Statistics and Probability Letters 88 (2014) 1–8

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the effect of perturbation of conditional probabilities in total variation

Alessandro Abate [a,b], Frank Redig [c], Ilya Tkachev [b,*]

[a] Department of Computer Science, University of Oxford, United Kingdom

[b] Delft Center for Systems & Control, Delft University of Technology, The Netherlands

[c] Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 13 January 2014
Accepted 14 January 2014
Available online 23 January 2014

Keywords:
Conditional probabilities
Total variation
Perturbation
Coupling

ABSTRACT

A celebrated result by A. Ionescu Tulcea provides a construction of a probability measure on a product space given a sequence of regular conditional probabilities. We study how the perturbations of the latter in the total variation metric affect the resulting product probability measure.

1. Introduction

The Ionescu Tulcea extension theorem (Ash, 1972, Section 2.7.2) states that given a sequence of stochastic kernels, there exists a unique probability measure on the product space generated by this sequence, that is a measure whose conditional probabilities equal to these kernels. Such a construction is often used in the theory of general Markov Decision Processes (Bertsekas and Shreve, 1978), and general Markov Chains (Revuz, 1984) in particular. Hence, it is of a certain interest to study how sensitive the resulting product measure is with respect to perturbations of the generating sequence of kernels. A possible direct application of such result concerns numerical methods, where characteristics of the original stochastic process are studied over its simpler approximations, often defined over a finite state space. Such approximations can be further regarded as a perturbation of the original sequence of kernels (Tkachev and Abate, 2013) which connects to the original problem.

Here we specifically focus on the metric between kernels and measures given by the total variation norm. Given the pairwise distances between corresponding transition kernels in this metric, we are interested in bounds on the distance between the resulting product measures. A similar study was given in Roberts and Rosenthal (2013) which used the Borel assumption, that is it assumed that spaces involved are (standard) Borel spaces. However, the bounds obtained in Roberts and Rosenthal (2013) grow linearly with the cardinality of the sequence and hence are not tight: recall that the total variation distance between two probability measures is always bounded from above by 2.

In this paper we elaborate on the result of Roberts and Rosenthal (2013) in the two following directions. First, we generalize linear bounds to the case of arbitrary measurable spaces. Second, we show that under the Borel assumption used in the paper Roberts and Rosenthal (2013) it is possible to derive sharper bounds, that appear to be precise in some special cases—e.g. in case of independent products of measures.

* Corresponding author. Tel.: +31 1527 87171.
E-mail addresses: alessandro.abate@cs.ox.ac.uk (A. Abate), f.h.j.redig@tudelft.nl (F. Redig), i.tkachev@tudelft.nl (I. Tkachev).

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2014.01.009

The rest of the paper is structured as follows. Section 2 gives a problem formulation together with statements of main results. Proofs are given in Section 3, which is followed by the discussion in Section 4 and an enlightening example in Section 5. With regards to the notation, terminology and conventions adopted in this paper, the readers should consult the Appendix.

2. Problem statement

Let us recall the construction of the product measure given the regular conditional probabilities. First of all, we need the following notion of a product of a probability measure and a stochastic kernel which extends a more usual product of two measures.

Proposition 1. *Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be arbitrary measurable spaces. For any probability measure $\mu \in \mathscr{P}(X, \mathfrak{X})$ and any stochastic kernel $K : X \to \mathscr{P}(Y, \mathfrak{Y})$ there exists a unique probability measure $\mathsf{Q} \in \mathscr{P}(X \times Y, \mathfrak{X} \otimes \mathfrak{Y})$, denoted by $\mathsf{Q} := \mu \otimes K$, such that*

$$\mathsf{Q}(A \times B) = \int_A K(x, B)\mu(\mathrm{d}x)$$

for any pair of sets $A \in \mathfrak{X}$ and $B \in \mathfrak{Y}$.

Proof. For a proof, see Ash (1972, Section 2.6.2). □

The construction above immediately extends to any finite sequence of spaces by induction, whereas for the countable products the following result holds true.

Proposition 2 (*Ionescu-Tulcea*)**.** *Let $\{(X_k, \mathfrak{X}_k)\}_{k\in\mathbb{N}_0}$ be a family of arbitrary measurable spaces and let $(\Omega_n, \mathscr{F}_n) = \prod_{k=0}^n (X_k, \mathfrak{X}_k)$ be product spaces for any $n \in \bar{\mathbb{N}}_0$. For any probability measure $P^0 \in \mathscr{P}(X_0, \mathfrak{X}_0)$ and any sequence of stochastic kernels $(P^k)_{k\in\mathbb{N}}$, where $P^k : \Omega_{k-1} \to \mathscr{P}(X_k, \mathfrak{X}_k)$, there exists a unique probability measure $\mathsf{P} \in \mathscr{P}(\Omega_\infty, \mathscr{F}_\infty)$, denoted by $\mathsf{P} := \bigotimes_{k=0}^\infty P^k$, such that the finite-dimensional marginal P^n of P on the measurable space $(\Omega_n, \mathscr{F}_n)$ is given by $\mathsf{P}^n = \bigotimes_{k=0}^n P^k$ for any $n \in \mathbb{N}_0$.*

Proof. For a proof, see (Ash, 1972, Section 2.7.2). □

In the setting of Proposition 2, suppose that we are given another sequence of kernels $(\tilde{P}^k)_{k=0}^\infty$ and let $\tilde{\mathsf{P}} := \bigotimes_{k=0}^\infty \tilde{P}^k$ be the corresponding product measure. Given the assumption that $\|P^k - \tilde{P}^k\| \le c_k$ for any $k \in \mathbb{N}_0$ and some sequence of reals $(c_k)_{k\in\mathbb{N}_0}$, we study how the distance $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\|$ can be bounded. For the general case of arbitrary measurable spaces, the following result holds true.

Theorem 1. *Let $\{(X_k, \mathfrak{X}_k)\}_{k\in\mathbb{N}_0}$ be any family of measurable spaces and let $\tilde{\mathfrak{X}}_k \subseteq \mathfrak{X}_k$ for any $k \in \mathbb{N}_0$. Denote by $(\Omega_n, \mathscr{F}_n) = \prod_{k=0}^n (X_k, \mathfrak{X}_k)$ and $(\Omega_n, \tilde{\mathscr{F}}_n) = \prod_{k=0}^n (X_k, \tilde{\mathfrak{X}}_k)$ the corresponding product spaces for any $n \in \bar{\mathbb{N}}_0$. Let $P^0 \in \mathscr{P}(X_0, \mathfrak{X}_0)$, $\tilde{P}^0 \in \mathscr{P}(X_0, \tilde{\mathfrak{X}}_0)$ and let kernels $P^k : \Omega_{k-1} \to \mathscr{P}(X_k, \mathfrak{X}_k)$ and $\tilde{P}^k : \Omega_{k-1} \to \mathscr{P}(X_k, \tilde{\mathfrak{X}}_k)$ for $k \in \mathbb{N}$ be $\mathscr{F}_{k-1}$- and $\tilde{\mathscr{F}}_{k-1}$-measurable respectively. If a sequence of reals $(c_k)_{k\in\mathbb{N}_0}$ is such that $\|P^k - \tilde{P}^k\| \le c_k$ for all $k \in \mathbb{N}_0$, then for any $n \in \mathbb{N}_0$ it holds that*

$$\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| \le \sum_{k=0}^n c_k. \tag{2.1}$$

Remark 1. Through this paper, and in particular in the statement of Theorem 1, we use the following convention. If the domain of one measure is a subset of the domain of another, the total variation distance between them is taken over the smaller domain. For example, in the setting of Theorem 1 we have $\|P^0 - \tilde{P}^0\| = 2 \cdot \sup_{A\in\tilde{\mathfrak{X}}_0} |P^0(A) - \tilde{P}^0(A)|$.

The validity of results of Theorem 1 in some special cases was previously established in Roberts and Rosenthal (2013) and Tkachev and Abate (2013): we discuss these connections in a greater detail in Section 4. As it has been mentioned in Introduction, the bounds (2.1) are not tight. For example, if $c_k = c > 0$ for all $k \in \mathbb{N}_0$ then the right-hand side of (2.1) is $c \cdot n$ and diverges to infinity as $n \to \infty$, whereas the left-hand side stays bounded above by 2. It appears, that under a rather mild assumption that all involved measurable spaces are (standard) Borel, a stronger result can be obtained.

Theorem 2. *Let $\{X_k\}_{k\in\mathbb{N}_0}$ be a family of Borel spaces and let $\Omega_n = \prod_{k=0}^n X_k$ be product spaces for any $n \in \bar{\mathbb{N}}_0$. Let further $P^0, \tilde{P}^0 \in \mathscr{P}(X_0)$ and $P^n, \tilde{P}^k : \Omega_{k-1} \to \mathscr{P}(X_k)$ for $k \in \mathbb{N}$. If a sequence of reals $(c_k)_{n\in\mathbb{N}_0}$ is such that $\|P^k - \tilde{P}^k\| \le c_k$ for all $k \in \mathbb{N}_0$, then for any $n \in \mathbb{N}_0$ it holds that*

$$\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| \le 2 - 2\prod_{k=0}^n \left(1 - \frac{1}{2}c_k\right). \tag{2.2}$$

The proofs of both theorems are given in the next section.

3. Proofs of the main results

3.1. *Proof of Theorem 1*

We prove both theorems by induction, by first studying how the perturbation of a measure and a kernel in Proposition 1 propagates to the product measure. The following lemma provides such study in the setting of Theorem 1.

Lemma 1. *Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be two measurable spaces, and let $\tilde{\mathfrak{X}} \subseteq \mathfrak{X}$ and $\tilde{\mathfrak{Y}} \subseteq \mathfrak{Y}$. Consider $\mu \in \mathscr{P}(X, \mathfrak{X})$ and $\tilde{\mu} \in \mathscr{P}(X, \tilde{\mathfrak{X}})$, and suppose that kernels $K : X \to \mathscr{P}(Y, \mathfrak{Y})$ and $\tilde{K} : X \to \mathscr{P}(Y, \tilde{\mathfrak{Y}})$ are $\mathfrak{X}$- and $\tilde{\mathfrak{X}}$-measurable, respectively. Denote by $\mathsf{Q} := \mu \otimes K$ and $\tilde{\mathsf{Q}} := \tilde{\mu} \otimes \tilde{K}$ the corresponding product measures. It holds that*

$$\|\mathsf{Q} - \tilde{\mathsf{Q}}\| \leq \|\mu - \tilde{\mu}\| + \|K - \tilde{K}\| \tag{3.1}$$

where in (3.1) we follow the convention in Remark 1.

Proof. Let the set $A \in \tilde{\mathfrak{X}} \otimes \tilde{\mathfrak{Y}}$ be arbitrary, and denote by $A_x = \{y \in Y : (x, y) \in A\}$ the x-section of A for any $x \in X$. It follows from Ash (1972, Section 2.6.2) that

$$\mathsf{Q}(A) = \int_X K_x(A_x)\mu(\mathrm{d}x), \qquad \tilde{\mathsf{Q}}(A) = \int_X \tilde{K}_x(A_x)\tilde{\mu}(\mathrm{d}x)$$

and as a result:

$$\begin{aligned} |\mathsf{Q}(A) - \tilde{\mathsf{Q}}(A)| &= \left| \int_X K_x(A_x)\mu(\mathrm{d}x) - \int_X \tilde{K}_x(A_x)\tilde{\mu}(\mathrm{d}x) \right| \\ &\leq \left| \int_X \left(K_x(A_x) - \tilde{K}_x(A_x) \right) \mu(\mathrm{d}x) \right| + \left| \int_X \tilde{K}_x(A_x)(\mu - \tilde{\mu})(\mathrm{d}x) \right| \\ &\leq \sup_{x\in X} \sup_{B\in\tilde{\mathfrak{Y}}} \left| K_x(B) - \tilde{K}_x(B) \right| + \sup_{B\in\tilde{\mathfrak{X}}} |\mu(B) - \tilde{\mu}(B)| \\ &= \frac{1}{2} \left(\|\mu - \tilde{\mu}\| + \|K - \tilde{K}\| \right) \end{aligned}$$

which together with (A.2) implies (3.1). □

To prove Theorem 1 we are only left to apply the result of Lemma 1 by induction to $\mu = \mathsf{P}^n$, $\tilde{\mu} = \tilde{\mathsf{P}}^n$, $K = P^{n+1}$ and $\tilde{K} = \tilde{P}^{n+1}$ for $n \in \mathbb{N}_0$.

3.2. *Proof of Theorem 2*

Again, we are going to apply induction, however in the current case the analogue of Lemma 1 requires a more intricate proof via the coupling techniques. Let us briefly recall some facts about the coupling. Given an arbitrary measurable space $(\Omega, \mathscr{F})$, a coupling of two probability measures $\mathsf{P}, \tilde{\mathsf{P}} \in \mathscr{P}(\Omega, \mathscr{F})$ is a probability measure $\mathbb{P} \in \mathscr{P}(\Omega^2, \mathscr{F}^2)$ such that the marginals of $\mathbb{P}$ are given by

$$\pi_*\mathbb{P} = \mathsf{P}, \qquad \tilde{\pi}_*\mathbb{P} = \tilde{\mathsf{P}}, \tag{3.2}$$

where $\pi(\omega, \tilde{\omega}) = \omega$ and $\tilde{\pi}(\omega, \tilde{\omega}) = \tilde{\omega}$ for all $(\omega, \tilde{\omega}) \in \Omega^2$ (Lindvall, 1992, Section 1.1). In particular, if Ω is a Borel space and $\mathscr{F} = \mathfrak{B}(\Omega)$, we have the following result:

$$\mathbb{P}(\pi = \tilde{\pi}) \leq \|\mathsf{P} \wedge \tilde{\mathsf{P}}\|.^{1} \tag{3.3}$$

The inequality (3.3) is called the coupling inequality (Lindvall, 1992, Section 1.2); it holds true for any coupling measure $\mathbb{P}$ as per (3.2). At the same time, thanks to the fact that Ω is a Borel space, there always exists a maximal coupling: the one for which the equality holds in (3.3). Some pairs of probability measures admit several choices of maximal coupling; here we focus on the γ-coupling (Lindvall, 1992, Section 1.5).

Definition 1 (*γ-coupling*)**.** Let Z be a Borel space and let $\nu, \tilde{\nu} \in \mathscr{P}(Z)$ be two probability measures on it. The γ-coupling of $(\nu, \tilde{\nu})$ is a measure $\gamma \in \mathscr{P}(Z^2)$ given by

$$\gamma(\nu, \tilde{\nu}) := (\psi_Z)_*(\nu \wedge \tilde{\nu}) + 1_{[0,1)}(\|\nu \wedge \tilde{\nu}\|) \cdot \frac{(\nu - \tilde{\nu})^+ \otimes (\nu - \tilde{\nu})^-}{1 - \|\nu \wedge \tilde{\nu}\|}$$

where $\psi_Z : Z \to Z^2$ is the diagonal map on Z given by $\psi_Z : z \mapsto (z, z)$.

[1] The precise definitions of $\wedge$ and π_* can be found in the Appendix, see (A.1) and (A.3) respectively.

The following lemma is the key in the proof of Theorem 2.

Lemma 2. *Let X and Y be two Borel spaces, and let $\mu, \tilde{\mu} \in \mathscr{P}(X)$ and $K, \tilde{K} : X \to \mathscr{P}(Y)$. Denote by $\mathrm{Q} := \mu \otimes K$ and $\tilde{\mathrm{Q}} := \tilde{\mu} \otimes \tilde{K}$ the corresponding product measures. It holds that*

$$\|\mathrm{Q} \wedge \tilde{\mathrm{Q}}\| \geq \|\mu \wedge \tilde{\mu}\| \cdot \inf_{x \in X} \|K_x \wedge \tilde{K}_x\|. \tag{3.4}$$

Proof. The proof is done via coupling of measures Q and $\tilde{\mathrm{Q}}$ in a sequential way.

Firstly, let $m = \gamma(\mu, \tilde{\mu}) \in \mathscr{P}(X^2)$ be the γ-coupling of $(\mu, \tilde{\mu})$. Secondly, we define a kernel $\kappa : X^2 \to \mathscr{P}(Y^2)$ by the following formula:

$$\kappa_{x\tilde{x}} = 1_{\Delta_X}(x, \tilde{x}) \cdot \gamma\left(K_x, \tilde{K}_{\tilde{x}}\right) + 1_{\Delta_X^c}(x, \tilde{x}) \cdot (K_x \otimes \tilde{K}_{\tilde{x}})$$

where Δ_X is the diagonal of X^2 as per Appendix. Clearly, the measure $\kappa_{x\tilde{x}}$ is a coupling of measures K_x and $\tilde{K}_{\tilde{x}}$ for any $(x, \tilde{x}) \in X^2$, which is a maximal coupling on the diagonal, and the independent (or product) coupling off the diagonal. Note that $\kappa : X^2 \to \mathscr{P}(Y^2)$ is indeed a kernel. The only non-trivial part of the latter statement concerns the measurability of the positive part and the negative part in the Hahn–Jordan decomposition of the kernel, which follows directly from Revuz (1984, Lemma 1.5, Chapter 6). Furthermore, the measurability of $K_x \otimes \tilde{K}_{\tilde{x}}$ obviously holds for any measurable rectangle $A \times \tilde{A} \subseteq Y^2$ and extends to the whole product σ-algebra by the monotone class theorem (see e.g. Ash (1972, Theorem 1.3.9)).

Define $\mathbb{Q} := m \otimes \kappa \in \mathscr{P}(X^2 \times Y^2)$ to be the product measure, and further denote by π_X, $\tilde{\pi}_X$, π_Y and $\tilde{\pi}_Y$ the obvious projection maps from that space, e.g.

$$\tilde{\pi}_X(x, \tilde{x}, y, \tilde{y}) = \tilde{x} \in X.$$

We claim that the random element (π_X, π_Y) is distributed according to Q and $(\tilde{\pi}_X, \tilde{\pi}_Y)$ is distributed according to $\tilde{\mathrm{Q}}$. Indeed, for any $A \in \mathfrak{B}(X)$ and $B \in \mathfrak{B}(Y)$ it holds that

$$\begin{aligned}
\mathbb{Q}(\pi_X \in A, \pi_Y \in B) &= \mathbb{Q}((A \times X) \times (B \times Y)) \quad \text{by definition of } \pi_X \text{ and } \pi_Y \\
&= \int_{A \times X} \kappa_{x\tilde{x}}(B \times Y) m(\mathrm{d}x \times \mathrm{d}\tilde{x}) \quad \text{by Proposition 1} \\
&= \int_{A \times X} K_x(B) m(\mathrm{d}x \times \mathrm{d}\tilde{x}) \quad \text{since } \kappa_{x\tilde{x}} \text{ is a coupling} \\
&= \int_A K_x(B) \mu(\mathrm{d}x) \quad \text{since } \mu \text{ is a marginal of } m \\
&= \mathrm{Q}(A \times B) \quad \text{by Proposition 1.}
\end{aligned}$$

Since both $(\pi_X, \pi_Y)_* \mathbb{Q}$ and Q are probability measures on a Borel space $X \times Y$, and they have been shown to agree on the class measurable rectangles which is closed under finite intersections, they are equal (Revuz, 1984, Proposition 3.6, Chapter 0). Similarly, it holds that $(\tilde{\pi}_X, \tilde{\pi}_Y)_* \mathbb{Q} = \tilde{\mathrm{Q}}$. Thus, by the coupling inequality (3.3)

$$\|\mathrm{Q} \wedge \tilde{\mathrm{Q}}\| \geq \mathbb{Q}(\pi_X = \tilde{\pi}_X, \pi_Y = \tilde{\pi}_Y). \tag{3.5}$$

On the other hand, since $\mathbb{Q} = m \otimes \kappa$, we can rewrite the right-hand side of (3.5) in terms of m and κ as follows:

$$\begin{aligned}
\mathbb{Q}(\pi_X = \tilde{\pi}_X, \pi_Y = \tilde{\pi}_Y) &= \mathbb{Q}\left((\Delta_X \times Y^2) \cap (X^2 \times \Delta_Y)\right) = \mathbb{Q}(\Delta_X \times \Delta_Y) \\
&= \int_{\Delta_X} \kappa_{x\tilde{x}}(\Delta_Y) m(\mathrm{d}x \times \mathrm{d}\tilde{x}).
\end{aligned}$$

Finally, we can then estimate the latter integral from below, and obtain

$$\begin{aligned}
\int_{\Delta_X} \kappa_{x\tilde{x}}(\Delta_Y) m(\mathrm{d}x \times \mathrm{d}\tilde{x}) &\geq m(\Delta_X) \cdot \inf_{(x,\tilde{x}) \in \Delta_X} \kappa_{x\tilde{x}}(\Delta_Y) = m(\Delta_X) \cdot \inf_{x \in X} \kappa_{xx}(\Delta_Y) \\
&= \|\mu \wedge \tilde{\mu}\| \cdot \inf_{x \in X} \|K_x \wedge \tilde{K}_x\|,
\end{aligned}$$

so the lemma is proved. □

As we have mentioned above, Lemma 2 is an analogue of Lemma 1. However, unlike in Section 3.1, here we cannot go directly from Lemma 2 to Theorem 2 and we need the following auxiliary result first:

Lemma 3. *Let $\{X_k\}_{k\in\mathbb{N}_0}$ be a family of Borel spaces and let $\Omega_n = \prod_{k=0}^n X_k$ be product spaces for any $n \in \bar{\mathbb{N}}_0$. Let further $P^0, \tilde{P}^0 \in \mathscr{P}(X_0)$ and $P^n, \tilde{P}^n : \Omega_{n-1} \to \mathscr{P}(X_n)$ for $n \in \mathbb{N}$ be initial probability measures and conditional stochastic kernels respectively. Denote:*

$$a_0 := \|P^0 \wedge \tilde{P}^0\|, \qquad a_k := \inf_{\omega_{k-1}\in\Omega_{k-1}} \|P^k_{\omega_{k-1}} \wedge \tilde{P}^k_{\omega_{k-1}}\|, \quad k \in \mathbb{N}$$

and further $\mathsf{P}^n := \bigotimes_{k=0}^n P^k$, $\tilde{\mathsf{P}}^n := \bigotimes_{k=0}^n \tilde{P}^k$. For any $n \in \mathbb{N}_0$ it holds that

$$\|\mathsf{P}^n \wedge \tilde{\mathsf{P}}^n\| \geq \prod_{k=0}^n a_k. \tag{3.6}$$

Proof. The inequality (3.6) can be proved by induction. Clearly, it holds true for $n = 0$. Suppose it holds true for some $n \in \mathbb{N}_0$, then in the setting of Lemma 2 put $X = \Omega_n$, $Y = X_{n+1}$, $\mu = \mathsf{P}^n$, $\tilde{\mu} = \tilde{\mathsf{P}}^n$, $K = P^{n+1}$ and $\tilde{K} = \tilde{P}^{n+1}$. For the product measures we obtain that $\mathsf{Q} = \mathsf{P}^{n+1}$ and $\tilde{\mathsf{Q}} = \tilde{\mathsf{P}}^{n+1}$, so (3.6) now follows immediately from (3.4). □

To prove Theorem 2 we are only left to notice that (2.2) is equivalent to (3.6) thanks to the duality argument in (A.2).

4. Discussion

Let us discuss the results obtained above and their connection to the literature. First of all, Theorem 1 extends the bounds on the total variation distance between the finite-dimensional marginals P^n and $\tilde{\mathsf{P}}^n$, obtained in Roberts and Rosenthal (2013, Theorem 1) under the Borel assumption, to the case of arbitrary measurable spaces. Its proof is inspired by the one of Tkachev and Abate (2013, Lemma 1) that focused on the Markovian case $P^1 = P^2 = \cdots$ exclusively. The work in Tkachev and Abate (2013) also emphasized the benefit of dealing with sub-σ-algebras as in Theorem 1: the perturbation constants c_k in such case are likely to be smaller than those in Theorem 2. Moreover, with focus on numerical methods, it allows dealing with kernels that may not have an integral expression. Although the bounds obtained in Roberts and Rosenthal (2013) is a special case of Theorem 1, the main focus of Roberts and Rosenthal (2013) was rather on the construction of the corresponding maximal coupling of the infinite-dimensional product measures P and $\tilde{\mathsf{P}}$. In the setting of Theorem 1 the existence of such coupling is unlikely, as even the measurability of the diagonal, required in the coupling inequality, may be violated in the case of arbitrary measurable spaces.

At the same time, in the setting of Roberts and Rosenthal (2013) (that is, under the Borel assumption) a stronger result in Theorem 2 holds true. Although the latter theorem is only focused on the bounds, Lemma 2 which is the core in the proof of Theorem 2, yields the coupling of finite-dimensional marginals P^n and $\tilde{\mathsf{P}}^n$ that trivially extends to the infinite-dimensional case by Proposition 2. With focus on bounds, to show that Theorem 2 indeed provides a less conservative estimate than Roberts and Rosenthal (2013) (or Theorem 1), let us mention that

$$2 - 2\prod_{k=0}^n \left(1 - \frac{1}{2}c_k\right) \leq \sum_{k=0}^n c_k \tag{4.1}$$

for any non-negative sequence $(c_k)_{k\in\mathbb{N}_0}$ and any $n \in \mathbb{N}_0$, which can be shown by induction over n.[2] In particular, there are several cases when bounds in Theorem 2 are exact: e.g. when kernels P^k and $\tilde{P}^k$ are just measures on X_k which correspond to dealing with a sequence of iid random variables; see also the example in Section 5. In contrast, the bounds in Theorem 1 can grow unboundedly e.g. in case $c_k = c > 0$ for all $k \in \mathbb{N}_0$.

Finally, let us mention that it is of interest whether Theorem 2 holds true in the general case of arbitrary measurable spaces and arbitrary stochastic kernels—recall however that even in such case the corresponding maximal coupling may not exist. Although we do not provide the answer to this general question, we show that under the assumption that kernels have densities, one can obtain a version of Lemma 2 for the case of arbitrary measurable spaces: this of course implies the validity of Theorem 2 in such case as well, provided that all the kernels involved have densities.

Lemma 4. *Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be two arbitrary measurable spaces, and let measures $\mu, \tilde{\mu} \in \mathscr{P}(X, \mathfrak{X})$ and $\lambda \in \mathscr{P}(Y, \mathfrak{Y})$. If kernels $K, \tilde{K} : X \to \mathscr{P}(Y, \mathfrak{Y})$ are given by*

$$K_x(\mathrm{d}y) := k(x, y)\lambda(\mathrm{d}y), \qquad \tilde{K}_x(\mathrm{d}y) := \tilde{k}(x, y)\lambda(\mathrm{d}y)$$

where $k, \tilde{k} : X \times Y \to [0, \infty)$ are $\mathfrak{X} \otimes \mathfrak{Y}$-measurable functions, then for the product measures $\mathsf{Q} := \mu \otimes K$ and $\tilde{\mathsf{Q}} := \tilde{\mu} \otimes \tilde{K}$ the inequality (3.4) holds true.

[2] Notice also that the right-hand side of (4.1) can be regarded as the first-order term of the expansion of the product in the left-hand side.

Proof. Note that the following expressions for Q and $\tilde{\mathsf{Q}}$ hold true:

$$\mathsf{Q}(\mathrm{d}x \times \mathrm{d}y) = k(x, y)[\mu \otimes \lambda](\mathrm{d}x \times \mathrm{d}y),$$
$$\tilde{\mathsf{Q}}(\mathrm{d}x \times \mathrm{d}y) = \tilde{k}(x, y)[\tilde{\mu} \otimes \lambda](\mathrm{d}x \times \mathrm{d}y).$$

To make them comparable, let us define $\nu := \mu + \tilde{\mu}$, so that

$$\begin{aligned} \|\mathsf{Q} \wedge \tilde{\mathsf{Q}}\| &= (\mathsf{Q} \wedge \tilde{\mathsf{Q}})(X \times Y) \\ &= \int_{X\times Y} \min\left(k(x,y)\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x), \tilde{k}(x,y)\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\nu}(x)\right)[\nu\otimes\lambda](\mathrm{d}x\times\mathrm{d}y) \\ &\geq \int_{X\times Y} \min\left(k(x,y), \tilde{k}(x,y)\right)\min\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x), \frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\nu}(x)\right)[\nu\otimes\lambda](\mathrm{d}x\times\mathrm{d}y) \\ &= \int_X (K_x \wedge \tilde{K}_x)(Y)(\mu\wedge\tilde{\mu})(\mathrm{d}x) \geq \|\mu\wedge\tilde{\mu}\| \cdot \inf_{x\in X}\|K_x \wedge \tilde{K}_x\| \end{aligned}$$

as desired. □

5. Example

To enlighten the theoretical results obtained above, we study how conservative are the bounds in Theorem 2 on a simple example for which the analytical expression for the total variation distance is available. For that purpose, we consider a Markov Chain with just two states, that is $X = \{0, 1\}$ and $P^k = P$ for all $k \in \mathbb{N}$. The kernel P is given by a stochastic matrix, which for simplicity we assume to be diagonal: $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The initial distribution, denoted by $P^0 = \mu$, is further given by $\mu(\{1\}) = p$, $\mu(\{0\}) = 1 - p$. We focus on the case when the perturbed stochastic process is a Markov Chain as well, that is $\tilde{P}^k = \tilde{P}$ for all $k \in \mathbb{N}$, and assume that $\tilde{P} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$ for some $\varepsilon \in (0, 1)$. The perturbed initial distribution we denote by $\tilde{P}^0 = \tilde{\mu}$, it is given by $\mu(\{1\}) = p - \delta$ and $\mu(\{0\}) = 1 - (p - \delta)$; here $\delta \in (p - 1, p)$.

Since the product space $\Omega_n = \{0, 1\}^{n+1}$ is finite, the precise value of the total variation distance between the product measures P^n and $\tilde{\mathsf{P}}^n$ can be computed as

$$\left\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\right\| = \sum_{\omega_n\in\Omega_n} \left|\mathsf{P}^n(\{\omega_n\}) - \tilde{\mathsf{P}}^n(\{\omega_n\})\right|.$$

Thanks to the special choice of P, the measure P^n is supported only on two points in Ω_n which makes it is easy to obtain an analytic expression for the sum above. Let us abbreviate: $f_n(\varepsilon) := 1 - (1 - \varepsilon)^n$. There are three possible cases depending on the interplay between parameters p, ε, δ and n:

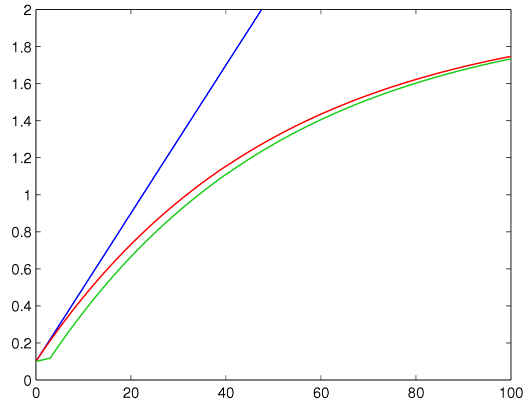
a. If $\delta < -\frac{pf_n(\varepsilon)}{1-f_n(\varepsilon)}$, then $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| = 2(1 - p)f_n(\varepsilon) - 2\delta(1 - f_n(\varepsilon))$.
b. If $-\frac{pf_n(\varepsilon)}{1-f_n(\varepsilon)} \leq \delta \leq \frac{(1-p)f_n(\varepsilon)}{1-f_n(\varepsilon)}$, then $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| = 2f_n(\varepsilon)$.
c. If $\delta > \frac{(1-p)f_n(\varepsilon)}{1-f_n(\varepsilon)}$, then $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| = 2pf_n(\varepsilon) + 2\delta(1 - f_n(\varepsilon))$.

According to Theorem 2, the bounds are $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\| \leq 2(f_n(\varepsilon) + \delta(1 - f_n(\varepsilon)))$. Clearly, in all cases [a.], [b.] and [c.] the bounds hold true. Moreover, if only the stochastic matrix is perturbed, that is $\delta = 0$ (case [b.]), then the total variation is $2f_n(\varepsilon)$ which precisely coincides with the bounds provided by Theorem 2. Similarly, if $p = 1$ and $\delta \in (0, 1)$ (case [c.]), then the total variation $2(f_n(\varepsilon) + \delta(1 - f_n(\varepsilon)))$ provides another example when bounds in Theorem 2 are exact. In all other cases it is easy to see that there is a strictly positive gap in the inequality (2.2).

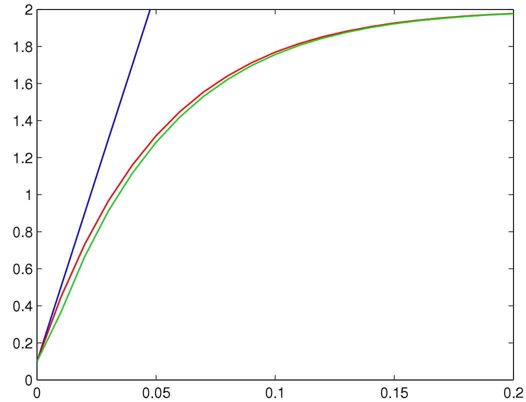
Unrelated to the precision of bounds in Theorem 2, it is interesting to further comment on the phenomenon in case [b.]. If p, ε and δ are fixed, then there exists $N \in \mathbb{N}$ such that for $n \geq N$ always [b.] holds: indeed, it follows directly from the fact that the denominator $1 - f_n(\varepsilon) = (1 - \varepsilon)^n \to 0$ monotonically as $n \to \infty$. As a result, regardless of the value of $\delta \in (0, 1)$, for all n big enough $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\|$ does not depend on δ: in particular, it is the same as in the case when $\delta = 0$.

To illustrate the discussion in this section more concretely, in Fig. 1 below we compare the error bounds in (2.1) and (2.2) with each other, and with the precise value of the total variation as functions of n, ε, δ and p based on the example above. The default values of the parameters are as follows: $n = 20$, $\varepsilon = 0.02$, $\delta = 0.05$ and $p = 0.2$. For example, in Fig. 1(c) we depict the dependence with respect to δ which varies in the interval $[-0.2, 0.2]$, whereas the other parameters n, ε and p are fixed to their default values.

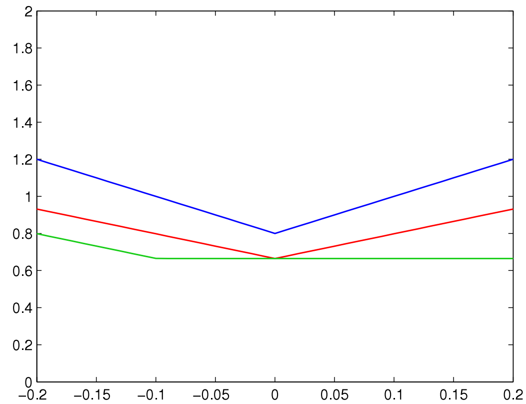
As expected, on all the figures the inequality (2.1) is more conservative than (2.2), which in turn provides an upper bound on the precise value of the total variation. In particular, Fig. 1(a) supports our observation that $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\|$ (the bottom green line) does not depend on δ for n big enough and would in fact coincide with the right-hand side of (2.2) computed for $\delta = 0$. In turn, Fig. 1(c) illustrates a somehow dual statement: for a fixed n the value of $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\|$ does not depend on δ if the latter is small enough and in particular it is the same as for $\delta = 0$, where it also coincides with the right-hand side of (2.2).



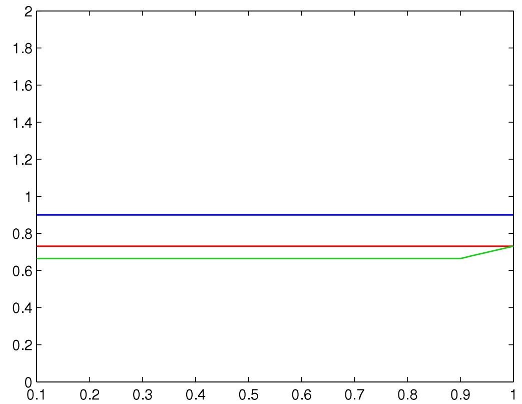
(a) Error bounds against the time horizon n.



(b) Error bounds against the value of ε.



(c) Error bounds against the value of δ.



(d) Error bounds against the value of p.

Fig. 1. Dependence of error bounds (2.1) (the top blue line) and (2.2) (the middle red line), and precise values of the total variation (the bottom green line) against n, ε, δ and p.

A similar effect can be found on Fig. 1(d): clearly, neither the right-hand side of (2.1) nor that (2.2) does depend on p, whereas the dependence of $\|\mathsf{P}^n - \tilde{\mathsf{P}}^n\|$ on p is only non-trivial when p is close to 1. Finally, Fig. 1(b) enlightens the fact that (2.1) is a linear approximation of (2.2) at $\varepsilon = 0$.

Acknowledgements

The last author would like to thank George Lowther for the idea of non-linear bounds, Michael Greinecker for his hints on dealing with kernels, and other users of MathStackexchange and MathOverflow for their valuable comments regarding measure theory and coupling. Authors are grateful to the anonymous referee for the insightful comments and suggestions.

Appendix

For any set X the corresponding diagonal is denoted by $\Delta_X := \{(x, x) : x \in X\} \subseteq X^2$. The set of all real numbers is denoted by $\mathbb{R}$ and the set of all natural numbers is denoted by $\mathbb{N}$. We further write $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\bar{\mathbb{N}}_0 := \mathbb{N}_0 \cup \{\infty\}$. For any $A \subseteq X$ we denote its indicator function by 1_A.

We say that X is a (standard) Borel space if X is a topological space homeomorphic to a Borel subset of a complete separable metric space. Any Borel space is assumed to be endowed with its Borel σ-algebra $\mathfrak{B}(X)$. An example of a Borel space is the space of real numbers $\mathbb{R}$ endowed with the Euclidean topology.

Given a measurable space $(X, \mathfrak{X})$ we denote the space of all σ-additive finite signed measures on it by $\mathrm{b}\mathscr{M}(X, \mathfrak{X})$. For any set $A \in \mathfrak{X}$ we introduce an evaluation map $\theta_A : \mathrm{b}\mathscr{M}(X, \mathfrak{X}) \to \mathbb{R}$ given by $\theta_A(\mu) := \mu(A)$. We always assume that $\mathrm{b}\mathscr{M}(X, \mathfrak{X})$ is endowed with the smallest σ-algebra that makes all evaluation maps measurable. The subspace of all probability measures in $\mathrm{b}\mathscr{M}(X, \mathfrak{X})$ is denoted by $\mathscr{P}(X, \mathfrak{X})$. In case X is a Borel space, we simply write $\mathrm{b}\mathscr{M}(X)$ and $\mathscr{P}(X)$ in place of $\mathrm{b}\mathscr{M}(X, \mathfrak{B}(X))$ and $\mathscr{P}(X, \mathfrak{B}(X))$ respectively. If $(Y, \mathfrak{Y})$ is another measurable space, by a bounded kernel we mean a measurable map $K : X \to \mathrm{b}\mathscr{M}(Y, \mathfrak{Y})$ such that

$$\sup_{x \in X} \sup_{A \in \mathfrak{Y}} |K_x(A)| < \infty$$

where we write K_x instead of a more cumbersome $K(x) \in \mathrm{b}\mathscr{M}(Y, \mathfrak{Y})$ for any $x \in X$. In case $K_x \in \mathscr{P}(Y, \mathfrak{Y})$ for any $x \in X$, we say that the kernel K is stochastic. The condition that $K : X \to \mathrm{b}\mathscr{M}(Y, \mathfrak{Y})$ is measurable is equivalent to $K_{(\cdot)}(A) : X \to \mathbb{R}$ being a measurable map for any $A \in \mathfrak{Y}$ (Kallenberg, 1997, Lemma 1.37). Clearly, any measure can be considered as a kernel which does not depend on its first argument. Furthermore, any measure $\nu \in \mathrm{b}\mathscr{M}(X, \mathfrak{X})$ admits the unique Hahn–Jordan decomposition given by $\nu = \nu^+ - \nu^-$, where ν^+ and ν^- are two mutually singular non-negative measures on $(X, \mathfrak{X})$ (Ash, 1972, Section 2.1.2). If $K : X \to \mathrm{b}\mathscr{M}(Y, \mathfrak{Y})$ is a kernel, then

$$\|K\| := \sup_{x \in X} \left(K_x^+(Y) + K_x^-(Y)\right)$$

defines the total variation of K. For any two measures $\nu, \tilde{\nu} \in \mathrm{b}\mathscr{M}(X, \mathfrak{X})$ we denote

$$\nu \wedge \tilde{\nu} := \nu - (\nu - \tilde{\nu})^- \in \mathrm{b}\mathscr{M}(X, \mathfrak{X}). \tag{A.1}$$

In particular, it holds that

$$\|\nu \wedge \tilde{\nu}\| = 1 - \frac{1}{2}\|\nu - \tilde{\nu}\| = 1 - \sup_{A \in \mathfrak{X}} |\nu(A) - \tilde{\nu}(A)|, \tag{A.2}$$

for any two probability measures $\nu, \tilde{\nu} \in \mathscr{P}(X, \mathfrak{X})$.

If $f : X \to Y$ is a measurable map between two measurable spaces $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$, for any $\nu \in \mathrm{b}\mathscr{M}(X, \mathfrak{X})$ the image measure $f_*\nu \in \mathrm{b}\mathscr{M}(Y, \mathfrak{Y})$ is given by

$$(f_*\nu)(A) := \nu\left(f^{-1}(A)\right) \tag{A.3}$$

for all $A \in \mathfrak{Y}$. Given a family of measurable spaces $\{(X_k, \mathfrak{X}_k)\}_{k \in \mathbb{N}_0}$ and an index set $I \subseteq \bar{\mathbb{N}}_0$, we denote the product measurable space by

$$\prod_{k \in I}(X_k, \mathfrak{X}_k) := \left(\prod_{k \in I} X_k, \bigotimes_{k \in I} \mathfrak{X}_k\right),$$

where $\bigotimes_{k \in I} \mathfrak{X}_k$ is the product σ-algebra. Let $\tilde{I} \subseteq I$ and denote $(\Omega, \mathscr{F}) := \prod_{k \in I}(X_k, \mathfrak{X}_k)$ and $(\tilde{\Omega}, \tilde{\mathscr{F}}) := \prod_{k \in \tilde{I}}(X_k, \mathfrak{X}_k)$. Let further $\pi : \Omega \to \tilde{\Omega}$ be an obvious projection map. For any $\nu \in \mathrm{b}\mathscr{M}(\Omega, \mathscr{F})$ we say that $\pi_*\mu$ is the marginal of μ on $(\tilde{\Omega}, \tilde{\mathscr{F}})$.

References

Ash, R., 1972. Real Analysis and Probability. In: Probability and Mathematical Statistics, No. 11. Academic Press, New York.

Bertsekas, D.P., Shreve, S.E., 1978. Stochastic Optimal Control: The Discrete Time Case, Vol. 139. Academic Press.

Kallenberg, O., 1997. Foundations of Modern Probability. In: Probability and its Applications, Springer-Verlag, New York.

Lindvall, T., 1992. Lectures on the Coupling Method. In: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Revuz, D., 1984. Markov Chains, second ed. North-Holland Publishing, Amsterdam.

Roberts, G.O., Rosenthal, J.S., 2013. A note on formal constructions of sequential conditional couplings. Statist. Probab. Lett. 83 (9), 2073–2076. http://dx.doi.org/10.1016/j.spl.2013.05.012.

Tkachev, I., Abate, A., 2013. Formula-free finite abstractions for linear temporal verification of stochastic hybrid systems. In: Proceedings of the 16th International Conference on Hybrid Systems: Computation and Control. HSCC'13. ACM, New York, NY, USA, pp. 283–292. http://dx.doi.org/10.1145/2461328.2461372.