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 $\pi_n$ is a polynomial of degree $n$ taking values
in a prescribed interval, $\psi$ is a given monotone function, and $\pi_n$ 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 $\psi$ 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 $\pi_n$ 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. © 1994 Academic Press, Inc.

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 Coun-
cil of Canada.
† email: aslewis @ orion.uwaterloo.ca

0022-247X/94 $6.00$
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

596
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 \( \tilde{x} \) 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 \( \tilde{x} \). 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 \( \tilde{x} \) 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 \( \pi_n \) 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 \(-\infty \leq p < q \leq +\infty, -\infty \leq \alpha < \beta \leq +\infty \), and \( \psi: (p, q) \rightarrow (\alpha, \beta) \) is a continuously invertible, strictly increasing function satisfying

\[
\lim \inf_{r \uparrow q} (q - r)\psi(r) > 0, \text{ if } q < +\infty, \quad (1)
\]

\[
\lim \sup_{r \downarrow p} (r - p)\psi(r) < 0, \text{ if } p > -\infty. \quad (2)
\]
(In particular, $\beta = +\infty$ if $q < +\infty$, and $\alpha = -\infty$ if $p > -\infty$.) Two cases of particular interest are
\begin{align*}
p = -\infty, & \quad q = +\infty, \quad \alpha = 0, \quad \beta = +\infty, \quad \text{and} \quad \psi(r) = e^r; \quad (3) \\
p = -\infty, & \quad q = 0, \quad \alpha = 0, \quad \beta = +\infty, \quad \text{and} \quad \psi(r) = -1/r. \quad (4)
\end{align*}

We consider a function $\bar{x}$ in $L_1[0, 1]$ satisfying
\begin{align*}
\alpha \leq \bar{x}(s) \leq \beta, & \quad \text{almost everywhere, and} \quad (5) \\
\bar{x}(s) \in (\alpha, \beta), & \quad \text{on a set of positive measure.} \quad (6)
\end{align*}

We wish to match the moments of $\bar{x}$ by the image of a polynomial under $\psi$. 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 $\pi_n$ of degree $n$ which satisfies
\begin{align*}
p < \pi_n(s) < q, & \quad \text{for all } s \in [0, 1], \text{ and} \quad (7) \\
\int_0^1 \psi(\pi_n(s))s^i \, ds = \int_0^1 \bar{x}(s) \, s^i \, ds, & \quad \text{for } i = 0, \ldots, n. \quad (8)
\end{align*}

**Lemma 2.2** Suppose $\bar{x}$ is piecewise continuous and $\pi_n$ 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 $\xi$ of degree $n$, with the same sign. However, (8) implies
\begin{equation*}
\int_0^1 \left( \psi(\pi_n(s)) - \bar{x}(s) \right) \xi(s) \, ds = 0,
\end{equation*}
whence the result. \hfill \Box

For any $d \geq 0$, 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
\begin{equation*}
p < \bar{y}(s) < q, \quad \text{for all } s \in [0, 1]. \quad (9)
\end{equation*}

Then if $\bar{x} = \psi(\bar{y})$, the polynomials $\pi_n$ of Theorem 2.1 converge uniformly to $\bar{y}$.
Proof. Lemma 2.2 shows that \( \pi_n \) interpolates \( \bar{y} \) at \( (n + 1) \) points, and if \( \bar{y} \) is analytic on \( \mathbb{R}_d \), 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 \( e^{\pi_n} \), these estimates converge uniformly if \( \bar{y} \) is analytic on \( \mathbb{R}_d \) 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 \( \pi_n \), we can weaken the assumptions on \( \psi \) in Theorem 2.1, simply requiring that it be continuous and increasing (not necessarily strictly). An interesting case is

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

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 \( \pi_n \) 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 \( \pi_n^\alpha - \bar{x} \) changes sign at \( (n + 1) \) points if \( \bar{x} \) is continuous, so in fact \( \pi_n \) interpolates \( \bar{x} \) at \( (n + 1) \) points. Thus if \( \bar{x} \) is actually analytic on \( \mathbb{R}_d \) for some \( d > 1 \) and non-negative, then \( \pi_n \) (and therefore \( \pi_n^\alpha \)) 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 \( \psi \) is the identity map on \( \mathbb{R} = (\alpha, \beta) \). In this case the polynomials \( \pi_n \) 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 \((\alpha, \beta)\). The conjugate function \( \phi^* \) is defined by

\[
\phi^*(u) := \sup_u \{uv - \phi(u)\},
\]
and is continuously differentiable on the interior of its domain \((p, q)\) (which is non-empty). We assume

\[
\lim \inf_{r \uparrow q} (q - r)(\phi^*)(r) > 0, \text{ if } q < +\infty, \text{ and}
\]

\[
\lim \sup_{r \downarrow p} (r - p)(\phi^*)(r) < 0, \text{ if } p > -\infty,
\]

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

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

the Burg entropy,

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

and the \(L_2\) entropy,

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

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

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

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 \(\pi_n\) 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 \(\phi^*\) is strictly convex on \((p, q)\) then \(\pi_n\) 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 \(\phi\) 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 [4, 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 \(\gamma\) is strictly convex on \((p, q)\) with \(\gamma' = \psi\), so \(\gamma\) is also essentially smooth in the sense of [22]. If we now define \(\phi := \gamma^*\) then \(\phi\) is strictly convex with \(\phi^* = \gamma\) [22]. Theorem 3.1 now shows the existence of \(\pi_n\) satisfying (7) and (8).

Finally, any \(\pi_n\) satisfying (7) and (8) is optimal for \((BE^*_n)\) as may be seen by differentiating with respect to each coefficient of \(\pi_n\). 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 \(\phi\) is essentially smooth (which is equivalent to \(\phi^*\) 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 Lebesgue 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 coefficients—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 $\phi$ required for attainment ((11) and (12)) to

$$\lim_{r \downarrow q} \inf (q - r)^{1/2}(\phi^*)'(r) > 0, \text{ if } q < +\infty, \text{ and}$$

$$\lim_{r \downarrow p} \sup (r - p)^{1/2}(\phi^*)'(r) < 0, \text{ if } p > -\infty.$$  \hspace{1cm} (16) \hspace{1cm} (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 $\tilde{x}$ we seek to estimate is even on $[-\pi, \pi]$ (and still satisfies (5) and (6)), so our problem becomes

$$(TE_n) \left\{ \begin{array}{l} \inf \int_{-\pi}^{\pi} \phi(x(\theta)) \, d\theta \\ \text{subject to} \\ \int_{-\pi}^{\pi} x(\theta) \cos (i\theta) \, d\theta = \int_{-\pi}^{\pi} \tilde{x}(\theta) \cos (i\theta) \, d\theta, \\ \text{for } i = 0, \ldots, n, \\ 0 \leq x \in L_1[-\pi, \pi], \end{array} \right.$$  \hspace{1cm} (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.)

$$(TE_n^*) \left\{ \begin{array}{l} \sup \int_{-\pi}^{\pi} (\tilde{x}(\theta)\omega(\cos \theta) - \phi^*(\omega(\cos \theta))) \, d\theta \\ \text{with} \omega \text{ a polynomial of degree } \leq n. \end{array} \right.$$  \hspace{1cm} (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 $\omega_n$ is any optimal solution of $(TE_n^*)$ then $p < \omega_n(\cos \theta) < q$ for all $\theta$ in $[-\pi, \pi]$ and the unique optimal solution of $(TE_n)$ is $x_n(\theta) := (\phi^*)'(\omega_n(\cos \theta))$. If $\phi^*$ is strictly convex on $(p, q)$ then $\omega_n$ is also unique.

Proof. See [4]. \hspace{1cm} \Box

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

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

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

References


