A characterization of probabilities with full support in metric spaces, and Laplace’s method

Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci

Working Paper n. 620

This Version: February 21, 2018
A characterization of probabilities with full support in metric spaces, and Laplace’s method

Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci

Università Bocconi and IGIER

February 21, 2018

Abstract

We show that a probability measure on a metric space $X$ has full support if and only if the set of all probability measures that are absolutely continuous with respect to it is dense in $\mathcal{P}(X)$. We illustrate the result through a general version of Laplace’s method, which in turn leads to a general stochastic convergence result to global maxima.

1 Introduction

A probability measure $\lambda$ on a finite space $X$ has full support if and only if the set $\mathcal{P}_\lambda(X)$ of all probability measures that are absolutely continuous with respect to $\lambda$ coincides with the set $\mathcal{P}(X)$ of all probability measures. Here we show that, on a metric space $X$, a probability measure $\lambda$ has full support if and only if $\mathcal{P}_\lambda(X)$ is dense in $\mathcal{P}(X)$.

Since the assumption of full support is equivalent to strict positivity on nonempty open sets, our result sheds light on the notion of strict positivity of a probability measure in the infinite case. In a functional analysis perspective, it can be regarded as a characterization of strictly positive continuous linear functionals in the dual pair $\langle C_b(X), ca(X) \rangle$.

To illustrate this result, we prove a general version of Laplace’s method. Specifically, if $\lambda$ is a full support measure on the compact metrizable set $K$ and $u \in C(K)$, then

$$v_n = \frac{1}{n} \log \int_K e^{nu(x)} d\lambda(x) \to v = \max_{x \in K} u(x)$$

By variational methods we show that, when the maximizer $x^u$ of $u$ on $K$ is unique, to the sequence $\{v_n\}$ corresponds a sequence $\{\mu_n\}$ of measures on $K$ that, eventually, concentrates on $x^u$. Moreover, if $K$ is contained in a reflexive and separable Banach space, the sequence of the barycenters of $\mu_n$ weakly converges to $x^u$. 


2 Setup and preliminaries

Let \( X \) be a topological space. We denote by \( C(X) \) (resp., \( C_b(X) \)) the vector space of all continuous (resp., continuous and bounded) functions \( f : X \to \mathbb{R} \), by \( \mathcal{B}(X) \) the Borel sigma-algebra of \( X \), and by \( \mathcal{P}(X) \) the set of all Borel probability measures on \( \mathcal{B}(X) \) with the topology \( \sigma(\mathcal{P}(X), C_b(X)) \) of weak convergence.

Given any \( \lambda \in \mathcal{P}(X) \), we denote by \( \mathcal{P}_{\lambda}(X) \) (resp., \( \mathcal{P}_{\lambda}^*(X) \)) the collection of all \( \mu \in \mathcal{P}(X) \) that are absolutely continuous with respect to \( \lambda \) (resp., that have continuous and bounded density with respect to \( \lambda \)), and by \( \ell_\lambda : C_b(X) \to \mathbb{R} \) the positive linear functional \( \ell_\lambda(f) = \int_X f d\lambda \).

**Definition 1** The *support* of \( \lambda \in \mathcal{P}(X) \), denoted by \( \text{supp} \lambda \), is (if it exists) a closed subset of \( X \) with \( \lambda \)-null complement and such that \( \lambda(G) > 0 \) for all open subsets \( G \) of \( X \) having nonempty intersection with it.

The probability measure \( \lambda \) has *full support* if \( \text{supp} \lambda = X \), that is, \( \lambda(G) > 0 \) for all nonempty open subsets \( G \) of \( X \).

If \( X \) is the dual of a separable normed space (for example, a reflexive and separable Banach space), we endow it with the weak* topology and consider the Borel sigma-algebra generated by this topology. With this topology, compact sets are metrizable and their closed and convex hulls are compact too.\(^1\) The next basic result is a slight modification of Proposition 1.1 of Phelps (2001).

**Proposition 1** If \( \mu \in \mathcal{P}(X) \) has bounded support, then there exists a unique element \( m \in X \) such that

\[
\langle \phi, m \rangle = \int_X \langle \phi, x \rangle d\mu(x)
\]

for all linear and continuous functionals \( \phi : X \to \mathbb{R} \).

The element \( m \), called *barycenter* of \( \mu \), belongs to the closed and convex hull of \( \text{supp} \mu \). When \( X \) is \( \mathbb{R}^n \), the barycenter of a Borel probability measure \( \mu \) on \( \mathbb{R}^n \) that has bounded support is easily seen to be the vector \( m = \int_X x d\mu(x) \).

3 Main result

We state and prove our main result. The equivalence between points (i) and (iv), i.e., between the strict positivity of \( \lambda \) and \( \ell_\lambda \), is essentially known and reported here for completeness and perspective.

\(^1\)Because of the Alaoglu's Theorem and of Theorem 6.30 of Aliprantis and Border, 2006 (henceforth, AB).
Theorem 1 Let $X$ be a metric space. The following conditions are equivalent for $\lambda \in \mathcal{P}(X)$:

(i) $\lambda$ has full support $X$;

(ii) $\text{cl} \left( \mathcal{P}_\lambda^*(X) \right) = \mathcal{P}(X)$;

(iii) $\text{cl} \left( \mathcal{P}_\lambda(X) \right) = \mathcal{P}(X)$;

(iv) $\ell_\lambda$ is strictly positive, i.e., $\int_X f d\lambda > 0$ for all $0 \neq f \in C_b^+(X)$.

Proof If $X$ is a singleton, the statement is trivial. Let us assume that $X$ contains more than one point.

(i) implies (ii). We first show that $\delta_x \in \text{cl} \left( \mathcal{P}_\lambda^*(X) \right)$ for all $x \in X$. Let $x \in X$, and, for each $n \in \mathbb{N}$, consider the sets $B_n$ and $C_n$ defined by

$$B_n = \left\{ x \in X : d(x, x) \leq \frac{1}{n} \right\} \text{ and } C_n = \left\{ x \in X : d(x, x) \geq \frac{2}{n} \right\}.$$ 

Both sets are closed and clearly $B_n \cap C_n = \emptyset$. If $n$ is large enough, say for all $n \geq \bar{n}$, both sets are nonempty because there exists $x \neq \bar{x}$ in $X$. By the Urysohn Lemma (e.g., [1, Theorem 2.46]), it follows that for each $n \geq \bar{n}$ there exists $\varphi_n \in C_b(X)$ such that $\varphi_n(x) \subseteq [0, 1]$, $\varphi_n(B_n) = 1$, and $\varphi_n(C_n) = 0$. Since $x \in \text{supp} \lambda$ and $\varphi_n(x) = 1$, it follows that

$$k_n = \int_X \varphi_n d\lambda > 0 \quad \forall n \geq \bar{n}.$$ 

Now, for each $n \geq \bar{n}$, set $\psi_n = \varphi_n / k_n$ and define the measure $\lambda_n : \mathcal{B} \to \mathbb{R}$ by $\lambda_n(B) = \int_B \psi_n d\lambda$. Notice that $\lambda_n \in \mathcal{P}_\lambda^*(X)$ because $\psi_n \in C_b(X)$.

We next show that $\lambda_n \to \delta_x$. Define $S_n = \{ x \in X : d(x, x) \leq 2/n \}$ for all $n \geq \bar{n}$. Notice that $S_n^c \subseteq C_n$ so that

$$1 = \int_{S_n} \psi_n d\lambda + \int_{S_n^c} \psi_n d\lambda = \int_{S_n} \psi_n d\lambda = \lambda_n(S_n)$$

for all $n \geq \bar{n}$. Consider an open subset $G$ of $X$. We have two cases:

1. $x \not\in G$. It follows that $\liminf \lambda_n(G) \geq 0 = \delta_x(G)$.

2. $x \in G$. For $n \geq \bar{n}$ large enough, say $n \geq \bar{m}$, we have that $S_n \subseteq G$. Then, for all $n \geq \bar{m}$, $\lambda_n(G) \geq \lambda_n(S_n) \geq 1$, yielding that $\liminf \lambda_n(G) \geq 1 = \delta_x(G)$.

In both cases, $\liminf \lambda_n(G) \geq \delta_x(G)$ holds. Since $G$ was an arbitrarily chosen open subset of $X$, by the Portmanteau Theorem (e.g., [1, Theorem 15.3]) it follows that $\lambda_n \to \delta_x$.

Since $x$ was arbitrarily chosen in $X$, we have that $\{ \delta_x \}_{x \in X} \subseteq \text{cl} \left( \mathcal{P}_\lambda^*(X) \right)$. Since $\mathcal{P}_\lambda^*(X)$ is convex, then $\text{cl} \left( \mathcal{P}_\lambda^*(X) \right)$ is closed and convex, it follows that $\text{cl} \left( \mathcal{P}_\lambda^*(X) \right) \supseteq \text{cl} \left( \text{co} \left( \{ \delta_x \}_{x \in X} \right) \right)$. But

\[\text{E.g., [1, Lemma 12.16].}\]
co \left( \{ \delta_x \}_{x \in X} \right) \) is dense in \( \mathcal{P} (X) \) (e.g., [1, Theorem 15.10]), we conclude that \( \mathcal{P} (X) \supseteq \text{cl} (\mathcal{P}_\lambda^* (X)) \supseteq \text{cl} (\text{co} \left( \{ \delta_x \}_{x \in X} \right)) = \mathcal{P} (X) \).

(ii) implies (iii). This follows from \( \mathcal{P}_\lambda^* (X) \subseteq \mathcal{P}_\lambda (X) \).

(iii) implies (iv). By contradiction, assume that \( \text{cl} (\mathcal{P}_\lambda (X)) = \mathcal{P} (X) \) and \( \ell_\lambda \) is not strictly positive. In this case, there exists \( g \in C^+_b (X) \setminus \{0\} \) such that \( \int_X gd\lambda = 0 \). Consider the open set \( G = \{ x \in X : g (x) > 0 \} \neq \emptyset \). Since \( \int_X gd\lambda = 0 \), then \( \lambda (\{ x \in X : g (x) > 0 \}) = 0 \), that is, \( \lambda (G) = 0 \). Consider \( \bar{x} \in G \). Since \( \text{cl} (\mathcal{P}_\lambda (X)) = \mathcal{P} (X) \), there exists a net \( \{ \lambda_\alpha \} \subseteq \mathcal{P}_\lambda (X) \) such that \( \lambda_\alpha \to \delta_{\bar{x}} \). For each \( \alpha \), since \( \lambda_\alpha \) is absolutely continuous with respect to \( \lambda \), we have that \( \lambda_\alpha (G) = 0 \). Since \( \lambda_\alpha \to \delta_{\bar{x}} \), by the Portmanteau Theorem, we have that \( 0 = \liminf \lambda_\alpha (G) \geq \delta_{\bar{x}} (G) = 1 \), a contradiction.

(iv) implies (i). By contradiction, assume that \( \ell_\lambda \) is strictly positive and there exists a nonempty open subset \( G \) of \( X \) with \( \lambda (G) = 0 \). Consider \( \bar{x} \in G \). By the Urysohn Lemma, and since \( G^c \) is closed and nonempty, there exists \( \varphi \in C_b (X) \) such that \( \varphi (X) \subseteq [0, 1] \), \( \varphi (\bar{x}) = 1 \), and \( \varphi (x) = 0 \) for all \( x \in G^c \). Since \( \varphi \in C^+_b (X) \setminus \{0\} \), it follows that

\[
0 < \ell_\lambda (\varphi) = \int_X \varphi d\lambda = \int_G \varphi d\lambda + \int_{G^c} \varphi d\lambda = 0,
\]

a contradiction.

Finally, observe that the result depends only on the topology of \( X \), so we could have used the term metrizable, rather than metric, throughout.

4 Illustration: Laplace’s Method

Consider the optimization problem

\[
\max_x u (x) \quad \text{sub} \ x \in K
\]

where \( u : X \to \mathbb{R} \) is a continuous function and \( K \) is a compact and metrizable set.

Laplace’s Method is a fundamental method to find maximum values and maximizers of this general optimization problem. For this reason, it plays an important role in many applications (see, e.g., Parpas and Rustem, 2009, for an introductory overview and some relevant references).

To illustrate the scope of our main result, here we establish a general abstract version of this classic method. A related result appears in Hwang (1980), though in a different setup and with an altogether different approach.

In the statement we denote by \( \Rightarrow \) the \( \sigma (\mathcal{P} (X), C_b (X)) \)-convergence and by \( \delta_x \) the Dirac probability measure concentrated on a point \( x \in X \).
Theorem 2 Let $X$ be a topological space, $u : X \to \mathbb{R}$ a continuous function, $\lambda$ a Borel probability measure with compact and metrizable support $K$, and $\{s_n\} \subseteq (0, \infty)$ a divergent sequence. Then

$$\frac{1}{s_n} \log \int_X e^{s_n u} d\lambda \to \max_K u \quad \text{as } n \to \infty$$

(3)

Moreover, if $u$ has a unique maximizer $x^u$ in $K$, then

$$\mu_n \xrightarrow{w} \delta_{x^u} \quad \text{as } n \to \infty$$

(4)

where $\mu_n$ is, for each $n \in \mathbb{N}$, defined by

$$\mu_n (B) = \frac{\int_B e^{s_n u} d\lambda}{\int_X e^{s_n u} d\lambda} \quad \forall B \in \mathcal{B} (X)$$

(5)

Proof First assume $K = X$, that is, $X$ is compact and metrizable, and $\lambda$ has full support. In this case, $\sigma (\mathcal{P} (X), C_b (X)) = \sigma (\mathcal{P} (X), C (X))$ is the relative weak* topology on $\mathcal{P} (X)$, and $\mathcal{P} (X)$ is compact and metrizable with respect to it (see Theorems 14.15 and 15.11 of AB). Denote

$$R (\mu \| \lambda) = \left\{ \begin{array}{ll} \int_X \frac{d\mu}{d\lambda} \log \left( \frac{d\mu}{d\lambda} \right) d\mu & \text{if } \mu \ll \lambda \\ \infty & \text{else} \end{array} \right.$$ 

the relative entropy of any $\mu$ in $\mathcal{P} (X)$ with respect to $\lambda$ (see Chapter 1.4 of Dupuis and Ellis, 1997, henceforth DE).

For each $n \in \mathbb{N}$, set $f_n = -s_n u$ and observe that, by Proposition 1.4.2 of DE,

$$- \log \int_X e^{-f_n} d\lambda = \min_{\mu \in \mathcal{P} (X)} \left\{ R (\mu \| \lambda) + \int_X f_n d\mu \right\}$$

and the minimum of this variational formula is uniquely attained at the element $\mu_n$ of $\mathcal{P} (X)$ given by

$$\mu_n (B) = \frac{\int_B e^{-f_n(x)} d\lambda (x)}{\int_X e^{-f_n(y)} d\lambda (y)}$$

for all Borel subsets $B$ of $X$. Recalling our substitution

$$\frac{1}{s_n} \log \int_X e^{s_n u} d\lambda = \frac{1}{s_n} \left[ - \log \int_X e^{-f_n} d\lambda \right] = \frac{1}{s_n} \min_{\mu \in \mathcal{P} (X)} \left\{ R (\mu \| \lambda) - s_n \int_X u d\mu \right\}$$

$$= \min_{\mu \in \mathcal{P} (X)} \left\{ \frac{1}{s_n} R (\mu \| \lambda) - \int_X u d\mu \right\}$$

For each $n \in \mathbb{N}$, the function $F_n : \mathcal{P} (X) \to (-\infty, \infty]$ defined by

$$F_n (\mu) = \frac{1}{s_n} R (\mu \| \lambda) - \int_X u d\mu \quad \forall \mu \in \mathcal{P} (X)$$

3A simple condition that ensures such uniqueness on convex sets is the strict quasi-concavity of $u$.  

5
is weak* lower semicontinuous on \( \mathcal{P} (X) \) (see Lemma 1.4.3 of DE and Proposition 1.9 of Dal Maso, 1993; henceforth, DM). Moreover, the sequence \( \{F_n\} \) is decreasing and pointwise converges to

\[
F_\infty (\mu) = \chi_{\text{dom } R(\cdot \| \lambda)} (\mu) - \int_X ud\mu \quad \forall \mu \in \mathcal{P} (X)
\]

By Proposition 5.7 of DM, this sequence \( \Gamma \)-converges to the weak* lower semicontinuous envelope \( \text{sc}^- F_\infty \) of \( F_\infty \). Since \( U : \mu \mapsto \int_X ud\mu \) is continuous and everywhere finite on \( \mathcal{P} (X) \), by Proposition 3.7 and Example 3.4 of DM

\[
\left( \text{sc} F_\infty \right) (\mu) = \left( \text{sc} \chi_{\text{dom } R(\cdot \| \lambda)} \right) (\mu) - \int_X ud\mu = \chi_{\text{cl(dom } R(\cdot \| \lambda)} \right) (\mu) - \int_X ud\mu
\]

For each \( \mu \in \mathcal{P}_X^* (X) \), \( d\mu/d\lambda \) is bounded and continuous, hence there exists \( k \geq 0 \) such that \( 0 \leq d\mu/d\lambda \leq k \) and so

\[
- \frac{1}{e} \leq \frac{d\mu}{d\lambda} \log \left( \frac{d\mu}{d\lambda} \right) \leq k^2 \implies R(\mu \| \lambda) < \infty \implies \mu \in \text{dom } R(\cdot \| \lambda)
\]

Therefore \( \mathcal{P}_X^* (X) \subseteq \text{dom } R(\cdot \| \lambda) \) and so, by Theorem 1, \( \mathcal{P} (X) = \text{cl}(\mathcal{P}_X^* (X)) \subseteq \text{cl(dom } R(\cdot \| \lambda) = \mathcal{P} (X) \). Summing up, \( F_n \) \( \Gamma \)-converges to \( -\int_X ud\mu \). By Theorem 7.4 of DM, this implies

\[
\lim_{n \to \infty} \min_{\mu \in \mathcal{P}(X)} \left\{ \frac{1}{s_n} R(\mu \| \lambda) - \int_X ud\mu \right\} = \min_{\mu \in \mathcal{P}(X)} \left\{ - \int_X ud\mu \right\} = - \max_{\mu \in \mathcal{P}(X)} \left\{ \int_X ud\mu \right\} = - \max_{x \in X} u(x)
\]

But, for all \( n \in \mathbb{N} \) we have

\[
\min_{\mu \in \mathcal{P}(X)} \left\{ \frac{1}{s_n} R(\mu \| \lambda) - \int_X ud\mu \right\} = - \frac{1}{s_n} \log \int_X e^{s_n u} d\lambda
\]

So, (3) holds.

Moreover, if \( u \) has a unique maximizer \( x^u \) in \( X \), then \( U \) has \( \delta_{x^u} \) as its unique maximizer. In fact, if \( \mu \in \mathcal{P} (X) \setminus \{ \delta_{x^u} \} \), then \( \mu (X \setminus \{ x^u \}) > 0 \), and so

\[
\int_X ud\delta_{x^u} - \int_X ud\mu = \int_X (u(x^u) - u(x)) d\mu (x)
\]

\[
= \int_{\{x^u\}} (u(x^u) - u(x)) d\mu (x) + \int_{X \setminus \{x^u\}} (u(x^u) - u(x)) d\mu (x)
\]

where the first summand is null, the second is strictly positive.\(^4\) Since \( \mathcal{P} (X) \) is compact, the sequence \( F_n \) is equi-coercive (see Definition 7.6 of DM); in addition, it \( \Gamma \)-converges to \( -U \) with unique minimum point \( \delta_{x^u} \) in \( \mathcal{P} (X) \). For each \( n \), the probability measure \( \mu_n \) is a minimizer for \( F_n \) in \( \mathcal{P} (X) \). By Corollary 7.24 of DM, \( \mu_n \) weak* converges to \( \delta_{x^u} \).

\(^4\) \( \int_{B \setminus \{x^u\}} (u(x^u) - u(x)) d\gamma (x) = 0 \) would imply \( \gamma (\{ x \in B \setminus \{ x^u \} : u(x^u) - u(x) > 0 \}) = 0 \), a contradiction because \( u(x^u) - u(x) > 0 \) for all \( x \in B \setminus \{ x^u \} \).
In the general case, consider the compact and metrizable space $K$, the continuous function $w = u |_K$, and the Borel probability measure $\nu = \lambda |_K$. It is easy to show that $\nu$ has full support on $K$. In facts, if $O$ is a nonempty open subset of $K$, there exists an open subset $G$ of $X$ such that $\emptyset = O = G \cap K = G \cap \text{supp } \lambda$; by definition of support, it follows $\lambda (G) > 0$, but then $\nu (O) = \lambda (G \cap \text{supp } \lambda) = \lambda (G \cap \text{supp } \lambda) + \lambda (G \cap (\text{supp } \lambda)^c) = \lambda (G) > 0$. The previous part of the proof implies

$$\frac{1}{s_n} \log \int_K e^{s_n w} d\nu \to \max_K w \quad \text{as } n \to \infty$$

But $s_n^{-1} \log \int_X e^{s_n u} d\lambda = s_n^{-1} \log \int_K e^{s_n w} d\nu$ for all $n \in \mathbb{N}$ and $\max_K u = \max_K w$, thus (3) holds.

Moreover, if $u$ has a unique maximizer $x^u$ in $K$, again by the previous part of the proof we can consider the sequence $\{\rho_n\}$ of probability measures defined by

$$\rho_n (L) = \frac{\int_L e^{s_n w} d\nu}{\int_K e^{s_n w} d\nu} \quad \forall n \in \mathbb{N}$$

for all Borel subsets $L$ of $K$, and have that, given any $g \in C (K)$,

$$\int_K g d\rho_n \to g (x^u) \quad \text{as } n \to \infty$$

But for each $f \in C_b (X)$, $f|_K \in C (K)$ and $\int_X f d\mu_n = \int_K f|_K d\rho_n$ for all $n \in \mathbb{N}$, then the sequence $\{\mu_n\}$, defined by (5), $\sigma (P (X), C_b (X))$ converges to $\delta_{x^u}$. \hfill \blacksquare

If $X$ is the dual of a separable normed space and is endowed with the weak* topology, then the boundedness of the support of $\lambda$ is equivalent to its compactness, and – as we observed in the previous section – each $\mu_n$ has a barycenter $m_n$ in the weak* closed and convex hull of $K = \text{supp } \lambda$. Next we show that these barycenters weak*-converge to the maximizer. Here $\xrightarrow{w^*}$ denotes weak*-convergence.

**Proposition 2** Let $X$ be the dual of a separable normed space. Under the assumptions of Theorem 2, we have

$$m_n \xrightarrow{w^*} x^u \quad \text{as } n \to \infty$$

where $m_n$ is, for each $n \in \mathbb{N}$, the barycenter of $\mu_n$.

In particular, if $X$ is a separable and reflexive Banach space, then its weak and weak* topologies coincide and so $m_n$ weakly converges to $x^u$. Clearly, the sequence of barycenters is included in $K$ if this set is convex.

When $X$ is $\mathbb{R}^n$ and $\ell$ is a sigma-finite Borel measure, we have

$$\frac{1}{s_n} \log \frac{1}{\ell (K)} \int_K e^{s_n u (x)} d\ell (x) \to \max_K u \quad \text{as } n \to \infty \quad (7)$$
and, if \( x^u \) is the unique maximizer of \( u \) on \( K \),

\[
m_n = \frac{\int_K e^{s_n u(x)} \, d\ell(x)}{\int_K e^{s_n u(y)} \, d\ell(y)} \to x^u
\]

This convergence in \( \mathbb{R}^n \) has been first established by Pincus (1968, 1970) (see Hiriart-Urruty, 1995, p. 22). The weak* convergence (4) thus substantially generalizes his results.

References


