Local convexity, adapted experiments and the wave equation: LeCam's randomization criterion and related results.

Heinz Weisshaupt
University of Vienna, Department of Statistics
A-1010 Vienna, Universitaetsstrasse 5/3, Austria
e-mail: heinz.weisshaupt@univie.ac.at

October 21, 2010

Abstract

We provide generalizations of LeCam's randomization criterion and present related results which we call black-box criteria. We investigate the general functional analytic structure behind randomization criteria and black-box criteria. We show by a general categorical argument that randomization criteria and corresponding black-box criteria are equivalent. By abstraction to locally convex spaces we eliminate the lattice structure usually involved. The applications of our general functional analytic results range from the comparison of adapted statistical experiments to the wave equation.

MSC 2000: 46N30, 46N20, 46A03, 62B15

Keywords: Randomized decisions, differential operators, locally convex spaces, comparison of statistical Experiments, LeCam Theory, adapted experiments, wave equation, separation of convex sets

1 Introduction

We provide generalizations of LeCam's randomization criterion and related results which we call black-box criteria. The emphasis lies on the mathematical structure of these theorems and the extension of the results to the non-stochastic case of operators on Hilbert spaces. Our framework is the general setting of locally convex spaces. The applications we present range from the comparison of adapted statistical experiments to the wave equation. Concrete applications of our results in theoretical statistics are not in the focus of the paper, but applicability of our results to hidden Markov models and asymptotic statistics is indicated by the examples and remarks in chapter 6. We regard the paper to be primary a contribution to (applied) functional analysis.

We start our investigations with Theorem 2.1 which we call the (classical) randomization criterion. We call theorems analogous to Theorem 2.1 (for example in the adapted case) randomization criteria. Theorem 2.1 converts comparisons of statistical experiments based on the risks of randomized decision rules with values in arbitrary decision spaces to a comparison based on the L^1 or variational distance of randomizations of these experiments. (This will be discussed in more detail at the end of section 2.)

In the literature the randomization criterion is stated as an equivalence of several (but not always the same number of) conditions including the converse of Theorem 2.1. (Compare with [1], [2] [3], [11], [6] and [12].) We remark that only the part of the randomization criterion displayed as Theorem 2.1 is non trivial. For the sake of completeness we prove the converse of Theorem 2.1 by proving the converse of the more general Theorem 5.2 in the end of section 5.

It is possible to formulate a statement (Theorem 2.2) that is mathematically equivalent to the randomization criterion (Theorem 2.1) which does not involve arbitrary decision spaces. We will call theorems that are analogous to Theorem 2.2 *black-box criteria.* We provide a categorical result (Theorem 3.1) revealing the general equivalence between randomization criteria and black-box criteria. We provide a version of the randomization criterion (Theorem 5.2) as well as a version of the black-box criterion (Theorem 5.1) for adapted experiments. For filtered experiments, which are special cases of adapted experiments, a randomization criterion has been proved in [5].

It is very useful to keep in mind the following interpretation when dealing with black-box criteria: Consider one statistical experiment as a collection of initial states of a system and the other experiment as a collection of final states. The black-box is the unknown mechanism governing the evolution of the system. Our aim is to obtain an approximation of this mechanism. In the case of statistical experiments the approximate mechanism is modelled by a stochastic operator.

We remark that the converse of the various black-box criteria is always trivial and included in the general functional analytic Theorem 4.3 as the implication $(v) \Rightarrow (iii)$. With the exception of Theorem 4.3 randomization criteria as well as black box criteria are formulated as theorems of the form $A \Rightarrow B$ with trivial converse and not as equivalences of various statements.

There also exist black-box criteria for operators between Hilbert spaces (Theorem 7.1) and thus for dynamical systems described by generalized wave equations or other partial differential equations. In these cases the black-box is an operator between Hilbert spaces that is related to the differential operator. We will solely consider the case of the generalized wave equations (Theorem 7.3). In this case the Hilbert space norm of the solution at time t has a natural interpretation related to the energy of the wave. We will discuss this and the "philosophical" interpretation of the black-box criterion at the end of section 7.).

All the results provided in this paper (except Theorem 8.1) are derived from a functional analytic result (Theorem 4.3) which can be considered as a very general version of the black-box criterion (Theorem 2.2). The proof of this result is based on the hyperplane separation theorem, which is equivalent with the Hahn-Banach theorem. The generalization we obtain is four-fold:

1. The general functional analytic black-box criteria are free from the lattice structure of L- and the dual M-spaces, which are usually involved in the randomization criterion. This shows that from a mathematical viewpoint a lattice structure is not in the heart of the randomization criterion. (Of course the lattice structure involved in the randomization criterion is in the heart of theoretical statics since it is a fundamental structure of L-spaces of statistical experiments.)

The fact that black-box criteria without lattice structure exist makes it possible to prove results for Operators on Hilbert spaces and thus to investigate the wave equation (and other partial differential equations) instead of statistical experiments. We regard this as one of the main findings of our paper. Further it is possible to reverse the role of the family of stochastic operators and the statistical experiment.

2. The approach applies to more general loss function spaces, so that the situation of unbounded loss functions can be handled in the setting of black-box criteria. This makes it possible to "compare" (in the sense of black-box criteria) statistical experiments with respect to their moments, or gives us the possibility to "compare" statistical experiments with respect to stochastic orders of distributions since such stochastic orders are often generated by function spaces consisting of unbounded functions. (See [9], [10] and the example following Theorem B.2 in the Appendix.)

3. The family of loss functions may depend on the parameter.

4. The approach applies to the case of adapted decision problems. In the case of adapted experiments with bounded loss functions, we obtain randomization criteria as well as black-box criteria. For the filtered situation (which is a special case of the adapted situation) the randomization criterion obtained is analogous to results obtained by Norberg ([5] Theorem 1; the result of Norberg includes various equivalent conditions which are important for theoretical

statistics. Some of these equivalent conditions can also be proved in the adapted case. We did not prove them since they are not important for the illumination of the general mathematical structure of randomization criteria and their corresponding black-box criteria).

The paper is organized as follows:

The following diagram illustrates the interdependence of the main theorems of this paper

$$
\begin{array}{ccccccc}
 & 3.1 & \longrightarrow & & \text{Remark } 6.1 & \longrightarrow & \text{Examples of Section } 6 \\
 & \Downarrow & & & \uparrow & & \\
5.2 & \Longleftrightarrow & 5.1 & \Longleftarrow & 4.3 & \Longrightarrow & \text{Results of appendix } B \text{ and } C \\
\Big\Downarrow & & \Big\Downarrow & & \Downarrow & & \\
 & & & & 4.2 & \Longrightarrow & 7.1 \Rightarrow 7.2 \Rightarrow 7.3 \\
\Big\Downarrow & & \Big\Downarrow & & \Downarrow & & \\
2.1 & \Longleftrightarrow & 2.2 & \Longleftarrow & 4.1 & &
\end{array}
$$

Diagram 1

In section 2 (The classical randomization criterion) we start with definitions and considerations concerning statistical experiments, stochastic operators, decision spaces and risk. We present the nontrivial part of LeCam's randomization criterion (Theorem 2.1). In our presentation we follow the lines of LeCam [1], LeCam and Yang [3], Strasser [11], Shiryaev and Spokoiny [6] and Torgersen [12]. All these presentations are essentially equivalent by the Kakutani-representation theorem for abstract L-spaces (see Torgersen [12] Theorem 5.7.4 and Schaefer [7] Chapter 5 Section 8.5) and by the equivalence of the description of generalized decision rules (see Torgersen [12] 4.5).

Next, we present the black-box criterion (Theorem 2.2) corresponding to Theorem 2.1. We prove the equivalence of Theorem 2.1 and Theorem 2.2. Theorem 2.2 is our starting point for the abstraction to locally convex spaces in section 4. Section 2 ends with a discussion of risk, deficiency, and the Δ-distance.

In section 3 (The equivalence principle) we prove a categorical result (Theorem 3.1) which serves as a general equivalence principle for black-box criteria and randomization criteria. In section 5 we demonstrate the broad applicability of Theorem 3.1; we use it to prove the equivalence of the adapted versions of the black-box criterion (Theorem 5.1) and the adapted version of the randomization criterion (Theorem 5.2). We do not discuss the relation of the Theorem 3.1 to the theorems 2.1 and 2.2 for two reasons: Theorems 2.1 and 2.2 are just special cases of 5.2 and 5.1; considering the proof of the equivalence of the theorems 2.1 and 2.2 given in section 2 it should be intuitively clear how these theorems relate to 3.1.

In section 4 (The setting of locally convex spaces) we introduce the notions of locally convex spaces, dual spaces and polar sets. Our first result in this section, Theorem 4.1, abstracts Theorem 2.2 to the situation of locally convex spaces without additional structure. Theorem 4.3 is the heart of the paper. It generalizes Theorem 4.1 and implies 5.1, i.e., Theorem 4.3 is an "adapted" version of Theorem 4.1. With the proof of Theorem 4.3 we complete the proof of the theorems 2.1, 2.2 and 4.1. The proof of Theorem 4.3 is based on the separation theorem for convex sets (see Schaefer [7] Chapter 2 Section 9.2).

In section 5 (The case of adapted experiments) we introduce the notion of adapted statistical experiment and adapted stochastic operator. We conclude the compactness of the space of adapted stochastic operators from the compactness of the space of stochastic operators (proved in Appendix A). Then we state and prove Theorem 5.1 which is an adapted version of Theorem 2.2, i.e., it is the black-box criterion in the adapted case. We state Theorem 5.2 which is an adapted version of Theorem 2.1, i.e., it is the randomization criterion in the adapted case. We investigate in Remark 5.2 why the statement of Theorem 5.2, as well as Norbergs theorem on filtered statistical experiments ([5] Theorem 1), have to be formulated with $ba(\Omega_2, \mathcal{B})$ as the possible image of the stochastic operator $\widetilde{S}$ (in contrast to Theorem 2.1 where $\widetilde{S}(L(E)) \subseteq L(F)$ can always be achieved). We then prove the converse of Theorem 5.2. Finally we prove the equivalence of the theorems 5.1 and 5.2 by our general categorical principle provided in section 3. Generalizations of Theorem 5.1 can be found in Appendix B. They can be proved analogously to Theorem 5.1. Theorem B.2 in Appendix B is also interesting in the case that the filtration consists only of one σ-algebra, since it generalizes the black-box criterion to the case that the family of loss functions depends on the parameter. Further, Appendix B provides an example of a "comparison" (in the black-box sense) of experiments based on stochastic orders.

In section 6 (Extensions of the theory beyond adaptedness) we investigate further properties which classes of stochastic operators may posses such that for these classes of stochastic operators randomization criteria and corresponding black-box criteria still can be proved. We provide two examples which show that an extension to other classes than adapted operators is important.

In section 7 (Hilbert spaces, diagonalizable operators and the wave equation) we state and derive from Theorem 4.2 Theorem 7.1. This is a version of the black-box criterion for Hilbert spaces. From this theorem, we derive Theorem 7.2, which together with Definition 7.1, provides the right background for our result on the generalized wave equation. Note that we adjust our definitions in such a way that we do not have to introduce unbounded operators. (For the usual development of the theory of self-adjoint differential operators and the generalized wave equation, which is equivalent to our definitions, consult [14] Chapter 5.) Further we discuss in section 7 the interpretation of the generalized wave equation for Laplace operators with variable coefficients and give an interpretation of the Hilbert space norm involved in our theorem as a wave energy. We conclude section 7 with a statement on the general "philosophical" meaning of black-box criteria.

In section 8 (The finite dimensional case and Helly's theorem) we state a theorem for the case of finite dimensional vector spaces. This theorem is an immediate consequence of Helly's Theorem (see Valentine [13]). It is not a consequence of our general functional analytic approach. We think that it is possible to apply the theorem to experiments with finite sample spaces or to approximations of the wave equation by a difference equation on a finite grid, but this is beyond the scope of the paper.

In Appendix A we state and prove compactness of the space of stochastic operators (see also [12] 4.5.13. or compare with [11] 42.3). In Appendix C, we show that in the black-box criterion the role of the space of operators and the experiment can be reversed.

2 The classical randomization criterion

We start with some definitions:

Let $(\Omega, \mathcal{A})$ be a measurable space. We denote by $ca(\Omega, \mathcal{A})$ the space of bounded countably additive real-valued set-functions on $(\Omega, \mathcal{A})$ (i.e. the space of σ-additive signed measures of bounded variation on $(\Omega, \mathcal{A})$). By $ba(\Omega, \mathcal{A})$ we denote the space of bounded finitely additive real-valued set-functions on $(\Omega, \mathcal{A})$. We denote by $\|.\|$ the variation norm on $ba(\Omega, \mathcal{A})$ defined by:

$$\|\mu\| := 2 \cdot [\sup_{A \in \mathcal{A}} |\mu(A)|] - |\mu(\Omega)| \ .$$

If we like to mention the σ-algebra $\mathcal{A}$ explicitly we denote the variation norm by $\|.\|_{\mathcal{A}}$.

Let further $(P_\vartheta)_{\vartheta \in \Theta}$ be a family of probability measures on $(\Omega, \mathcal{A})$. Then we call $E := (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta \in \Theta})$ a statistical experiment. We denote by $L(E)$ the L-space of the experiment E which is the vector-space of measures defined by

$$L(E) := \{\mu \in ca(\Omega, \mathcal{A}) \mid \ [\nu \in ca(\Omega, \mathcal{A}) \ \text{ and } \ P_\vartheta \perp \nu \text{ for all } \vartheta \in \Theta] \Longrightarrow \mu \perp \nu\} \ .$$

We denote by $ba^+(\Omega, \mathcal{A})$ the positive cone of $ba(\Omega, \mathcal{A})$ defined by

$$ba^+(\Omega, \mathcal{A}) := \{\mu \in ba(\Omega, \mathcal{A}) \mid \mu(A) \geq 0 \ \text{ for all } \ A \in \mathcal{A}\} \ .$$

We further define the positive cones of $ca(\Omega, \mathcal{A})$ and $L(E)$ by

$$ca^+(\Omega, \mathcal{A}) = ca(\Omega, \mathcal{A}) \cap ba^+(\Omega, \mathcal{A}) \ \text{ and } \ L^+(E) := L(E) \cap ba^+(\Omega, \mathcal{A}) \ .$$

Let $E := (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta \in \Theta})$ and $F := (\Omega_2, \mathcal{B}, (Q_\vartheta)_{\vartheta \in \Theta})$ be statistical experiments. We say that a linear mapping $T : L(E) \mapsto L(F)$ is a stochastic operator or a transition (see also [11] 55.2) if

$$T(L^+(E)) \subseteq L^+(F) \ \text{ and } \mu \in L(E)^+ \Rightarrow T(\mu)(\Omega_2) = \mu(\Omega).$$

Analogous we say that a linear mapping $T : L(E) \to ba(\Omega_2, \mathcal{B})$ [respectively $T : ba(\Omega, \mathcal{A}) \to ba(\Omega_2, \mathcal{B})$] is a stochastic operator if $T(L^+(E)) \subseteq ba^+(\Omega_2, \mathcal{B})$

and $\mu \in L^+(E) \Rightarrow T(\mu)(\Omega_2) = \mu(\Omega)$ [respectively $T(ba^+(\Omega, \mathcal{A})) \subseteq ba^+(\Omega_2, \mathcal{B})$ and $\mu \in ba^+(\Omega, \mathcal{A}) \Rightarrow T(\mu)(\Omega_2) = \mu(\Omega)$].

Note that any stochastic operator fulfills $\|T\| \leq 1$ with $\|T\| := \sup\{|T(\mu)| \mid \|\mu\| \leq 1\}$. This fact can be obtained using the Jordan decomposition (see [8]) of the finitely [respectively countably] additive measure μ and the properties of stochastic operators.

We say that a stochastic operator M_K is induced by a Markov-kernel

$$K : \Omega \times \mathcal{B} \mapsto \mathbb{R} \quad \text{if} \quad [M_K(\mu)](B) = \int K(x, B) d\mu(x) \ .$$

So far we have introduced the notion of statistical experiment and some abstract spaces related to this notion. Since LeCam-Theory is concerned with the comparison of statistical experiment based on decisions and risks we have to introduce further the notions of decision space, decision, loss function and risk.

Remark 2.1 If we observe the outcome $x \in \Omega$ of the experiment E we would like to base a decision $d(x)$ in some measurable space $(\Omega_3, \mathcal{D})$ solely on x. So a deterministic decision rule is an $\mathcal{A}/\mathcal{D}$-measurable mapping $d : (\Omega, \mathcal{A}) \to (\Omega_3, \mathcal{D})$ and the space $(\Omega_3, \mathcal{D})$ is called a decision space. More general we will consider randomized decision rules. A randomized decision rule makes a decision not by selecting a single point $d(x) \in \Omega_3$ for $x \in \Omega$, but by selecting a probability measure P_x on $(\Omega_3, \mathcal{D})$. Formally such a randomized decision rule is given by a Markov-kernel $K : \Omega \times \mathcal{D} \mapsto \mathbb{R}$ by $P_x(D) = K(x, D)$.

Definition 2.1 Given a $\mathcal{D}$-measurable function $f : (\Omega_3, \mathcal{D}) \to \mathbb{R}$ and a randomized decision rule M_K given by a Markov-kernel K, we say that if

$$R_{f,P}(M_K) := \int_{x \in \Omega} \int_{\Omega_3} f(y)\ K(x, dy) dP(x)$$

exists, then $R_{f,P}(M_K)$ is the risk (expected loss) of the decision rule M_K given the loss function f and the probability P.

By further abstraction we define the risk (risk-function) of a general stochastic operator T to be $R_{f,P}(T) := \int f\ dT(P)$. We note that any stochastic operator can be weakly approximated by (a net of) decision rules induced by Markov kernels (see [12] 4.5.17. or compare with [11] 42.5, or [2] Chapter 1.4 Theorem 1) and thus a stochastic operator can be viewed as a generalized decision rule. (We will in the next chapter generalize the notion of decision rule further.)

Before we state the classical randomization criterion of LeCam we sketch the general structure of randomization criteria by a diagram and some comments:

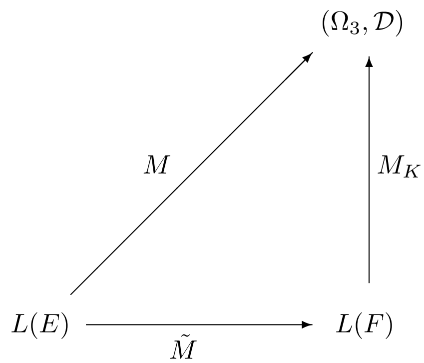

Diagram 2

If we are given a specific set $\mathcal{P}$ of properties p which apply to stochastic operators $M_K : L(F) \to (\Omega_3, \mathcal{D})$ induced by Markov kernels K as well as to general stochastic operators $M : L(F) \to (\Omega_3, \mathcal{D})$ then the following assertions are equivalent:

(i) To any $p \in \mathcal{P}$ and any $M_K : L(F) \to (\Omega_3, \mathcal{D})$ fulfilling p there exists an $M : L(E) \to (\Omega_3, \mathcal{D})$ which also fulfills p.

(ii) There exists a stochastic operator $\widetilde{M} : L(E) \to L(F)$ and a specific property $\tilde{p}(\notin \mathcal{P}!)$ such that $\widetilde{M}$ fulfills $\tilde{p}$.

We will continue the discussion of the abstract structure of the randomization criterion and its relation to the abstract structure of the black-box criterion in section 3. We apply the result obtained in section 3 to the proof of the equivalence of the theorems 5.1 and 5.2 in section 5.

Note that in the concrete situation of the classical randomization criterion as well as of the adapted randomization criterion (which includes the filtered case) the implication $(ii) \Rightarrow (i)$ of our abstract discussion above is trivial. We therefore display only the implications $(i) \Rightarrow (ii)$ and call these implications randomization criteria (theoremes 2.1 and 5.2). For the sake of completeness we prove in section 5 also the converse (corresponding to $(ii) \Rightarrow (i)$ in the abstract discussion above) of Theorem 5.2 which implies the converse of Theorem 2.1.

The above discussion also applies to the results on e-stochastic operators indicated in section 6.

We state now (the nontrivial part of) the classical randomization criterion of LeCam (Theorem 2.1).

Theorem 2.1 *Let*

$$\begin{aligned} E &:= (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta\in\Theta}) \quad \textit{and} \\ F &:= (\Omega_2, \mathcal{B}, (Q_\vartheta)_{\vartheta\in\Theta}) \end{aligned}$$

be statistical experiments indexed by the same set Θ *and let* $(\varepsilon_\vartheta)_{\vartheta\in\Theta}$ *be an indexed family of reals* ≥ 0*. Suppose that for any measurable space* $(\Omega_3, \mathcal{D})$ *(the decision space), any stochastic operator* $M_K : L(F) \mapsto ba(\Omega_3, \mathcal{D})$ *which is induced by a Markov-kernel* K *(the randomized decision rule) and*

(I) *for any parameterized family* $(f_\vartheta)_{\vartheta\in\Theta}$ *of* $\mathcal{D}$*-measurable functions* $f_\vartheta : \Omega_3 \mapsto [-1, +1]$ *(the loss functions)*

there exists a stochastic operator $M : L(E) \mapsto ba(\Omega_3, \mathcal{D})$ *such that*

$$\int f_\vartheta \, dM(P_\vartheta) \leq \int f_\vartheta \, dM_K(Q_\vartheta) + \varepsilon_\vartheta \quad \textit{for all} \quad \vartheta \in \Theta \, .$$

Then there exists a stochastic operator $\widetilde{M} : L(E) \mapsto L(F)$ *such that*

$$\|Q_\vartheta - \widetilde{M}(P_\vartheta)\| \leq \varepsilon_\vartheta \quad \textit{for all} \quad \vartheta \in \Theta \, .$$

A mathematically equivalent formulation of Theorem 2.1 is the following Theorem 2.2. Theorem 2.2 does not involve arbitrary decision spaces; instead it uses $(\Omega_2, \mathcal{B})$ itself as a decision space. We call theorems analogous to Theorem 2.2 black-box criteria.

Theorem 2.2 *Let* $E := (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta\in\Theta})$ *and* $F := (\Omega_2, \mathcal{B}, (Q_\vartheta)_{\vartheta\in\Theta})$ *be statistical experiments indexed by the same set* Θ *and let* $(\varepsilon_\vartheta)_{\vartheta\in\Theta}$ *be a family of nonnegative real numbers. Suppose that*

(II) *for any parameterized family* $(g_\vartheta)_{\vartheta\in\Theta}$ *of* $\mathcal{B}$*-measurable functions* $g_\vartheta : \Omega_2 \to [-1, +1]$

there exists a stochastic operator $S : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ *such that*

$$\int g_\vartheta \, dS(P_\vartheta) \leq \int g_\vartheta \, dQ_\vartheta + \varepsilon_\vartheta \quad \textit{for all} \quad \vartheta \in \Theta \, .$$

Then there exists a stochastic operator $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ *such that*

$$\|\widetilde{S}(P_\vartheta) - Q_\vartheta\| \leq \varepsilon_\vartheta \quad \textit{for all} \quad \vartheta \in \Theta \, .$$

We prove now that the theorems 2.1 and 2.2 imply each other. We do this in both directions by the following general principle: If one wants to show that a theorem A implies a theorem B one simply shows that the hypotheses of B imply the hypotheses of A and that the conclusions of A imply the conclusions of B.

We show [Theorem 2.2 $\Rightarrow$ Theorem 2.1] first:

Let $(\Omega_3, \mathcal{D}) := (\Omega_2, \mathcal{B})$, and let M_K be the identical imbedding of $L(E)$ into $ba(\Omega_2, \mathcal{B}_2)$. In this special case the hypotheses of Theorem 2.1 say that:

$$\left\{ \begin{array}{l} \text{For any parametrized family } (f_\vartheta)_{\vartheta\in\Theta} \text{ of } \mathcal{B}\text{-measurable functions} \\ f_\vartheta : \Omega_2 \mapsto [-1+1] \text{ there exists a stochastic operator } M \text{ such that} \\ \displaystyle \int f_\vartheta \, dM(P_\vartheta) \leq \int f_\vartheta \, dQ_\vartheta + \varepsilon_\vartheta \quad \text{for all} \quad \vartheta \in \Theta \, . \end{array} \right.$$

By a change of notation (i.e. $g_\vartheta = f_\vartheta$ and $S = M$) we see that these are exactly the hypotheses of Theorem 2.2. Thus by assumption of the truth of Theorem 2.2 we get that there exists a stochastic operator $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ such that

$$\|\widetilde{S}(P_\vartheta) - Q_\vartheta\| \leq \varepsilon_\vartheta \quad \text{for all} \quad \vartheta \in \Theta \, .$$

By [11] 41.7 or [12] 4.5.11. there exists a stochastic operator $T : ba(\Omega_2, \mathcal{B}) \mapsto L(F)$ such that $T \mid_{L(F)} = id_{L(F)}$. Since T (being a stochastic operator) fulfills $\|T\| \leq 1$ and the concatenation of stochastic operators is again a stochastic operator we obtain a stochastic operator $\widetilde{M} : L(E) \mapsto L(F)$ given by $\widetilde{M} = T \circ \widetilde{S}$ such that

$$\|\widetilde{M}(P_\vartheta) - Q_\vartheta\| = \|[T \circ \widetilde{S}](P_\vartheta) - T(Q_\vartheta)\| \leq \|T\| \cdot \|\widetilde{S}(P_\vartheta) - Q_\vartheta\| \leq \varepsilon_\vartheta$$

for all $\vartheta \in \Theta$. $\square$

We prove now [Theorem 2.1 $\Rightarrow$ Theorem 2.2].

Let a measurable space $(\Omega_3, \mathcal{D})$, a stochastic operator $M_K : L(F) \mapsto (\Omega_3, \mathcal{D})$ and a parameterized family $(f_\vartheta)_{\vartheta\in\Theta}$ of $\mathcal{D}$-measurable functions $f_\vartheta : \Omega_3 \mapsto [-1, 1]$ be given. Suppose that M_K is induced by a Markov kernel $K : \Omega_2 \times \mathcal{D} \mapsto [0, 1]$. Define functions g_ϑ for all $\vartheta \in \Theta$ by

$$g_\vartheta(x) := \int f_\vartheta(y) \, K(x, dy)$$

The g_ϑ are $\mathcal{B}$-measurable functions on Ω_2 with ranges contained in $[-1, +1]$. Thus by the hypotheses of Theorem 2.2 there exists for our parameterized family $(g_\vartheta)_{\vartheta\in\Theta}$ a stochastic operator $S : L(E) \mapsto (\Omega_2, \mathcal{B})$ such that

$$\int g_\vartheta \, dS(P_\vartheta) \leq \int g_\vartheta \, dQ_\vartheta + \varepsilon_\vartheta \quad \text{for all} \quad \vartheta \in \Theta$$

If we let $M := M_K \circ S$ then we obtain for all $\vartheta \in \Theta$

$$\int_{\Omega_3} f_\vartheta \, dM(P_\vartheta) = \int_{\Omega_3} f_\vartheta \, d[M_K \circ S](P_\vartheta) =$$

$$\int_{x\in\Omega_2}\int_{y\in\Omega_3} f_\vartheta(y)\ K(x,dy)\ [dS(P_\vartheta)](x) = \int_{\Omega_2} g_\vartheta\ dS(P_\vartheta) \leq$$

$$\int_{\Omega_2} g_\vartheta\ dQ + \varepsilon_\vartheta = \int_{x\in\Omega_2}\int_{y\in\Omega_3} f_\vartheta(y)\ K(x,dy)\ dQ_\vartheta(x) + \varepsilon_\vartheta =$$

$$\int_{\Omega_3} f_\vartheta\ dM_K(Q_\vartheta) + \varepsilon_\vartheta\ .$$

By the arbitrary choice of $(\Omega_3, \mathcal{D})$, $(f_\vartheta)_{\vartheta\in\Theta}$ and M_K we see that the hypotheses of Theorem 2.1 are fulfilled. Since we assumed Theorem 2.1 to be true, the conclusion of Theorem 2.1 also holds. That the conclusion of Theorem 2.1 implies the conclusion of Theorem 2.2 is trivial and thus the implication has been shown. □

Remark 2.2 In Theorems 2.1 statement (I) can be replaced by

(I') for any finite set $\Theta_0 \subseteq \Theta$ and any parameterized family $(f_\vartheta)_{\vartheta\in\Theta_0}$ of $\mathcal{D}$-measurable functions $f_\vartheta : \Omega_3 \mapsto [-1,+1]$.

In theorem 2.2 statemtn (II) can be replaced by

(II') for any finite set $\Theta_0 \subseteq \Theta$ and any parameterized family $(g_\vartheta)_{\vartheta\in\Theta}$ of $\mathcal{B}$-measurable functions $g_\vartheta : \Omega_2 \to [-1,+1]$.

This fact can be interpreted as follows: To consider arbitrary finite sub experiments (based on an arbitrary finite parameter space $\Theta_0 \subseteq \Theta$) suffices for our investigations on risk and decision. We will therefore proceed by stating all further results including a conditioning on arbitrary finite sub experiments as indicated by (I') and (II'). This is in accordance with the presentation of the results for filtered experiments in [5].

Usually one defines the deficiency $\delta(E,F)$ of the experiment E with respect to the experiment F and afterwords the Δ-distance to obtain a quantitative measure of the closeness of experiments. The deficiency δ is defined by

$$\delta(E,F) := \sup_{(\Omega_3,\mathcal{D})}\ \sup_{M_K}\ \sup_{(f_\vartheta)_{\vartheta\in\Theta}}\ \inf_{M}\ \sup_{\vartheta}\ [R_{f_\vartheta,Q_\vartheta}(M_K) - R_{f_\vartheta,P_\vartheta}(M)]$$

where the first "sup" ranges over all decision spaces $(\Omega_3, \mathcal{D})$ the second "sup" ranges over all stochastic operators induced by Markov kernels K on $\Omega_2 \times \mathcal{D}$ the third "sup" over all selections $(f_\vartheta)_{\vartheta\in\Theta}$ of loss functions f_ϑ on $\mathcal{D}$, the "inf" ranges over all stochastic operators from $L(E)$ to $ba(\Omega_3, \mathcal{D})$ and the last "sup" over the parameters $\vartheta \in \Theta$. ($R_{f,P}(.)$ denotes the risk introduced in Definition 2.1.)

By Theorem 2.1 and its (trivial) converse we obtain that the deficiency of E with respect to F is alternatively given by

$$(iii) \qquad \delta(E,F) = \inf_{\widetilde{M}}\ \sup_{\vartheta\in\Theta} \|Q_\vartheta - \widetilde{M}(P_\vartheta)\|\ .$$

Considering (I') we also obtain

$$\delta(E,F) := \sup_{(\Omega_3,\mathcal{D})} \sup_{M_K} \sup_{\Theta_0\subset\Theta} \sup_{(f_\vartheta)_{\vartheta\in\Theta_0}} \inf_{M} \sup_{\vartheta\in\Theta_0} [R_{f_\vartheta,Q_\vartheta}(M_K) - R_{f_\vartheta,P_\vartheta}(M)]$$

with the "sups" and the "inf" ranging over appropriate spaces.

Considering Theorem 2.2 and (*iii*) we obtain that another expression for $\delta(E,F)$ is provided by

$$\delta(E,F) = \sup_{(g_\vartheta)_{\vartheta\in\Theta}} \inf_{S} \sup_{\theta\in\Theta} [R_{g_\vartheta,Q_\vartheta}(id) - R_{g_\vartheta,P_\vartheta}(S)]$$

where the first "sup" is taken over all selections of loss functions on $(\Omega_2,\mathcal{B})$ and the "inf" taken over all stochastic operators from $L(E)$ to $L(F)$.

After one has defined deficiencies one defines the Δ-distance (on the space of experiments with the same index set Θ) by $\Delta(E,F) := \max\{\delta(E,F),\delta(F,E)\}$. With respect to this distance Δ one considers convergence of experiments and defines the limit of a sequence of experiments. A generalization of the deficiencies and the Δ-distance to the filtered case can be found in [5]. A similar distance concept could be introduced for adapted experiments. We will introduce and discuss a generalization of $\delta(.,.)$ within the context of experiments and decision rules endowed with even a richer structure than adaptedness after 6.2. Although we are not going to discuss asymptotic theory note that a connection between asymptotic theory and adapted experiments is indicated by example 6.2.

3 The equivalence principle

We prove in this section a categorical theorem that reveals a general principle concerning the equivalence between randomization criteria and black-box criteria. This categorical result does not only enlighten the equivalence principle but will also be applied to prove the equivalence in the case of adapted experiments in section 5.

Definition 3.1 Let a category C consisting of a "family" $\mathbf{Ob}(C)$ of objects and a "family" $\mathbf{Mor}(C)$ of morphisms be given. We denote by $\text{ob}_\circ$ with $\circ$ replaced by any other subscript "elements" of $\mathbf{Ob}(C)$ and by $\text{mor}_\circ$ with $\circ$ replaced by any other subscript "elements" of $\mathbf{Mor}(C)$. If mor_S is a Morphism then we denote by $\mathbf{Ran}(\text{mor}_S)$ the object which forms the range of mor_S and by $\mathbf{Dom}(\text{mor}_S)$ the object which forms the domain of mor_S. Suppose that with any $\text{ob}_F \in \mathbf{Ob}(C)$ there is associated a "family" $\mathbf{Prop}(\text{ob}_F)$ of properties $p(.) \in \mathbf{Prop}(\text{ob}_F)$ which an $\text{ob}_\bullet \in \mathbf{Ob}(C)$ may possess or may not possess, i.e. $p(\text{ob}_\bullet)$ is true or false. Suppose further, that the properties p (with $\text{ob}_\bullet \in \mathbf{Ob}(C)$ arbitrary and $p \in \mathbf{Prop}(\text{ob}_\bullet)$) are the Objects of a category $\vec{C}$, i.e. $p \in \mathbf{Ob}(\vec{C})$ which is dual to C in the sense that:

a) $\begin{cases} \text{with any } \mathrm{mor}_K \text{ and any } p \in \mathbf{Prop}(\mathbf{Ran}(\mathrm{mor}_K)) \text{ there is associated a morphism } \mathrm{mor}_K^p \in \mathbf{Mor}(\vec{C}) \text{ such that } \mathbf{Dom}(\mathrm{mor}_K^p) = p \text{ and } \mathbf{Ran}(\mathrm{mor}_K^p) \in \mathbf{Prop}(\mathbf{Dom}(\mathrm{mor}_K)) \end{cases}$

b) $\begin{cases} \text{for any } \mathrm{mor}_K \text{ and } \mathrm{ob}_\bullet \text{ there exists a morphism } \mathrm{mor}_{\tilde{K}} \text{ with } \mathbf{Dom}(\mathrm{mor}_{\tilde{K}}) = \mathrm{ob}_\bullet \text{ such that for } p \in \mathbf{Prop}(\mathbf{Ran}(\mathrm{mor}_K)) \text{ we have:} \\ [\mathbf{Ran}(\mathrm{mor}_K^p)](\mathbf{Dom}(\mathrm{mor}_{\tilde{K}})) \implies [\mathbf{Dom}(\mathrm{mor}_K^p)](\mathbf{Ran}(\mathrm{mor}_{\tilde{K}})). \end{cases}$

Theorem 3.1 *Given Categories C and $\vec{C}$ which fulfill a) and b) (Definition 3.1) and $ob_E, ob_F \in \mathbf{Ob}(C)$ then the following assertions are equivalent:*

i) $\begin{cases} \text{for any } \hat{p} \in \mathbf{Prop}(ob_F) \text{ there exists a morphism } mor_S \text{ with } ob_E = \mathbf{Dom}(mor_S) \text{ such that } \hat{p}(\mathbf{Ran}(mor_S)) \\ \text{implies that} \\ \text{there exists a morphism } mor_{\tilde{S}} \text{ with } ob_E = \mathbf{Dom}(mor_{\tilde{S}}) \text{ such that for all } \tilde{p} \in \mathbf{Prop}(ob_F) \text{ we have } \tilde{p}(\mathbf{Ran}(mor_{\tilde{S}})). \end{cases}$

ii) $\begin{cases} \text{for any } mor_K \text{ with } Dom(mor_K) = ob_F \text{ and any } p \in \mathbf{Prop}(\mathbf{Ran}(mor_K)) \text{ there exists a morphism } mor_M \text{ such that } \mathbf{Dom}(mor_M) = ob_E \text{ and } p(\mathbf{Ran}(mor_M)) \\ \text{implies that} \\ \text{there exists a morphism } mor_{\tilde{S}} \text{ with } ob_E = \mathbf{Dom}(mor_{\tilde{S}}) \text{ such that for all } \tilde{p} \in \mathbf{Prop}(ob_F) \text{ we have } \tilde{p}(\mathbf{Ran}(mor_{\tilde{s}})). \end{cases}$

Proof: [i) $\Rightarrow$ ii)]: If we let $\mathrm{mor}_K = id$, $p = \hat{p}$ and $\mathrm{mor}_M = \mathrm{mor}_S$ we see that the hypothesis of ii) implies the hypothesis of i). Since the conclusions of i) and ii) are equal [i) $\Rightarrow$ ii)] has been proved.

[ii) $\Rightarrow$ i)] (The proof is illustrated by the Diagram below). Since the conclusion of ii) equals the conclusion of i) it suffices to show that the hypothesis of i) implies the hypothesis of ii). Thus let mor_K be an arbitrary morphism which fulfills $\mathbf{Dom}(\mathrm{mor}_K) = \mathrm{ob}_F$ and let $p \in \mathbf{Prop}(\mathbf{Ran}(\mathrm{mor}_K))$ be arbitrary. Let $\hat{p} = \mathbf{Ran}(\mathrm{mor}_K^p)$, let mor_S fulfill the hypothesis of i), set $\mathrm{ob}_\bullet := \mathbf{Ran}(\mathrm{mor}_S)$ and let $\mathrm{mor}_{\tilde{K}}$ be the morphism granted by b). We let $\mathrm{mor}_M = \mathrm{mor}_{\tilde{K}} \star \mathrm{mor}_S$. Then we get from the hypothesis of i) that

$$true = \hat{p}(\mathbf{Ran}(\mathrm{mor}_S)) = \hat{p}(\mathbf{Dom}(\mathrm{mor}_{\tilde{K}})) = [\mathbf{Ran}(\mathrm{mor}_K^p)](\mathbf{Dom}(\mathrm{mor}_{\tilde{K}})).$$

Thus by b)

$$true = [\mathbf{Dom}(\mathrm{mor}_K^p)](\mathbf{Ran}(\mathrm{mor}_{\tilde{K}})) = p(\mathbf{Ran}(\mathrm{mor}_{\tilde{K}})) = p(\mathbf{Ran}(\mathrm{mor}_M))$$

and the hypothesis of ii) is established. [ii) $\Rightarrow$ i)] has been proved. □

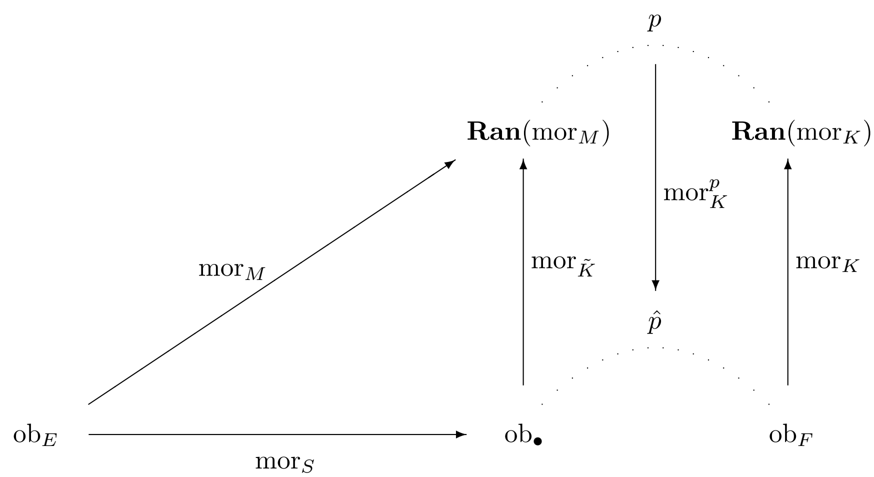

Diagram 3

4 The setting of locally convex spaces

In this section we free the black-box criterion from the setting of L-spaces.

To do this it is most comfortable to use the language of the theory of locally convex topological vector spaces. So we will state and prove in this section rather abstract theorems but we gain by this abstraction a flexible theory which can be easily applied to several problems in the next sections. Applicability of the theorems 4.1 and 4.3 depends only on compactness properties of the space of mappings (abstracted randomization rules) involved. Thus it is very easy to apply our locally convex theory to the cases of adapted stochastic operators, general loss function spaces and operators on Hilbert spaces (theorems 5.1, B.2 and 7.1).

The theorems 4.1 and 4.2 introduce the reader to the black-box criterion in the setting of locally convex spaces. Theorem 2.2 is derived from Theorem 4.1. This shows that compactness of the space of abstracted randomization rules is the property in the heart of Theorem 2.2 as well as in the heart of the equivalent Theorem 2.1. It further shows in how far the order structure of the L-spaces is really involved in black-box and randomization criteria.

Next, we state and prove Theorem 4.3. The theorems 4.3 and 4.4 are the most general versions of the black-box criterion provided in this paper. They make it possible to deal with parameterized families of loss functions. The theorems 4.1 and 4.2 are special cases of Theorem 4.3. So altogether we provided a proof of a version (Theorem 2.1) of the classical randomization criterion of LeCam.

We introduce some notations and definitions first (See also [7]):

Given a topological space (X, τ) and a subset Y of X we denote by (Y, τ) the topological space Y endowed with the relative topology Y inherits from τ. The term vector space always denotes a vector space over the real field.

Given a vector space W and a vector space topology τ which possesses a basis consisting of convex sets, then we call (W, τ) a locally convex space. If we do not want to mention τ explicitly, we will also write W instead of (W, τ) for the topological vector space (W, τ). We denote by W' the topological dual of (W, τ). (i.e. the vector space of all τ-continuous linear functionals on W).

We denote by (W, σ) the vector space W endowed with the weakest topology σ making all the elements of the topological dual W' continuous. Symmetrically we denote by (W', σ') the vector space W' endowed with the weakest topology making all the elements of W to continuous linear functionals on (W', σ'). We note that (W, σ) and (W', σ') are locally convex spaces. We further note that the topological dual of (W', σ') is the space W (See [7] Chapter 4 Section 1.2). Thus the relation between (W, σ) and (W', σ') is completely symmetric and we write $\langle w', w \rangle$ for the value the functional $w'(.)$ takes on at w, or equivalently the functional $w(.)$ takes on at w'. To ease notation we will often not mention the vector space topology explicitly and will write W instead of (W, τ) and W' instead of (W', σ'). We will speak of the locally convex space W and its weak dual W'.

Let $M \subseteq W$ or $N \subseteq W'$, then we let

$$M^\circ := \{w' \in W' \mid \langle w', w \rangle \leq 1 \text{ if } w \in M\} \quad \text{and}$$

$$N^\circ := \{w \in W \mid \langle w', w \rangle \leq 1 \text{ if } w' \in N\} .$$

We call M° the polar of M and N° the polar of N. By $N^{\circ\circ}$ we denote the polar of the polar of N (also called the bipolar of N).

Further we denote by $\prod_{\xi \in \Xi} W_\xi$ the product of the topological vector spaces W_ξ and by $\bigoplus_{\xi \in \Xi} W_\xi$ the direct sum of the vector spaces W_ξ. By $card(\Xi)$ we denote the cardinality of a set Ξ.

Theorem 4.1 *Let V be a set and let W be a locally convex space. Let $G \subseteq W$ be given such that G is closed, convex and contains 0. Let $\mathcal{L}$ be a compact convex subset of $(W')^V$ endowed with the product topology. Let Θ be an index set. Let $(v_\vartheta)_{\vartheta \in \Theta}$ be a family of points in V and let $(w_\vartheta)_{\vartheta \in \Theta}$ be a family of points in W'. Let $(\varepsilon_\vartheta)_{\vartheta \in \Theta}$ be a family of reals $\varepsilon_\vartheta \in [0, \infty)$. Suppose that for any finite subset $\Theta_0 \subseteq \Theta$ and for any family $(g_\vartheta)_{\vartheta \in \Theta_0}$ of functionals $g_\vartheta \in G$ there exists an $l \in \mathcal{L}$ such that for all $\vartheta \in \Theta_0$ we have:*

$$\langle\, l(v_\vartheta) - w_\vartheta \;,\; g_\vartheta \,\rangle \;\leq\; \varepsilon_\vartheta \;.$$

Then there exists an $l \in \mathcal{L}$ such that for all $\vartheta \in \Theta$ we have:

$$\sup_{g_\vartheta \in G} \langle\, l(v_\vartheta) - w_\vartheta \;,\; g_\vartheta \,\rangle \;\leq\; \varepsilon_\vartheta \;.$$

The theorem remains true if we replace the points w_ϑ by compact convex sets K_ϑ; i.e. we have

Theorem 4.2 *Let V be a set and let W be a locally convex space. Let $G \subseteq W$ be given such that G is closed, convex and contains 0. Let $\mathcal{L}$ be a compact convex subset of $(W')^V$ endowed with the product topology. Let Θ be an index set and let $(v_\vartheta)_{\vartheta\in\Theta}$ be a family of points in V and $(K_\vartheta)_{\vartheta\in\Theta}$ be a family of compact convex subsets of W'. Let $(\varepsilon_\vartheta)_{\vartheta\in\Theta}$ be a family of reals $\varepsilon_\vartheta \in [0,\infty)$. Suppose that for any finite subset $\Theta_0 \subseteq \Theta$ and for any family $(g_\vartheta)_{\vartheta\in\Theta_0}$ of functionals $g_\vartheta \in G$ there exists an $l \in \mathcal{L}$ such that for all $\vartheta \in \Theta_0$ we have:*

$$\inf_{w_\vartheta \in K_\vartheta} \langle\, l(v_\vartheta) - w_\vartheta \ ,\ g_\vartheta \,\rangle \ \le\ \varepsilon_\vartheta \ .$$

Then there exists an $l \in \mathcal{L}$ such that for all $\vartheta \in \Theta$ we have:

$$\inf_{w_\vartheta \in K_\vartheta} \big(\sup_{g_\vartheta \in G} \langle\, l(v_\vartheta) - w_\vartheta \ ,\ g_\vartheta \,\rangle \big) \ \le\ \varepsilon_\vartheta \ .$$

We prove now [Theorem 4.1 $\Rightarrow$ Theorem 2.2]

For this let $\mathcal{M}_b$ denote the vector space of bounded $\mathcal{B}$-measurable real valued functions on $(\Omega_2, \mathcal{B})$ endowed with (the topology induced by) the supremum norm. Let σ' be the topology which $\mathcal{M}_b$ induces on $ba(\Omega_2,\mathcal{B})$ by the mappings $\mu \mapsto \int m\, d\mu$ with $m \in \mathcal{M}_b$. Then $ba(\Omega,\mathcal{B})$ endowed with σ' is the weak dual of the locally convex space $\mathcal{M}_b$. From Lemma A2 of the appendix we obtain that the space of stochastic operators from $L(E)$ to $ba(\Omega_2,\mathcal{B})$ is a compact convex subset of $(ba(\Omega_2,\mathcal{B}),\sigma')^{L(E)}$. The set $\{m \in \mathcal{M}_b \mid -1 \le m \le 1\}$ is a closed convex subset of $\mathcal{M}_b$.

Consider now Theorem 4.1 in the following special case:

Let V denote the space $L(E)$, let W be the space $\mathcal{M}_b$ endowed with the supremum norm. Let $G := \{m \in \mathcal{M}_b \mid -1 \le m \le 1\}$, let $v_\vartheta = P_\vartheta$ and let $w_\vartheta = Q_\vartheta$. Denote by $\mathcal{L}$ the space of stochastic operators from $L(E)$ to $ba(\Omega_2,\mathcal{B})$.

In the special case we consider the hypotheses of Theorem 2.2 imply the hypotheses of Theorem 4.1. The conclusion of Theorem 4.1 is in this special case equivalent with the conclusion of Theorem 2.2. Thus Theorem 4.1 in fact implies Theorem 2.2. □

Theorem 4.3 *Let Ξ be a set and let $(W_\xi)_{\xi\in\Xi}$ be a family of locally convex vector spaces. Let $(G_\xi)_{\xi\in\Xi}$ be a family of sets with $G_\xi \subseteq W_\xi$ and let $(\varepsilon_\xi)_{\xi\in\Xi}$ be a family of real numbers $\varepsilon_\xi \in [0,\infty)$. Let $\mathcal{K}, \mathcal{J} \subset \prod_{\xi\in\Xi} W'_\xi$.*

Suppose that the following hypotheses are fulfilled:

(i) $\mathcal{K}$ and $\mathcal{J}$ are compact, convex subsets of $\prod_{\xi\in\Xi} W'_\xi$.

(ii) *The sets* G_ξ *are closed, convex and contain* 0.

Then the following hypotheses are equivalent:

(iii) *For any finite* $\Xi_0 \subseteq \Xi$ *and for any selection* $(g_\xi)_{\xi\in\Xi_0}$ *of functionals* $g_\xi \in G_\xi$ *there exists* $(k,j) \in \mathcal{K}\times\mathcal{J}$ *such that for all* $\xi \in \Xi_0$ *we have*

$$\langle\, j(\xi) - k(\xi) \;,\; g_\xi \,\rangle \;\le\; \varepsilon_\xi \;.$$

(iv) *For any finite* $\Xi_0 \subseteq \Xi$*, for any selection* $(g_\xi)_{\xi\in\Xi_0}$ *of functionals* $g_\xi \in G_\xi$ *there exists* $(k,j) \in \mathcal{K}\times\mathcal{J}$ *such that for any indexed family* $(\alpha_\xi)_{\xi\in\Xi_0}$ *of real numbers with* $\alpha_\xi \ge 0$ *and* $\sum_{\xi\in\Xi_0}\alpha_\xi \le 1$ *we have*

$$\sum_{\xi\in\Xi_0} \alpha_\xi \cdot \langle\, j(\xi) - k(\xi) \;,\; g_\xi \,\rangle \;\le\; \sum_{\xi\in\Xi_0} \alpha_\xi \cdot \varepsilon_\xi \;.$$

(v) *There exists* $(k,j) \in \mathcal{K}\times\mathcal{J}$ *such that for all* $\xi \in \Xi$

$$\sup_{g_\xi\in G_\xi} \left(\langle\, j(\xi) - k(\xi) \;,\; g_\xi \,\rangle\right) \;\le\; \varepsilon_\xi \;.$$

Proof of Theorem 4.3 We define for all $\xi \in \Xi$ sets $\widetilde{G_\xi}$ by

$$\widetilde{G_\xi} := \frac{1}{\varepsilon_\xi} G_\xi \qquad \text{if} \qquad \varepsilon_\xi > 0$$
$$\widetilde{G_\xi} := \overline{\bigcup_{n\in\mathbb{N}} n\, G_\xi} \qquad \text{if} \qquad \varepsilon_\xi = 0 \;.$$

With these definitions the hypothesis (iii), [resp. (iv) or (v)] becomes equivalent with the following hypothesis (iii') [resp. (iv') or (v')].

(iii') For any finite $\Xi_0 \subseteq \Xi$ and for any selection $(\widetilde{g_\xi})_{\xi\in\Xi_0}$ of functionals $\widetilde{g_\xi} \in \widetilde{G_\xi}$ there exists $(k,j) \in \mathcal{K}\times\mathcal{J}$ such that for all $\xi \in \Xi_0$ we have

$$\langle\, j(\xi) - k(\xi) \;,\; \widetilde{g_\xi} \,\rangle \;\le\; 1 \;.$$

(iv') For any finite $\Xi_0 \subseteq \Xi$, for any selection $(\widetilde{g_\xi})_{\xi\in\Xi_0}$ of functionals $\widetilde{g_\xi} \in \widetilde{G_\xi}$ there exists $(k,j) \in \mathcal{K}\times\mathcal{J}$ such that for any indexed family $(\alpha_\xi)_{\xi\in\Xi_0}$ of real numbers with $\alpha_\xi \ge 0$ and $\sum_{\xi\in\Xi_0}\alpha_\xi \le 1$ we have

$$\sum_{\xi\in\Xi_0} \alpha_\xi \cdot \langle\, j(\xi) - k(\xi) \;,\; \widetilde{g_\xi} \,\rangle \;\le\; 1 \;.$$

(v') There exists $(k,j) \in \mathcal{K}\times\mathcal{J}$ such that for all $\xi \in \Xi$

$$\sup_{\widetilde{g_\xi}\in\widetilde{G_\xi}} \left(\langle\, j(\xi) - k(\xi) \;,\; \widetilde{g_\xi} \,\rangle\right) \;\le\; 1 \;.$$

So the statement of Theorem 4.3 is equivalent with the statement that under the hypotheses (i) - (ii) the hypotheses (iii') - (v') are equivalent.

It is immediate that (iii) implies (iv) [or equivalently that (iii') implies (iv')] and also that (v) implies (iii) [or equivalently that (v') implies (iii')]. So to prove the theorem it remains only to show that (iv) implies (v). This is done indirect by proving that the negation of (v') implies the negation of (iv').

The negation of (v') is the following statement:

$$\left.\begin{array}{l} \text{For any } (k,j) \in \mathcal{K} \times \mathcal{J} \text{ there exists a } \xi \in \Xi \text{ such that} \\ \qquad\qquad \sup\limits_{\widetilde{g_\xi} \in \widetilde{G_\xi}} (\langle\, j(\xi) - k(\xi) \;,\; \widetilde{g_\xi} \,\rangle) \;>\; 1 \;. \end{array}\right\} \quad (1)$$

By hypothesis (i) we have that.

$$(\mathcal{J} - \mathcal{K}) \;\text{ is a compact convex subset of } \prod_{\xi\in\Xi} W'_\xi. \tag{2}$$

Since the sets $\widetilde{G_\xi}^\circ$ are closed convex subsets of W'_ξ, the set

$$\prod_{\xi\in\Xi} \widetilde{G_\xi}^\circ \;\text{ is a closed convex subset of } \prod_{\xi\in\Xi} W'_\xi. \tag{3}$$

With the notations introduced (1) can be reformulated as

$$\prod_{\xi\in\Xi} \widetilde{G_\xi}^\circ \;\cap\; (\mathcal{J} - \mathcal{K}) = \emptyset. \tag{4}$$

From (2) - (4) and the separation theorem for convex sets (see [7] chapter II, 9.2) we conclude that there exists a continuous linear functional $f \neq 0$ on $\prod_{\xi\in\Xi} W'_\xi$ and a constant γ such that

$$f(\prod_{\xi\in\Xi} \widetilde{G_\xi}^\circ) < \gamma < f(\mathcal{J} - \mathcal{K}) \;\text{ and} \tag{5}$$

$$\text{since } \; 0 \in \prod_{\xi\in\Xi} \widetilde{G_\xi}^\circ \;\text{ we get in addition that } \; \gamma > 0. \tag{6}$$

The dual of $\prod_{\xi\in\Xi} W'_\xi$ is algebraically isomorphic with $\bigoplus_{\xi\in\Xi} W_\xi$ (see [7] chapter IV, 4.3). Thus (and since $f \neq 0$) our functional f can be represented in the form

$$f((w_\xi)_{\xi\in\Xi}) = \sum_{\xi\in\Xi_0} f_\xi(w_\xi) \tag{7}$$

for some finite nonempty set $\Xi_0 \subseteq \Xi$ and a family of continuous linear functionals $f_\xi \in W_\xi$ with $f_\xi \neq 0$ for $\xi \in \Xi_0$.

Let $\Xi'_0 := \{\xi \in \Xi_0 \mid \sup_{g\in\widetilde{G_\xi}^\circ} f_\xi(g) > 0\}$ and define reals α_ξ by

$$\alpha_\xi := \sup_{g \in \widetilde{G_\xi}^\circ} \frac{f_\xi(g)}{\gamma} \quad \text{for} \quad \xi \in \Xi_0'. \tag{8}$$

From (5) - (8) we obtain that

$$\alpha_\xi > 0 \ \text{ for } \ \xi \in \Xi_0' \ \text{ and that } \sum_{\xi \in \Xi_0'} \alpha_\xi < 1. \tag{9}$$

Further, if $\Xi_0' \neq \Xi_0$, we define

$$\alpha_\xi := \frac{1 - \sum_{\xi \in \Xi_0'} \alpha_\xi}{card(\Xi_0 \setminus \Xi_0')} \quad \text{for} \quad \xi \in \Xi_0 \setminus \Xi_0' \ . \tag{10}$$

By (9) and (10) we get that

$$\alpha_\xi > 0 \ \text{ for } \xi \in \Xi_0 \ \text{ and } \sum_{\xi \in \Xi_0} \alpha_\xi \leq 1. \tag{11}$$

We define linear functionals g_ξ by

$$g_\xi(.) := \frac{f_\xi(.)}{\gamma \cdot \alpha_\xi}. \tag{12}$$

We obtain from (8) and (12) that $g_\xi \in \widetilde{G_\xi}^{\circ\circ}$ in the case that $\xi \in \Xi_0'$. From the fact that $\xi \in \Xi_0 \setminus \Xi_0'$ implies that $\sup_{g \in \widetilde{G_\xi}^\circ} g_\xi(g) = \sup_{g \in \widetilde{G_\xi}^\circ} \frac{f_\xi(g)}{\gamma \cdot \alpha_\xi} = 0$ we get that $g_\xi \in \widetilde{G_\xi}^{\circ\circ}$ in the case that $\xi \in \Xi_0 \setminus \Xi_0'$. So together with (ii) and the bipolar theorem (see [7] chapt IV, 1.5) we obtain in any case that

$$g_\xi \in \widetilde{G_\xi}^{\circ\circ} = \widetilde{G_\xi} \ \text{ for } \ \xi \in \Xi_0. \tag{13}$$

From (5), (7) and (12) we obtain that

$$\sum_{\xi \in \Xi_0} \alpha_\xi \cdot g_\xi(\mathcal{J} - \mathcal{K}) = \sum_{\xi \in \Xi_0} \frac{f_\xi(\mathcal{J} - \mathcal{K})}{\gamma} \ > \frac{\gamma}{\gamma} = 1 \ . \tag{14}$$

By (11), (13) and (14) we thus have found that

there exists a finite set $\Xi_0 \subseteq \Xi$, a selection $(\widetilde{g_\xi})_{\xi \in \Xi_0}$ of functionals $\widetilde{g_\xi} \in \widetilde{G_\xi}$ and there exists an indexed family $(\alpha_\xi)_{\xi \in \Xi_0}$ of real numbers with $\alpha_\xi \geq 0$ and $\sum_{\xi \in \Xi_0} \alpha_\xi \leq 1$ such that for any $(k, j) \in \mathcal{K} \times \mathcal{J}$ we have

$$\Big(\sum_{\xi \in \Xi_0} \alpha_\xi \cdot \langle\, j(\xi) - k(\xi) \ , \ \widetilde{g_\xi} \,\rangle\Big) \ > \ 1 \ . \tag{15}$$

The statement (15) is the negation of (iv') as well as (1) is the negation of (v'). Since (15) was concluded from (1) we see that (iv') implies (v'). Thus the theorem has been proved. □

Replacing $\mathcal{K} \times \mathcal{J}$ by $\widetilde{\mathcal{K}}$ our proof of Theorem 4.3 proves also the following theorem 4.4.

Theorem 4.4 *Theorem 4.3 remains true if we replace the compact convex set* $\mathcal{K} \times \mathcal{J} \subset \prod_{\xi\in\Xi} W'_\xi \times \prod_{\xi\in\Xi} W'_\xi$ *by an arbitrary compact convex set* $\widetilde{\mathcal{K}} \subset \prod_{\xi\in\Xi} W'_\xi \times \prod_{\xi\in\Xi} W'_\xi$.

We obtain the theorems 4.2 and 4.1 by a proof of:

[Theorem 4.3 $\Rightarrow$ Theorem 4.2]

We assume that the hypotheses of Theorem 4.2 are true and consider Theorem 4.3 in the following special case:

Let Θ, V, W, G, $\mathcal{L}$, $(K_\vartheta)_{\vartheta\in\Theta}$, $(\varepsilon_\vartheta)_{\vartheta\in\Theta}$ and $(v_\vartheta)_{\vartheta\in\Theta}$ denote the same mathematical objects as in Theorem 4.2. Suppose without loss of generality that Θ and V are disjoint. Let $\Xi = \Theta \cup V$. Let $G_\xi = G$ for $\xi \in \Theta$ and $G_\xi = \{0\}$ for $\xi \in V$. Let $\varepsilon_\xi \geq 0$ be arbitrary for $\xi \in V$ and let $W_\xi = W$ for all $\xi \in \Xi$. Let $\mathcal{K} := \prod_{v\in V}\{0\} \times \prod_{\vartheta\in\Theta} K_\vartheta$ and let

$$\mathcal{J} := \{j \in (W')^\Xi \mid \exists l \in \mathcal{L} \text{ s.t. } v \in V, \vartheta \in \Theta \Rightarrow [j(v) = l(v) \text{ and } j(\vartheta) = l(v_\vartheta)]\}$$

(Note that $\mathcal{J}$ is compact since it is a continuous image of $\mathcal{L}$.)

In this special case the hypotheses of Theorem 4.2 are equivalent with the conjunction of the hypotheses (i),(ii) and (iii) of Theorem 4.3. Thus by Theorem 4.3 the hypothesis (v) of Theorem 4.3 also holds. But the conjunction of the hypotheses (i),(ii) and (v) of Theorem 4.3 implies in the case we consider the conclusion of Theorem 4.2. Thus Theorem 4.2 has been proved. $\square$

Remark 4.1 The theorems 4.1 – 4.4 remain true if we replace the dual space (W', σ') resp. the dual spaces (W'_ξ, σ') by generalized dual spaces. Given a topological vector-space (W, τ) we call a topological vector-space (W^*, σ^*) a generalized dual if

- any $w^* \in W^*$ acts as a continuous linear functional $w^* : W \to \mathbb{R}$
- $\forall w' \in W' \;\; \exists w^* \in W^*$ such that $\forall w \in W \;\; w^*(w) = w'(w)$,
- the topology σ^* is the weakest topology such that for any $w \in W$ the mapping $w^* \mapsto w^*(w)$ is continuous on W^*.

We could have introduced instead of dual spaces and generalized duals the notion of dual pairing (also called dual system or duality [see [7] chapter IV, 1.]). But we think that our approach is more intuitive and helpful for the reader not familiar with topological vector spaces.

5 The case of adapted experiments

We define now adapted experiments and adapted decision rules and generalize the classical randomization criterion (Theorem 2.1) as well as the black-box criterion (Theorem 2.2) to the adapted setting. This makes it possible to apply the black-box criterion and the randomization criterion to stochastic processes. (See [5] and compare with section 6 of this paper).

Let a measurable space $(\Omega, \mathcal{A})$, a family $(P_\vartheta)_{\vartheta\in\Theta}$ of probability measures on $(\Omega, \mathcal{A})$ and a family $\mathcal{A}_t$ of sub-σ-algebras of $\mathcal{A}$ be given. We call $E := (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta})$ an adapted statistical experiment. (Note that $(\mathcal{A}_t)_{t\in T}$ does not need to be a filtration. But of course $(\mathcal{A}_t)_{t\in T}$ can be any filtration. Therefore our theory of the comparison of adapted experiments includes the filtered situation as a special case. Our definition coincides in the filtered case with the definition given in [6] 1.10 and [5].) The L-space $L(E)$ of the adapted experiment E is defined to coincide with the L-space of $(\Omega, \mathcal{A}, (P_t)_{t\in T})$.

Definition 5.1 Let $(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T})$ and $(\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T})$ be adapted measurable spaces and let $K : \Omega \times \mathcal{B} \to [0, 1]$ be a Markov kernel. We say that K is an $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$ adapted Markov kernel if for any $t \in T$ and any $B_t \in \mathcal{B}_t$ the mapping $\omega \mapsto K(\omega, B_t)$ is $\mathcal{A}_t$ measurable. We denote the space of all $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$ adapted Markov kernels by $\kappa_{(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}}$.

Let two adapted statistical experiments $E := (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta})$ and $F := (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta})$ be given. We denote by $\mathcal{S} \subseteq ba(\Omega_2, \mathcal{B})^{L(E)}$ the space of stochastic operators from $L(E)$ to $ba(\Omega_2, \mathcal{B})$. We let $\mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$ be the closure of $\kappa_{(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}} \cap \mathcal{S}$ in $\mathcal{S} \subseteq ba(\Omega_2, \mathcal{B})^{L(E)}$ (endowed with the topology of point wise convergence). We say that $\mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$ is the space of $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operators. A mapping $S : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ is called a $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operator if $S \in \mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$.

Let $\mathcal{M}_b(\Omega_2, \mathcal{B}_t)$ denote the vector space of bounded $\mathcal{B}_t$-measurable real valued functions on $(\Omega_2, \mathcal{B})$ endowed with the supremum norm. Let σ'_t be the topology which $\mathcal{M}_b(\Omega_2, \mathcal{B}_t)$ induces on $ba(\Omega_2, \mathcal{B})$ by the mappings $\mu \mapsto \int m \, d\mu$ with $m \in \mathcal{M}_b(\Omega_2, \mathcal{B}_t)$.

Remark 5.1 From Lemma A2 of the appendix we conclude that the space $\mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$ of $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operators is compact and convex. It is easy to prove that $S \in \mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$ implies that

$$\mu, \nu \in L(E) \text{ implies } [\mu \mid_{\mathcal{A}_t} = \nu \mid_{\mathcal{A}_t} \Longrightarrow S(\mu) \mid_{\mathcal{B}_t} = S(\nu) \mid_{\mathcal{B}_t}] .$$

We can now derive an adapted version (Theorem 5.1) of the black-box criterion from Theorem 4.3 with nearly no additional effort:

Theorem 5.1 *Let*

$$E := (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta}) \quad \textit{and}$$
$$F := (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta})$$

be adapted statistical experiments indexed by the same sets Θ *and* T *and let* $(\varepsilon_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ *be a family of nonnegative real numbers. Suppose that for any finite set* $\Theta_0 \times T_0 \subset \Theta \times T$ *and any selection* $(g_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0}$ *of functions* $g_{\vartheta,t} : \Omega_2 \mapsto [-1,+1]$ *such that* $g_{\vartheta,t}$ *is* $\mathcal{B}_t$*-measurable there exists a* $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$*-adapted stochastic operator* $S : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ *such that*

$$\int g_{\vartheta,t}\, dS(P_\vartheta) \le \int g_{\vartheta,t}\, dQ_\vartheta + \varepsilon_{\vartheta,t} \quad \textit{for all} \quad \vartheta \in \Theta_0 \quad \textit{and} \quad t \in T_0\ .$$

Then there exists a $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}-$*adapted stochastic operator* $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ *such that*

$$\|\widetilde{S}(P_\vartheta) - Q_\vartheta\|_{\mathcal{B}_t} := \sup_{\substack{g\in\mathcal{M}_b(\Omega_2,\mathcal{B}_t)\\ -1\le g\le 1}} (\int g\, d[\widetilde{S}(P_\vartheta) - Q_\vartheta]) \le \varepsilon_{\vartheta,t}$$

for all $\vartheta \in \Theta$ *and* $t \in T$.

Proof of Theorem 5.1 By Remark 5.1 we know that the space $\mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}$ of $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operators from $L(E)$ to $ba(\Omega_2, \mathcal{B})$ is a compact convex subset of $(ba(\Omega_2, \mathcal{B}), \sigma')^{L(E)}$. The set $\{m \in \mathcal{M}_b(\Omega_2, \mathcal{B}_t), | -1 \le m \le 1\}$ is a closed convex subset of $\mathcal{M}_b(\Omega_2, \mathcal{B})$.

Consider Theorem 4.3 together with Remark 4.1 in the following special case:

Let $\Xi := (\Theta \cup L(E)) \times T$. For $\xi = (x,t)$ let $W_\xi := \mathcal{M}_b(\Omega_2, \mathcal{B}_t)$ and let $(W^*_\xi, \sigma^*) := (ba(\Omega_2, \mathcal{B}), \sigma'_t)$. Let

$$G_\xi := \begin{cases} \{m \in \mathcal{M}_b(\Omega_2, \mathcal{B}_t) \mid -1 \le m \le 1\} & \text{if} \quad \xi = (\vartheta, t) \in \Theta \times T \\ \{0\} & \text{if} \quad \xi \in L(E) \times T \ \text{ and} \end{cases}$$

let $\varepsilon_\xi > 0$ be arbitrary if $\xi \in L(E) \times T$.

Let $K = \{k\}$ with $k : \Xi \mapsto ba(\Omega_2, \mathcal{B})$ defined by

$$k(\xi) := \begin{cases} Q_\vartheta & \text{if} \quad \xi = (\vartheta, t) \in \Theta \times T \\ 0 & \text{if} \quad \xi \in L(E) \times T. \end{cases}$$

Let $\mathcal{J} := \{s \circ h \mid s \in \mathcal{S}_{(L(E),(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T})}\}$ with h given by

$$h(\xi) := \begin{cases} P_\vartheta & \text{if} \quad \xi = (\vartheta, t) \in \Theta \times T \\ 0 & \text{if} \quad \xi \in L(E) \times T. \end{cases}$$

In the special case under consideration the hypotheses of Theorem 5.1 are equivalent with the conjunction of the hypotheses (i),(ii) and (iii) of the Remark 4.1 modification of Theorem 4.3. Thus by Theorem 4.3 hypothesis (v) of Theorem 4.3 also holds in the special case. But the conjunction of the hypotheses (i),(ii) and (v) of the Remark 4.1 modification of Theorem 4.3 implies in the special case we consider the conclusion of Theorem 5.1. Thus Theorem 5.1 has been proved. $\square$

Next we state the adapted case of the randomization criterion (Theorem 5.2). We will prove the converse of Theorem 5.2 after Remark 5.2. The equivalence of the theorems 5.1 and 5.2 will be proved in the end of the section.

Theorem 5.2 *Let*

$$\begin{aligned} E &:= (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta}) \quad \text{and} \\ F &:= (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta}) \end{aligned}$$

be adapted statistical experiments indexed by the same sets Θ *and* T *let* $(\varepsilon_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ *be an indexed family of reals* ≥ 0. *Suppose that for any adapted measurable space* $(\Omega_3, \mathcal{D}, (\mathcal{D}_t)_{t\in T})$ *(the adapted decision space), any finite subset* $\Theta_0 \times T_0 \subseteq \Theta \times T$, *any stochastic operator* $M_K : L(F) \to ba(\Omega_3, \mathcal{D})$ *which is induced by a* $(\mathcal{B}_t)_{t\in T}/(\mathcal{D}_t)_{t\in T}$*-adapted Markov-kernel* K *and any selection* $(f_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ *of functions* $f_{\vartheta,t} : \Omega_3 \mapsto [-1,+1]$ *such that* $f_{\vartheta,t}$ *is* $\mathcal{D}_t$*-measurable there exists an* $(\mathcal{A}_t)_{t\in T}/(\mathcal{D}_t)_{t\in T}$*-adapted stochastic operator* $M : L(E) \mapsto ba(\Omega_3, \mathcal{D})$ *such that*

$$\int f_{\vartheta,t}\, dM(P_\vartheta) \leq \int f_{\vartheta,t}\, dM_K(Q_\vartheta) + \varepsilon_{\vartheta,t} \quad \text{for all} \quad (\vartheta, t) \in \Theta_0 \times T_0 \, .$$

Then there exists an $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$*-adapted stochastic operator* $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ *such that*

$$\|Q_\vartheta - \widetilde{S}(P_\vartheta)\|_{\mathcal{B}_t} = \sup_{\substack{g\in\mathcal{M}_b(\Omega_2,\mathcal{B}_t)\\ -1\leq g\leq 1}} (\int g\; d[\widetilde{S}(P_\vartheta) - Q_\vartheta]) \leq \varepsilon_{\vartheta,t}$$

for all $\vartheta \in \Theta$ *and* $t \in T$.

Remark 5.2 In general it is not possible to define the operator $\widetilde{S}$ in Theorem 5.2 in such a way that $\widetilde{S}(L(E)) \subseteq L(F)$. Therefore we find in the formulation of Theorem 5.2 of this paper (as well as in the formulation of Theorem 1 of [5]) $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ instead of $\widetilde{S} : L(E) \mapsto L(F)$. This is in contrast to the situation of Theorem 2.1. An analysis of the proof of Theorem 2.1 shows that the key ingredient (41.7 of [11] or 4.5.11. of [12] which is stated below) for proving that $\widetilde{S}(L(E) \subseteq L(F)$ is not available in the adapted situation:

For adapted experiments there exists no result analogous to [11] 41.7 or [12] 4.5.11., which says that given a statistical experiment $F = (\Omega_2, \mathcal{B}, (Q_\vartheta)_{\vartheta\in\Theta})$

there exists a stochastic operator $T : ba(\Omega_2, \mathcal{B}) \mapsto L(F)$ such that $T |_{L(F)} = id_{L(F)}$. To see this let us consider the following situation: Let $\mathcal{B}$ denote the σ-algebra of Borel-sets on $[0, 1]$ and let $(\mathcal{B}_n)_{n\in\mathbb{N}}$ denote an arbitrary filtration of finite algebras which fulfills:

$\sigma(\bigcup_{n\in\mathbb{N}} \mathcal{B}_n) = \mathcal{B}$ and

for any nonempty $B \in \bigcup_{n\in\mathbb{N}} \mathcal{B}_n$ we have $\lambda(B) > 0$.

Let $F := ([0, 1], \mathcal{B}, (\mathcal{B}_n)_{n\in\mathbb{N}}, (Q_\vartheta)_{\vartheta\in\Theta})$ be an adapted experiment such that $L(F)$ consists of the space of all finite signed measures on $[0, 1]$ which are absolutely continuous with respect to Lebesgue measure. There exists no $(\mathcal{B}_n)_{n\in\mathbb{N}}/(\mathcal{B}_n)_{n\in\mathbb{N}}$-adapted stochastic operator $T : ba(\Omega_2, \mathcal{B}) \to L(F)$ such that $T |_{L(F)} = id_{L(F)}$. To see this we proceed indirect. Let $x \in [0, 1]$ be arbitrary and let δ_x denote the Dirac measure at x. Choose now a falling sequence $(B_n)_{n\in\mathbb{N}}$ of sets $B_n \in \mathcal{B}_n$ such that

B_n is an atom of $\mathcal{B}_n$ and $\{x\} = \bigcap_{n\in\mathbb{N}} B_n$

Let μ_n be the uniform probability measure on B_n. Then we obtain from the fact that T is filtered and the identity on $L(F)$ that

$$T(\delta_x)(B_m) = \lim_{n\to\infty} T(\mu_n)(B_m) = \lim_{n\to\infty} \mu_n(B_m) = 1.$$

This forces $T(\delta_x) = \delta_x$ which contradicts our assumption that all elements of $L(F)$ are absolutely continuous with respect to Lebesgue measure.

Proof of the converse of Theorem 5.2. Let $M_K : L(F) \to ba(\Omega_3, \mathcal{D})$ be a stochastic operator induced by a $(\mathcal{B}_t)_{t\in T}/(\mathcal{D}_t)_{t\in T}$-adapted Markov-kernel K. Let $f_{\vartheta,t} : \Omega_3 \mapsto [-1, +1]$ be a $\mathcal{D}_t$-measurable function. Let $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ be an $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operator such that

$$\|Q_\vartheta - \widetilde{S}(P_\vartheta)\|_{\mathcal{B}_t} = \sup_{\substack{g\in\mathcal{M}_b(\Omega_2,\mathcal{B}_t)\\ -1\le g\le 1}} (\int g \; d[\widetilde{S}(P_\vartheta) - Q_\vartheta]) \le \varepsilon_{\vartheta,t}$$

Define an an $(\mathcal{A}_t)_{t\in T}/(\mathcal{D}_t)_{t\in T}$-adapted stochastic operator $M : L(E) \mapsto ba(\Omega_3, \mathcal{D})$ for $\mu \in L(E)$ by

$$[M(\mu)](.) := \int K(x, .) \; d[\tilde{S}(\mu)](x).$$

and let $g(x) := \int f_{\vartheta,t}(y)K(x, dy)$. Then

$$\int f_{\vartheta,t} \; dM(P_\vartheta) = \int f_{\vartheta,t}(y)K(x, dy) \; d[\tilde{S}(P_\vartheta)](x) = \int g \; d[\tilde{S}(P_\vartheta)] \le$$

$$\int g \; dQ_\vartheta + \varepsilon_{\vartheta,t} = \int f_{\vartheta,t}(y)K(x, dy) \; dQ_\vartheta(x) + \varepsilon_{\vartheta,t} = \int f_{\vartheta,t} \; dM_K(Q_\vartheta) + \varepsilon_{\vartheta,t} \; \Box.$$

Proof of the equivalence of the theorems 5.1 and 5.2. We use Theorem 3.1 for the proof of this equivalence. The stochastic operators involved in the theorems 5.1 and 5.2 are in one one correspondence to morphisms in theorem 3.1 as follows: $\text{mor}_S = S$, $\text{mor}_K = M_K$, $\text{mor}_{\widetilde{S}} = \widetilde{S}$, $M = \widetilde{K} \circ S$ and $\text{mor}_M = M$.

The Experiments of the theorems 5.1 and 5.2 correspond to the objects of theorem 3.1 in the following way $\text{ob}_E = E$, $\text{ob}_F = F$, $ob_\bullet = \mathbf{Ran}(\text{mor}_S) = (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (S(P_\vartheta))_{\vartheta\in\Theta})$, $\mathbf{Ran}(\text{mor}_K) = (\Omega_3, \mathcal{D}, (\mathcal{D}_t)_{t\in T}, (M_K(Q_\vartheta))_{\vartheta\in\Theta})$, $\mathbf{Ran}(\text{mor}_M) = (\Omega_3, \mathcal{D}, (\mathcal{D}_t)_{t\in T}, (M(P_\vartheta))_{\vartheta\in\Theta})$.

The families of properties are specified in such a way that they correspond to the conditions the stochastic operators must fulfill in the theorems 5.1 and 5.2 which are given by the selection of functions $(g_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0}$ and the real numbers $\varepsilon_{\vartheta,t}$:

$$\hat{p} \in \mathbf{Prop}(ob_F) \Leftrightarrow \begin{cases} \text{there exist finite sets } \Theta_0 \subseteq \Theta,\ T_0 \subseteq T \text{ and a selection } (g_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0} \text{ of functions } g_{\vartheta,t} \in \mathcal{M}_b(\Omega_2,\mathcal{B}_t), \\ -1 \le g_{\vartheta,t} \le 1 \text{ such that } \hat{p}(\text{ob}_{F'}) = \text{true} \\ \\ \text{if and only if} \\ \\ \text{ob}_{F'} \text{ is of the form } (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (\hat{R}_\vartheta)_{\vartheta\in\Theta}) \text{ and for} \\ \text{all } \vartheta \in \Theta_0, t \in T_0 \text{ we have } \int g_{\vartheta,t}\, d[\hat{R}_\vartheta - Q_\vartheta] \le \varepsilon_{t,\vartheta}, \end{cases}$$

$$p \in \mathbf{Prop}(\mathbf{Ran}(mor_K)) \Leftrightarrow \begin{cases} \text{there exist finite sets } \Theta_0 \subseteq \Theta,\ T_0 \subseteq T \\ \text{and a selection } (f_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0} \text{ of functions} \\ f_{\vartheta,t} \in \mathcal{M}_b(\Omega_3,\mathcal{D}_t),\ -1 \le f_{\vartheta,t} \le 1 \text{ such} \\ \text{that } p(\text{ob}_{D'}) = \text{true} \\ \\ \text{if and only if} \\ \\ \text{ob}_{D'} \text{ is of the form } (\Omega_2, \mathcal{D}, (\mathcal{D}_t)_{t\in T}, R_\vartheta) \\ \text{and for all } \vartheta \in \Theta_0, t \in T_0\ \int f_{\vartheta,t}\, d[R_\vartheta - \\ M_K(Q_\vartheta)] \le \varepsilon_{t,\vartheta} \end{cases}$$

With these definitions the equivalence of the theorems 5.1 and 5.2 is a consequence of Theorem 3.1.

Remark 5.3 Although we did not need it in the proof stated above, we would like to mention the following fact:

The morphism $\text{mor}_{\widetilde{K}}$ occurring in the proof of 3.1 corresponds to the adapted stochastic operator

$$\tilde{K} : L(\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (S(P_\vartheta))_{\vartheta\in\Theta}) \to L(\Omega_3, \mathcal{D}, (\mathcal{D}_t)_{t\in T}, ([M_K \circ S](P_\vartheta))_{\vartheta\in\Theta})$$

that is induced by K.

6 Extensions of the theory beyond adaptedness

Definition 6.1 Let e be a property which a stochastic operator $S \in \mathcal{S} \subset ba(\Omega_2, \mathcal{B})^{L(E)}$ may posses or may not posses, i.e. $e(S) =$ true or $e(S) =$ false. We say that $S \in \mathcal{S}$ is e-stochastic if $e(S) =$ true. We call the property e a compact convex property if the set of e-stochastic operators is compact and convex, i.e., if $\{S \in \mathcal{S} \mid e(S) = true\}$ is compact and convex in $ba(\Omega_2, \mathcal{B})^{L(E)}$.

Remark 6.1 An example of a compact convex property is $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adaptedness. Another example is the following:

Let us denote by $\mathcal{R}$ the Borel σ-algebra on $\mathbb{R}$ and by $\mathcal{R}^{\mathbb{N}}$ the Borel σ-algebra on $\mathbb{R}^{\mathbb{N}}$. Let x_i denote the i-th coordinate of x. Let $K : \mathbb{R}^{\mathbb{N}} \times \mathcal{R}^{\mathbb{N}} \to [0,1]$ be a Markov Kernel such that K is of the form $K := K_1 \otimes K_2 \otimes \ldots$ with $K_i : \mathbb{R} \times \mathcal{R} \to [0,1]$ Markov kernels and $\bigotimes_{i=1}^{\mathbb{N}} K_i(x, R) = \prod_{i=1}^{\mathbb{N}} K_i(x_i, R_i)$ for $R = \prod R_i$. Then we say that the Markov Kernel K possesses product form. We denote by π the space of all Markov kernels which posses product form. We say that a Markov kernel is of multi linear form, if it is in the convex hull of π. Let $E = (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}, (P_\vartheta)_{\vartheta\in\Theta})$. We say that a stochastic operator $S : L(E) \to ba(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ is of multi linear form if S is in the closure of the convex hull of $\pi \cap \mathcal{S}$ in $\mathcal{S}$. It is easy to see that the property of being of multi linear form is convex and compact.

A further example of a compact convex property an operator may posses is provided by the property of being a convolution operator. If $\Omega = \Omega_2$ is a measurable group we call a stochastic operator a convolution operator if it is in the closure of the set of all stochastic operators induced by Markov kernels K for which $K(x + y, A + y) = K(x, A)$ holds.

The conjunction of two compact convex properties is of course also a compact convex property.

Remark 6.2 Theorem 5.1 remains valid, if we replace (in the hypotheses and in the conclusion of Theorem 5.1) the requirement that the stochastic operators are adapted by the requirement that the operators are e-stochastic for some compact convex property e. This can be deduced from Theorem 4.3 in the same way as 5.1 has been deduced. If property e or property $\neg e$ can also be assigned to any stochastic operator from $L(E)$ or $L(F)$ to $ba(\Omega_3, \mathcal{D})$ for arbitrary decision spaces $(\Omega_3, \mathcal{D})$ and the concatenation of two stochastic operators with property e is again a stochastic operator with property e, then Theorem 5.2 remains also valid with adaptednes replaced by property e. This can be established using theorem 3.1 in the same way as 5.2 has been established.

One can also deduce the following example from a property e-Version of Theorem 5.1. The assertion in the following example can be derived from Remark 6.2 by an argument completely analogous to the argument given in the proof of the equivalence of the theorems 5.1 and 5.2. (Note that in this case we consider just one decision space $(\Omega_3, \mathcal{D}) = (\mathbb{R}^T, \mathcal{R}^T)$.)

Example 6.1 Let $pr_t : \mathbb{R}^T \to \mathbb{R}$ denote the i-th coordinate projection. Let $\mathcal{R}_t$ be the σ-algebras induced by pr_t, i.e., $\mathcal{R}_t := \sigma(\{pr_t^{-1}(O) \mid O \text{ open in } \mathbb{R}\})$. Let $E := (\mathbb{R}^T, \mathcal{R}^T, (\mathcal{R}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta})$ and $F := (\mathbb{R}^T, \mathcal{R}^T, (\mathcal{R}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta})$ be two adapted statistical experiments. Let $(\varepsilon_t)_{t\in T}$ be a sequence of positive real numbers. Suppose that for any finite set $\Theta_0 \times T_0 \subset \Theta \times T$ for any stochastic operator $M_K : L(F) \mapsto ba(\mathbb{R}^T, \mathcal{R}^T)$ which is induced by a $(\mathcal{R}_t)_{t\in T}/(\mathcal{R}_t)_{t\in T}$-adapted Markov-kernel K of multi linear form and any selection $(h_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ of Borel measurable functions $h_{\vartheta,t} : \mathbb{R} \to [-1, +1]$ there exists a $(\mathcal{R}_t)_{t\in T}/(\mathcal{R}_t)_{t\in T}$-adapted stochastic operator $M : L(E) \mapsto ba(\mathbb{R}^T, \mathcal{R}^T)$ of multi linear form such that

$$\int h_{\vartheta,t} \circ pr_t \, dM(P_\vartheta) \le \int h_{\vartheta,t} \circ pr_t \, dM_K(Q_\vartheta) + \varepsilon_{\vartheta,t} \quad \text{for all} \quad (\vartheta, t) \in \Theta_0 \times T_0 \, .$$

Then there exists a $(\mathcal{R}_t)_{t\in T}/(\mathcal{R}_t)_{t\in T}$-adapted stochastic operator $\widetilde{S} : L(E) \mapsto ba(\mathbb{R}^T, \mathcal{R}^T)$ of multi linear form such that

$$\|Q_\vartheta - \widetilde{S}(P_\vartheta)\|_{\mathcal{R}_t} = \sup_{\substack{h\in\mathcal{M}_b(\mathbb{R},\mathcal{R})\\ -1\le h\le 1}} (\int h \circ pr_t \, d[\widetilde{S}(P_\vartheta) - Q_\vartheta]) \le \varepsilon_{\vartheta,t}$$

for all $\vartheta \in \Theta$ and $t \in T$. (Here $\mathcal{M}_b(\mathbb{R}, \mathcal{R})$ denotes the vector space of bounded, Borel-measurable, real valued functions.)

Remark 6.3 Example can be interpreted in the special case that $T = \mathbb{N}$ and P_ϑ, Q_ϑ are Markov chains. Then the example says: Compare arbitrary hidden Markov models derived from the Markov model Q_ϑ with generalized hidden Markov models derived from the Markov model P_ϑ by the use of loss-functions based on coordinate projections. If the models are close with respect to this comparison, then there exists a generalized hidden Markov model derived from P_ϑ which is close to the Markov model Q_ϑ in a uniform sense based on the coordinate projections. (The simple hidden Markov model we have in mind consists of a stochastic process on $\mathbb{R}^\mathbb{N}$ which is the image of a Markov chain on $\mathbb{R}^\mathbb{N}$ under a stochastic operator S of multi linear form induced by a Markov kernel. We speak of a generalized hidden Markov model if the stochastic operator S is of multi linear form, but is not necessarily induced by a Markov kernel.)

Before we give another example of an adapted non-filtered situation we state the following proposition, which is almost trivial. (See [2] chapter 9 end of section 3)

Proposition 6.1 *Let ν_1, ν_2 be two finite measures on an interval $(a, b]$ and let Φ_{η^2} denote the measure of the $N(0, \eta^2)$-distribution. Then*

$$\|\nu_1 \star \Phi_{\eta^2} - \nu_2 \star \Phi_{\eta^2}\| \le 2\frac{b-a}{\eta} \cdot \min(\nu_1((a, b]), \nu_2((a, b])) + |\nu_1((a, b]) - \nu_2((a, b])|.$$

Example 6.2 Let $pr_i : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ denote the i-th coordinate projection. Let $\mathcal{Y}_n$ be the σ-algebras induced by the functions $Y_n(x) := \sum_{i=1}^{n} pr_i(x)$, i.e., $\mathcal{Y}_n := \sigma(\{Y_n^{-1}(O) \mid O \text{ open in } \mathbb{R}\})$. Let Φ_{σ^2} be the probability measure of the $N(0, \sigma^2)$ distribution and μ a probability measure with expectation 0 and variance σ^2. Let $E := (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}, (\mathcal{Y}_n)_{n\in\mathbb{N}}, P_\vartheta)_{\vartheta\in\mathbb{R}})$ and $F := (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}, (\mathcal{Y}_n)_{n\in\mathbb{N}}, Q_\vartheta)_{\vartheta\in\mathbb{R}})$ be two adapted statistical experiments with

$$P_\vartheta := \bigotimes_{i=1}^{\infty} (\mu \star \delta_\vartheta) \quad \text{and} \quad Q_\vartheta := \bigotimes_{i=1}^{\infty} (\Phi_{\sigma^2} \star \delta_\vartheta).$$

Then there exists a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ of positive real numbers with $\lim_{n\to\infty} \varepsilon_n \downarrow 0$ and a $(L(E), (\mathcal{Y}_n)_{n\in\mathbb{N}}/(\mathcal{Y}_n)_{n\in\mathbb{N}})$-adapted convolution operator $\widetilde{S}$ of multi linear form such that $\|\widetilde{S}(P_\vartheta) - Q_\vartheta\|_{Y_n} \le \varepsilon_n$. An example of such a stochastic operator is given by the stochastic operator $\widetilde{S}_K$ induced by the convolution kernel $K := \bigotimes_{i=1}^{\infty} K_i$ with $K_i(x, [a, b]) = N(x, c_n^2)$ with c_n a sequence of positive real numbers which tends with the right velocity to 0.

That such a Markov kernel K is indeed doing the job can be proved by application of the central limit theorem, the description of convergence in distribution provided by the portmanteau theorem (See [4] Theorem 11.1.1 (d)) and by Proposition 6.1 as follows:

Let $\mu^{\star n}$ be the n-fold convolution product of μ and let $\mu_n = \frac{\mu^{\star n}}{\sqrt{n}}$. Then μ_n converges by the central limit theorem in distribution to Φ_{σ^2}. Let a finite sequence $(z_j^m)_{j=0}^{2^{m+1}}$ of real numbers z_j^m be recursively defined by

$$z_0^m = -m \quad \text{and} \quad z_{j+1}^m - z_j^m = \frac{m}{2^m}.$$

Let $\mathcal{Z}_m$ be the partition defined by

$$\mathcal{Z}_m := \{\, \mathbb{R} \setminus (-m, m] \,\} \cup \{\, (z_j^m, z_{j+1}^m] \mid j = 0, \ldots, 2^{m+1} \,\}$$

For any number $\eta \in (0, 1)$ we chose $m(\eta) \in \mathbb{N}$ such that $\Phi_{\sigma^2}(Z) \le \frac{\eta^2}{m(\eta)}$ if $Z \in \mathcal{Z}_{m(\eta)}$. By Theorem 11.1.1 (d) of [4] there exists an $N_\eta \in \mathbb{N}$ such that

(i) $\quad \forall n \ge N_\eta, \forall Z \in \mathcal{Z}_{m(\eta)}$ we have $|\mu_n(Z) - \Phi_{\sigma^2}(Z)| \le \frac{\eta^2}{m(\eta) \cdot 2^{m(\eta)+1}}$.

Let μ_n^i and $\Phi_{\sigma^2}^i$ be the restriction of the measures μ_n and Φ_{σ^2} to the Borel measurable subsets of $(z_{i-1}^{m(\eta)}, z_i^{m(\eta)}]$ if $1 \le i \le 2^{m+1}$ and let μ_n^0 and $\Phi_{\sigma^2}^0$ be the restriction of the measures μ_n and Φ_{σ^2} to the Borel measurable subsets of $\mathbb{R} \setminus (-m, m]$. Note that our assumptions imply that

(ii) $\quad \forall Z \in \mathcal{Z}_{m(\eta)}, \forall n \ge N_\eta$ we have $\mu_n(Z) \le \frac{2\eta^2}{m(\eta)}$ and $\|\mu_n^0 - \Phi_{\sigma^2}^0\| \le 3\eta$.

Thus we get by an application of (i), (ii) and Proposition 6.1 that for $n \ge N_\eta$

$$\|\mu_n \star \Phi_{\eta^2} - \Phi_{\sigma^2+\eta^2}\| \le \sum_{i=0}^{2^{m(\eta)+1}} \|\mu_n^i \star \Phi_{\eta^2} - \Phi_{\sigma^2}^i \star \Phi_{\eta^2}\| \le$$

$$\|\mu_n^0 - \Phi^0_{\sigma^2}\| + \sum_{i=1}^{2^{m(\eta)+1}} 2\frac{z_j^{m(\eta)} - z_{j-1}^{m(\eta)}}{\eta} \cdot \frac{2\eta^2}{m(\eta)} +$$

$$\sum_{i=1}^{2^{m(\eta)+1}} |\mu_n^i((z_{i-1}^{m(\eta)}, z_i^{m(\eta)}]) - \Phi^i_{\sigma^2}((z_{j-1}^{m(\eta)}, z_j^{m(\eta)}])| \leq$$

$$3\eta + 8\eta + \sum_{i=1}^{2^{m(\eta)+1}} |\mu_n((z_{i-1}^{m(\eta)}, z_i^{m(\eta)}]) - \Phi_{\sigma^2}((z_{j-1}^{m(\eta)}, z_j^{m(\eta)}])| \leq$$

$$11\eta + \sum_{i=0}^{2^{m(\eta)+1}} \frac{\eta}{2^{m(\eta)+1}} \leq 12\eta.$$

Thus there exists a function $\eta \mapsto N(\eta)$ such that $N(\eta) \in \mathbb{N}$ is the smallest natural number such that

$$\|\mu_n \star \Phi_{\eta^2} - \Phi_{\sigma^2+\eta^2}\| \leq 12 \cdot n \quad \text{for all} \quad n \geq N(\eta)$$

Define real numbers c_n for all $n \in \mathbb{N}$ by $c_n := 12 \cdot \inf\{\eta > 0 \mid N(\eta) \leq n\}$. Then for all $\eta > c_n$

$$\|\mu_n \star \Phi_{\eta^2} - \Phi_{\sigma^2+\eta^2}\| \leq \|\mu_n \star \Phi_{c_n^2} - \Phi_{\sigma^2+c_n^2}\| \leq c_n$$

Note that $c_n \downarrow 0$ and $\sqrt{[\sum_{i=1}^n c_i^2] \cdot n^{-1}} \geq c_n$. Thus

$$\|\mu_n \star \Phi_{[\sum_{i=1}^n c_i^2] \cdot n^{-1}} - \Phi_{\sigma^2+[\sum_{i=1}^n c_i^2] \cdot n^{-1}}\| \leq c_n$$

Let

$$\varepsilon_n := c_n + \|\Phi_{\sigma^2+[\sum_{i=1}^n c_i^2] \cdot n^{-1}} - \Phi_{\sigma^2}\|.$$

Since $c_n \downarrow 0$ we get that also $\varepsilon_n \downarrow 0$. Let $\widetilde{S}_K$ be the stochastic operator induced by the convolution kernel $K := \bigotimes_{i=1}^\infty K_i$ with $K_i(x, [a,b]) = N(x, c_n^2)$. We calculate finally

$$\|\widetilde{S}_K(P_\vartheta) - Q_\vartheta\|_{\mathcal{Y}_n} = \|\widetilde{S}_K(P_\vartheta) \circ Y_n^{-1} - Q_\vartheta \circ Y_n^{-1}\| =$$

$$\|\widetilde{S}_K(P_0) \circ Y_n^{-1} - Q_0 \circ Y_n^{-1}\| = \|\mu_n \star \Phi_{[\sum_{i=1}^n c_i^2] \cdot n^{-1}} - \Phi_{\sigma^2}\| =$$

$$\|\mu_n \star \Phi_{[\sum_{i=1}^n c_i^2] \cdot n^{-1}} - \Phi_{\sigma^2+[\sum_{i=1}^n c_i^2] \cdot n^{-1}}\| + \|\Phi_{\sigma^2+[\sum_{i=1}^n c_i^2] \cdot n^{-1}} - \Phi_{\sigma^2}\| \leq \varepsilon_n.$$

Since $c_n \downarrow 0$ and $\varepsilon_n \downarrow 0$ this proves the assertion. □

Remark: Given a convex compact property e which is more or equal restrictive than adaptedness and two adapted experiments E, F indexed by the same set $T \times \Theta$ (as defined in Theorem 5.1), then the generalized deficiency $\delta_e(E, F)$ of E with respect to F can be defined by $\delta_e(E, F) := t \mapsto \inf_S \sup_{\vartheta \in \Theta} \|Q_\vartheta - S(P_\vartheta)\|_{\mathcal{B}_t}$ (with the "inf" taken over all S with property e). In the case of Example 6.2 we have $T = \mathbb{N}$, e is the property of being an adapted convolution operator of multi linear form, and $[\delta_e(E, F)](n) \leq \varepsilon_n$. Further $\lim_{n\to\infty}[\delta_e(E, F)](n) = 0$. This means that by the use of adapted convolution operators of multi linear form we can (for large n) extract out of E nearly as much information as we can extract out of F. (Note that the concatenation of adapted convolution operators of multi linear form gives again an adapted convolution operator of multi linear form and thus remark 6.2 applies.)

7 Hilbert spaces, Diagonalizable operators and the Wave equation

The following theorem is the Hilbert space version of the black-box criterion. It is an easy consequence of Theorem 4.2 but it can not be established using the L-space version of the black-box criterion.

Theorem 7.1 *Let $\mathcal{H}$, $\widetilde{\mathcal{H}}$ be a Hilbert spaces and let Ξ be a set. Let $(v_\xi)_{\xi\in\Xi}$ be a indexed family of vectors in $\mathcal{H}$ and let $(K_\xi)_{\xi\in\Xi}$ be a family of convex norm-closed norm-bounded subsets of $\widetilde{\mathcal{H}}$. Let $\mathcal{L}$ be a norm-bounded norm-closed convex family of continuous linear operators $l : \mathcal{H} \to \widetilde{\mathcal{H}}$ such that for any finite set $\Xi_0 \subseteq \Xi$ and any family $(g_\xi)_{\xi\in\Xi_0}$ of elements of $\widetilde{\mathcal{H}}'$ with $\|g_\xi\|_{\widetilde{\mathcal{H}}'} = 1$ there exists an $l \in \mathcal{L}$ such that*

$$\inf_{w_\xi\in K_\xi} \langle\, l(v_\xi) - w_\xi \;,\; g_\xi \,\rangle \;\le\; \varepsilon_\xi \;.$$

Then there exists an $l \in \mathcal{L}$ such that for all $\xi \in \Xi$ we have:

$$\inf_{w_\xi\in K_\xi} \|l(v_\xi) - w_\xi\|_{\widetilde{\mathcal{H}}} \;\le\; \varepsilon_\xi \;.$$

Proof of Theorem 7.1 Note that the hypothesis $\|g_\xi\|_{\widetilde{\mathcal{H}}'} = 1$ is equivalent with the hypothesis $\|g_\xi\|_{\widetilde{\mathcal{H}}'} \le 1$. Note further, that the sets K_ξ and the set $\mathcal{L}$ are weakly compact. (This is the theorem of Alaoglu-Bourbaki in the case of the sets K_ξ [see [7] Chapter 3 Section 4.3] and can in the case of the set $\mathcal{L}$ be concluded using the Tychonoff product Theorem analogous to the theorem of Alaoglou-Bourbaki). Apply now Theorem 4.2 with $V = \mathcal{H}, W = \widetilde{\mathcal{H}}$, $\Theta = \Xi$ and $G = \{g \in \widetilde{\mathcal{H}}' \mid \|g\|_{\widetilde{\mathcal{H}}'} \le 1\}$. $\Box$

Definition 7.1 Let H_E and H be Hilbert spaces. Let H_E' denote the dual of H_E and let the dual H' of H be identified with H itself. Suppose that there exists an identification of the spaces H_E, H and H_E' in such a way that $H_E \subset H \subset H_E'$ as sets and that the identities $id_{H_E} : H_E \to H$ and $id_H : H \to H_E'$ are continuous. If we are given H_E and H identified as above, we say that H_E is canonically embedded into H. Given a measure space $(\Omega, \mathcal{B}, \mu)$ we denote by $\mathbb{L}^2(\Omega, \mathcal{B}, \mu)$ the space of square integrable real valued functions on $(\Omega, \mathcal{B}, \mu)$. We say that a continuous linear operator $\mathbb{A} : H_E \to H_E'$ is positive and diagonalizable if there exists a measurable space $(\Omega, \mathcal{B})$, measures μ_{H_E}, $\mu_{H_E'}$ on $(\Omega, \mathcal{B})$ and an isometric isomorphisms $U : H_E' \to \mathbb{L}^2(\Omega, \mathcal{B}, \mu'_{H_E})$ such that the restriction $U\,|_{H_E}$ of U to H_E is an isometry between H_E and $\mathbb{L}^2(\Omega, \mathcal{B}, \mu_{H_E})$ and such that the operator

$$\Psi : \mathbb{L}^2(\Omega, \mathcal{B}, \mu_{H_E}) \to \mathbb{L}^2(\Omega, \mathcal{B}, \mu_{H'_E}) \text{ defined by } \Psi := U \circ \mathbb{A} \circ U^{-1}$$

$$\text{fulfills } [\Psi(\phi)](\omega) = [\psi \circ \phi](\omega) \text{ almost everywhere } \mu_{H'_E}$$

for any $\phi \in \mathbb{L}^2(\Omega, \mathcal{B}, \mu_{H_E})$ and a $[0, \infty)$-valued $\mathcal{B}$-measurable function ψ.

Theorem 7.2 *Let H_E and H be Hilbert spaces and suppose that H_E is canonically embedded into H. Let $\aleph$ be a norm-bounded norm-closed convex family of positive, diagonalizable, continuous linear operators $\mathbb{A} : H_E \to H'_E$. Let Θ be a set. For any $\vartheta \in \Theta$ let $v_\vartheta = \begin{pmatrix} v^1_\vartheta \\ v^2_\vartheta \end{pmatrix} \in H_E \times H_E$ be such that for any $\mathbb{A} \in \aleph$ we have $\mathbb{A}(v^1_\vartheta) \in H$. Let $T = [0, \rho)$ be some non-degenerate interval. For any $\vartheta \in \Theta$ and $t \in T$ let $w_{\vartheta,t} = \begin{pmatrix} w^1_{\vartheta,t} \\ w^2_{\vartheta,t} \end{pmatrix} \in H_E \times H$.*

Suppose that for any finite set $\Theta_0 \times T_0 \subset \Theta \times T$ and any family $(g_{\vartheta,t})_{(\vartheta,t) \in \Theta_0 \times T_0}$ of elements $g_{\vartheta,t} = \begin{pmatrix} g^1_{\vartheta,t} \\ g^2_{\vartheta,t} \end{pmatrix} \in H'_E \times H' = (H_E \times H)'$ with $\|g_\vartheta\|_{(H_E \times H)'} = 1$ there exists an $\mathbb{A} \in \aleph$ such that

$$\left\langle \begin{pmatrix} v^1_\vartheta + t \cdot v^2_\vartheta \\ v^2_\vartheta - t \cdot \mathbb{A}(v^1_\vartheta) \end{pmatrix} - \begin{pmatrix} w^1_{\vartheta,t} \\ w^2_{\vartheta,t} \end{pmatrix} , \begin{pmatrix} g^1_{\vartheta,t} \\ g^2_{\vartheta,t} \end{pmatrix} \right\rangle \leq \varepsilon_{\vartheta,t} \cdot t \, .$$

Then there exists an $\mathbb{A} \in \aleph$ such that for all $\vartheta \in \Theta$ and $t \in T$ we have:

(i) $$\left\| \begin{pmatrix} v^1_\vartheta + t \cdot v^2_\vartheta \\ v^2_\vartheta - t \cdot \mathbb{A}(v^1_\vartheta) \end{pmatrix} - \begin{pmatrix} w^1_{\vartheta,t} \\ w^2_{\vartheta,t} \end{pmatrix} \right\| \leq \varepsilon_{\vartheta,t} \cdot t \, .$$

Proof: The theorem is easily derived from Theorem 7.1. To see this we let $\Sigma = \Theta \times [0, \rho)$, $\mathcal{H} = (H_E \times H_E)^2$, $\widetilde{\mathcal{H}} = H_E \times H$, $v_{\vartheta,t} = \left(\begin{pmatrix} v^1_\vartheta \\ v^2_\vartheta \end{pmatrix}, \begin{pmatrix} t \cdot v^1_\vartheta \\ t \cdot v^2_\vartheta \end{pmatrix} \right) \in (H_E \times H_E)^2$ and $K_\xi = \{w_{\vartheta,t}\}$ (for $\xi = (\vartheta, t)$). Let further $\mathcal{L}$ be the set of all operators $L : (H_E \times H_E)^2 \to H_E \times H'_E$ of the form

$$L\left(\begin{pmatrix} h^1_1 \\ h^2_1 \end{pmatrix}, \begin{pmatrix} h^1_2 \\ h^2_2 \end{pmatrix} \right) = \begin{pmatrix} h^1_1 + h^2_2 \\ h^1_2 - \mathbb{A}(h^2_1) \end{pmatrix} \text{ with } \mathbb{A} \in \aleph.$$

Since the compactness of $\aleph$ implies the compactness of $\mathcal{L}$ we see now that Theorem 7.2 follows from Theorem 7.1.

Theorem 7.3 *Suppose that the hypothesis of Theorem 7.2 hold. Suppose additionally that $\mathbb{B} : H_E \to H'_E$ is a positive, diagonalizable, continuous operator. Let $\{v^1_\vartheta \mid \vartheta \in \Theta\}$ be a subset of H_E which is dense in the unit ball of H_E, such that $\mathbb{B}(v^1_\vartheta) \in H$ and such that for all $\mathbb{A} \in \aleph$ also $\mathbb{A}(v^1_\vartheta) \in H$. Let $v^2_\vartheta \in H_E$. Let $t \to w^1_{\vartheta,t}$ be the solution of the generalized wave equation $[D^2_t u](t) + \mathbb{B}(u(t)) = 0$*

with initial conditions $u(0) = v_\vartheta^1$ *and* $[D_t u](0) = v_\vartheta^2$ *and let* $t \mapsto w_{\vartheta,t}^2$ *be the derivative with respect to* t *of* $t \mapsto w_{\vartheta,t}^1$. *Let* $\varepsilon_{\vartheta,t}$ *be positive real numbers such that for any* $\vartheta \in \Theta$ *we have* $\lim_{t\to 0} \varepsilon_{\vartheta,t} = \varepsilon \geq 0$. *Then for the operator* $\mathbb{A}$ *occurring in the conclusion of Theorem 7.2* $\|\mathbb{A} - \mathbb{B}\| \leq \varepsilon$ *holds. Especially, if* $\varepsilon = 0$, *then the operator* $\mathbb{A}$ *occurring in the conclusion of Theorem 7.2 is uniquely determined and equals* $\mathbb{B}$.

Proof: Since $t \mapsto w_{\vartheta,t}^1$ is the solution of the wave equation and $t \mapsto w_{\vartheta,t}^2$ is its first derivative we obtain

$$\begin{pmatrix} w_{\vartheta,0}^1 \\ w_{\vartheta,0}^2 \end{pmatrix} = \begin{pmatrix} v_\vartheta^1 \\ v_\vartheta^2 \end{pmatrix} \quad \text{and} \quad D_t \left(t \mapsto \begin{pmatrix} w_{\vartheta,t}^1 \\ w_{\vartheta,t}^2 \end{pmatrix} \right) \Bigg|_{t=0} = \begin{pmatrix} v_\vartheta^2 \\ \mathbb{B}(v_\vartheta^1) \end{pmatrix}$$

Let us denote by $\mathbb{A}$ the operator occuring in the conclusion of Theorem 7.2. From (i) in Theorem 7.2 and the equation above we obtain that for all $\vartheta \in \Theta$

$$0 \leq \lim_{t\to 0} \left\| \begin{pmatrix} v_\vartheta^2 \\ \mathbb{A}(v_\vartheta^1) \end{pmatrix} - \begin{pmatrix} v_\vartheta^2 \\ \mathbb{B}(v_\vartheta^1) \end{pmatrix} \right\| \leq \lim_{t\to 0} \varepsilon_{\vartheta,t} = \varepsilon$$

holds. Thus $\|\mathbb{B}(v_\vartheta^1) - \mathbb{A}(v_\vartheta^1)\| \leq \varepsilon$ for all $\vartheta \in \Theta$ and since one of our hypothesis is that $\{v_\vartheta^1 \mid \vartheta \in \Theta\}$ is dense in the unit ball of H_E we obtain $\|\mathbb{A} - \mathbb{B}\| \leq \varepsilon$. In the case $\varepsilon = 0$ we obtain $\mathbb{A} = \mathbb{B}$

Discussion: To say it in a different way. If we are given for a sufficiently rich class of initial conditions the solution curves of the generalized wave equation and if we are given for any finite subset of $T \times \Theta$ an operator $\mathbb{A} \in \aleph$ such that the affinity $t \mapsto \begin{pmatrix} v_\vartheta^1 + t \cdot v_\vartheta^2 \\ v_\vartheta^2 - t \cdot \mathbb{A}(v_\vartheta^1) \end{pmatrix}$ approximates the solution curve $t \mapsto \begin{pmatrix} w_{\vartheta,t}^1 \\ w_{\vartheta,t}^2 \end{pmatrix}$ up to $\varepsilon_{\vartheta,t}$ then there exists an operator $\mathbb{B}$ such that the affinity $t \mapsto \begin{pmatrix} v_\vartheta^1 + t \cdot v_\vartheta^2 \\ v_\vartheta^2 - t \cdot \mathbb{B}(v_\vartheta^1) \end{pmatrix}$ approximates the solution for all $\varepsilon_{\vartheta,t}$ with $\vartheta \in \Theta$ and $t \in T$. Further $\mathbb{B}$ is the operator which occurs in the generalized wave equation. This holds especially true, if $\aleph$ consists of the operators $\sum_{i=1}^n \alpha_i(x) \cdot \frac{\partial^2}{\partial^2 x_i}$ with $\alpha_i : \overline{U} \mapsto [0, \infty)$ strictly positive functions defined on the closure of a bounded open set $U \subset \mathbb{R}^n$ and $\{x \mapsto (\alpha_1(x), \ldots, \alpha_n(x)) \mid \sum_{i=1}^n \alpha_i(x) \cdot \frac{\partial^2}{\partial^2 x_i} \in \aleph\}$ convex and compact with respect to the topology of uniform convergence. Let $H = \mathbb{L}^2(U)$ and $H_E = \overset{\circ}{W}{}_2^1$ be the Sobolev space defined in [14] 2.5.4. Then H_E is the energetic space of the Laplace operator (see [14] 5.3). In this case $\|.\|_{H_E \times H}$ is the energy the wave at time t would posses with respect to $\triangle$ (see [14] 5.11). The true energy of the wave (with respect to the true operator $\mathbb{B} = \sum_{i=1}^n \beta_i \frac{\partial^2}{\partial^2 x_i} \in \aleph$) is given by a norm which is equivalent to $\|.\|_{H_E \times H}$. Thus $\|v - w\|_{H_E \times H}$ is a natural distance for waves v, w. It is however an $\mathbb{L}^2$-distance and thus there is a need for our extension of the black-box criterion to a non-L-space (non-L^1) context.

Finally we give the following interpretation of the black-box criterion. Suppose that one wants to establish that a certain type of linear differential equation (in our example the wave equation with the differential operator varying in a certain compact convex set) describes a dynamical system correctly up to an error ε. Then it suffices to check that for any finite set of pairs of initial conditions and empirical measurements (revealing only part of the information of the state of the dynamical system), there exists a differential equation of the specified type (with differential operator in a certain compact convex set) such that the empirical measurement is sufficiently close to the predicted value of the measurement. Philosophically speaking this is actually what one does to establish laws of nature up to the precision of the measurement.

8 The finite dimensional case and Helly's Theorem

We consider now the case that the spaces $L(E)$ and $L(F)$ or two Hilbert spaces or more general two arbitrary vector spaces are finite dimensional. This makes it possible to prove a theorem which seems closely related to the black-box criterion, but which involves in its hypotheses only a constant finite number (depending on the dimension of the vector spaces involved) of constraints.

Theorem 8.1 *Let $\mathcal{L}$ be a compact convex family of linear mappings $l : \mathbb{R}^m \to \mathbb{R}^n$, let Θ be an index set, let $(Z_\vartheta)_{\vartheta\in\Theta}$ be a family of convex subsets of $\mathbb{R}^n$ and let $(\varepsilon_\vartheta)_{\vartheta\in\Theta}$ be a family of real numbers. Let further $(x_\vartheta)_{\vartheta\in\Theta}$ with $x_\vartheta \in \mathbb{R}^m$ and $(y_\vartheta)_{\vartheta\in\Theta}$ with $y_\vartheta \in \mathbb{R}^n$ be given. Let I be an index set consisting of $(m \cdot n) + 1$ points. Suppose that for any indexed family $(\vartheta_i)_{i\in I}$ with $\vartheta_i \in \Theta$ and any indexed family $(z_{\vartheta_i})_{i\in I}$ with $z_{\vartheta_i} \in Z_{\vartheta_i}$ there exists an $l \in \mathcal{L}$ such that*

$$\langle\, l(x_{\vartheta_i}) - y_{\vartheta_i}\ ,\ z_{\vartheta_i}\,\rangle \le \varepsilon_{\vartheta_i}.$$

Then there exists an $l \in \mathcal{L}$ such that for all $\vartheta \in \Theta$ we have

$$\sup_{z\in Z_\vartheta} \langle\, l(x_\vartheta) - y_\vartheta\ ,\ z\,\rangle \le \varepsilon_\vartheta.$$

Proof of Theorem 8.1 The sets

$$\mathcal{L}_{\vartheta,z} := \{\ l \in \mathcal{L}\ \text{ such that }\ \langle\, l(x_\vartheta) - y_\vartheta\ ,\ z\,\rangle \le \varepsilon_\vartheta\ \}$$

with $z \in Z_\vartheta$ are compact convex subsets of $\mathcal{L}$ and thus subsets of the $m \cdot n$-dimensional space of all linear mappings from $\mathbb{R}^m$ to $\mathbb{R}^n$. Our Hypothesis says that the intersection of any $(m \cdot n) + 1$ sets $\mathcal{L}_{\vartheta,z}$ wit $z \in Z_\vartheta$ is nonempty. Therefore by Helly's Theorem (see [13] Part VI) the intersection $\bigcap_{\vartheta\in\Theta,z\in Z_\vartheta} \mathcal{L}_{\vartheta,z}$ is nonempty which is precisely the conclusion of our Theorem. $\square$

Appendix

A Compactness of the spaces $ba(\Omega_2, \mathcal{B})$ and $\mathcal{S}$

Let $\mathcal{M}_b$ denote the vector space of $\mathcal{B}$-measurable real valued functions on $(\Omega_2, \mathcal{B})$. In the following Lemmata we denote by σ' the topology on $ba(\Omega_2, \mathcal{B})$ induced by the integrals $I_m : (\Omega_2, \mathcal{B}) \to \mathbb{R}$ defined by $I_m := \int m \, d\mu$. By $E := (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta \in \Theta})$ we denote a statistical experiment.

Lemma A.1 *The set $\{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 1\}$ is σ'-compact.*

Sketch of proof: Let $\mathcal{G} := \{m \in \mathcal{M}_b \mid -1 \leq m(\omega) \leq 1 \text{ for all } \omega \in \Omega_2\}$ and let $I : (\Omega_2, \mathcal{B}) \to \mathbb{R}^G$ be the unique mapping such that $I_m = pr_m \circ I$ for all $m \in \mathcal{G}$. The topology induced by I equals the topology σ' induced by the family of integrals $\{I_m \mid m \in G\}$. Thus $(\{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 1\}, \sigma')$ is homeomorphic with the set $I(\{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 1\}) \subseteq [-1, +1]^G$. Since $[-1, +1]^G$ is by the Tychonoff product theorem compact, it suffices to prove that $I(\{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 1\})$ is closed in $[-1, +1]^G$. But this is clear since

$$
\begin{aligned}
& I(\{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 1\}) = \\
= \{f \in [-1, +1]^G \quad \mid \quad & -1 \leq pr_{m_1+m_2}(f) = pr_{m_1}(f) + pr_{m_2}(f) \leq 1 \\
& \forall\, m_1, m_2 \in \mathcal{M}_b \text{ with } -1 \leq m_1, m_2, m_1 + m_2 \leq 1\}. \ \Box
\end{aligned}
$$

Lemma A.2 *The space $\mathcal{S}$ of stochastic operators $S : L(E) \to ba(\Omega_2, \mathcal{B})$ is a compact subset of $(ba(\Omega_2, \mathcal{B}))^{L(E)}$.*

Sketch of proof: We have

$$
\mathcal{S} \subseteq \prod_{\mu \in L(E)} \{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 2\|\mu\|\}
$$

and we know by the preceding Lemma and the Tychonoff product theorem that $\prod_{\mu \in L(E)} \{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 2\|\mu\|\}$ is compact. It therefore suffices to show that S is closed in $\prod_{\mu \in L(E)} \{\nu \in ba(\Omega_2, \mathcal{B}) \mid \|\nu\| \leq 2\|\mu\|\}$. But this is clear since the space of stochastic operators can be described by

$$
\begin{aligned}
\mathcal{S} := \{S \quad \mid \quad & S \in ba(\Omega_2, \mathcal{B})^{L(E)}, S \text{ is linear and} \\
& [\mu \in L^+(E) \wedge \|\mu\| = 1] \Longrightarrow \\
& [\forall m \in \mathcal{M}_b (m \geq 0 \Rightarrow I_m(S(\mu)) \geq 0) \text{ and } I_1(S(\mu)) = 1]\}. \ \Box
\end{aligned}
$$

B Some generalized black-box criteria

We can use the fact that the sets G_ξ of Theorem 4.3 can vary with ξ and can be different from $\{m \in \mathcal{M}_b \mid -1 \leq m \leq +1\}$ to formulate a generalization of the black-box criterion (in the sense of Theorem 2.2) as follows:

Theorem B.1 *Let*

$$\begin{aligned} E &:= (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta}) \quad \text{and} \\ F &:= (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta}) \end{aligned}$$

be adapted statistical experiments indexed by the same index sets Θ and T. Let $(\varepsilon_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ be a family of nonnegative real numbers and let $(f_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ be a family of $\mathcal{B}$-measurable functions $f_{\vartheta,t} : \Omega_2 \to I\!R$. Suppose that for any finite set $\Theta_0 \times T_0 \subseteq \Theta \times T$ and any selection $(g_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0}$ of functions $g_{\vartheta,t} \in \{g \in M_b(\Omega_2, \mathcal{B}_t) \mid -f_{\vartheta,t} \leq g \leq f_{\vartheta,t}\}$ there exists an $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operator $S : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ such that

$$\int g_{\vartheta,t}\, dS(P_\vartheta) \leq \int g_{\vartheta,t}\, dQ_\vartheta + \varepsilon_{\vartheta,t} \quad \text{for all} \quad \vartheta \in \Theta_0 \quad \text{and} \quad t \in T_0\ .$$

Then there exists an $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operator $\widetilde{S} : L(E) \mapsto ba(\Omega_2, \mathcal{B})$ such that

$$\sup_{\substack{g\in M_b(\Omega_2,\mathcal{B}_t)\\ -f_{\vartheta,t}\leq g\leq f_{\vartheta,t}}} \left(\int g\, d[\widetilde{S}(P_\vartheta) - Q_\vartheta]\right) \leq \varepsilon_{\vartheta,t}$$

for all $\vartheta \in \Theta$ and $t \in T$.

Remark B.1 Note that even the case $T := \{t_0\}$ and $f_{\vartheta,t_0} = x^2$ for all ϑ was not covered by Theorem 2.2.

More general we formulate:

Theorem B.2 *Let*

$$\begin{aligned} E &:= (\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\in T}, (P_\vartheta)_{\vartheta\in\Theta}) \quad \text{and} \\ F &:= (\Omega_2, \mathcal{B}, (\mathcal{B}_t)_{t\in T}, (Q_\vartheta)_{\vartheta\in\Theta}) \end{aligned}$$

be adapted statistical experiments indexed by the same index sets Θ and T, let $(\varepsilon_{\vartheta,t})_{\vartheta\in\Theta,t\in T}$ be a family of nonnegative real numbers and let $(G_{\vartheta,t})_{(\vartheta,t)\in\Theta\times T}$ be a family of closed convex sets of bounded $\mathcal{B}$-measurable functions, such that $0 \in G_{\vartheta,t} \subseteq M_b(\Omega_2, \mathcal{B}_t)$. Let $\mathcal{L}$ be a compact convex set of $(\mathcal{A}_t)_{t\in T}/(\mathcal{B}_t)_{t\in T}$-adapted stochastic operator $S : L(E) \mapsto ba(\Omega_2, \mathcal{B})$. Suppose that for any finite set $\Theta_0 \times T_0 \subseteq \Theta \times T$ and any selection $(g_{\vartheta,t})_{\vartheta\in\Theta_0,t\in T_0}$ of functions $g_{\vartheta,t} \in G_{\vartheta,t}$ there exists an $S \in \mathcal{L}$ such that

$$\int g_{\vartheta,t}\, dS(P_\vartheta) \leq \int g_{\vartheta,t}\, dQ_\vartheta + \varepsilon_{\vartheta,t} \quad \text{for all} \quad \vartheta \in \Theta_0 \quad \text{and} \quad t \in T_0\ .$$

Then there exists an $\widetilde{S} \in \mathcal{L}$ *such that*

$$\sup_{g \in G_{\vartheta,t}} (\int g \; d[\widetilde{S}(P_\vartheta) - Q_\vartheta]) \leq \varepsilon_{\vartheta,t}$$

for all $\vartheta \in \Theta$ *and* $t \in T$.

Remark B.2 Theorem B.1 is an immediate consequence of Theorem B.2 and Theorem B.2 can be proved in a completely analogous way as Theorem 5.1 has been proved. Even in the non adapted case (i.e. $T = \{t_0\}$) and even if we assume that the loss function space is independent of the parameter (i.e. $G_\vartheta = G$) Theorem B.2 still remains very interesting as is shown by the following example on stochastic orders:

Example B.1 Let F denote the family of all measurable monotone increasing functions $f : \mathbb{R} \to [0,1]$. Suppose that we are given two families $(P_\vartheta)_{\vartheta \in \Theta}$ and $(Q_\vartheta)_{\vartheta \in \Theta}$ of probability measures on $(\mathbb{R}, \mathcal{B})$ (with $\mathcal{B}$ the Borel σ-algebra on $\mathbb{R}$). Suppose further that for any finite set $\Theta_0 \subset \Theta$ and any family $(f_\vartheta)_{\vartheta \in \Theta_0}$ of functions $f_\vartheta \in F$ there exists a stochastic operator M such that for all $\vartheta \in \Theta_0$

$$\int f_\vartheta \; d[M(P_\vartheta) - Q_\vartheta] \leq 0$$

and $\forall \vartheta \in \Theta$

$$\int \arctan(x) \; d[M(P_\vartheta)](x) = \int arctan(x) \; dP_\vartheta(x)$$

Then there exists a stochastic operator $\tilde{M}$ such that

$$\sup_{\vartheta \in \Theta} \sup_{f \in F} \int f \; d[\tilde{M}(P_\vartheta) - Q_\vartheta] \leq 0;$$

and $\forall \vartheta \in \Theta$

$$\int \arctan(x) \; d[\tilde{M}(P_\vartheta)](x) = \int arctan(x) \; dP_\vartheta(x)$$

i.e. there exists a stochastic operator $\tilde{M}$ such that $\tilde{M}(P_\vartheta) \overset{(1)}{\leq} Q_\vartheta$ for all $\vartheta \in \Theta$, with $\overset{(1)}{\leq}$ the stochastic order defined in Section 1.2 of [9] and $\tilde{M}$ preserves for all θ the integral over the arctan function.

To see that the example holds simply apply Theorem B.2 in the case that there is no filtration (i.e. $T = \{t_0\}$), that $G_{\vartheta,t_0} = F$ for all $\vartheta \in \Theta$, that $\mathcal{L}$ is the set of all stochastic operators M with

$$\int \arctan(x) \; d[M(P_\vartheta)](x) = \int arctan(x) \; dP_\vartheta(x) \; \forall \vartheta \in \Theta$$

and note that F induces the relation $\overset{(1)}{\leq}$ in the sense of [9]; i.e. $\mu_1 \overset{(1)}{\leq} \mu_2 \Leftrightarrow \int f d\mu_1 \leq \int f d\mu_2$ for all $f \in F$.

C Reversing the role of stochastic operators and experiments

It is possible to reverse the role of the space of stochastic operators and the experiment $E := (\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta\in\Theta})$ in Theorem 2.2. To be more precise we state and prove the following Theorem:

Theorem C.1 *Let Θ be an index set. Let $\Upsilon \subset ba(\Omega, \mathcal{A})$ be compact and convex. Let $(f_\vartheta)_{\vartheta\in\Theta}$ be a family of functions from $L(\Omega, \mathcal{A}, \Upsilon)$ to $ba(\Omega_2, \mathcal{B})$ and let $(Q_\vartheta)_{\vartheta\in\Theta}$ be a family of elements of $ba(\Omega_2, \mathcal{B})$. Denote by $\mathcal{M}_b(\Omega_2, \mathcal{B})$ the space of bounded measurable real valued functions endowed with the supremum norm.*

Suppose that for any finite set $\Theta_0 \subseteq \Theta$ and any family $(g_\vartheta)_{\vartheta\in\Theta}$ of functions $g_\vartheta \in \mathcal{M}_b(\Omega_2, \mathcal{B})$ with $\sup_{x\in\Omega_2} |g_\vartheta| \le 1$ there exists a $P \in \Upsilon$ such that

$$\int g_\vartheta \; d[f_\vartheta(P)] \le \int g_\vartheta \; dQ_\vartheta + \varepsilon_\vartheta \; .$$

Then there exists a $P \in \Upsilon$ such that $\forall \vartheta \in \Theta$

$$\|f_\vartheta(P) - Q_\vartheta\| \le \varepsilon_\vartheta.$$

Remark C.1 The functions $f_\vartheta : L(\Omega, \mathcal{A}, \Upsilon) \mapsto ba(\Omega_2, \mathcal{B})$ can be arbitrary. They can of course be of the form $f_\vartheta = S_\vartheta$ or $f_\vartheta = S_\vartheta - id$, with id the identity on $L(\Omega, \mathcal{A})$ and $(S_\vartheta)_{\vartheta\in\Theta}$ a family of stochastic operators. If we let $f_\vartheta = S_\vartheta - id$ and $Q_\vartheta = 0$ for all $\vartheta \in \Theta$ then we obtain the following corollary:

Corollary C.1 *Let $\Upsilon \subset ba(\Omega, \mathcal{A})$ be compact and convex and let $(S_\vartheta)_{\vartheta\in\Theta}$ be a family of stochastic operators from $L(\Omega, \mathcal{A}, \Upsilon)$ to $ba(\Omega_2, \mathcal{B})$. Suppose that for any finite $\Theta_0 \subseteq \Theta$ and any family $(g_\vartheta)_{\vartheta\in\Theta_0}$ with $g_\vartheta \in \mathcal{M}_b(\Omega_2, \mathcal{B})$ and $\sup_{x\in\Omega_2} |g_\vartheta| \le 1$ there exists a $\mu \in \Upsilon$ with*

$$\int g_\vartheta \; d[S_\vartheta(\mu) - \mu] \le \varepsilon_\vartheta \; .$$

Then there exists a μ in Υ such that for all $\vartheta \in \Theta$ we have

$$\|S_\vartheta(\mu) - \mu\| \le \varepsilon_\vartheta \; .$$

Proof of Theorem C.1 To obtain Theorem C.1 simply apply Theorem 4.1 with $v_\vartheta = f_\vartheta$, $w_\vartheta = Q_\vartheta$, $\mathcal{L} = \Upsilon$, $G := \{g \in \mathcal{M}_b(\Omega_2, \mathcal{B}) \mid |g| \le 1\}$ in a completely analogous way as in the derivation of Theorem 2.2 from Theorem 4.1.

Acknowledgement: I would like to thank E. Makai and O. Boxma for reading through a preliminary version of the paper. I would further like to thank E. Makai for advices concerning section 8. I would like to thank R. Gill, G. Pflug and A. van der Vaart for Discussions on various aspects of LeCam's randomization criterion.

References

[1] Le Cam, L. Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** (1964) 1419–1455. **34** #6909

[2] Le Cam, L. *Asymptotic Methods in Statistical Decision Theory* Springer Series in Statistics. Springer-Verlag 1986

[3] Le Cam, Lucien; Yang, Grace Lo *Asymptotics in statistics. Some basic concepts. Second edition.* Springer Series in Statistics. Springer-Verlag, New York, 2000. **2001f**:62019

[4] Dudley, R. M. *Real Analysis and Probability* Cambridge studies in advanced mathematics. Cambridge University Press

[5] E. Norberg Comparison of statistical experiments with filtered probability spaces. *Stat. Decis.* **20**, No.1, 1-27 (2002).

[6] Shiryaev A.N., Spokoiny V. G. *Statistical Experiments and Decisions : Asymptotic Theory* Advanced Series on Statistical Science and Applied Probability Volume **8** World Scientific 2001.

[7] Schaefer, H. H.; Wolff, M. P. *Topological vector spaces. Second edition.* Graduate Texts in Mathematics, **3**. Springer-Verlag, New York, 1999. **2000j**:46001

[8] Segal, Irving E.; Kunze, Ray A. *Integrals and operators. Second revised and enlarged edition.* Grundlehren der Mathematischen Wissenschaften, Band **228**. Springer-Verlag, Berlin-New York, 1978. **58** #6126

[9] Stoyan, Dietrich *Qualitative Eigenschaften und Abschätzungen stochastischer Modelle. (German)* Akademie-Verlag, Berlin, 1977. **56** #13397

[10] Stoyan, Dietrich *Comparison methods for queues and other stochastic models. Translation from the German edited by Daryl J. Daley.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1983. **85f**:60147

[11] Strasser, Helmut *Mathematical theory of statistics. Statistical experiments and asymptotic decision theory.* de Gruyter Studies in Mathematics, 7. Walter de Gruyter & Co., Berlin, 1985. **87h**:62034

[12] Torgersen, Erik *Comparison of statistical experiments.* Encyclopedia of Mathematics and its Applications, **36**. Cambridge University Press, Cambridge, 1991. **92i**:62007

[13] Valentine, Frederick A. *Convex sets.* McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Co., New York-Toronto-London 1964 **30** #503

[14] Zeidler, Eberhart *Applied Functional Analysis. Applications to Mathematical Physics* Applied Mathematical Sciences **108**. Springer Verlag