pi_n(s)) - \bar{x}(s))\, \xi(s) \, ds = 0,$$

whence the result. ∎

For any $d \ge 0$, let $R_d \subset \mathbb{C}$ denote those complex numbers whose distance from $[0, 1]$ is no larger than d.

THEOREM 2.3 *Suppose $\bar{y}$ is an analytic function on R_d for some $d > 1$, and satisfies*

$$p < \bar{y}(s) < q, \text{ for all } s \text{ in } [0, 1]. \tag{9}$$

Then if $\bar{x} = \psi(\bar{y})$, the polynomials π_n of Theorem 2.1 *converge uniformly to $\bar{y}$.*

Proof. Lemma 2.2 shows that π_n interpolates $\bar{y}$ at $(n+1)$ points, and if $\bar{y}$ is analytic on R_d, any such interpolation scheme converges uniformly (see [6], Section 4.3, Example 1). (Clearly (9) implies that $\bar{x}$ satisfies (6).) ∎

Under the above conditions we thus see that the reconstructions $\psi(\pi_n)$ in Theorem 2.1 converge uniformly to $\bar{x}$. Case (3) above therefore shows that if we match the moments of $e^{\bar{y}}$ with e^{π_n}, these estimates converge uniformly if $\bar{y}$ is analytic on R_d with $d > 1$, while case (4) shows the same result if we match the moments of $1/\bar{y}$ with $1/\pi_n$ (assuming $\bar{y}$ is strictly positive on [0, 1]).

In fact, as we shall see in the next section, if we drop the uniqueness requirement on π_n, we can weaken the assumptions on ψ in Theorem 2.1, simply requiring that it be continuous and increasing (not necessarily strictly). An interesting case is

$$p = -\infty,\ q = +\infty,\ \alpha = 0,\ \beta = +\infty, \text{ and } \psi(r) = r^+, \tag{10}$$

where r^+ denotes the positive part of r. Theorem 2.1 then shows that if $\bar{x}$ is non-negative and not identically zero then there is a polynomial π_n of degree n whose positive part matches the first n moments of $\bar{x}$ (c.f. [10]). The argument of Lemma 2.2 then shows that $\pi_n^+ - \bar{x}$ changes sign at $(n+1)$ points if $\bar{x}$ is continuous, so in fact π_n interpolates $\bar{x}$ at $(n+1)$ points. Thus if $\bar{x}$ is actually analytic on R_d for some $d > 1$ and non-negative, then π_n (and therefore π_n^+) converges uniformly to $\bar{x}$. The same argument works with $\psi(r) = (r^+)^\gamma$, for any $\gamma > 0$.

One way to see that a uniform convergence result like Theorem 2.3 is not surprising is to consider the case where ψ is the identity map on $\mathbb{R} = (\alpha, \beta)$. In this case the polynomials π_n are simply the partial sums of the expansion of $\bar{x}$ in appropriate orthogonal polynomials on [0, 1]: these are well-known to converge uniformly to sufficiently smooth $\bar{x}$ (see for example [6]).

3. Best Entropy Estimation

In this section we derive the connection between best entropy estimation and the moment-matching problems of the previous section. We will use the convex analysis notation of [22]. We suppose $\phi : \mathbb{R} \to (-\infty, +\infty]$ is a closed convex function whose domain has non-empty interior (α, β). The conjugate function ϕ^* is defined by

$$\phi^*(v) := \sup_u \{uv - \phi(u)\},$$

and is continuously differentiable on the interior of its domain (p, q) (which is non-empty). We assume

$$\liminf_{r \uparrow q} (q - r)(\phi^*)'(r) > 0, \text{ if } q < +\infty, \text{ and} \tag{11}$$

$$\limsup_{r \downarrow p} (r - p)(\phi^*)'(r) < 0, \text{ if } p > -\infty, \tag{12}$$

which are simply (1) and (2) with $\psi := (\phi^*)'$. The special cases (3), (4), and (10) correspond to the Boltzmann–Shannon entropy,

$$\phi(u) := \begin{cases} u \log u - u, & \text{if } u > 0, \\ 0, & \text{if } u = 0, \\ +\infty, & \text{if } u < 0, \end{cases} \tag{13}$$

the Burg entropy,

$$\phi(u) := \begin{cases} -\log u, & \text{if } u > 0, \\ +\infty, & \text{if } u \leq 0, \end{cases} \tag{14}$$

and the L_2 entropy,

$$\phi(u) := \begin{cases} u^2/2, & \text{if } u \geq 0, \\ +\infty, & \text{if } u < 0, \end{cases} \tag{15}$$

respectively.

Best entropy estimation seeks to estimate the L_1 function $\bar{x}$, which we suppose satisfies (5) and (6), by that L_1 function x which matches the first n moments of $\bar{x}$ and minimizes the entropy functional $\int_0^1 \phi(x(s))\, ds$ (well-defined as a "normal convex integral" [21]). The optimization problem we therefore consider is

$$(BE_n) \begin{cases} \inf & \int_0^1 \phi(x(s))\, ds \\ \text{subject to} & \int_0^1 x(s)s^i\, ds = \int_0^1 \bar{x}(s)s^i\, ds,\ i = 0, \ldots, n, \\ & x \in L_1[0, 1]. \end{cases}$$

As usual in convex optimization there is a natural dual problem:

$$(BE_n^*) \begin{cases} \sup & \int_0^1 (\bar{x}(s)\pi(s) - \phi^*(\pi(s)))ds \\ \text{with} & \pi \text{ a polynomial of degree} \leq n. \end{cases}$$

THEOREM 3.1. *(Strong Duality) The problems* (BE_n) *and* (BE_n^*) *have equal, finite value, which is attained in both problems. If* π_n *is any optimal solution of* (BE_n^*) *then* $p < \pi_n(s) < q$ *for all* s *in* $[0, 1]$ *and the unique optimal solution of* (BE_n) *is* $x_n(s) := (\phi^*)'(\pi_n(s))$. *If* ϕ^* *is strictly convex on* (p, q) *then* π_n *is also unique.*

Proof. The case $\alpha = 0$ and $\beta = +\infty$ is a special case of results in [2]. The argument for this case is analogous: a general result may be found in [4]. The necessary constraint qualification is ensured by (5) and (6). ▮

The special cases of the above result when ϕ is given by (13), (14), and (15) are well-known. The Boltzmann–Shannon and L_2 cases are covered in [3] (see also [10]). For the Burg case see for example [14], [13], and [17].

Using this strong duality theorem we can now prove Theorem 2.1:

Proof of Theorem 2.1. Choose any r_0 in (p, q) and define a function $\gamma : \mathbb{R} \to (-\infty, +\infty]$ by $\gamma(v) := \int_{r_0}^{v} \psi(r)dr$. Then γ is strictly convex on (p, q) with $\gamma' = \psi$, so γ is also essentially smooth in the sense of [22]. If we now define $\phi := \gamma^*$ then ϕ is strictly convex with $\phi^* = \gamma$ [22]. Theorem 3.1 now shows the existence of π_n satisfying (7) and (8).

Finally, any π_n satisfying (7) and (8) is optimal for (BE_n^*) as may be seen by differentiating with respect to each coefficient of π_n. The uniqueness follows. This could also be proved more directly by the argument of Lemma 2.2. ▮

Suppose, in addition to our previous assumptions, that ϕ is essentially smooth (which is equivalent to ϕ^* being strictly convex on (p, q)), as is the case for the Boltzmann–Shannon and Burg entropies. Theorems 2.3 and 3.1 then show that the optimal solution x_n of the best entropy estimation problem (BE_n) converges uniformly to $\bar{x}$ providing $((\phi^*)')^{-1}(\bar{x})$ is analytic on R_d for some $d > 1$. The argument at the end of the previous section shows the same for the L_2 entropy providing $\bar{x}$ is analytic on R_d for some $d > 1$ and non-negative.

4. THE TRIGONOMETRIC CASE

Theorem 3.1 is a special case of a much more general result (see [4]). For example, all of the results in this paper will remain true if we replace Lebesgue measure throughout by any positive regular Borel measure on $[0, 1]$, providing it dominates a positive multiple of Lebesque measure. Furthermore there is nothing special about the functions with respect to which we take the moments: we could replace $1, s, \ldots, s^n$ with other Lipschitz functions in Theorem 3.1 (see [4]).

Probably the most important special case in practice is the trigonometric moment problem, as [12] and [16] will testify. The moment constraints are then Fourier coefficiants—moments with respect to cost $i\theta$ and sin $i\theta$ on $[-\pi, \pi]$. In this case, due to the periodicity, we can weaken the conditions on ϕ required for attainment ((11) and (12)) to

$$\liminf_{r \uparrow q} (q - r)^{1/2}(\phi^*)'(r) > 0, \text{ if } q < +\infty, \text{ and} \tag{16}$$

$$\limsup_{r \downarrow p} (r - p)^{1/2}(\phi^*)'(r) < 0, \text{ if } p > -\infty. \tag{17}$$

The same type of convergence we have seen for algebraic moment problems will occur for trigonometric problems. To illustrate, let us assume for simplicity that the function $\bar{x}$ we seek to estimate is even on $[-\pi, \pi]$ (and still satisfies (5) and (6)), so our problem becomes

$$(TE_n) \begin{cases} \inf & \int_{-\pi}^{\pi} \phi(x(\theta))\, d\theta \\ \text{subject to} & \int_{-\pi}^{\pi} x(\theta) \cos(i\theta)\, d\theta = \int_{-\pi}^{\pi} \bar{x}(\theta) \cos(i\theta)\, d\theta, \\ & \qquad\qquad \text{for } i = 0, \ldots, n, \\ & 0 \le x \in L_1[-\pi, \pi], \end{cases}$$

with dual

$$(TE_n^*) \begin{cases} \sup & \int_{-\pi}^{\pi} (\bar{x}(\theta)\omega(\cos\theta) - \phi^*(\omega(\cos\theta)))\, d\theta \\ \text{with} & \omega \text{ a polynomial of degree} \le n. \end{cases}$$

(Note that any linear combination of $1, \cos\theta, \ldots, \cos n\theta$ is a polynomial of degree n in $\cos\theta$, and vice versa.)

THEOREM 4.1 *Assuming* (16) *and* (17) *hold (in place of* (11) *and* (12)), *the problems* (TE_n) *and* (TE_n^*) *have equal, finite value, which is attained in both problems. If* ω_n *is any optimal solution of* (TE_n^*) *then* $p < \omega_n(\cos\theta) < q$ *for all* θ *in* $[-\pi, \pi]$ *and the unique optimal solution of* (TE_n) *is* $x_n(\theta) := (\phi^*)'(\omega_n(\cos\theta))$. *If* ϕ^* *is strictly convex on* (p, q) *then* ω_n *is also unique.*

Proof. See [4]. ∎

In an exactly analogous fashion to the previous sections, ω_n will be the unique solution of the moment-matching problem

$$p < \omega(\cos\theta) < q, \qquad \text{for all } \theta \text{ in } [-\pi, \pi], \text{ and}$$

$$\int_{-\pi}^{\pi} \psi(\omega(\cos\theta)) \cos(i\theta)\, d\theta = \int_{-\pi}^{\pi} \bar{x}(\theta) \cos(i\theta)\, d\theta, \text{ for } i = 0, \ldots, n,$$

if $\psi = (\phi^*)'$ and ϕ^* is strictly convex. The same argument as Lemma 2.2 shows that $\omega_n(\cos\theta) - \psi^{-1}(\bar{x}(\theta))$ has at least $(n+1)$ sign changes in $(0, \pi)$ (using polynomials in $\cos\theta$ rather than s). Thus, using the change of variables $t := \cos\theta$, $\omega_n(\cdot)$ interpolates $\psi^{-1}(\bar{x}(\cos^{-1}(\cdot)))$ $(n+1)$ times in $(-1, 1)$, and hence converges uniformly to it providing $\psi^{-1}(\bar{x}(\cos^{-1}(\cdot)))$ is analytic on R_d', the set of complex numbers whose distance fom $[-1, 1]$ is no larger than d, for some $d > 2$.

REFERENCES

1. J. M. BORWEIN AND A. S. LEWIS, Convergence of best entropy estimates, *SIAM J. Optim.* **1** (1991), 191–205.
2. J. M. BORWEIN AND A. S. LEWIS, Duality relationships for entropy-like minimization problems, *SIAM J. Control Optim.* **29** (1991), 325–338.
3. J. M. BORWEIN AND A. S. LEWIS. On the convergence of moment problems, *Trans. Amer. Math. Soc.* **325** (1991), 249–271.
4. J. M. BORWEIN AND A. S. LEWIS, Partially-finite programming in L_1 and the existence of maximum entropy estimates, *SIAM J. Optim.* **3,** 1993, 248–267.
5. D. DACUNHA-CASTELLE AND F. GAMBOA, Maximum d'entropie et problème des moments, *Ann. Inst. H. Poincaré* **26** (1990), 567–596.
6. P. J. DAVIS, "Interpolation and Approximation," Dover, New York, 1975.
7. B. FORTE, W. HUGHES, AND Z. PALES, Maximum entropy estimators and the problem of moments, *Rend. Mat. Appl. (7)* **9** (1989), 689–699.
8. F. GAMBOA, "Methode du Maximum d'Entropie sur la Moyenne et Applications." Ph.D. thesis, Universite Paris Sud, Centre d'Orsay, 1989.
9. E. GASSIAT, Probléme sommatoire par maximum d'entropie, *C. R. Acad. Sci. Paris Sér. I Math* **303** (1986). 675–680.
10. R. K. GOODRICH AND A. STEINHARDT, L_2 spectral estimation, *SIAM J. Appl. Math.* **46** (1986), 417–428.
11. L. K. JONES AND V. TRUTZER, Computationally feasible high-resolution minimum distance procedures which extend the maximum-entropy method, *Inverse Problems* **5** (1989), 749–766.
12. S. M. KAY AND S. L. MARPLE, Spectrum analysis—a modern perspective. *IEEE Proc.* **69** (1981), 1380–1419.
13. S. W. LANG AND J. H. MCCLELLAN, Spectral estimation for sensor arrays, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-31** (1983), 349–358.
14. S. W. LANG AND J. H. MCCLELLAN, Multidimensional MEM spectral estimation. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-30** (1984), 880–887.
15. A. S. LEWIS, The convergence of entropic estimates for moment problems, *In* "Workshop/Miniconference on Functional Analysis/Optimization" (S. Fitzpatrick and J. Giles, Eds.), pp. 100–115, Centre for Mathematical Analysis, Australian National University, Canberra, 1989.
16. D. M. LIN AND E. K. WONG, A survey on the maximum entropy method and parameter spectral estimation. *Phys. Rep.* **193** (1990), 41–135.
17. J. H. MCCLELLAN AND S. W. LANG, Multi-dimensional MEM spectral estimation, *in* "Spectral Analysis and Its Use in Underwater Acoustics," pp. 10.1–10.8, Imperial College, Institute of Acoustics. London, 1982.
18. L. R. MEAD AND N. PAPANICOLAOU, Maximum entropy in the problem of moments, *J. Math. Phys.* **25** (1984), 2404–2417.

19. J. NAVAZA, The use of non-local constraints in maximum-entropy electron density reconstruction, *Acta Cryst. Sect. A* **42** (1986), 212–223.
20. R. NITYANANDA AND R. NARAYAN, Maximum entropy image reconstruction—a practical non-information-theoretic approach, *J. Astrophys. Astronomy* **3** (1982), 419–450.
21. R. T. ROCKAFELLAR, Integrals which are convex functionals, *Pacific. J. Math.* **24** (1968), 525–539.
22. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, N.J., 1970.
23. A. SEGHIER, Reconstruction de la densité spectrale par maximum d'entropie cas d-dimensionnel, *C. R. Acad. Sci. Paris, Sér. I Math* **305** (1987), 517–520.
24. J. SKILLING AND S. F. GULL, The entropy of an image, *SIAM–AMS Proc.* **14** (1984), 167–189.