SÉMINAIRE D'ANALYSE CONVEXE

Montpellier, 19xx, exposé n° y

ON PROHOROV'S THEOREM FOR TRANSITION PROBABILITIES *

Erik BALDER

1 Introduction

In this paper we settle the open problem of proving Prohorov's theorem for transition probabilities (alias Young measures) in its most general form, viz. in the case where the transition probabilities act into a completely regular Suslin space. Among others, the present author and C. Castaing have considered this problem (in [9, p. 5.17] it is said to seem "difficult").

Versions of Prohorov's theorem for transition probabilities, giving a sufficient condition for the relative weak compactness of a set of transition probabilities acting from a fixed measure space into a topological space S, have been known for some time. The first such versions can be found in the work of Young [24] and McShane [17] in the calculus of variations, and of Wald [22] and LeCam [15] in statistics. Later developments include work of Warga [23], Renyi [19] and Berliocchi-Lasry [8] and the author. A recent treatment of these developments was given by Valadier [21].

In recent years the present author has extended the scope of this theorem by developing a theory of weak convergence for transition prob-

*Written while visiting the Dipartimento di Matematica, Università di Genova in May, 1988. The author is grateful to the department, and in particular to Professor J.P. Cecconi, for generous support

abilities around it. When applied to limit and existence problems in optimal control theory, mathematical economics and game theory, this has produced superior results in several instances ([1, 2, 5, 6]. Yet for some applications the degree of generality of the method, while substantial when compared to more conventional methods of analysis, was not quite satisfactory. For instance, for his proof of the Prohorov-type theorem in [1, 2] the present author required S to be a metrizable Lusin space, and later Luschgy [16], Jawhar [14] and Castaing [9] followed this example. Subsequently, to handle certain applications with nonmetrizable S, the author gave in [4, 6] an ad hoc extension of [1, 2] which requires that S is a countable union of metrizable Lusin spaces. In contrast, important surrounding parts of the theory were clearly and naturally valid for a completely regular Suslin space S. This led to the question whether Prohorov's theorem for transition probabilities could also be valid under these more basic conditions ([3, p. 468], [9, p. 5.17]).

The author has recently already proven that for *sequential* relative compactness this question has an affirmative answer. In a development which is quite unrelated to the work discussed thus far, he used Komlós' theorem [7]. This resulted in what is recapitulated here as Theorem 2.2, a result which is actually of a more general nature than the common Prohorov-type theorem. However, the nature of this approach is strictly confined to the sequential case. In this paper we shall pursue the more traditional approach, and show that also for ordinary, nonsequential relative compactness a Prohorov-type result holds with the intended degree of generality. The analysis used for this is based on the following viewpoint. Rather than using the defining property of the Suslin space S (S being the continuous image of a Polish space), we exploit the well-known – if rarely used – fact that on the completely regular Suslin space S a "weak" metric d can be introduced, whose topology is not finer than the original topology. We then show that, thanks to the available inf-compactness provided by the tightness condition, the initial problem can be carried over completely to the framework where the *metric* Suslin space (S, d) replaces S. With one modification the results of [2] then apply.

2 Main result

Let S be a completely regular Hausdorff space, and let $\mathcal{P}(S)$ be the set of all Radon probability measures on $(S, \mathcal{B}(S))$ [11, 20]. Recall that the *narrow* (alias *weak*) topology on $\mathcal{P}(S)$ [11, III.54] is the initial topology with respect to all functionals

$$\nu \to \int_S c(s)\, \nu(ds), c \in \mathcal{C}_b(S).$$

Here $\mathcal{C}_b(S)$ denotes the set of all bounded continuous functions $c : S \to$ R. Let $\mathcal{K}(S)$ denote the collection of all compact subsets of S. Recall that a subset $\mathcal{P}_0$ of $\mathcal{P}(S)$ is said to be *tight* if for every $\epsilon > 0$ there exists $K_\epsilon \in \mathcal{K}(S)$ such that

$$\sup_{\nu \in \mathcal{P}_0} \nu(S \backslash K_\epsilon) \leq \epsilon.$$

Prohorov's classical theorem states the following ([11, III.59], [20, Appendix,Thm. 3]):

Theorem 2.1 (Prohorov) *Suppose that $\mathcal{P}_0 \subset \mathcal{P}(S)$ is tight. Then $\mathcal{P}_0$ is relatively narrowly compact.*

Let us remark that if S is a Polish (separable, metric and complete) space, the converse (i.e., relative narrow compactness implies tightness) is also true [20, Appendix, Thm. 4]. This same fact continues – trivially – to hold for the Prohorov-type theorem below (see [13, 2]).

Let $(T, \mathcal{T}, \mu)$ be an abstract complete σ-finite measure space. We define the set $\mathcal{R}(T; S)$ to consist of all functions $\delta : T \to \mathcal{P}(S)$ which are *transition probabilities* from $(T, \mathcal{T})$ into $(S, \mathcal{B}(S))$ [18, III.2], i.e., which are such that for every $B \in \mathcal{B}(S)$ the function $t \mapsto \delta(t)(B)$ is $\mathcal{T}$-measurable.

From now on we shall suppose that the completely regular Hausdorff space S is also a *Suslin* space, i.e. a Hausdorff space which is the continuous image of some Polish space ([11, III.67], [20, II]). This entails that every probability measure on $(S, \mathcal{B}(S))$ is Radon [11, III.69]. It is not hard to verify that in this framework $\mathcal{R}(T; S)$ is precisely the set of all functions from T into $\mathcal{P}(S)$ that are measurable with respect to $\mathcal{T}$ and the Borel σ-algebra on $\mathcal{P}(S)$ for the narrow topology (cf. [10, p.

103]). (It is of some interest to note that $\mathcal{P}(S)$ itself is a Suslin space for the narrow topology [20, Appendix, Thm. 7].)

Let $\mathcal{G}_C(T;S)$ be the set of all *Carathéodory integrands* on $T \times S$, i.e., the set of all $\mathcal{T} \times \mathcal{B}(S)$-measurable functions $g : T \times S \to \mathrm{R}$ such that $g(t,\cdot)$ is continuous on S for every $t \in T$ and such that for some μ-integrable function $\phi : T \to \mathrm{R}$

$$\sup_{s\in S} |g(t,s)| \leq \phi(t) \text{ for all } t \in T. \tag{2.1}$$

The functional $I_g : \mathcal{R}(T;S) \to \mathrm{R}$ corresponding to $g \in \mathcal{G}_C(T;S)$ is defined by

$$I_g(\delta) := \int_T [\int_S g(t,s)\, \delta(t)(ds)] \mu(dt). \tag{2.2}$$

The integral in (2.2) is well-defined [18, III.2]. The *weak* topology on $\mathcal{R}(T;S)$ is defined to be the initial topology with respect to all functionals $I_g : \mathcal{R}(T;S) \to \mathrm{R}, g \in \mathcal{G}_C(T;S)$. Note that in case T is a singleton (and $\mu(T) > 0$), the space $\mathcal{R}(T;S)$ with the weak topology and the space $\mathcal{P}(S)$ with the narrow topology can be identified.

We shall now state our Prohorov-type theorem, which provides a sufficient condition for the relative weak compactness of subsets of $\mathcal{R}(T;S)$. Let $\mathcal{H}(T;S)$ be the set of all $\mathcal{T} \times \mathcal{B}(S)$-measurable functions $h : T \times S \to [0,+\infty]$ such that for every $t \in T$ the function $h(t,\cdot)$ is inf-compact on S (i.e., for every $\beta \in \mathrm{R}$ the set $\{s \in S : h(t,s) \leq \beta\}$ belongs to $\mathcal{K}(S)$). A subset $\mathcal{R}_0$ of $\mathcal{R}(T;S)$ is defined to be *tight* if

$$\sup_{\delta\in\mathcal{R}_0} I_h(\delta) < +\infty, \tag{2.3}$$

where $I_h(\delta)$ is defined in complete analogy to (2.2).

Theorem 2.2 *Suppose that $\mathcal{R}_0 \subset \mathcal{R}(T;S)$ is tight. Then $\mathcal{R}_0$ is relatively weakly compact and relatively weakly sequentially compact.*

For purposes of comparison we also state a generalization of the sequential part of the above theorem, obtained in [7, Thm. 5.1]. In turn, this result follows as a special case from an abstract sequential compactness result for scalarly integrable functions [7, Thm. 2.1].

Theorem 2.3 ([7]) *Suppose that for* $\mathcal{R}_0 \subset \mathcal{R}(T;S)$ *the following tightness condition holds: there exists a function* $h : T \times S \to [0, +\infty]$ *such that* $h(t, \cdot)$ *is inf-compact on* S *and*

$$\sup_{\delta \in \mathcal{R}_0} \int_T^* [\int_S h(t,s)\, \delta(t)(ds)] \mu(dt) < +\infty,$$

where $\int_T^*$ *stands for outer integration over* T. *Then to every sequence* (δ_k) *in* $\mathcal{R}_0$ *there correspond a subsequence* (δ_{k_j}) *and* $\delta_* \in \mathcal{R}(T;S)$ *such that for every sub-subsequence* (δ_{k_i}) *of* (δ_{k_j})

$$\textit{narrow-}\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{k_i}(t) = \delta_*(t)\ \mu - a.e.\ in\ T.$$

In particular, this implies that (δ_{k_j}) *converges weakly to* δ_* *in* $\mathcal{R}(T;S)$. *So* $\mathcal{R}_0$ *is relatively weakly sequentially compact.*

Note that in Theorem 2.3 no measurability is required for the function h, which explains why outer integration is employed there.

From our next remark it will become obvious that Theorem 2.1 generalizes Prohorov's theorem (note that if $(T, \mathcal{T}, \mu)$ is trivial, the Suslin condition for S becomes extraneous).

Remark 2.4 *The definition of tightness for* $\mathcal{R}_0$ *given in (2.3) is equivalent to the following: for every* $\epsilon > 0$ *there exists a multifunction* $\Gamma_\epsilon : T \to \mathcal{K}(S)$, *having a* $\mathcal{T} \times \mathcal{B}(S)$*-measurable graph* $\{(t,s) \in T \times S : s \in \Gamma_\epsilon(t)\}$, *such that*

$$\sup_{\delta \in \mathcal{R}_0} \int_T \delta(t)(S \backslash \Gamma_\epsilon(t)) \mu(dt) \leq \epsilon$$

(cf. [14, Prop. 2.2]). Indeed, the above condition entails the existence of a sequence (Γ_n) *of multifunctions* $\Gamma_n : T \to \mathcal{K}(S)$ *corresponding to* $\epsilon = 3^{-n}$. *Without loss of generality we can take* (Γ_n) *nondecreasing (pointwise). Then setting*

$$h(t,s) := \begin{cases} 2^n & \text{if } s \in \Gamma_\epsilon(t) \backslash (\cup_{m=1}^{n-1} \Gamma_m(t)), n \in N \\ +\infty & \text{if } s \in S \backslash (\cup_{n=1}^{\infty} \Gamma_n(t)) \end{cases}$$

defines a function $h \in \mathcal{H}(T;S)$ *for which (2.3) obviously holds. Conversely, if* $h \in \mathcal{H}(T;S)$ *satisfies the original tightness condition (2.3),* *then let $\sigma \in R$ be the value of the supremum in (2.3). Define for $\epsilon > 0$ the multifunction $\Gamma_\epsilon : T \to \mathcal{K}(S)$ by*

$$\Gamma_\epsilon(t) := \{s \in S : h(t,s) \leq \frac{\sigma}{\epsilon}\}.$$

Then evidently the alternative tightness condition also holds.

3 Proof of the main result

Since S is a completely regular Suslin space, it follows from [10, III.31] that $\mathcal{C}_b(S)$ contains a countable subset (c_j) of functions taking values in $[0,1]$, which separates the points of S (i.e., if $x \neq y$ then $c_j(x) \neq c_j(y)$ for some j). Therefore, a *weak* metric d on S is defined by

$$d(x,y) := \sum_{j=1}^{\infty} 2^{-j} |c_j(x) - c_j(y)|.$$

Of course, the d-topology is weaker than the original topology on S. From now on we shall denote S by (S,o) when we want to emphasize the fact that S is equipped with its original topology, and by (S,d) when S is provided with the metric d. Clearly, (S,d) is a (metric) Suslin space. Since its topology is weaker than the original one, it follows that for the Borel σ-algebras

$$\mathcal{B}(S,o) = \mathcal{B}(S,d) \tag{3.1}$$

by [20, Cor. 2 of Thm. II.10]. Since (S,d) is at least of the same type as (S,o), all the aforegoing definitions and notations carry naturally over to the framework in which (S,o) is replaced by (S,d). The following inclusions are obvious consequences of the fact that the topology on (S,d) is weaker than the one on (S,o):

$$\mathcal{K}(S,o) \subset \mathcal{K}(S,d), \tag{3.2}$$

$$\mathcal{G}_C(T;(S,d)) \subset \mathcal{G}_C(T;(S,o)). \tag{3.3}$$

From (3.1)-(3.2) it also follows that, with a converse orientation,

$$\mathcal{H}(T;(S,o)) \subset \mathcal{H}(T;(S,d)). \tag{3.4}$$

From (3.1) we have that, as a *set*, $\mathcal{R}(T;(S,o))$ coincides with $\mathcal{R}(T;(S,d))$. Also, it follows from (3.3) that the weak topology on $\mathcal{R}(T;(S,o))$ is not coarser than the weak topology on $\mathcal{R}(T;(S,d))$.

Let $h \in H(T;(S,o))$ be as in the definition of tightness of the set $\mathcal{R}_0$; in particular, let σ be the supremum in (2.3). Define $\mathcal{R}_1$ to be the set of all $\delta \in \mathcal{R}_1$ such that $I_h(\delta) \leq \sigma$; then $\mathcal{R}_0 \subset \mathcal{R}_1$, so it is enough to prove weak (sequential) compactness of $\mathcal{R}_1$ in $\mathcal{R}(T;(S,o))$. We claim now that on $\mathcal{R}_1$ the weak topologies of $\mathcal{R}(T;(S,o))$ and $\mathcal{R}(T;(S,d))$ coincide. We have already seen that one inclusion is true. To show the other, recall from [2, Lemma A.2] that the weak topology on $\mathcal{R}(T;(S,d))$ is the coarsest topology for which all functionals $I_g : \mathcal{R}(T;(S,d)) \to [0,+\infty], g \in \mathcal{G}^+(T;(S,d))$, are lower semicontinuous. Here $\mathcal{G}^+(T;(S,d))$ stands for the set of all nonnegative *normal integrands* on $T \times (S,d)$, i.e. the set of all $T \times \mathcal{B}(S)$-measurable functions $g : T \times S \to [0,+\infty]$ such that $g(t,\cdot)$ is lower semicontinuous on (S,d) for every $t \in T$. (Note that [2, Lemmas A.1-A.2] continue to hold for the metrizable Suslin space (S,d).) Now given an arbitrary $g \in \mathcal{G}_C(T;(S,o))$, with associated ϕ as in (2.1), observe that for every $\delta \in \mathcal{R}_1$

$$I_g(\delta) = \sup_{\epsilon>0}[I_{g_\epsilon}(\delta) - \int_T \phi d\mu - \epsilon\sigma], \tag{3.5}$$

where $g_\epsilon \in \mathcal{H}(T;(S,o)) \subset \mathcal{H}(T;(S,d)) \subset \mathcal{G}^+(T;(S,d))$ (see (3.3-(3.4)) is defined by

$$g_\epsilon(t,s) := g(t,s) + \epsilon h(t,s) + \phi(t).$$

Therefore, it follows from the above that the right hand side of (3.5) forms a functional on $\mathcal{R}_1$ which is semicontinuous for the weak topology of $\mathcal{R}(T;(S,d))$. This argument can be repeated for $-g$ instead of g, and leads to the conclusion that I_g is also continuous on $\mathcal{R}_1$ for the weak topology of $\mathcal{R}(T;(S,d))$. This shows that on $\mathcal{R}_1$ the weak topology of $\mathcal{R}(T,(S,o))$ is not finer than the weak topology of $\mathcal{R}(T;(S,d))$. Therefore, they must coincide.

It is now obviously sufficient to prove the weak (sequential) compactness of $\mathcal{R}_1$ with respect to the weak topology for $\mathcal{R}(T;(S,d))$. If the metric space (S,d) were *Lusin*, this would directly follow from [2, Appendix] (or from [1]), in view of (3.4). However, since this is not so we must extend the argument of [1b] slightly. Note that in [2, Appendix A] (S,d) is embedded in a compact metric space, denoted here by (K,ρ). Since (S,d) is metric and Suslin this is still possible [11, III]. The extension $\hat{h} : T \times K \to [0,+\infty]$ of h in [1] and [2, p. 593] is given by

$$\hat{h} := \begin{cases} h & \text{on } T \times S \\ +\infty & \text{on } T \times (K \backslash S) \end{cases}$$

We wish to prove that $\hat{h}$ is $\mathcal{T} \times \mathcal{B}(K)$-measurable (for then it follows immediately that $\hat{h} \in \mathcal{H}(T;K)$, and the proof of [2, Thm. I] takes over). Let $\beta \in \mathrm{R}$ be arbitrary. Denote the set $\{(t,s) \in T \times K : \hat{h}(t,s) \leq \beta\}$ by C. Then, evidently, $C = \{(t,s) \in T \times S : h(t,s) \leq \beta\}$. Define $C(t)$ to be the section of C at t, $t \in T$; then clearly $C(t) \in \mathcal{K}(S)$ by our second expression for C. Since the embedding of (S,d) into (K,ρ) is continuous, it follows that $C(t) \in \mathcal{K}(K)$ for every $t \in T$, so in particular $C(t)$ is ρ-closed in K. Define $u : T \times K \to [0, +\infty]$ by

$$u(t,s) := \rho\text{-dist}(s, C(t)) := \inf_{s' \in C(t)} \rho(s,s').$$

(The ρ-distance to the empty set equals $+\infty$ by definition.) Evidently, $u(t,\cdot)$ is continuous on K for every $t \in T$. Also, for every $s \in K$ the function $u(\cdot,s)$ is $\mathcal{T}$-measurable on T, because for every $\alpha \in \mathrm{R}$

$$\{t \in T : u(t,s) < \alpha\} = \text{proj}_T\{(t,s') \in C : \rho(s,s') < \alpha\}.$$

The set of which the projection is taken on the right belongs to $\mathcal{T} \times \mathcal{B}(S)$, so it follows from a well-known projection theorem [10, III.23] that the set on the left is $\mathcal{T}$-measurable, where we use the facts that S is Suslin and $\mathcal{T}$ is complete for μ. The $\mathcal{T} \times \mathcal{B}(K)$-measurability of u follows now directly from applying [10, III.14]. By ρ-closedness of the sections $C(t)$, shown above, it follows that $C = \{(t,s) \in T \times K : u(t,s) = 0\}$, so we conclude that $\hat{h}$ is $T \times \mathcal{B}(K)$-measurable. Thus, the proof of [2, Thm. I] continues to go through in our case of a metrizable *Suslin* space (S,d), and this gives that $\mathcal{R}_1$ is weakly compact and weakly sequentially compact in $\mathcal{R}(T;(S,d))$. (Observe that the fact that in [2] μ is finite – as opposed to σ-finite – offers no obstacle whatsoever. The only reason why in [2] μ is taken to be finite is the frequent occurrence of uniform integrability.) Thus, Theorem 2.2 has been proven.

Let us just mention a few applications of Theorem 2.2, next to the obvious applications it has to the problems considered in [2, 4, 5, 6]. The relative compactness criterion in generalized Köthe spaces of [10, Thm. V.13] follows directly from Theorem 2.2, as does Diestel's theorem [12]. In [7] the latter result (and actually a rather substantial extension of it) was shown to follow from Theorem 2.3. Also, in [7] a sequential variant of the former result was derived from Theorem 2.3. In both cases a nonsequential adaptation can now be derived using Theorem 2.2. We refer to [7] and forthcoming work of the author for further details.

References

[1] BALDER, E.J., On a useful compactification for optimal control problems, J. Math. Anal. Appl. **72** (1979), 391-398.

[2] BALDER, E.J. A general approach to lower semicontinuity and lower closure in optimal control theory, SIAM J. Control Optim. **22** (1984), 570-598.

[3] BALDER, E.J. A general denseness result for relaxed control theory, Bull. Austral. Math. Soc. **30** (1984), 463-475.

[4] BALDER, E.J. An extension of Prohorov's theorem for transition probabilities with applications to infinite-dimensional lower closure problems, Rend. Circ. Mat. Palermo (II) **34** (1985), 427-447.

[5] BALDER, E.J. Generalized equilibrium results for games with incomplete information, Math. Oper. Res. **13** (1988), 265-276.

[6] BALDER, E.J. Fatou's lemma in infinite dimensions, J. Math. Anal. Appl. **136** (1988), 450-465.

[7] BALDER, E.J. New sequential compactness results for spaces of scalarly integrable functions, Preprint no. 488, Department of Mathematics, Utrecht, 1987, J. Math. Anal. Appl. (1990), to appear.

[8] BERLIOCCHI, H. and LASRY, J.-M. Intégrandes normales et mesures paramétreés en calcul des variations, Bull. Soc. Math. France **101** (1973), 129-184.

[9] CASTAING, C. Compacité dans l'espace des mesures de probabilité de transition, Séminaire d'Analyse Convexe Montpellier, **16** (1986), 5.1-5.19.

[10] CASTAING, C. and VALADIER, M. Convex Analysis and Measurable Multifunctions, Lecture Notes in Mathematics **580**, Springer-Verlag, Berlin, 1977.

[11] DELLACHERIE, C. and MEYER, P.-A. Probabilités et Potentiel, Hermann, Paris, 1975.

[12] DIESTEL, J. Remarks on weak compactness in $L_1(\mu, X)$, Glasgow Math. J. **18** (1977), 87-91.

[13] JACOD, J. and MEMIN, J. Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité, Séminaire de Probabilités XV. Lecture Notes in Mathematics **850**, Springer-Verlag, Berlin, 1981.

[14] JAWHAR, A. Mesures de transition et applications, Séminaire d'Analyse Convexe Montpellier, **14** (1984), 13.1-13.61.

[15] LECAM, L.M. An extension of Wald's theory of statistical decision functions, Ann. Math. Stat. **26** (1955), 69-81.

[16] LUSCHGY, H. Invariante Statistische Entscheidungsprobleme, Habilitationsschrift, University of Münster, 1984.

[17] MCSHANE, E.J. Generalized curves, Duke Math. J. **6** (1940), 513-536.

[18] NEVEU, J. Bases Mathématiques du Calcul des Probabilités, Masson, Paris, 1964.

[19] RENYI, A. On stable sequences of events, Sankhya, Ser. A. **25** (1963), 293-302.

[20] SCHWARTZ, L. Radon Measures, Oxford University Press, London, 1973.

[21] VALADIER, M. Young Measures, C.I.M.E. Lecture Notes, Springer-Verlag, Berlin, to appear.

[22] WALD, A. Statistical Decision Functions, Wiley, New York, 1950.

[23] WARGA, J. Optimal Control of Differential and Functional Equations, Academic Press, New York, 1972.

[24] YOUNG, L.C. Lectures on the Calculus of Variations and Optimal Control Theory, Saunders, Philadelphia, 1969.