Extended Laplace Principle for Empirical Measures of a Markov Chain - Supplementary Material

Stephan Eckstein[*]

December 9, 2018

In this additional material, the choice of β as a transport cost is discussed. The main purpose is to showcase a standard technique to obtain applicability of the main Theorem 1.1.

For $c : E^2 \to [0,\infty]$ let $\beta(\nu,\mu) := \inf_{\tau \in \Pi(\nu,\mu)} \int_{E^2} c d\tau$ which is simply a transport cost between $\nu \in \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$. Recall the definition of β_2 which reads for $\theta \in \mathcal{P}(E)$ and $\nu \in \mathcal{P}(E^2)$:

$$\beta_2^\theta(\nu) = \beta(\nu_{0,1},\theta) + \int_E \beta(\nu_{1,2}(x),\pi(x))\nu_{0,1}(dx)$$

For this choice of β, the upper bound of Theorem 1.1. can be applied in compact spaces, as Lemma 0.1 below establishes. For the lower bound on the other hand, condition (B.3), i.e. if $\nu \not\ll \mu$, then $\beta(\nu,\mu) = \infty$, is in general not satisfied. Intuitively the reason for this is that transport distances do not agree with absolute continuity conditions. There is a simple workaround however in discrete spaces if the Markov chain satisfies $\pi_0(\{x\}) > 0$ and $\pi(x)(\{y\}) > 0$ for all $x, y \in E$. Then, one can simple define $\underline{\beta}(\nu,\mu) := \beta(\nu,\mu)$, if $\nu \ll \mu$ and $\underline{\beta}(\nu,\mu) = \infty$, else. This is highly discontinuous of course, but the only important thing for the applicability of the lower bound is that it is still measurable and convex, which it is. The idea behind this choice of $\underline{\beta}$ is that with π_0 and π strictly positive, all terms in the main theorem read precisely the same as for the functional β, and hence one obtains fitting upper and lower bounds. This workaround is obviously not limited to the choice of β as presented here, but can be applied whenever one works in discrete spaces with π_0 and π strictly positive.

Lemma 0.1. *Let (E,d) be a compact polish space, and $c : E^2 \to [0,\infty]$ be lower semi-continuous. Assume π satisfies the Feller property, i.e. $x \mapsto \pi(x)$ is continuous. For $\beta(\nu,\mu) := \inf_{\tau\in\Pi(\nu,\mu)} \int c d\tau$ conditions (B.1) and (B.2) are satisfied, i.e. the mapping*

$$\mathcal{P}(E) \times \mathcal{P}(E^2) \ni (\theta,\nu) \mapsto \beta_2^\theta(\nu)$$

is convex and lower semi-continuous.

Proof. First, it suffices to show that the mapping

$$\mathcal{P}(E^2) \ni \nu \mapsto \varphi(\nu) := \int_E \beta(\nu_{1,2}(x),\pi(x))\nu_{0,1}(dx)$$

is convex and lower semi-continuous, since the other summand in the expression of $\beta_2^\theta(\nu)$ is a simplified version of this mapping (simplified in the sense that ν is a product measure and π is constant).

[*]Department of Mathematics, University of Konstanz, 78464 Konstanz Germany, stephan.eckstein@uni-konstanz.de

By using the dual representation of β (see e.g. [3, Theorem 5.10.]) one obtains

$$\varphi(\nu) = \int_E \left(\sup_{\substack{f_1, f_2 \in C_b(E): \\ f_1 \oplus f_2 \leq c}} \int_E f_1(y) \nu_{1,2}(x, dy) + \int_E f_2(y) \pi(x, dy) \right) \nu_{0,1}(dx)$$

where we use the notation $(f_1 \oplus f_2)(x, y) = f_1(x) + f_2(y)$. Since by compactness of E the space $C_b(E)$ endowed with the sup-metric is Polish, we can apply a measurable selection argument [1, Proposition 7.50] to obtain

$$\varphi(\nu) =$$
$$\sup_{\substack{F_1, F_2 : E \to C_b(E): \\ F_1(x) \oplus F_2(x) \leq c}} \int_E \left(\int_E F_1(x)(y) \nu_{1,2}(x, dy) + \int_E F_2(x)(y) \pi(x, dy) \right) \nu_{0,1}(dx),$$

where one first obtains that the supremum is over all universally measurable functions. A very slight adaptation of the arguments of [1, Lemma 7.27 and 7.28] yields that Borel functions are sufficient, see Lemma 0.2. From there, Lemma 0.3 yields

$$\varphi(\nu) = \sup_{\substack{F_1, F_2 \in C_b(E^2): \\ F_1(x, \cdot) \oplus F_2(x, \cdot) \leq c}} \left(\int_{E^2} F_1 d\nu + \int_{E^2} F_2 d\nu_{0,1} \otimes \pi \right)$$

which implies lower semi-continuity of φ, as by the Feller property of π the term inside the brackets is continuous in ν (this follows by [2, Lemma 8.3.2.]). Since the term inside the bracket is linear in ν, this representation implies convexity of φ as well. $\square$

Lemma 0.2. *Let (E, d) be a compact polish space and $C_b(E)$ be endowed with the* sup*-metric. For $F : E \to C_b(E)$ universally measurable and $\nu \in \mathcal{P}(E)$ there exists a Borel measurable function $F_\nu : E \to C_b(E)$ such that $F(x) = F_\nu(x)$ for ν almost all $x \in E$ (and for example $F_\nu(x) = 0$, else).*

Proof. The proof is a slight adaptation of arguments in [1, Lemma 7.27 and 7.28].

First, fix a countable dense subset $\mathcal{Q} \subseteq E$. For $y \in \mathcal{Q}$ let $f_y : E \to \mathbb{R}$ be given by $f_y(x) := F(x)(y)$. Then f_y is universally measurable since it can be written as $x \mapsto F(x) \mapsto F(x)(y)$ and is hence universally measurable as a concatenation of universally measurable functions. By [1, Lemma 7.27] there exists a Borel set $A \subseteq E$ with $\nu(A) = 1$ and functions $f_y^\nu : E \to \mathbb{R}$ Borel such that for all $x \in A$ and for all $y \in \mathcal{Q}$

$$f_y^\nu(x) = f_y(x).$$

We define $F_\nu(x) := F(x)$ for all $x \in A$ and $F_\nu(x) = 0$, else. F_ν is $C_b(E)$-valued and ν almost surely equal to F. It remains to show that F_ν is Borel:

The Borel sigma algebra on the polish space $C_b(E)$ with sup-metric is generated by the family

$$\bigcup_{S \subseteq \mathcal{Q} \text{ finite}} \pi_F^{-1}(Bor(E)^S)$$

where $\pi_F : C_b(E) \to E^S,\ f \mapsto (f(x))_{x \in S}$. Hence it suffices to show that for $S \subseteq \mathcal{Q}$ and $a \in \mathbb{R}^S$ the pre-image of $\{f \in C_b(E) : f(y) \leq a_y \text{ for all } y \in \mathcal{S}\}$ under F_ν is Borel. One calculates

$$\begin{aligned} &F_\nu^{-1}(\{f \in C_b(E) : f(y) \leq a_y \text{ for all } y \in \mathcal{S}\}) \\ =&\{x \in A : f_y^\nu(x) \leq a_y \text{ for all } y \in \mathcal{S}\} \cup \{x \in A^C : 0 \leq a_y \text{ for all } y \in \mathcal{S}\} \end{aligned}$$

which is Borel. $\square$

Lemma 0.3. *Let (E,d) be a compact polish space and $C_b(E)$ be endowed with the* sup*-metric. Let $c : E^2 \to [0,\infty]$ be lower semi-continuous. Let $\nu \in \mathcal{P}(E^2)$. Define*

$$(A) := \sup_{\substack{F_1,F_2: E\to C_b(E) \text{ Borel} \\ F_1(x)\oplus F_2(x)\le c}} \int_E \left(\int_E F_1(x)(y)\nu_{1,2}(x,dy) + \int_E F_2(x)(y)\pi(x,dy) \right) \nu_{0,1}(dx),$$

$$(B) := \sup_{\substack{G_1,G_2 \in C_b(E^2): \\ G_1(x,\cdot)\oplus G_2(x,\cdot)\le c}} \int_{E^2} G_1 d\nu + \int_{E^2} G_2 d\nu_{0,1}\otimes\pi.$$

Then it holds $(A) = (B)$.

Proof. **First step:** It holds $(B) \le (A)$ as for given $G \in C_b(E^2)$ the mapping $E \ni x \mapsto G(x,\cdot)$ is Borel as it is continuous (by uniform continuity of G, one obtains $\sup_{y\in E}|G(x_n,y) - G(x,y)| \to 0$ for $d(x_n,x)\to 0$).

Second step: It is $(A) \le (C)$ for

$$(C) := \sup_{\substack{G_1,G_2: E^2\to\mathbb{R} \text{ Borel}: \\ G_1(x,\cdot)\oplus G_2(x,\cdot)\le c}} \int_{E^2} G_1 d\nu + \int_{E^2} G_2 d\nu_{0,1}\otimes\pi,$$

since for given $F : E \to C_b(E)$ Borel, the mapping $E^2 \ni (x,y) \mapsto F(x)(y)$ is Borel as well (obvious if rewritten as $(x,y) \mapsto (F(x),y) \mapsto F(x)(y)$).

Third step: It holds $(C) = (B)$, where $(B) \le (C)$ is obvious. To show $(B) = (C)$, first note that by standard duality theory for optimal transport (e.g. [3, Theorem 5.10.]), where one can regard c as a function on $E^2 \times E^2$ that is constant in two components, one obtains for $\mu^1 = \nu,\ \mu^2 = \nu_{0,1}\otimes\pi$

$$(B) = \sup_{\substack{G_i\in L_1(\mu^i): \\ G_1(x,\cdot)\oplus G_2(x,\cdot)\le c}} \int_{E^2} G_1 d\nu + \int_{E^2} G_2 d\nu_{0,1}\otimes\pi$$

On the other hand, in the definition of (C) we might as well impose that $\min\{G_i, 0\} \in L_1(\mu^i)$ due to the convention that $\infty - \infty = -\infty$. For such a G_i, we can define $G_i^n := \min\{G_i, n\} \in L_1(\mu^i)$ and obtain $(B) = (C)$ by monotone convergence. Notably, by approximating functions G_1 and G_2 in the definition of (C) from below, the inequality constraint $G_1^n(x,\cdot)\oplus G_2^n(x,\cdot) \le c$ is always satisfied. $\square$

References

[1] Bertsekas, D. P. and Shreve, S. (1996). *Stochastic optimal control: the discrete-time case.* Athena Scientific.

[2] Dupuis, P. and Ellis, R. S. (2011). *A weak convergence approach to the theory of large deviations* vol. 902. John Wiley & Sons.

[3] Villani, C. (2008). *Optimal transport: old and new* vol. 338. Springer Science & Business Media.