On Strongly Equivalent Nonrandomized Transition Probabilities

Eugene A. Feinberg¹ and Alexey B. Piunovskiy²

Abstract

Dvoretzky, Wald and Wolfowitz proved in 1951 the existence of equivalent and strongly equivalent mappings for a given transition probability when the number of nonatomic measures is finite and the decision set is finite. This paper introduces a notion of strongly equivalent transition probabilities with respect to a finite collection of functions. This notion contains the notions of equivalent and strongly equivalent transition probabilities as particular cases. This paper shows that a strongly equivalent mapping with respect to a finite collection of functions exists for a finite number of nonatomic distributions and finite decision set. It also provides a condition when this is true for a countable decision set. According to a recent example by Loeb and Sun, a strongly equivalent mapping may not exist under these conditions when the decision set is uncountable. This paper also provides two additional counterexamples and shows that strongly equivalent mappings exist for homogeneous transition probabilities.

1 Introduction

Dvoretzky, Wald and Wolfowitz [2, 3] studied the following problem. Consider a standard Borel space $\mathbf{X}$, i.e. $\mathbf{X}$ is a one-to-one measurable image of a Borel subset of a Polish space, and consider another standard Borel space $\mathbf{A}$. Sometimes $\mathbf{X}$ is called a state space (or set) and $\mathbf{A}$ is called a decision space (or set). Let a finite number of nonatomic probability measures $\mu_1, \dots, \mu_K$ on $\mathbf{X}$ be given. We recall that a measure μ on a standard Borel space $\mathbf{X}$ is called nonatomic if $\mu(\{x\}) = 0$ for all $x \in \mathbf{X}$. We follow the terminology that a finite set is countable. Of course, if $\mathbf{A}$ is countable then $\int_{\mathbf{A}} f(a)\nu(da) = \sum_{a \in \mathbf{A}} f(a)\nu(a)$ for any measure ν and for any function f on $\mathbf{A}$. Denote by $\pi(da|x)$ a regular transition probability from $\mathbf{X}$ to $\mathbf{A}$. Since we consider only regular transition probabilities and only measurable functions, we sometimes omit the terms “regular” and “measurable.” The following two facts were proved in [3].

Proposition 1. ([3, Theorem 3.1]) *If $\mathbf{A}$ is a finite set then for any finite number K , for any measurable functions $r_k(x, a)$, $k = 1, \dots, K$, and for any transition probability π from $\mathbf{X}$ to $\mathbf{A}$, there exists a measurable mapping φ of $\mathbf{X}$ to $\mathbf{A}$ such that*

$$\int_{\mathbf{X}} \int_{\mathbf{A}} r_k(x, a) \pi(da|x) \mu_k(dx) = \int_{\mathbf{X}} r_k(x, \varphi(x)) \mu_k(dx), \quad k = 1, \dots, K.$$

¹Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, NY 11794-3600, USA.

²Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK.

Proposition 2. ([3, Theorem 3.2]) *If $\mathbf{A}$ is a finite set then for any transition probability π from $\mathbf{X}$ to $\mathbf{A}$ there exists a measurable mapping $\varphi : \mathbf{X} \rightarrow \mathbf{A}$ such that*

$$\int_{\mathbf{X}} \pi(a|x) \mu_k(dx) = \mu_k(\{x : \varphi(x) = a\}), \quad a \in \mathbf{A}, \quad k = 1, \dots, K.$$

We remark that Dvoretzky, Wald, and Wolfowitz [2, 3] considered $\mathbf{X} = \mathbb{R}^n$ but their proofs are valid for a standard Borel space $\mathbf{X}$. Let the finite integer K and measurable functions $r_k(x, a)$ on $\mathbf{X} \times \mathbf{A}$ be fixed, $k = 1, \dots, K$. According to [3], two transition probabilities π^1 and π^2 are called equivalent if

$$\int_{\mathbf{X}} \int_{\mathbf{A}} r_k(x, a) \pi^1(da|x) \mu_k(dx) = \int_{\mathbf{X}} \int_{\mathbf{A}} r_k(x, a) \pi^2(da|x) \mu_k(dx), \quad k = 1, \dots, K; \quad (1)$$

and π^1 and π^2 are called strongly equivalent if for any measurable subset B of $\mathbf{A}$

$$\int_{\mathbf{X}} \pi^1(B|x) \mu_k(dx) = \int_{\mathbf{X}} \pi^2(B|x) \mu_k(dx), \quad k = 1, \dots, K. \quad (2)$$

Thus, for any transition probability, Proposition 1 states the existence of an equivalent measurable mapping when the set $\mathbf{A}$ is finite. Proposition 2 claims the existence of a strongly equivalent measurable mapping when $\mathbf{A}$ is finite. As mentioned in [3], strongly equivalent transition probabilities are equivalent for any set of measurable functions $\{r_k\}$ on $\mathbf{A}$, but they may not be equivalent for a given set of measurable functions on $\mathbf{X} \times \mathbf{A}$.

The natural question is whether, for a given transition probability, equivalent and strongly equivalent measurable mappings exist for infinite $\mathbf{A}$. Dvoretzky, Wald, and Wolfowitz [3, Section 4] studied this question by considering a compact $\mathbf{A}$ and approximating it by finite sets. Feinberg and Piunovskiy [4, Theorem 1] proved the existence of an equivalent measurable mapping for an arbitrary Borel space $\mathbf{A}$. Loeb and Sun [6, Example 2.7] provided a surprising example when a strongly equivalent measurable mapping does not exist when $K = 2$ and $\mathbf{A} = [-1, 1]$. In particular, in this example the transition probability, for which there is no equivalent mapping, is a stochastic mixture of two mappings.

In this paper we introduce a definition of strongly equivalent transition probabilities with respect to (wrt) a finite collection of functions on $\mathbf{X} \times \mathbf{A}$. By selecting an appropriate collection of functions, strongly equivalent transition probabilities wrt to a collection become strongly equivalent in the sense of (2). By selecting another collection, strongly equivalent transition probabilities wrt to that collection become equivalent. Theorem 1 below states that if the set $\mathbf{A}$ is finite, then, for a transition probability and a finite collection of measurable functions on $\mathbf{X} \times \mathbf{A}$, there exists a mapping that is strongly equivalent wrt to that collection. Theorem 2 states that, if the transition probability does not depend on x , then the corresponding generalization of Proposition 2 holds for an arbitrary Borel space $\mathbf{A}$. Theorem 3 provides a sufficient condition for a countable set $\mathbf{A}$ when strongly equivalent wrt to a given finite collection of functions mapping exists for a given transition probability. Corollary 1 states, for a countable $\mathbf{A}$, the existence of a strongly equivalent mapping wrt a given collection of functions for a transition probability being a finite mixture of deterministic mappings. In addition, we construct examples that illustrate, for a given transition probability, the following possibilities: (i) there is no strongly equivalent mapping wrt to a function $f(x, a)$ on $\mathbf{X} \times \mathbf{A}$ when $K = 1$ (if $f = f(x)$, such a mapping exists), and (ii) equivalent and strongly equivalent mappings do not exist for a countable number of nonatomic measures.

2 Definitions and preliminary result

Consider nonempty standard Borel spaces $\mathbf{X}$ and $\mathbf{A}$. Let P be a probability measure on $\mathbf{X}$ and π be a transition probability from $\mathbf{X}$ to $\mathbf{A}$. A transition probability π is deterministic if for each $x \in \mathbf{X}$ there exists $a \in \mathbf{A}$ such that $\pi(\{a\}|x) = 1$. Each deterministic transition probability π is identified by a measurable mapping $\varphi : \mathbf{X} \rightarrow \mathbf{A}$ such that $\pi(\{\varphi(x)\}|x) = 1$ for all $x \in \mathbf{X}$.

Definition 1. Suppose a natural K and a collection of measurable functions $\{f_k(x, a)\}_{k=1,2,\dots,K}$ on $\mathbf{X} \times \mathbf{A}$ are given. Two transition probabilities π^1 and π^2 are called *strongly equivalent wrt* $\{f_k\}_{k=1,2,\dots,K}$ if for each $k = 1, 2, \dots, K$ and for any measurable function $\rho(a)$ on $\mathbf{A}$

$$\int_{\mathbf{X}} \int_{\mathbf{A}} \rho(a) f_k(x, a) \pi^1(da|x) P(dx) = \int_{\mathbf{X}} \int_{\mathbf{A}} \rho(a) f_k(x, a) \pi^2(da|x) P(dx).$$

We remark that in this paper we follow a standard agreement that a finite or infinite Lebesgue - Stieltjes integral $\int g dP$ is defined for any nonnegative measurable function g . In general, $\int g dP = \int g^+ dP - \int g^- dP$, and this integral is defined only if either $\int g^+ dP < \infty$ or $\int g^- dP < \infty$, where $g^+(x) = \max\{0, g(x)\}$ and $g^-(x) = \max\{0, -g(x)\}$. If two integrals are not defined, they are considered to be equal.

Notice that if π is a deterministic transition probability defined by a measurable mapping φ then

$$\int_{\mathbf{X}} \int_{\mathbf{A}} \rho(a) f_k(x, a) \pi(da|x) P(dx) = \int_{\mathbf{X}} \rho(\varphi(x)) f_k(x, \varphi(x)) P(dx).$$

We observe that, for $P = K^{-1} \sum_{k=1}^K \mu_k$, the strong equivalence of transition probabilities wrt $\{f_k\}_{k=1,2,\dots,K}$ for the properly selected functions f_k implies the described in the Introduction strong equivalence and equivalence in the sense of Dvoretzky, Wald and Wolfowitz [2, 3]. Indeed, since $\mu_k \ll P$, $k = 1, \dots, K$, by the Radon-Nykodým theorem there are nonnegative measurable functions g_k , $k = 1, \dots, K$, such that $\mu_k(dx) = g_k(x) P(dx)$. Thus, if transition probabilities π^1 and π^2 are strongly equivalent wrt $\{f_k\}_{k=1,2,\dots,K}$ with $f_k(x, a) = g_k(x) r_k(x, a)$ then, by setting $\rho(a) = 1$ we have that they are equivalent in the sense of Dvoretzky, Wald and Wolfowitz [3]; (see (1)). Transition probabilities π^1 and π^2 are strongly equivalent wrt $\{g_k\}_{k=1,2,\dots,K}$ if and only if they are strongly equivalent in the sense of Dvoretzky, Wald and Wolfowitz [3]; see (2).

The following theorem, whose proof is based on the method introduced in [3], states the existence of strongly equivalent mappings wrt to a given finite collection of functions.

Theorem 1. *Let $P(dx)$ be a non-atomic probability measure, $\mathbf{A}$ be finite, K be a natural number, and a collection $\{f_k(x, a)\}_{k=1,2,\dots,K}$ of measurable functions on $\mathbf{X} \times \mathbf{A}$ be given. Then, for any transition probability π from $\mathbf{X}$ to $\mathbf{A}$, there exists a measurable mapping φ such that π and φ are strongly equivalent wrt $\{f_k\}_{k=1,2,\dots,K}$.*

Proof. For any $b \in \mathbf{A}$ consider the functions $r_{bk}(x, a) = I\{a = b\} f_k(x, a)$, $k = 1, \dots, K$. Then $\{r_{bk}\}_{\{b \in \mathbf{A}, k=1,\dots,K\}}$ is a finite collections of functions and, therefore, Proposition 1 implies that $\exists \varphi$

$$\int_{\mathbf{X}} I\{\varphi(x) = b\} f_k(x, \varphi(x)) P(dx) = \int_{\mathbf{X}} f_k(x, b) \pi(b|x) P(dx), \quad b \in \mathbf{A}, k = 1, \dots, K.$$

By multiplying each side of this equality by $\rho(b)$, adding them in $b \in \mathbf{A}$, and replacing the notation b with a , we have

$$\int_{\mathbf{X}} \rho(\varphi(x)) f_k(x, \varphi(x)) P(dx) = \int_{\mathbf{X}} \sum_{a \in \mathbf{A}} \rho(a) f_k(x, a) \pi(a|x) P(dx), \quad k = 1, \dots, K.$$

□

Remark 1. $\mathbf{X}$ can be considered as a state space and $\mathbf{A}$ can be considered as an action space. Sometimes it is natural to consider the set of available actions $\mathbf{A}(x) \subseteq \mathbf{A}$ for each state $x \in \mathbf{X}$. A measurable mapping $\varphi : \mathbf{X} \rightarrow \mathbf{A}$ is called a selector if $\varphi(x) \in \mathbf{A}(x)$ for all $x \in \mathbf{X}$. The natural assumptions are that the graph $\text{Gr}_{\mathbf{X}}(\mathbf{A}) = \{(x, a) : x \in \mathbf{X}, a \in \mathbf{A}(x)\}$ of the set-valued mapping $\mathbf{A}(x)$ is a measurable subset of $\mathbf{X} \times \mathbf{A}$ and there exists a selector. A transition probability π must satisfy $\pi(\mathbf{A}(x)|x) = 1$ for all $x \in \mathbf{X}$. Functions $f_k(x, a)$, with respect to which a strong equivalence is considered, should satisfy $f_k(x, a) = 0$ for $a \notin \mathbf{A}(x)$. Though such a model looks more general, in fact, it is not more general for the purpose of this paper, because, given a transition probability π , the existence of a strongly equivalent measurable function φ wrt $\{f_k\}_{k=1,2,\dots,K}$ implies the existence of a strongly equivalent selector $\tilde{\varphi}$ wrt $\{f_k\}_{k=1,2,\dots,K}$ for the new model. Let $Y = \{x \in \mathbf{X} | \varphi(x) \notin \mathbf{A}(x)\}$. Since the mapping $g(x) = (x, \varphi(x))$ is Borel for a Borel mapping φ and the graph of a Borel function φ is a Borel set, Y is a measurable subset of $\mathbf{X}$. Let ϕ be a selector. Define the selector $\tilde{\varphi}$: $\tilde{\varphi}(x) = \varphi(x)$ for $x \in \mathbf{X} \setminus Y$ and $\tilde{\varphi}(x) = \phi(x)$ for $x \in Y$. Then the transition probability π and selector $\tilde{\varphi}$ are strongly equivalent wrt $\{f_k\}_{k=1,2,\dots,K}$.

3 Uncountable set $\mathbf{A}$

Let $\mathbf{A}$ be an uncountable standard Borel space. The following example shows that, for a given transition probability π , the strongly equivalent function may not exist. Therefore, Theorem 1 does not hold if $\mathbf{A}$ is uncountable.

Example 1. (Loeb and Sun [6, Example 2.7]) For $K > 1$, Proposition 2 cannot be extended to an uncountable $\mathbf{A}$. Let $\mathbf{X} = [0, 1]$, $\mathbf{A} = [-1, 1]$, $\mu_1(dx) = dx$, $\mu_2(dx) = 2xdx$, and $\pi(\Gamma|x) \triangleq \frac{1}{2}[I\{x \in \Gamma\} + I\{-x \in \Gamma\}]$ for any Borel subset Γ of $[-1, 1]$. In this example there is no measurable mapping $\varphi : \mathbf{X} \rightarrow \mathbf{A}$ strongly equivalent to π ; see Loeb and Sun [6, Example 2.7] for detail. Therefore, Proposition 2 cannot be extended to a continuum $\mathbf{A}$. □

Thus, when $\mathbf{A}$ is not countable, Proposition 2 and Theorem 1 do not hold for $K > 1$. Consider the case $K = 1$. Let $f = f_1$. If f does not depend on a , i.e. $f = f(x)$, then Theorem 1 holds for any standard Borel space $\mathbf{A}$. Indeed, since all Borel spaces are isomorphic to Borel subsets of $[0, 1]$, it is sufficient to consider $\mathbf{X} = \mathbf{A} = [0, 1]$. By replacing P with $\mu(dx) = f(x)P(dx)$, the statement of Theorem 1 for $K = 1$ and an arbitrary standard Borel space $\mathbf{A}$ means that any random variable on $[0, 1]$ can be presented as a measurable mapping φ of a random variable with the nonatomic distribution μ . This is a well-known fact broadly used in simulation. In particular, when μ is the Lebesgue measure, the existence of such mapping, that we denote by $\tilde{\varphi}$, was proved by Aumann [1, Lemma F]. For a general random variable X defined by a nonatomic measure μ , observe that $F(X)$ is uniformly distributed on $[0, 1]$, where F is the distribution function of X . Thus, $\varphi(x) = \tilde{\varphi}(F(x))$ is the required mapping. However, the following example shows that Theorem 1 does not hold for $K = 1$ when $f = f(x, a)$ and $\mathbf{A}$ is uncountable.

Example 2. For $K = 1$ there is no strongly equivalent mapping wrt to f depending on a . Consider the same $\mathbf{X}$, $\mathbf{A}$, and π as in Example 1. Let $f(x, a) \triangleq 2x - |a|$, and P be the Lebesgue measure on $[0, 1]$. Then, $\int_{-1}^1 \rho(a)f(x, a)\pi(da|x) = 0.5(\rho(x) \cdot x + \rho(-x) \cdot x)$ for any measurable function

$\rho(a)$ and therefore

$$\int_0^1 \int_{-1}^1 \rho(a) f(x, a) \pi(da|x) dx = \frac{1}{2} \int_0^1 [\rho(x) \cdot x + \rho(-x) \cdot x] dx. \quad (3)$$

For any measurable function φ on $[0, 1]$

$$\begin{aligned} \int_0^1 \rho(\varphi(x)) f(x, \varphi(x)) dx &= \int_0^1 \rho(\varphi(x)) I\{\varphi(x) > 0\} [2x - \varphi(x)] dx \\ &+ \int_0^1 \rho(\varphi(x)) I\{\varphi(x) \leq 0\} [2x + \varphi(x)] dx. \end{aligned} \quad (4)$$

Let φ be strongly equivalent to π wrt to $\{f\}$. Consider $\rho_1(a) \triangleq a \cdot I\{a > 0\}$. Since the lefthand parts of (3) and (4) are qual,

$$\frac{1}{2} \int_0^1 x^2 dx = \int_0^1 \varphi(x) I\{\varphi(x) > 0\} [2x - \varphi(x)] dx.$$

Hence

$$\int_0^1 I\{\varphi(x) > 0\} [x - \varphi(x)]^2 dx = \int_0^1 I\{\varphi(x) > 0\} x^2 dx - \frac{1}{2} \int_0^1 x^2 dx. \quad (5)$$

Consider $\rho_2(a) \triangleq a \cdot I\{a \leq 0\}$. Then

$$-\frac{1}{2} \int_0^1 x^2 dx = \int_0^1 \varphi(x) I\{\varphi(x) \leq 0\} [2x + \varphi(x)] dx.$$

Hence

$$\int_0^1 I\{\varphi(x) \leq 0\} [x + \varphi(x)]^2 dx = \int_0^1 I\{\varphi(x) \leq 0\} x^2 dx - \frac{1}{2} \int_0^1 x^2 dx. \quad (6)$$

Sum (5) and (6). The righthand side of this sum is 0. The lefthand side is the sum of two nonnegative numbers. Therefore,

$$\int_0^1 I\{\varphi(x) > 0\} [x - \varphi(x)]^2 dx = \int_0^1 I\{\varphi(x) \leq 0\} [x + \varphi(x)]^2 dx = 0$$

and

$$\varphi(x) = \begin{cases} x, & \text{if } \varphi(x) > 0, \\ -x, & \text{if } \varphi(x) \leq 0, \end{cases} \quad (7)$$

where in (7) and in all the relations for $\varphi(x)$ in this example we mean that they hold P -a.s. Observe that $P(\{\varphi(x) = 0\}) = P(\{x = 0\}) = 0$.

Consider

$$\rho_3(a) \triangleq \begin{cases} I\{\varphi(a) = a, a > 0\}/a, & \text{if } a \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Since the lefthand sides of (3) and (4) are qual,

$$\begin{aligned} \frac{1}{2} \int_0^1 I\{\varphi(x) = x\} dx &= \int_0^1 I\{\varphi(\varphi(x)) = \varphi(x), \varphi(x) > 0\} \left[\frac{2x}{\varphi(x)} - 1 \right] dx \\ &= \int_0^1 I\{\varphi(x) = x\} dx, \end{aligned} \quad (8)$$

where the last equality holds because, in view of (7), the following is true: (i) $\varphi(x) = x$, when $\varphi(x) \geq 0$, (ii) $I\{\varphi(x) = x\} = I\{\varphi(\varphi(x)) = \varphi(x), \varphi(x) \geq 0\}$, and (iii) $P(\{\varphi(x) = 0\}) = 0$.

Equality (8) implies that $P(\{\varphi(x) = x\}) = 0$ and, in view of (7), $P(\{\varphi(x) = -x\}) = 1$. This is impossible, because for $\rho_4(a) = I\{a < 0\}$ we have $\int_0^1 \int_{-1}^1 \rho_4(a) f(x, a) \pi(da|x) dx = \frac{1}{4}$ and $\int_0^1 \rho_4(\varphi(x)) f(x, \varphi(x)) dx = \frac{1}{2}$, where the former follows from (3) and the latter follows from (4). $\square$

Consider a particular situation when the transition probability π does not depend on x , i.e., $\pi(da|x) = \pi(da|y)$ for all $x, y \in \mathbf{X}$. Though, according to Example 1, strongly equivalent policies may not exist, they exist when the transition probability does not depend on x .

Theorem 2. *If all the measures $\mu_1, \dots, \mu_K$ are nonatomic and the transition probability π does not depend on $x \in \mathbf{X}$ then for π there exists a strongly equivalent mapping.*

Proof. Let $\pi(B) = \pi(B|x)$, where $x \in \mathbf{X}$ and B is a measurable subset of $\mathbf{A}$. We observe that $\int_{\mathbf{X}} \pi(B|x) \mu_k(dx) = \pi(B)$, $k = 1, \dots, K$. Therefore, we need to prove that there exists a measurable mapping φ of $\mathbf{X}$ to $\mathbf{A}$ such that for any $k = 1, \dots, K$ and for any measurable subset B of $\mathbf{A}$

$$\mu_k(\{x : \varphi(x) \in B\}) = \pi(B). \quad (9)$$

Since all uncountable Borel spaces are isomorphic, we assume without loss of generality that $\mathbf{A} = (0, 1]$. We shall construct a strongly equivalent mapping $\varphi : \mathbf{X} \rightarrow (0, 1]$. In particular, we shall construct a sequence of measurable step functions φ^n , $n = 1, 2, \dots$, such that $\varphi^n \searrow \varphi$ as $n \rightarrow \infty$.

First, we shall construct the function φ^1 . To do this, we consider a decision rule π^1 that takes only two values: $\frac{1}{2}$ and 1; $\pi^1(\frac{1}{2}|x) = \pi((0, \frac{1}{2}])$ and $\pi^1(1|x) = \pi((\frac{1}{2}, 1])$. According to Proposition 2, for the decision rule π^1 there exists a strongly equivalent mapping φ^1 . Let

$$X_{1,i} = \{x \in \mathbf{X} | \varphi^1(x) = \frac{i}{2}\}, \quad i = 1, 2.$$

We observe that, for $i = 1, 2$

$$\mu_k(X_{1,i}) = \pi((\frac{i-1}{2}, \frac{i}{2}]), \quad k = 1, \dots, K.$$

Let for some $n = 1, 2, \dots$, a partition $X_{n,1}, \dots, X_{n,2^n}$ of $\mathbf{X}$ be given such that all elements of this partition are measurable sets and for $i = 1, \dots, 2^n$

$$\mu_k(X_{n,i}) = \pi((\frac{i-1}{2^n}, \frac{i}{2^n}]), \quad k = 1, \dots, K. \quad (10)$$

We construct a partition $X_{n+1,1}, \dots, X_{n+1,2^{n+1}}$ of $\mathbf{X}$ such that: (i) all elements of this partition are measurable, (ii) $X_{n,i} = X_{n+1,2i-1} \cup X_{n+1,2i}$, and (iii) for $i = 1, \dots, 2^{n+1}$

$$\mu_k(X_{n+1,i}) = \pi((\frac{i-1}{2^{n+1}}, \frac{i}{2^{n+1}}]), \quad k = 1, \dots, K. \quad (11)$$

To construct such a partition, we consider an arbitrary set $X_{n,j}$, $j = 1, \dots, 2^n$. If $\pi((\frac{j-1}{2^n}, \frac{j}{2^n}]) = 0$, we set $X_{n+1,2j-1} = X_{n,j}$ and $X_{n+1,2j} = \emptyset$. If $\pi((\frac{j-1}{2^n}, \frac{j}{2^n}]) > 0$, we consider the probability measures $\mu_1^{n,j}, \dots, \mu_K^{n,j}$ on $X_{n,j}$:

$$\mu_k^{n,j}(B) = \mu_k(B \cap X_{n,j}) / \pi((\frac{j-1}{2^n}, \frac{j}{2^n}])$$

for measurable subsets B of $X_{n,j}$. We also define the probability measure $\pi^{n,j}$ on $(\frac{j-1}{2^n}, \frac{j}{2^n}]$:

$$\pi^{n,j}(D) = \pi(D \cap (\frac{j-1}{2^n}, \frac{j}{2^n}]) / \pi((\frac{j-1}{2^n}, \frac{j}{2^n}])$$

for measurable subsets D of $(\frac{j-1}{2^n}, \frac{j}{2^n}]$. The arguments used to partition the set X into $X_{1,1}$ and $X_{1,2}$ imply that the set $X_{n,j}$ can be partitioned into two measurable subsets $X_{n+1,2j-1}$ and $X_{n+1,2j}$ such that (11) holds for $i = 2j-1, 2j$. Thus, the construction of $X_{n+1,1}, \dots, X_{n+1,2^{n+1}}$ is completed.

For each $n = 1, 2, \dots$ we set

$$\varphi^n(x) = \frac{i}{2^n} \quad \text{when} \quad x \in X_{n,i}, \quad i = 1, \dots, 2^n.$$

For $n = 1$ this definition is consistent with the previously given definitions. Since $0 < \varphi^{n+1}(x) \leq \varphi^n(x)$, we can define $\varphi(x) = \lim_{n \rightarrow \infty} \varphi^n(x)$, $x \in \mathbf{X}$.

To complete the proof of the theorem, we need to prove (9). To verify (9), it is sufficient to show that for any $n = 1, 2, \dots$ and for any $i = 1, \dots, 2^n$

$$\mu_k\{\varphi(x) < \frac{i}{2^n}\} = \pi((0, \frac{i}{2^n})), \quad k = 1, \dots, K. \quad (12)$$

We observe that for $n = 1, 2, \dots$ and for $i = 1, \dots, 2^n$

$$\{x \in \mathbf{X} \mid \varphi^n(x) < \frac{i}{2^n}\} = \cup_{j=1}^{\infty} \{x \in \mathbf{X} \mid \varphi^{n+j}(x) \leq \frac{i2^j - 1}{2^{n+j}}\}$$

and for $j = 1, 2, \dots$

$$\{x \in \mathbf{X} \mid \varphi^{n+j}(x) \leq \frac{i2^j - 1}{2^{n+j}}\} \subseteq \{x \in \mathbf{X} \mid \varphi^{n+j+1}(x) \leq \frac{i2^{j+1} - 1}{2^{n+j+1}}\}.$$

Therefore, for $k = 1, \dots, K$

$$\begin{aligned} \mu_k\{\varphi(x) < \frac{i}{2^n}\} &= \lim_{j \rightarrow \infty} \mu_k\{\varphi^{n+j}(x) \leq \frac{i2^j - 1}{2^{n+j}}\} = \lim_{j \rightarrow \infty} \pi((0, \frac{i2^j - 1}{2^{n+j}}]) \\ &= \pi((0, \frac{i}{2^n})) - \lim_{j \rightarrow \infty} \pi((\frac{i2^j - 1}{2^{n+j}}, \frac{i}{2^n})) = \pi((0, \frac{i}{2^n})), \end{aligned}$$

where the second equality follows from (10). Thus, (12) is proved. $\square$

4 Countable set $\mathbf{A}$

In this section, we prove a particular case of Theorem 1 when $\mathbf{A}$ is an infinite countable set. Since, in view of Example 1, strongly equivalent mappings may not exist when $\mathbf{A}$ is uncountable, the natural question is whether strongly equivalent mappings exist for a countable set $\mathbf{A}$. We do not know the answer to this question. The following Theorem 3 states the existence of strongly equivalent mappings wrt to a finite family of functions under the additional assumption that the transition probability $\pi(\cdot|x)$ is concentrated on a finite set $B(x)$ for every $x \in X$. We remark that, if $\mathbf{A}$ is uncountable, then Example 1 demonstrates that the statements of neither Theorem 3 nor Corollary 1 are correct.

Theorem 3. Let $P(dx)$ be a non-atomic probability measure, $\mathbf{A}$ be a countable set, K be a natural number, and $\{f_k(x, a)\}_{k=1,2,\dots,K}$ be a collection of measurable functions on $\mathbf{X} \times \mathbf{A}$. If for P -almost all $x \in \mathbf{X}$ the set $B(x) \triangleq \{a \in \mathbf{A} | \pi(a|x) > 0\}$ is finite then there exists a measurable mapping strongly equivalent to π wrt $\{f_k(x, a)\}_{k=1,2,\dots,K}$.

Proof. For each finite subset B of $\mathbf{A}$, consider the measurable set $\mathbf{X}_B \triangleq \{x \in \mathbf{X} | B(x) = B\}$. Theorem 1 implies that, for the problem with the set $\mathbf{X}$ replaced with $\mathbf{X}_B$, there exists a strongly equivalent mapping wrt $\{f_k(x, a)\}_{k=1,2,\dots,K}$. Therefore, the strongly equivalent mapping φ wrt $\{f_k(x, a)\}_{k=1,2,\dots,K}$ is defined on the measurable set $Y = \cup_B \mathbf{X}_B$. This set is measurable, because there is only a countable number of terms in this union. Since $P(\mathbf{X} \setminus Y) = 0$, we can set $\varphi(x) \equiv b$ for $x \in \mathbf{X} \setminus Y$, where b is an arbitrary element of $\mathbf{A}$. $\square$

We recall that a transition probability π from a measurable set $\mathbf{X}$ to a measurable set $\mathbf{A}$ is called a finite mixture of mappings if there exist a finite number of measurable mappings $\phi_1, \dots, \phi_n$ of $\mathbf{X}$ to $\mathbf{A}$ such that, for some nonnegative numbers $\alpha_1, \dots, \alpha_n$ with their sum equal to 1, the equality $\pi(\Gamma|x) = \sum_{i=1}^n \alpha_i I\{\phi_i(x) \in \Gamma\}$ holds for all measurable subsets Γ of $\mathbf{A}$ and for all $x \in \mathbf{X}$.

Corollary 1. Let $\mathbf{A}$ be a countable set, K be a natural number, $\{f_k\}_{k=1,\dots,K}$ be measurable functions on $\mathbf{X} \times \mathbf{A}$, and π be a finite mixture of measurable mappings. Then there exists a measurable mapping strongly equivalent to π wrt $\{f_k\}_{k=1,\dots,K}$.

5 Counterexample when the number of measures μ_k is countable

Dvoretzky, Wald, and Wolfowitz [3, Examples 1 and 2] provide examples when Propositions 1 and 2 do not hold when the set of parameters k , on which the functions r_k and measures μ_k depend, is continuum; see also Ferguson [5, p. 142]. As we see from Example 1, Theorem 3, and Corollary 1, some facts on the existence of equivalent mappings may hold for countable sets and do not hold for continuum sets. The following example demonstrates that Propositions 1, 2 and Theorem 1 do not hold when k belongs to a countable set.

Example 3. Let $\mathbf{X} = [0, 1]$, $\mathbf{A} = \{0, 1\}$, and $\pi(0|x) = \pi(1|x) \equiv 1/2$. According to [7, Theorem 6.6 in Chapter II], there exists a sequence $\{f_k(x)\}_{k=1,2,\dots}$ of nonnegative bounded uniformly continuous functions on $\mathbf{X}$ such that, if for all $k = 1, 2, \dots$

$$\int_{\mathbf{X}} f_k(x) \nu_1(dx) = \int_{\mathbf{X}} f_k(x) \nu_2(dx), \quad (13)$$

then measures ν_1 and ν_2 coincide.

(i) Consider $r_k(x, a) = f_k(x) I\{a = 1\}$, $k = 1, \dots$, and suppose there exists a measurable mapping φ such that

$$\int_{\mathbf{X}} \int_{\mathbf{A}} r_k(x, a) \pi(da|x) dx = \int_{\mathbf{X}} r_k(x, \varphi(x)) dx, \quad k = 1, 2, \dots$$

i.e. in this example all μ_k are Lebesgue measures. This is equivalent to

$$\int_{\mathbf{X}} f_k(x) \frac{1}{2} dx = \int_{\mathbf{X}} f_k(x) I\{\varphi(x) = 1\} dx, \quad k = 1, 2, \dots \quad (14)$$

Therefore, the measures $\nu_1(dx) = \frac{1}{2}dx$ and $\nu_2(dx) = I\{\varphi(x) = 1\}dx$ are equal. This is impossible, because their densities $\frac{1}{2}$ and $I\{\varphi(x) = 1\}$ are not equal for all x . Thus, Proposition 1 does not hold when K is not finite.

(ii) Since (13) remains valid if f_k is multiplied by any constant c_k , we select f_k in the way $\int_{\mathbf{X}} f_k(x)dx = 1, k = 1, 2, \dots$. Then (14) is equivalent to

$$\int_{\mathbf{X}} \pi(1|x)\mu_k(dx) = \int_{\mathbf{X}} I\{\varphi(x) = 1\}\mu_k(dx), \quad k = 1, 2, \dots, \quad (15)$$

where $\mu_k(dx) = f_k(x)dx$. Since (14) is impossible, Proposition 2 does not hold when K is not finite.

(iii) Let $P(dx) = \sum_{k=1}^{\infty} (\frac{1}{2})^k \mu_k(dx)$ and $g_k(x) = \mu_k(dx)/P(dx)$. Then (15) becomes

$$\int_{\mathbf{X}} \int_{\mathbf{A}} a g_k(x) \pi(da|x) P(dx) = \int_{\mathbf{X}} \varphi(x) g_k(x) P(dx), \quad k = 1, 2, \dots$$

Since (15) is impossible, Theorem 1 does not hold either when K is countable. $\square$

Acknowledgement. This research was partially supported by NSF grant DMI-0600538.

References

- [1] R.J. Aumann, Mixed and behavior strategies in infinite games. *Ann. Math. Studies*, **53**(1964), pp. 627-650.
- [2] A. Dvoretzky, A. Wald, J. Wolfowitz, Elimination of randomization in certain problems of statistics and of the theory of games, *Proc. Nat. Acad. Sci.*, 36(1950), pp. 256-260.
- [3] A. Dvoretzky, A. Wald, J. Wolfowitz, Elimination of randomization in certain statistical decision procedures and zero-sum two-person games, *Ann. Math. Stat.*, 22(1951), pp. 1-21.
- [4] E.A. Feinberg and A.B. Piunovskiy. On Dvoretzky-Wald-Wolfowitz theorem on nonrandomized statistical decisions. *SIAM Theory Probability Appl.*, 50(2006), pp. 463-466.
- [5] T.S. Ferguson. *Mathematical Statistics. A Decision Theoretic Approach*. Academic Press, New York, 1967.
- [6] P. Loeb and Y. Sun. Purification of measure-valued maps. *Illinois J. of Mathematics*, 50(2006), pp. 747-762.
- [7] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, Providence, Rhode Island, 2005.