Implicit Markov kernels in probability theory

Daniel Hlubinka

Abstract. Having Polish spaces $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ we shall discuss the existence of an $\mathbb{X}\times\mathbb{Y}$-valued random vector (ξ,η) such that its conditional distributions $\mathrm{K}_x=\mathcal{L}(\eta \mid \xi=x)$ satisfy $e(x,\mathrm{K}_x)=c(x)$ or $e(x,\mathrm{K}_x)\in C(x)$ for some maps $e:\mathbb{X}\times\mathcal{M}_1(\mathbb{Y})\to\mathbb{Z}$, $c:\mathbb{X}\to\mathbb{Z}$ or multifunction $C:\mathbb{X}\to 2^{\mathbb{Z}}$ respectively. The problem is equivalent to the existence of universally measurable Markov kernel $\mathrm{K}:\mathbb{X}\to\mathcal{M}_1(\mathbb{Y})$ defined implicitly by $e(x,\mathrm{K}_x)=c(x)$ or $e(x,\mathrm{K}_x)\in C(x)$ respectively. In the paper we shall provide sufficient conditions for the existence of the desired Markov kernel. We shall discuss some special solutions of the (e,c)- or (e,C)-problem and illustrate the theory on the generalized moment problem.

Keywords: Markov kernels, universal measurability, selections, moment problems, extreme points

Classification: 28A35, 28B20, 46A55, 60A10, 60B05

1. Introduction

Markov kernels are an important tool of modern probability theory. They can be used, for example, as a model of the conditional structure of a random vector. Since it is possible to construct a two-dimensional distribution from a given properly measurable Markov kernel and a one-dimensional distribution, a natural question arises: "Under which conditions is it possible to construct a two-dimensional random vector with one given marginal distribution such that its conditional distributions satisfy given requirements?". We shall consider in what follows, that the requirements on the sought Markov kernel can be written implicitly. In our study we shall show that the general existence conditions are quite mild then, and that under some additional assumptions it is possible to look for a special solution which exists as well.

Consider two Polish (separable, completely metrizable) spaces $\mathbb{Y}$ and $\mathbb{Z}$. The problem is to find a Borel probability measure $\mu\in\mathcal{M}_1(\mathbb{Y})$ satisfying an implicit condition $e(\mu)=c$, where the *evaluating map* $e:\mathcal{M}_1(\mathbb{Y})\to\mathbb{Z}$ and the *control value* $c\in\mathbb{Z}$ are given. A typical example is the moment problem

$$\left\{\mu\in\mathcal{M}_1(\mathbb{Y}) : e(\mu)=\textstyle\int_{\mathbb{Y}} g(y)\mu(dy)=c\right\}.$$

Supported by the project MSM 113200008. The author would like to appreciate the help and support of Josef Štěpán.

Consider a third Polish space $\mathbb{X}$ which will be the space of initial conditions. Then we can assume that both evaluating and control values depend on the value $x \in \mathbb{X}$ which represents a deterministic initial condition here. We obtain sets

$$\mathcal{P}_x := \{\mu \in \mathcal{M}_1(\mathbb{Y}) : e(x,\mu) = c(x)\} \tag{1}$$

$$\mathcal{P} := \{(x,\mu) \in \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) : e(x,\mu) = c(x)\}. \tag{2}$$

The set $\mathcal{P}_x$ contains all probability measures obeying (e,c) requirement for fixed initial condition x and it is called *set of admissible solutions for fixed* x, while the set $\mathcal{P}$ is called *set of admissible solutions*. It is clear that the set $\mathcal{P}_x$ is Borel whenever the map $e(x,\cdot) : \mu \mapsto e(x,\mu)$ is Borel, or it is measurable in the same sense as the map e is measurable in general. The measurability of $\mathcal{P}$ is not clear at all, but it will prove to be a crucial question in the sequel, since we need to find a measurable selection for the set $\mathcal{P}$.

It is quite natural to consider $\mathcal{P}_x$ as a set of possible conditional distributions of a random variable η given x and, for a distribution $\lambda \in \mathcal{M}_1(\mathbb{X})$ representing the stochastic initial condition, to ask about the existence of a random vector (ξ,η) such that for the distributions of the vector it holds

$$\mathcal{L}(\xi) = \lambda;\ \mathcal{L}(\eta \,|\, \xi = x) \in \mathcal{P}_x. \tag{3}$$

This is equivalent to the construction of a probability measure P^λ on $\mathbb{X} \times \mathbb{Y}$ such that

$$\begin{aligned} P^\lambda(B \times \mathbb{Y}) &= \lambda(B),\ B \in \mathcal{B}(\mathbb{X}) \\ P^\lambda(A \,|\, x) &= P_x(A), \quad \text{for some} \quad P_x \in \mathcal{P}_x,\ A \in \mathcal{B}(\mathbb{Y}). \end{aligned} \tag{4}$$

Under the assumption that $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are Polish the existence of P^λ is equivalent to the existence of a universally measurable Markov kernel (UMK) $\mathrm{K} : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$ such that $\mathrm{K}(x) = P_x \in \mathcal{P}_x$ a.s. $[\lambda]$. Then the measure P^λ constructed by

$$P^\lambda(C) = \int_{\mathbb{X}} P_x(C_x)\lambda(dx), C \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

where $C_x = \{y \in \mathbb{Y} : (x,y) \in C\}$ are the sections of C, satisfies conditions (4).

Note that the definition of UMK as a universally measurable map $\mathrm{K} : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$ is equivalent to the usual one (see e.g. Lemma 1.37 of [3]):

$\mathrm{K} : \mathbb{X} \times \mathcal{B}(\mathbb{Y}) \to [0,1]$ such that $x \mapsto \mathrm{K}(x,\cdot)$ is a universally measurable map and $B \mapsto \mathrm{K}(\cdot,B)$ is a Borel probability measure on $\mathbb{Y}$.

We shall not distinguish these two definitions in what follows.

Recall that both, the Markov kernel $\mathrm{K} : x \mapsto P_x$ together with distribution λ, and measure P^λ induced by P_x and λ, define the same $\mathbb{X} \times \mathbb{Y}$-valued random vector (ξ, η) with joint distribution $\mathcal{L}(\xi, \eta) = P^{\lambda}$. It holds further that any two-dimensional random vector (ξ, η) specifies a probability measure P on $\mathbb{X} \times \mathbb{Y}$ and $\lambda = \mathcal{L}(\xi)$ on $\mathbb{X}$, and a Markov kernel $\mathrm{K}(x) = \mathcal{L}(\eta \,|\, \xi = x)$ as well. It is therefore equivalent to speak about a Markov kernel K, probability measure P^{λ}, or a two-dimensional random vector (ξ, η) being a solution to a given (e, c)-problem with initial condition $\lambda \in \mathcal{M}_1(\mathbb{X})$.

In Section 2 we will start with the general theory of multi-valued maps and measurable selections. The theory is used to prove existence theorems for implicitly defined measurable maps in Section 3. In Corollary 5.1 of Section 3, a sufficient condition for the existence of a two-dimensional random element (ξ, η) in the Polish space $\mathbb{X} \times \mathbb{Y}$ with conditional distributions satisfying

$$e(x, \mathcal{L}(\eta \,|\, \xi = x)) = c(x), \text{ a.s. } [\mathcal{L}(\xi)],$$

is provided. The result is generalized for a multi-valued map $C : \mathbb{X} \to 2^{\mathbb{Y}}$ and conditions

$$e(x, \mathcal{L}(\eta \,|\, \xi = x)) \in C(x), \text{ a.s. } [\mathcal{L}(\xi)]$$

in Corollary 6.1. In Section 4, the existence of solutions with largest possible support is studied using the CS-condition of [6]. Finally the existence and possible extremal solution is discussed for an affine map e in Section 5. The discussed problem is illustrated in Section 6 on generalized moment and barycentre problems, but many other applications are possible.

2. Multifunctions and selection theorems

In this section we shall recall basic facts about techniques which are useful for the proof of our main results, namely the existence theorem for measurable selections. We profit mainly from the theory of multi-valued maps, and the theory of Souslin spaces (see [4], [2], [1]).

In what follows $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$ will always be Polish spaces, and $\mathcal{F}, \mathcal{G}, \mathcal{B}, \mathcal{U}$ will denote classes of all closed, open, Borel and universally measurable sets, respectively. We shall denote by $\mathcal{A}$ the class of all analytic subsets of $\mathbb{X}$, i.e., of all projections $\pi_{\mathbb{X}}(B)$ of Borel sets $B \in \mathcal{B}(\mathbb{X} \times \mathbb{X})$ (see [4]). Recall that the space $\mathcal{M}_1(\mathbb{X})$ of all Borel probability measures with weak topology is Polish again.

A *multi-valued map* (MVM) is a map $\Phi : \mathbb{X} \to 2^{\mathbb{Z}}$; we shall also write $\Phi : \mathbb{X} \rightrightarrows \mathbb{Z}$. If $\Phi(x) \neq \emptyset$, $\forall x \in \mathbb{X}$ we shall call Φ a *multifunction*. The notions of continuity and measurability of MVM are needed. A multifunction $\Phi : \mathbb{X} \rightrightarrows \mathbb{Z}$ is called *upper semicontinuous* (USC), *lower semicontinuous* (LSC) if

$$\Phi^w(F) := \{x : \Phi(x) \cap F \neq \emptyset\} \in \mathcal{F}(\mathbb{X}) \ \ \forall F \in \mathcal{F}(\mathbb{Z}), \text{ or} \tag{5}$$

$$\Phi^w(G) := \{x : \Phi(x) \cap G \neq \emptyset\} \in \mathcal{G}(\mathbb{X}) \ \ \forall G \in \mathcal{G}(\mathbb{Z}), \text{ respectively.} \tag{6}$$

The map Φ^w is called the *weak inverse* of Φ, and we can define the *strong inverse* Φ^s as

$$\Phi^s(A) := \complement\Phi^w(\complement A) = \{x : \Phi(x) \subset A\}, \quad A \subset \mathbb{Z}.$$

We shall say that a multifunction $\Phi : \mathbb{X} \rightrightarrows \mathbb{Z}$ is *strongly B-*, *U-measurable* if

$$\Phi^w(B) \in \mathcal{B}(\mathbb{X}) \quad \forall B \in \mathcal{B}(\mathbb{Z}), \text{ or} \tag{7}$$

$$\Phi^w(B) \in \mathcal{U}(\mathbb{X}) \quad \forall B \in \mathcal{B}(\mathbb{Z}), \text{ respectively,} \tag{8}$$

upper B- (U)-measurable if

$$\Phi^w(F) \in \mathcal{B}(\mathbb{X}) \big(\mathcal{U}(\mathbb{X})\big) \quad \forall F \in \mathcal{F}(\mathbb{Z}),$$

and *lower B- (U-measurable* if

$$\Phi^w(G) \in \mathcal{B}(\mathbb{X}) \big(\mathcal{U}(\mathbb{X})\big) \quad \forall G \in \mathcal{G}(\mathbb{Z}).$$

The definitions above can be used for a multi-valued map Ψ as well. In such a case we should replace the space $\mathbb{X}$ by $\mathrm{Dom}(\Psi)$, where the domain of a MVM is defined as usually: $\mathrm{Dom}(\Psi) := \{x \in \mathbb{X} : \Psi(x) \neq \emptyset\}$. In this sense it is possible to consider only multifunctions, since any multi-valued map restricted to its domain is multifunction. We work then with sets open, closed or Borel relatively with respect to $\mathrm{Dom}(\Psi)$ and $\mathrm{Dom}(\Psi) \times \mathbb{Y}$, and continuity or measurability of a MVM Ψ is always relative to its domain $\mathrm{Dom}(\Psi)$.

Closed valued multifunctions, i.e. maps $\Phi : \mathbb{X} \rightrightarrows \mathbb{Z}$, $\Phi(x) \in \mathcal{F}(\mathbb{Z})$, $\forall x$ play the key role in the multifunction theory. We shall speak about *correspondences* (CVC) in such a case.

2.1 Multifunction and its graph. Two lemmas concerning the relation between measurability of multifunction and its graph will be useful later in Section 3. Recall that the *graph of a multifunction* Ψ is the set

$$\mathrm{Gr}\,\Psi = \{(x, y) \in \mathbb{X} \times \mathbb{Y} : y \in \Psi(x)\}.$$

Lemma 1. *The graph of an upper semicontinuous correspondence $\Psi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a closed subset of $\mathbb{X} \times \mathbb{Y}$.*
The graph of a lower semicontinuous correspondence $\Psi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a Borel subset of $\mathbb{X} \times \mathbb{Y}$.
Suppose that the correspondence $\Psi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is (strongly, upper, lower) U-measurable. Then $\mathrm{Gr}\,\Psi \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$.
Suppose that the correspondence $\Psi : \mathbb{X} \rightrightarrows \mathbb{Y}$ is (strongly, upper, lower) B-measurable. Then $\mathrm{Gr}\,\Psi \in \mathcal{B}(\mathbb{X} \times \mathbb{Y})$.

Lemma 2. *Consider a measurable multifunction $\Psi : \mathbb{X} \rightrightarrows \mathbb{Z}$. Then*

$$\begin{gathered}\Psi \textit{ is strongly B-measurable}\\ \Downarrow\\ \Psi \textit{ is upper B-measurable}\\ \Downarrow\\ \Psi \textit{ is lower B-measurable.}\end{gathered}$$

If Ψ is a U-measurable correspondence then all the above implications can be changed to equivalences.

Similar results are true for an *open valued correspondence*, where graphs of upper semicontinuous multifunctions are "only" Borel subsets of $\mathbb{X} \times \mathbb{Y}$. On the other hand, we cannot drop the assumption that all images $\Psi(x)$ are closed (or all are open). A counterexample is provided by the function $\Psi(x) = G, x \in A$ and $\Psi(x) = \overline{G}$ for $x \in \complement A$, where $G \in \mathcal{G}(\mathbb{X})$ and $A \notin \mathcal{U}(\mathbb{X})$. The multifunction is clearly lower U-measurable but its graph is not in $\mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$.

2.2 Sections and selections. The proofs employ the celebrated *Cross-sections theorem* (8.5.3, 8.5.4 of [2]) which reads:

Lemma 3 (Cross-sections theorem)**.** *Let $\mathbb{X}$, $\mathbb{Y}$ be Polish spaces. Consider that either*

(a) $A \in \mathcal{A}(\mathbb{X} \times \mathbb{Y})$, *or*
(b) $A \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$,

and $A_{\mathbb{X}}$ is the projection of A to X. Then there exists a map $f : A_{\mathbb{X}} \to \mathbb{Y}$ such that

(i) $\mathrm{Gr}(f) \subset A$ *and*
(ii) f *is* $\mathcal{U}(\mathbb{X})/\mathcal{B}(\mathbb{Y})$ *measurable.*

The "(a)" part of the lemma is called von Neumann's theorem.

Other useful theorems on selection from multifunction can be found in [1] or [4]. We state a special form of the well known Kuratowski and Ryll-Nardzewski's theorem which reads

Lemma 4 (Kuratowski and Ryll-Nardzewski)**.** *Let $\mathbb{Y}$ be a Polish space and $\mathbb{X}$ be a nonempty set. Then every U-measurable CVC $\Psi : \mathbb{X} \to \mathbb{Y}$ admits a universally measurable selection, and every B-measurable CVC admits a Borel measurable selection.*

It is now clear that we need to prove measurability of the set of admissible solutions $\mathcal{P}$, or measurability of the multi-valued map $x \rightrightarrows \mathcal{P}_x$ provided $\mathcal{P}_x$ is closed for all x, in order to answer our question.

3. Implicit measurable maps

Having Polish spaces $\mathbb{X}$ and $\mathbb{Y}$ we study conditions sufficient for the existence of a two-dimensional random element (ξ, η) such that the (e, c)-condition $e(x, \mathrm{K}_x) = c(x)$ holds for its conditional distributions $\mathrm{K}_x = \mathcal{L}(\eta \,|\, \xi = x)$ a.s. $\mathcal{L}(\xi)$. The $\mathbb{Z}$-valued maps e and c are given and $\mathbb{Z}$ is assumed to be Polish. We shall consider a more general (e, C)-condition $e(x, \mathrm{K}_x) \in C(x)$ a.s. $\mathcal{L}(\xi)$ as well.

We will start with a more general question. Consider three Polish spaces $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$, and a pair of maps $e : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ and $c : \mathbb{X} \to \mathbb{Z}$. We seek a map $k : \mathbb{X} \to \mathbb{Y}$ called here (e, c)-*selection* such that

$$e(x, k(x)) = c(x) \quad \forall\ x \in D := \{u \in \mathbb{X} : \exists y \in \mathbb{Y}, e(u, y) = c(u)\},$$

provided the set D is nonempty. The nonemptyness of D is assumed throughout the section.

Theorem 5 (Implicit measurable map)**.** *Let $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ be Polish spaces. Suppose that one of the following conditions holds:*

(i) *$e : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ is Borel measurable and continuous in second argument and $c : \mathbb{X} \to \mathbb{Z}$ is universally measurable;*
(ii) *$e : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ and $c : \mathbb{X} \to \mathbb{Z}$ are both Borel measurable.*

Then there exists a universally measurable (e, c)-selection k.

Proof: (i) We shall prove that the correspondence $\Phi(x) := \{y \in \mathbb{Y} : e(x, y) = c(x)\}$ is U-measurable, or equivalently that $\mathrm{Gr}(\Phi) \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$. Note that Φ is a correspondence since the map e is continuous in y. For $F \in \mathcal{F}(\mathbb{Y})$ it holds

$$\begin{aligned} \Phi^w(F) &= \{x : \Phi(x) \cap F \neq \emptyset\} = \{x : \exists y \in F, e(x, y) = c(x)\} \\ &= \{x : e(x, F) \ni c(x)\}, \end{aligned} \tag{9}$$

where $e(x, F) = \Psi(x) := \{e(x, y) : y \in F\}$ is a correspondence $\Psi : \mathbb{X} \rightrightarrows \mathbb{Z}$ for fixed closed F since e is continuous in y. It holds for an arbitrary fixed open $G \subset \mathbb{Z}$

$$\begin{aligned} \Psi^w(G) &= \{x : \Psi(x) \cap G \neq \emptyset\} = \{x : \exists y \in F, e(x, y) \in G\} \\ &= \mathrm{pr}_{\mathbb{X}} \big[\underbrace{\{(x, y) : e(x, y) \in G\} \cap \mathbb{X} \times F}_{\in \mathcal{B}(\mathbb{X} \times \mathbb{Y})} \big] \in \mathcal{A}(\mathbb{X}) \subset \mathcal{U}(\mathbb{X}), \end{aligned}$$

since e is Borel measurable, and the correspondence Ψ is U-measurable. Hence $\mathrm{Gr}(\Psi) \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Z})$.

Using (9) and universal measurability of c we can conclude that $\Phi^w(F) \in \mathcal{U}(\mathbb{X})$, and hence $\mathrm{Gr}(\Phi) \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$ and there exists a universally measurable map $g : \mathbb{X} \to \mathbb{Y}$ solving the (e, c)-problem.

(ii) Since the map e is Borel measurable, the map $E : (x, y) \mapsto \big(x, e(x, y)\big)$ is Borel as well. It holds $\mathrm{Gr}(c) \in \mathcal{B}(\mathbb{X}\times\mathbb{Z})$ for the graph of c since c is a Borel map. Hence

$$\begin{aligned} E^{-1}\big(\mathrm{Gr}(c)\big) &= \big\{(x, y) : \big(x, e(x, y)\big) = \big(x, c(x)\big)\big\} \\ &= \big\{(x, y) : e(x, y) = c(x)\big\} \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}). \end{aligned}$$

The existence of a universally measurable (e, c)-selection follows from the Cross-section theorem in both cases.

□

Corollary 5.1. *Let $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ be Polish spaces. Suppose that one of the following conditions holds:*

(i) *$e : \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ is Borel measurable and continuous in second argument and $c : \mathbb{X} \to \mathbb{Z}$ is universally measurable;*
(ii) *$e : \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ and $c : \mathbb{X} \to \mathbb{Z}$ are Borel measurable.*

Then there exists a universally measurable Markov kernel solving the (e, c)-problem.

We shall now generalize the results above admitting the control map to be multi-valued. We can see that the only restriction is that the multi-valued map is closed valued.

Theorem 6 (Generalized IMM)**.** *Let $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ be Polish spaces. Suppose that one of the following conditions holds:*

(i) *$e : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ is Borel measurable and continuous in second argument and $C : \mathbb{X} \rightrightarrows \mathbb{Z}$ is a U-measurable correspondence;*
(ii) *$e : \mathbb{X} \times \mathbb{Y} \to \mathbb{Z}$ is Borel measurable and $C : \mathbb{X} \rightrightarrows \mathbb{Z}$ is a B-measurable correspondence.*

Then there exists a universally measurable (e, C)-selection k.

Proof: (i) Consider the multifunction $\Psi(x) := \{y : e(x, y) \in C(x)\}$. Fix a closed subset $E \subset \mathbb{Y}$. Then

$$\begin{aligned} \Psi^w(E) &= \{x : \Psi(x) \cap E \neq \emptyset\} = \{x : \exists y \in E, e(x, y) \in C(x)\} \\ &= \{x : e(x, E) \cap C(x) \neq \emptyset\}. \end{aligned} \tag{10}$$

We shall use the proof of Theorem 5 to show that for a closed set E, the multifunction $e(x, E) := \{e(x, y) : y \in E\}$ is a U-measurable correspondence. We conclude from (10) that

$$\Psi^w(E) = \mathrm{pr}_{\mathbb{X}}\big[\mathrm{Gr}\big(e(\cdot, E)\big) \cap \mathrm{Gr}(C)\big] \in \mathcal{U}(\mathbb{X}),$$

hence Ψ is a U-measurable correspondence and $\mathrm{Gr}(\Psi) \in \mathcal{U}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$. The claim follows by the Cross-section theorem again.

(ii) The graph of a B-measurable correspondence is a Borel set. Thus

$$\{(x,y) : e(x,y) \in C(x)\} = \big\{(x,y) : \big(x, e(x,y)\big) \in \mathrm{Gr}(C)\big\} \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}),$$

since it is the inverse image of a Borel set under the Borel map $(x,y) \mapsto \big(x, e(x,y)\big)$. The assertion then follows from the Cross-section theorem. $\square$

Corollary 6.1. *Let $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ be Polish spaces. Suppose that one of the following conditions holds:*

(i) *$e : \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ is Borel measurable and continuous in second argument and $C : \mathbb{X} \rightrightarrows \mathbb{Z}$ is a U-measurable correspondence;*

(ii) *$e : \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ is Borel measurable and $C : \mathbb{X} \rightrightarrows \mathbb{Z}$ is a B-measurable correspondence.*

Then there exists a universally measurable Markov kernel solving the (e,C)-problem.

Note that from the parts (ii) of both proofs it follows immediately that in such a case

$$D = \{x \in \mathbb{X} : \exists y \in \mathbb{Y}, e(x,y) = c(x) \ [\text{ or } \ \in C(x)]\} \in \mathcal{U}(\mathbb{X}).$$

Remark 1. The random element or the Markov kernel representing its conditional structure satisfying the (e,c)- or (e,C)-condition a.s. $[\lambda]$ will be called *solution of the (e,c)-problem*, or *(e,C)-problem* respectively, with initial condition λ. Using Theorems 5 and 6 and their corollaries we can conclude that for any $\lambda \in \mathcal{M}_1(\mathbb{X})$ with $\lambda^*(\complement D) = 0$ there exists a solution of the (e,c)-problem under mild conditions on measurability of e and c. Any such measure $\lambda \in \mathcal{M}_1(\mathbb{X})$ will be called an *admissible initial condition* for the given (e,c)-problem.

The preceding theorems should be read as
"If there is a subset $D \subset \mathbb{X}$ such that $\mathcal{P}_x := \{\mu \in \mathcal{M}_1(\mathbb{Y}) : e(x,\mu) = c(x)\} \neq \emptyset$ for $x \in D$, then there exists a UMK solving the (e,c)-problem and the kernel is defined on the set D, being universally measurable w.r.t. $\mathcal{U}(D) := D \cap \mathcal{U}(\mathbb{X})$."
Unfortunately, we are not able to solve the problem of existence of such a nonempty set D — *set of admissible initial conditions* — in the general case. This question has to be answered for given e, c and C separately.

4. Solution with largest support

In papers [6] and [5], *rich $\mathcal{P}$-vectors* are studied. Recall that in our notation a *rich solution* is a solution (ξ,η) of the (e,c)-problem for which

$$\mathrm{supp}\big(\mathcal{L}(\eta \,|\, \xi = x)\big) \supset \mathrm{supp}\big(\mathcal{L}(\eta' \,|\, \xi' = x)\big) \text{ a.s. } [\mathcal{L}(\xi)]$$

holds for any random element $(\xi',\eta') : \mathcal{L}(\xi') = \mathcal{L}(\xi)$ solving the (e,c)-problem. We have denoted by $\mathrm{supp}(\mu)$ the *support* of μ, which is the smallest closed set with μ measure 1.

It is proved in Theorem 2 of [6] that the *CS-condition*

$$\forall\Big(x \in \mathbb{X}, (\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_x\Big)\ \exists\Big(\alpha_n > 0, \sum_1^\infty \alpha_n = 1\Big) : \sum_1^\infty \alpha_n\mu_n \in \mathcal{P}_x$$

is sufficient for the existence of an (e,c)-selection which is a rich solution of the given (e,c)-problem for any admissible initial condition λ. We exhibit sufficient conditions on e and c (C) such that the set of admissible solutions $\mathcal{P}$ satisfies the CS-condition.

Theorem 7. (i) *Let maps* $e : \mathbb{X} \times \mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ *and* $c : \mathbb{X} \to \mathbb{Z}$ *be given. If* e *is continuous and affine in the second variable for any* x, *then the set of admissible solutions* $\mathcal{P}$ *of the given* (e,c)*-problem satisfies the CS-condition, and hence there exists a rich solution.*

(ii) *Let a map* $e : \mathbb{X}\times\mathcal{M}_1(\mathbb{Y}) \to \mathbb{Z}$ *and a multifunction* $C : \mathbb{X} \rightrightarrows \mathbb{Z}$ *be given. If* e *is continuous and affine in the second variable for any* x, *and* $C(x)$ *is a closed convex set for any* x, *then the set of admissible solutions* $\mathcal{P}$ *of the given* (e,C)*-problem satisfies the CS-condition, and hence there exists a rich solution.*

Proof: (i) Consider an arbitrary sequence (μ_n) such that for some fixed $x \in \mathbb{X}$ and for all n it holds $e(x,\mu_n) = c(x)$. Then

$$\begin{aligned} e(x,\mu) &= e\Big(x, \textstyle\sum_{n=1}^\infty \alpha_n\mu_n\Big) = \lim_{N\to\infty} e\Big(x, \textstyle\sum_{n=1}^N \alpha_n\mu_n\Big) \\ &= \lim_{N\to\infty} \textstyle\sum_{n=1}^N \alpha_n e(x,\mu_n) = \lim_{N\to\infty} \textstyle\sum_{n=1}^N \alpha_n c(x) = c(x) \end{aligned}$$

holds for $\mu = \sum_{n=1}^\infty \alpha_n\mu_n$, where $\alpha_n \geq 0\ \forall n$ and $\sum_{n=1}^\infty \alpha_n = 1$.

(ii) We can repeat the argument of the preceding proof. The only difference is that $e(x,\mu_n) = c_n(x) \in C(x)$ and

$$\begin{aligned} e(x,\mu) &= e\Big(x, \textstyle\sum_{n=1}^\infty \alpha_n\mu_n\Big) = \lim_{N\to\infty} e\Big(x, \textstyle\sum_{n=1}^N \alpha_n^N \mu_n\Big) \\ &= \lim_{N\to\infty} \underbrace{\textstyle\sum_{n=1}^N \alpha_n^N \overbrace{e(x,\mu_n)}^{\in C(x)}}_{\in C(x)} \in C(x), \end{aligned}$$

where $\alpha_n^N = \alpha_n / \sum_{i=1}^N \alpha_n$ for $n \leq N$. □

Corollary 7.1. *Suppose that a UMK* $\mathrm{K} : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$ *is a rich solution a.s.* $[\lambda]$ *for a given* (e,c)*- or* (e,C)*-problem. Then for any other solution* $\mathrm{L} : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$ *it holds* $\operatorname{supp}(\mathrm{L}^\lambda) \subset \operatorname{supp}(\mathrm{K}^\lambda)$.

Proof: Denote $N := \{x : \operatorname{supp}(\mathrm{L}(x)) \not\subset \operatorname{supp}(\mathrm{K}(x))\}$. Note that $\lambda(N) = 0$. Fix an arbitrary $(x, y) \in \operatorname{supp}(\mathrm{L}^\lambda)$. Then for any open balls U_x, U_y it holds

$$(11) \qquad 0 < \mathrm{L}^\lambda(U_x \times U_y) = \int_{U_x \setminus N} \mathrm{L}(z)(U_y)\lambda(dz) + \underbrace{\int_{U_x \cap N} \mathrm{L}(z)(U_y)\lambda(dz)}_{0}.$$

We need to prove

$$\mathrm{K}^\lambda(U_x \times U_y) = \int_{U_x \setminus N} \mathrm{K}(z)(U_y)\lambda(dz) > 0,$$

but this holds since $\mathrm{L}(z)(G) > 0 \Rightarrow \mathrm{K}(z)(G) > 0$ for open $G \subset \mathbb{Y}$ and $z \notin N$ and since $\mathrm{L}(z)(U_y) > 0$ for all $z \in S \subset U_x \setminus N$ and $\lambda(S) > 0$, as follows from (11). $\square$

We have seen that any $[\lambda]$ rich solution (ξ, η) inherits the richness property also for the two-dimensional distribution whenever $\mathcal{L}(\xi) \ll \lambda$. The property can be useful when looking for solutions with distributions covering an area as large as possible.

5. Affine evaluating map

We have seen in the previous sections that the affinity of the evaluating map is sufficient for the existence of a rich solution of the (e, c)-problem. We will show that the affinity can also provide conditions under which the set of admissible solutions is nonempty. We profit from convexity of the space of Borel probability measures on a Polish space.

5.1 Existence of solution. It is not hard to prove that the set $\mathcal{P}_x := \{\mu : e(x, \mu) = c(x)\}$ is convex provided the evaluating map $e : (x, \mu) \mapsto z$ is affine in the second variable. Indeed, if

$$e(x, \alpha\mu + (1 - \alpha)\nu) = \alpha e(x, \mu) + (1 - \alpha)e(x, \nu), \quad \forall \mu, \nu \in \mathcal{M}_1(\mathbb{Y}), \quad \forall \alpha \in [0, 1],$$

then it easily follows that

$$\mu \in \mathcal{P}_x \;\&\; \nu \in \mathcal{P}_x \Rightarrow \alpha\mu + (1 - \alpha)\nu \in \mathcal{P}_x \quad \forall \alpha \in [0, 1].$$

An extreme point of a convex set A is any point $x \in A$ such that $A \setminus \{x\}$ is again convex. It is not surprising that extreme points can characterize a closed compact convex set in a Polish space. On the other hand, some convex sets have no extreme points at all, the open unit sphere in $\mathbb{R}^3$ for example. We shall denote the set of all extreme points of a convex set A by $\operatorname{ex} A$.

Let us denote by $\operatorname{co} A$ the *convex hull* of a set A, which is the smallest convex set containing A. The *closed convex hull* $\overline{\operatorname{co}} A$ is the smallest closed convex set containing the set A. It is clear that

$$\operatorname{co} \operatorname{ex} A \subset \operatorname{co} A = A$$

for any convex set A. For any closed convex set A it holds even

$$\overline{\operatorname{co}} \operatorname{ex} A \subset \overline{\operatorname{co}} A = \overline{\operatorname{co} A} = A,$$

with equality for closed compact convex sets.

A nice and useful example of a convex set is the space of probability measures $\mathcal{M}_1(\mathbb{Y})$ on Polish space $\mathbb{Y}$. The set of extreme points is the set $\{\delta_y : y \in \mathbb{Y}\}$ of all Dirac measures. It holds (see e.g. [7]) that

$$\mathcal{M}_1(\mathbb{Y}) = \overline{\operatorname{co}}\{\delta_y : y \in \mathbb{Y}\} = \overline{\operatorname{co}} \operatorname{ex} \mathcal{M}_1(\mathbb{Y}). \tag{12}$$

The preceding part can suggest sufficient conditions for $\mathcal{P}_x \neq \emptyset$. First of all note that the set of results of the evaluating map

$$E(x) := \{e(x, \mu) : \mu \in \mathcal{M}_1(\mathbb{Y})\} \tag{13}$$

is convex. This follows immediately from the convexity of $\mathcal{M}_1$ and affinity of e in μ. It is clear that in order to $\mathcal{P}_x \neq \emptyset$ it must hold $c(x) \in E(x)$. It is easy to see that

$$c(x) \in \operatorname{co}\{e(x, \delta_y) : y \in \mathbb{Y}\} \tag{14}$$

is a sufficient condition since

$$\operatorname{co}\{e(x, \delta_y) : y \in \mathbb{Y}\} = \{e(x, \mu) : \mu \in \operatorname{co}\{\delta_y : y \in \mathbb{Y}\}\} \subset E(x).$$

Indeed $\operatorname{co}\{e(x, \delta_y) : y \in \mathbb{Y}\} \subset \{e(x, \mu) : \mu \in \operatorname{co}\{\delta_y : y \in \mathbb{Y}\}\}$ since both sets are convex. For the reverse inclusion consider $\mu \in \operatorname{co}\{\delta_y\}$. Then

$$\exists \{\alpha_i\}_{i=1}^n, \{y_i\}_{i=1}^n : \alpha_i > 0, \sum_{i=1}^n \alpha_i = 1, \mu = \sum_{i=1}^n \alpha_i \delta_{y_i},$$

and since e is affine, $e(x, \mu) = \sum_i \alpha_i e(x, \delta_{y_i}) \in \operatorname{co}\{e(x, \delta_y)\}$. The last inclusion is obvious.

We have proved that it is sufficient to explore the set $\{e(x, \delta_y), y \in \mathbb{Y}\}$ and its convex hull which is according to (12) dense in $E(x)$. It follows that provided e is *affine and continuous* in μ then it is possible to consider weaker condition

$$c(x) \in \overline{\operatorname{co}}\{e(x, \delta_y) : y \in \mathbb{Y}\} \ \big(= \{e(x, \mu) : \mu \in \overline{\operatorname{co}}\{\delta_y : y \in \mathbb{Y}\}\} = E(x) \ \big).$$

Let us denote $E_{\text{ex}}(x) := \{e(x, \delta_y) : y \in \mathbb{Y}\}$. We can conclude the following existence theorem.

Theorem 8. *Assume that for a given (e, c)-problem one of the following conditions holds.*

1. *The maps $e(x, \mu)$ and $c(x)$ are Borel measurable, e is affine in μ and $c(x) \in \operatorname{co} E_{\mathrm{ex}}(x)$ for all $x \in \mathbb{X}$.*
2. *The maps $e(x, \mu)$ and $c(x)$ are Borel measurable, e is affine and continuous in μ and $c(x) \in \overline{\operatorname{co}} E_{\mathrm{ex}}(x)$ for all $x \in \mathbb{X}$.*
3. *The map $e(x, \mu)$ is Borel measurable and it is affine and continuous in μ and the map $c(x)$ is universally measurable and $c(x) \in \overline{\operatorname{co}} E_{\mathrm{ex}}(x)$ for all $x \in \mathbb{X}$.*

Then the sets $\mathcal{P}_x$ are nonempty for all $x \in \mathbb{X}$ and, hence, for an arbitrary initial condition $\lambda \in \mathcal{M}_1(\mathbb{X})$ there exists a solution of the (e, c)-problem.

Remark 2. We have restricted our attention to the family of (e, c)-problems. However, the result can be easily generalized to the family of (e, C)-problems. It is just sufficient to replace

$$c(x) \in \operatorname{co} E_{\mathrm{ex}}(x) \quad \text{by the condition} \quad C(x) \cap \operatorname{co} E_{\mathrm{ex}}(x) \neq \emptyset.$$

5.2 Extremal solution. Since we know that the sets $\mathcal{P}_x$ are convex provided e is affine in μ, we can be interested in the extreme points of these sets. This suggests also the problem of an extremal solution of the (e, c)-problem. We shall consider only (e, c)-problems obeying the assumptions of Theorem 8, so we can assume that $\mathcal{P}_x \neq \emptyset$ for all x.

Let us denote the set of all UMK's solving the (e, c)-problem by

$$\mathcal{J} := \Big\{ K : \big(\mathbb{X}, \mathcal{U}(\mathbb{X})\big) \to \big(\mathcal{M}_1(\mathbb{Y}), \mathcal{B}(\mathcal{M}_1(\mathbb{Y}))\big) : e(x, K(x)) = c(x) \Big\}.$$

Note that provided e is affine in the second variable, the set $\mathcal{J}$ is a convex set of universally measurable kernels. The natural question arises "is there any relation between extreme points of $\mathcal{P}_x$ and $\mathcal{J}$?".

Assume that the solution J is an extreme point of $\mathcal{J}$. Then

$$J = \alpha K + (1 - \alpha) L \quad \text{for some} \quad \alpha \in (0, 1), K, L \in \mathcal{J} \Rightarrow K = L = J.$$

Consider a point $z \in \mathbb{X}$ such that $J(z)$ is not an extreme point of $\mathcal{P}_z$. Then there exist two solutions K_z and L_z in $\mathcal{P}_z$ and $\alpha \in (0, 1)$ such that

$$\begin{aligned} J(z) &= \alpha K(z) + (1 - \alpha) L(z) \\ J &= \alpha K + (1 - \alpha) L, \end{aligned} \tag{15}$$

where $K(x) = L(x) = J(x), x \neq z, K(z) = K_z \neq L(z) = L_z$. The maps K and L are clearly measurable solutions of the (e, c)-problem and J is not an extreme point of $\mathcal{J}$ in contrary to the assumption.

Suppose on the other hand that $J(x) \in \operatorname{ex} \mathcal{P}_x$ for all x. Then for any $\alpha \in (0, 1)$

$$\begin{aligned} J = \alpha K + (1 - \alpha) L &\Rightarrow J(x) = \alpha K(x) + (1 - \alpha) L(x) \\ &\Rightarrow L(x) = K(x) = J(x) \quad \forall x \in \mathbb{X} \Rightarrow J = K = L. \end{aligned} \tag{16}$$

Using (15) and (16) we can conclude the following result.

Proposition 9. *Assume that for an (e, c)-problem satisfying the assumptions of Theorem* 8 *there exists a solution $J : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$. Then the solution J is an extreme point of the set of all solutions $\mathcal{J}$ if and only if $J(x)$ is an extreme point of the sets $\mathcal{P}_x$ of admissible solutions for all $x \in \mathbb{X}$.*

We have provided quite a natural characterization of extremal solutions. Since we know already that the extreme points of the space $\mathcal{M}_1$ of probability measures are Dirac measures, we can try to find another characterization of extreme points of $\mathcal{P}_x$, the sets of admissible solutions given x. A Dirac measure δ_y is the only measure which has support $\{y\}$. It means that the extreme points of $\mathcal{M}_1$ are exactly those with smallest possible support. Let us observe any solution $\mu \in \mathcal{P}_x$ with smallest support, more precisely any solution

$$\mu \in \mathcal{P}_x \quad \text{such that} \quad \forall \nu \in \mathcal{P}_x, \nu \neq \mu \Rightarrow \operatorname{supp}(\nu) \nsubseteq \operatorname{supp}(\mu).$$

We claim that any such solution is an extreme point of $\mathcal{P}_x$. Assume that μ is not. Then there exist ν_1, ν_2 in $\mathcal{P}_x$ such that for some $\alpha \in (0, 1)$

$$\mu = \alpha\nu_1 + (1 = \alpha)\nu_2 \Rightarrow \operatorname{supp}(\mu) = \operatorname{supp}(\nu_1) \cup \operatorname{supp}(\nu_2) \Rightarrow \operatorname{supp}(\nu_1) \subset \operatorname{supp}(\mu).$$

Since we have assumed that μ has minimal support we have a contradiction.

We have proved

Proposition 10. *Consider an (e, c)-problem satisfying the assumptions of Theorem* 8. *Suppose that for some $\mu \in \mathcal{P}_x$ it holds* $\operatorname{supp}(\nu) \nsubseteq \operatorname{supp}(\mu)$ *for all $\nu \in \mathcal{P}_x$. Then $\mu \in \operatorname{ex} \mathcal{P}_x$.*

6. Examples

Some examples of implicitly defined Markov kernels are provided in this section. A generalized moment problem is a typical example which leads to the implicit definition of a Markov kernel. There are other examples in barycentre or quantile problems. Extreme points of the respective problems are also studied.

6.1 Moment problems. We shall start with the modified moment problem (MMP)

$$\begin{aligned} &e(x, \mu) = \left(\textstyle\int_{\mathbb{Y}} g_i(x, y)\mu(dy)\right)_{i \in I}, \ c(x) = (c_i(x))_{i \in I}, \\ &\Rightarrow \mathcal{P}_x = \{\mu \in \mathcal{M}_1(\mathbb{Y}) : \mathrm{E}_\mu\, g_i(x, \eta) = c_i(x) \ \ \forall i \in I\}, \end{aligned} \tag{17}$$

where g_i and c_i are proper measurable for all $i \in I$, I being some index set. It is not difficult to see that according to Theorem 5 and its corollary there exists a UMK solving the generalized moment problem whenever the index set I is at most countable and both g_i and c_i are Borel measurable. There is another sufficient condition, namely c_i being universally measurable and g_i Borel measurable in

x and bounded continuous in y. The second condition on g's follows from the continuity assumption on the map $\mu \mapsto e(\cdot, \mu)$.

Under these mild conditions there exists a *measurable selection* solving the (g, c) MMP. On the other hand, it is clear that for $g(x, y) < k$ and $c(x) = k$ there is no probability measure μ such that $\mathrm{E}_\mu\, g(x, \eta) = k$. Hence we need to find conditions under which the measurable selection *does exist*.

Note that for any MMP the map $e(x, \mu)$ is affine in μ as follows from the linearity of the integral, and that we can employ the results of Section 5. From Theorem 8 it follows that there exists a nonempty solution whenever

$$c(x) \in \mathrm{co}\{e(x, \delta_y) : y \in \mathbb{Y}\} = \mathrm{co}\{g(x, y) : y \in \mathbb{Y}\},$$

and c and g are Borel measurable, or if

$$c(x) \in \overline{\mathrm{co}}\{e(x, \delta_y) : y \in \mathbb{Y}\} = \overline{\mathrm{co}}\{g(x, y) : y \in \mathbb{Y}\},$$

and $c(x)$ is universally measurable, $g(x, y)$ is Borel measurable and bounded continuous in y.

Example. Consider the modified moment problem for $\mathbb{Z} = \mathbb{Y} = \mathbb{R}$

$$(\mathrm{E}(\eta \,|\, \xi = x), \mathrm{E}(\eta^2 \,|\, \xi = x)) = (c_1(x), c_2(x)) \tag{18}$$

which is defined by maps $g(x, y) = (y, y^2)$ and $c(x) = (c_1(x), c_2(x))$, assuming c is Borel measurable. The map g is obviously Borel and hence there exists a UMK solving the MMP provided the set of admissible probability measures is nonempty. The set will be nonempty if

$$(c_1(x), c_2(x)) \in \mathrm{co}\{(y, y^2) : y \in \mathbb{R}\}.$$

Since the convex hull of $\{(y, y^2) : y \in \mathbb{R}\}$ is the interior of the parabola $z = y^2$, it follows that the set of admissible solutions for the initial condition x is nonempty if and only if $c_2(x) \geq c_1^2(x)$. We claim the "only if" part since $c_1^2(x) > c_2(x)$ means that η given $[\xi = x]$ has negative variance and this is clearly impossible. Note also that the images of extreme points of $\mathcal{M}_1(\mathbb{Y})$ are exactly extreme points of $\{e(x, \mu) : \mu \in \mathcal{M}_1(\mathbb{Y})\}$.

Let us replace the second moment by the third moment now and see what happens. The MMP is changed to

$$(\mathrm{E}(\eta \,|\, \xi = x), \mathrm{E}(\eta^3 \,|\, \xi = x)) = (c_1(x), c_2(x)) \tag{19}$$

defined by $g(x, y) = (y, y^3)$ and the convex hull of $\{(y, y^2) : y \in \mathbb{R}\}$ is obviously $\mathbb{R}^2$. We can conclude that for any Borel measurable $c(x) : \mathbb{X} \to \mathbb{R}^2$ defined on $\mathbb{X}$, the set of admissible solutions is nonempty for any initial condition x, hence for any initial condition $\lambda \in \mathcal{M}_1(\mathbb{X})$ there exists a UMK solving the problem. It should be also noted that the set of images of extremal probability measures

$$\{e(x,\delta_y) : y \in \mathbb{R}\} = \{(y,y^3) : y \in \mathbb{R}\} \neq \mathrm{ex}\{e(x,\mu) : \mu \in \mathcal{M}_1(\mathbb{R})\},$$

in other words the extremality is not preserved. □

It is not difficult to see that the *generalized modified moment problem*

$$\begin{aligned} &e(x,\mu) = \left(\textstyle\int_{\mathbb{Y}} g_i(x,y)\mu(dy)\right)_{i\in I}, \ C(x) = (C_i(x))_{i\in I}, \\ &\Rightarrow \mathcal{P}_x = \{\mu \in \mathcal{M}_1(\mathbb{Y}) : \mathrm{E}_\mu\, g_i(x,\eta) \in C_i(x) \ \ \forall i \in I\}, \end{aligned} \tag{20}$$

can be studied in a very similar way.

Extreme points of moment sets were studied in [9]. The main result is

Lemma 11 (Theorem 2.1 of [9]). *Consider measurable functions $g_1,\dots,g_n$ defined on $\mathbb{Y}$ and real numbers $c_1,\dots,c_n$. Consider the set*

$$\mathcal{Q} := \left\{\mu \in \mathcal{M}_1(\mathbb{Y}) : g_i \ \text{ is } \mu \text{ integrable, and } \textstyle\int_{\mathbb{Y}} g_i d\mu = c_i\right\}.$$

Then the set $\mathcal{Q}$ is convex and

$$\begin{aligned} \mathrm{ex}\,\mathcal{Q} = \Big\{&\mu \in \mathcal{Q} : \mu = \sum_{i=1}^m \alpha_i\delta_{y_i}, \alpha_i > 0, \sum_{i=1}^m \alpha_i = 1, y_i \in \mathbb{Y}, 1 \le m \le n+1, \\ &\text{vectors } (g_1(y_i),\dots,g_n(y_i),1), 1 \le i \le m, \ \text{ are lin. independent}\Big\}. \end{aligned} \tag{21}$$

Example (cont.). We can now use the result of the preceding lemma and study extreme points of the two moment sets defined above.

According to the lemma the support of any extreme point for two moment conditions has at most three points. Let us start with the moment problem defined by (18).

First of all note that for three different points y_1, y_2, y_3

$$\text{the rank of } \begin{pmatrix} y_1 & y_1^2 & 1 \\ y_2 & y_2^2 & 1 \\ y_3 & y_3^2 & 1 \end{pmatrix} \text{ is equal to 3,} \tag{22}$$

and hence any solution of the moment problem with exactly three points of support is extremal. Note also that for any $c_1^2 < c_2$ there exists a solution supported by the two-point set $\{-\sqrt{c_2}, \sqrt{c_2}\}$. However, it holds $-\sqrt{c_2} \notin \mathrm{supp}(\mu)$ or $\sqrt{c_2} \notin \mathrm{supp}(\mu)$ for any "three-point" solution μ. If $c_1^2 = c_2$ then there exists only the trivial solution δ_{c_1} and for $c_1^2 > c_2$ there is no solution at all.

It is possible to conclude that in this special case any solution μ is extremal if and only if

$$\forall \nu \in \mathcal{P}_x : \operatorname{supp}(\nu) \not\subset \operatorname{supp}(\mu)$$

and, hence, there is equivalence in Proposition 10. The reason is that the triple of different points (y_1, y_1^2) forms a simplex and hence there is (at most) a unique convex combination of these points resulting in (c_1, c_2).

An extremal solution of the MMP for the first and third moment is any solution with exactly three-point support $\operatorname{supp}(\mu) = \{y_1, y_2, y_3\}$ such that

$$(y_1, y_1^3), (y_2, y_2^3), \quad \text{and} \quad (y_3, y_3^3) \quad \text{do not lie on a line,}$$

and any solution with two-point supports. As a special case there is a solution with one-point support $\{c_1\}$ if $c_2(x) = c_1^3(x)$.

There is no solution with four and more points in the support in both examples. □

6.2 Barycentres. Recall that a *barycentre* $r(\mu)$ of a probability measure $\mu \in \mathcal{M}_1(\mathbb{X})$ is defined by

$$l(r(\mu)) = \int_{\mathbb{X}} l(x)\mu(dx), \quad \forall l : \mathbb{X} \to \mathbb{R}, l \in \mathbb{X}^*.$$

Any x is the barycentre of the Dirac measure δ_x concentrated on x. On the other hand, the barycentre need not exist for all probability measures.

A special case of probability measures $\mathcal{M}_1(\mathbb{Y})$, where $\mathbb{Y} = \mathcal{M}_1(\mathbb{Y}')$ and $\mathbb{Y}'$ is Polish space, is described in [8]. Any $\mu \in \mathcal{M}_1(\mathbb{Y})$ possesses the barycentre $r(\mu) \in \mathbb{Y}$. The barycentrical map

$$r : \mu \mapsto r(\mu)$$

is affine and continuous in the weak topology of $\mathcal{M}_1(\mathbb{Y})$. Using Corollary 5.1 it can be proved that for a given universally measurable map $c : \mathbb{X} \to \mathbb{Y}$ there exists a UMK $K : \mathbb{X} \to \mathcal{M}_1(\mathbb{Y})$ such that the barycentre of $K(x)$ is $c(x)$.

Note that we do not need Corollary 5.1 for the proof since in this simplest case it is sufficient to take $K(x) = \delta_{c(x)}$. The need of Corollaries 5.1 and 6.1 comes when combining two or more conditions together. An example is a measure with given barycentre and second moment. It is possible to study extremal solutions in this case as well. We use the fact that conditions on moments and barycentre are affine.

Consider a closed convex set $H \subset \mathbb{X}$. Then $\operatorname{ex} H$ is a $\mathcal{G}_\delta$ set, and any measure $\mu \in \mathcal{M}_1(\mathbb{X})$ such that $\mu(\operatorname{ex} H) = 1$ is called an *extremal measure* on H.

Proposition 12. *Consider a weakly closed set $H \subset \mathbb{Y} = \mathcal{M}_1(\mathbb{Y}')$ and a universally measurable map $c : \mathbb{X} \to H$. Then there exists a UMK K such that $r(K(x)) = c(x)$ and $K(x)$ is an extremal measure on H.*

Proof: Since $\operatorname{ex} H$ is a $\mathcal{G}_\delta$ set, the set of extremal measures is Borel. It follows that

$$\begin{aligned}\{(x,\mu) : r(\mu) = c(x), \mu(\operatorname{ex} H) = 1\}& \\ = \underbrace{\{(x,\mu) : r(\mu) = c(x)\}}_{\in\mathcal{U}(\mathbb{X})\otimes\mathcal{B}(\mathcal{M}_1(\mathbb{Y}))} \cap \underbrace{\mathbb{X}\times\mathcal{M}_1(\operatorname{ex} H)}_{\in\mathcal{B}(\mathbb{X}\times\mathcal{M}_1(\mathbb{Y}))}& \\ \in \mathcal{U}(\mathbb{X})\otimes\mathcal{B}\big(\mathcal{M}_1(\mathbb{Y})\big).&\end{aligned}$$

We have used Corollary 5.1 for the measurability of the first set. □

Example. Assume $\mathbb{Y}' = [-1,1]$, $\mathbb{Y} = \mathcal{M}_1(\mathbb{Y}')$, and let $H \subset \mathbb{Y}$ be the set of all probability measures with zero mean. Extreme points of H are δ_0 and all probability measures

$$\alpha\delta_x + (1-\alpha)\delta_y, \quad \text{where} \quad -1 \le x,y \le 1, xy < 0, \alpha = \frac{y}{y-x} \in (0,1). \tag{24}$$

Consider the map $c : \mathbb{X} \to H$ assigning to each x the required barycentre. Then, as a result of Proposition 12 there exists a UMK K such that $K(x)(\operatorname{ex} H) = 1$ and $c(x)$ is a "$K(x)$-mixture" of points of $\operatorname{ex} H$ being the barycentre of $K(x)$. Note also that H is not a simplex. Indeed, the measure

$$\begin{aligned}\frac{1}{6}\delta_{-1} + \frac{1}{3}\delta_{-1/2} + \frac{1}{3}\delta_{1/2} + \frac{1}{6}\delta_1 &= \frac{1}{3}\left(\frac{1}{2}(\delta_{-1}+\delta_1)\right) + \frac{2}{3}\left(\frac{1}{2}(\delta_{-1/2}+\delta_{1/2})\right) \\ &= \frac{1}{2}\left(\frac{1}{3}(\delta_{-1}+2\delta_{1/2})\right) + \frac{1}{2}\left(\frac{1}{3}(2\delta_{-1/2}+\delta_1)\right)\end{aligned}$$

can be written as two different convex combinations of extreme points of H.

Using (24) we can find a bijection between $\operatorname{ex} H$ and a set $D \subset \mathbb{R}^3$. We assign $(\frac{1}{2},0,0)$ to δ_0 and (α,x,y) to other extreme points. Then the Markov kernel on $\operatorname{ex} H$ can be replaced by a kernel on D. It is possible to repeat this idea for more moment conditions and for n moment conditions to find a Markov kernel on $D' \subset \mathbb{R}^{n+2}$. □

6.3 Quantiles. We have seen that for the usual moment the function $e : \mu \mapsto \int_{\mathbb{Y}} g(y)\mu(dy)$ does not depend on x. It is dependent on x for quantiles. Assume that $\mathbb{Y} = \mathbb{Z} = \mathbb{R}$ for simplicity. Then μ_α is called an α *quantile of a measure* μ if $\int_{-\infty}^{\mu_\alpha} \mu(dy) = \mu(-\infty,\mu_\alpha] = \alpha$. Note that the quantile does not need to be unique.

Proposition 13. *Consider a Polish space $\mathbb{X}$ and Borel measurable functions $g : \mathbb{X} \to \mathbb{R}$ and $c : \mathbb{X} \to [0,1]$. Then there exists a UMK $K : \mathbb{X} \to \mathcal{M}_1(\mathbb{R})$ such that a $c(x)$ quantile of K_x is equal to $g(x)$.*

Proof: We need to check measurability of the map $e : (x,\mu) \mapsto \mu(I_{(-\infty,g(x)]})$, where I is an indicator function, in order to use Theorem 9. The map $h : (x,y) \mapsto I_{(-\infty,g(x)]}$ is Borel measurable for a Borel function g, since the subgraph $\{(x,y) : y \leq g(x)\}$ of a measurable function is measurable set. Hence h is a Borel map as the indicator of a Borel set.

To check that $(x,\mu) \mapsto \int_{\mathbb{Y}} h(x,y)\mu(dy)$ is Borel measurable use the fact, that this is true for indicator functions $h = I_{A\times B}$, and that the subset of Borel measurable maps

$$\left\{f : \mathbb{X} \times \mathbb{Y} \to [0,+\infty] : (x,\mu) \mapsto \int_{\mathbb{Y}} f(x,y)\mu(dy) \ \text{ is measurable}\right\}$$

satisfies the hypothesis of the Functional Sierpińsky lemma. These two facts imply the proposition. □

References

[1] Aubin J.-A., Frankowska H., *Set-valued Analysis*, Birkhäuser, Boston, 1990.

[2] Cohn D.L., *Measure Theory*, Birkhäuser, Boston, second edition, 1993.

[3] Kallenberg O., *Foundations of modern probability*, Probab. Appl., Springer Verlag, New York, 1997.

[4] Srivastava S.M., *A Course on Borel Sets*, Graduate Texts in Mathematics, vol. 180, Springer Verlag, New York, 1998.

[5] Štěpán J., Hlubinka D., *Two-dimensional probabilities with a given conditional structure*, Kybernetika **35(3)** (1999), 367–381.

[6] Štěpán J., *How to construct a two-dimensional random vector with a given conditional structure*, in: Viktor Beneš and Josef Štěpán, Eds, *Distribution with given marginals and moment problems*, 1997.

[7] Topsøe F., *Topology and Measure*, Lecture Notes in Mathematics 133, Springer Verlag, Berlin, 1970.

[8] Winkler G., *Choquet Order and Simplices*, Lecture Notes in Mathematics 1145, Springer Verlag, Berlin, 1985.

[9] Winkler G., *Extreme points of moment sets*, Mathematics of operational research **13(4)** (1988), 581–587.

Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic

E-mail: hlubinka@karlin.mff.cuni.cz

(*Received* July 30, 2001)