On the convolution theorem for infinite-dimensional parameter spaces

K. Pötzelberger, W. Schachermayer and H.Strasser

Revised version, 20.9.2000

Dedicated to the memory of Lucien LeCam

Abstract

In this paper we give examples which show that the convolution theorem (Boll, [1], Hajek, [2]) cannot be extended to infinite-dimensional shift experiments. This answers a question posed by van der Vaart, [9], and which has been considered also by LeCam, [5].

1 Introduction

Let H be a finite dimensional vector space. Assume that P is a probability measure on the Borel-σ-field $\mathcal{B}$ of H, which is absolutely continuous with respect to the Lebesgue-measure, and that $f : H \to \mathbb{R}$ is a linear function.

Let $T : H \to \mathbb{R}$ be a measurable function. The image of P under T is denoted by $\mathcal{L}(T|P)$. The symbol δ_h denotes the point measure at $h \in H$. The measurable function $T : H \to \mathbb{R}$ is called an equivariant estimator of the function f, if

$$\mathcal{L}(T - f(h)|P * \delta_h) = \mathcal{L}(T|P) \quad \text{for } h \in H.$$

The assertion of the convolution theorem states that under some regularity conditions there exists a probability measure μ on $\mathcal{B}$ such that

$$\mathcal{L}(T|P) = \mathcal{L}(f|P) * \mu.$$

[1] *AMS 1991 subject classifications.* Primary: 62A05, secondary: 28C20.

[2] *Key words and phrases:* convolution theorem, Hilbert space, cylindrical measures.

In this form the convolution theorem has been proved by Boll, [1]. For historical remarks on the convolution theorem we refer to LeCam, [5]. The convolution theorem can be extended in a more or less straightforward manner to locally compact groups. We refer to the exposition of the convolution theorem in Strasser, [7], sections 38 and 39.

The family of probability measures $P_h = P * \delta_h$, $h \in H$ is a so-called shift experiment. In his paper, [5], LeCam posed the question whether the assertion of the convolution theorem remains true for shift experiments if the parameter space H is an infinite dimensional vector space. This question is important for the application of the convolution theorem to statistical problems where the data are paths of stochastic processes.

In the thesis by Moussatat, [6], a version of the convolution theorem is proved where H is a Hilbert space and $(P_h)_{h \in H}$ is a standard Gaussian shift experiment. Related versions of the convolution theorem can be found in Strasser, [7], Theorem 72.15, and in van der Vaart, [9].

The question, whether an infinite dimensional version of the convolution theorem remains true if the family $(P_h)_{h \in H}$ is not a standard Gaussian shift experiment, is regarded as unsettled by van der Vaart, [9]. The same problem is the subject of LeCam, [5]. In view of that paper and in view of the remarks made by LeCam in Yang, [10], pp. 236-237, this seems to have been an open problem.

In the present paper we give examples which show that the infinite-dimensional version of the convolution theorem does not hold true, in general.

For expository reasons we will present our examples in two different but equivalent ways corresponding to different levels of mathematical abstraction. We consider this procedure as justified since the past literature on the convolution theorem is also written on very different levels of mathematical sophistication. Moreover, our great respect for and our admiration of L. LeCam encourages us to present our results also in that abstract mathematical framework which was used by LeCam.

In section 2 we present versions of the counterexamples in terms of classical shift experiments which are defined via embedding the parameter space into the sample space by a continuous operation. However, this mathematical representation of shift experiments was not that used by LeCam. LeCam liked to represent shift experiments as families of cylindrical measures on the parameter space. Therefore, section 3 contains the explicit construction of the counterexamples to the problem posed by LeCam, [5], using the framework of functional analysis and cylindrical measures which was used LeCam himself (see LeCam, [4] and [5]).

2 The case of classical experiments

In this section we will consider only classical experiments, i.e. experiments defined as a family of probability measures on a sigma-field. The more general situation of experiments defined by cylindrical measures will be treated in section 3.

2.1 Basic concepts

Let H be a Hilbert space and let Ω be a Banach space. Suppose that the Hilbert space H is the parameter space of a statistical experiment $E = (P_h)_{h \in H}$ with sample space $(\Omega, \mathcal{A})$, where $\mathcal{A}$ denotes the Borel-σ-field of Ω. A so-called shift experiment deals with the situation where the parameter space H is embedded into the sample space by an injection $\tau : H \to \Omega$ and operates on Ω by the shift operation $x \mapsto x + \tau(h)$. If we define the probability measures P_h by applying the shift operation $x \mapsto x + \tau(h)$ to P_0, i.e. if

$$P_h := P_0 * \delta_{\tau(h)}, \ h \in H,$$

then the experiment E is called the shift experiment generated by P_0 and (H, τ).

Let us illustrate the concept of shift experiments by some familiar examples.

(2.1) EXAMPLE Let $\Omega = \mathbb{R}^n$ and $H = \mathbb{R}$. Define the embedding of H into Ω by

$$\tau : h \mapsto (h, h, \ldots, h)', \quad h \in H = \mathbb{R}.$$

The arising shift experiment is then a so-called univariate location family.

Let $\Omega = H = \mathbb{R}^n$. In this case the parameter space H can be embedded into the sample space in a most simple way by the identity $\tau : h \mapsto h$. The arising shift experiment is then a so-called multivariate location family.

But for $\Omega = H = \mathbb{R}^n$ there are also other embeddings feasible. In order to prepare the ground for later examples with infinite dimensional parameter spaces let us consider the embedding

$$\tau : (h_1, h_2, \ldots, h_n) \mapsto (h_1, h_1 + h_2, \ldots, h_1 + \cdots + h_n),$$

where $h = (h_1, h_2, \ldots, h_n) \in H = \mathbb{R}^n$. This is a parametrization of a multivariate location family by the first differences of the location parameter. $\Box$

If the embedding $\tau : H \to \Omega$ is surjective, then the shift experiment is a full shift experiment. In such a case the embedding is a bijection.

Let us consider an example with an infinite dimensional parameter space.

(2.2) EXAMPLE Let $H = L^2([0,1])$ and $\Omega = \mathcal{C}_0([0,1]) = \{x \in \mathcal{C}([0,1]) : x(0) = 0\}$. and define

$$\tau(h) := \Big(\int_0^t h(s)\, ds \Big)_{0 \le t \le 1}. \tag{1}$$

Then $\tau : H \to \Omega$ is a continuous embedding, and $\tau(H)$ is dense in Ω. By X_t we denote the coordinate function $x \mapsto x(t)$, $x \in \mathcal{C}([0,1])$.

Let P_0 be a Borel measure on $(\Omega, \mathcal{A})$ such that (X_t) is the standard Wiener process. Then the family of probability measures $P_h := P_0 * \delta_{\tau(h)}$, $h \in H$, is a shift experiment. For every $h \in H$ the process

$$X_t - \int_0^t h(s)\, ds$$

is the standard Wiener process under P_h. The probability measures P_h are mutually absolutely continuous and we have Cameron-Martin-Girsanov formula

$$\frac{dP_h}{dP_0} = \exp\Big(\int_0^1 h(t)\, dX_t - \frac{1}{2} \int_0^1 h^2(t)\, dt \Big).$$

Any shift experiment having this likelihood structure is called a standard Gaussian shift. □

The convolution theorem deals with the structure of equivariant estimators for shift experiments. For simplicity we consider only the case of estimating a continuous linear function $f : H \to \mathbb{R}$, the case of a continuous linear function f from H to a general topological vector space being similar.

(2.3) DEFINITION *Let $(P_h)_{h \in H}$ be a shift experiment.*

A measurable function $T : \Omega \to \mathbb{R}$ is called an equivariant estimator (of the linear function f), if

$$\mathcal{L}(T - f(h) | P_h) = \mathcal{L}(T | P_0), \quad \text{for all } h \in H.$$

A Markov kernel $\rho : \Omega \times \mathcal{B} \to \mathbb{R}$ is called a (randomized) equivariant estimator (for the function f), if

$$\int \rho(., B + f(h))\, dP_h = \int \rho(., B)\, dP_0, \quad \text{for all } h \in H \text{ and } B \in \mathcal{B},$$

where $\mathcal{B}$ denotes the Borel field of $\mathbb{R}$.

For an arbitrary Markov kernel ρ we define the quadratic risk under P_h, $h \in H$, to be

$$\int \int (\xi - f(h))^2\, \rho(., d\xi)\, dP_h.$$

The quadratic risk will play an important role in our counterexamples.

If we add a constant to an equivariant estimator then the resulting estimator is again equivariant. If an equivariant estimator has a finite first moment then by adding a uniquely determined constant we may obtain an unbiased equivariant estimator.

The equivariant estimators constitute a large class of estimators. A considerably smaller and simpler class are the so-called strictly equivariant estimators.

(2.4) DEFINITION *Let $(P_h)_{h\in H}$ be a shift experiment.*

A measurable function $T : \Omega \to \mathbb{R}$ is called a strictly equivariant estimator (for the function f), if

$$T(x+\tau(h)) = T(x) + f(h), \quad \text{for all } x \in \Omega \text{ and } h \in H. \tag{2}$$

A Markov kernel $\rho : \Omega \times \mathcal{B} \to \mathbb{R}$ is called a (randomized) strictly equivariant estimator (for the function f), if

$$\rho(x+\tau(h), B+f(h)) = \rho(x, B) \text{ for all } x \in \Omega,\ h \in H \text{ and } B \in \mathcal{B}.$$

Given a full shift experiment there is (up to an additive constant) only one estimator which is strictly equivariant and non-randomized. Indeed, if in equation (2) we put $h = -\tau^{-1}(x)$, then we obtain

$$T(x) = T(0) + (f \circ \tau^{-1})(x). \tag{3}$$

Let us call these estimators canonical estimators. All canonical estimators are strictly equiviriant and non-randomized. The estimator $T_0 := f \circ \tau^{-1}$ is a particularly simple canonical estimator.

If the canonical estimators have a finite first moment then there is a uniquely determined unbiased canonical estimator. Moreover, if the second moment is finite, then this unbiased canonical estimator minimizes the quadratic risk among all strictly equivariant estimators.

Let us illustrate canonical estimators by examples.

(2.5) EXAMPLE In case of $\Omega = H = \mathbb{R}^n$ and $\tau = \mathrm{Id}_{\mathbb{R}^n}$ we have $T_0(x) = f(x)$. Thus, in this case the linear function f is a canonical estimator.

Let $\Omega = H = \mathbb{R}^n$ and

$$\tau : (h_1, h_2, \ldots, h_n) \mapsto (h_1, h_1+h_2, \ldots, h_1+\cdots+h_n),$$

where $h = (h_1, h_2, \ldots, h_n) \in H = \mathbb{R}^n$. Then we have

$$\tau^{-1}(x) = (x_1, x_2 - x_1, \ldots, x_n - x_{n-1}).$$

If we want to estimate the special linear function

$$f(h) := \sum_{i=1}^{n} h_i$$

then we obtain $T_0(x) = x_n$ as the canonical unbiased equivariant estimator. □

(2.6) EXAMPLE In the case of an infinite dimensional Gaussian shift like Example ((2.2)) we may define a canonical estimator in a similar way. The case is slightly more subtle than for a full shift experiment since the embedding τ is not surjective and therefore we obtain an equation like (3) only on a dense subset of Ω. But if we impose continuity of the estimator then we may extend (3) to the whole sample space and arrive at a similar situation as in full shift case.

As an illustration let us consider the estimation of the linear function

$$f(h) := \int_0^1 h(s)\, ds,\ h \in H$$

in the model of Example (2.2). Recall the definition of τ in (1). For all $x \in \tau(H)$ we have $\tau^{-1}(x) = x'$ which implies

$$T(x) = T(0) + \int_0^1 x'(s)\, ds = T(0) + x(1).$$

If T is assumed to be continuous, then this equation must hold on Ω. In this particular sense the function $T_0 = X_1$ is the canonical unbiased estimator of f. □

2.2 The convolution property

We are going to introduce some definitions which will simplify the discussion of the convolution theorem and related questions.

(2.7) DEFINITION *Two equivariant estimators are called equivalent to each other, if their distributions (under P_0 and then also under each P_h, $h \in H$) differ by a one-point measure which is independent of the parameter $h \in H$.*

(2.8) DEFINITION *Let us say that an estimator ρ_0 has the convolution property, if for any further equivariant estimator ρ there exists a probability measure μ on $(\mathbb{R}, \mathcal{B})$, such that the distribution of ρ is the convolution of the distribution of ρ_0 with μ.*

It is obvious that the convolution property is valid or not valid for all members of an equivalence class of equivariant estimators simultaneously.

There are several decision theoretical consequences of the convolution property. The most simple type of such an assertion is the following trivial lemma.

(2.9) LEMMA *Suppose that T is equivariant and has the convolution property. If T is in $L^2(P_0)$ and unbiased, then it has minimal quadratic risk among all equivariant estimators.*

The mathematical theorem which is known as convolution theorem states that under certain assumptions canonical estimators have the convolution property. If the canonical estimators are also in $L^2(P_0)$ then it follows from the convolution theorem that the unbiased canonical estimator has minimal quadratic risk among all equivariant estimators.

In this paper we will present some examples which show that, in general, canonical estimators need not have the convolution property. These examples are infinite-dimensional shift experiments.

It is a natural question to ask whether at least the decision theoretic optimality (with respect to the quadratic risk) of the unbiased canonical estimator remains valid. Therefore we will present two different counterexamples. In the first counterexample the unbiased canonical estimator does not have the convolution property but still has minimal quadratic risk among all equivariant estimators. In the second example a non-canonical estimator has minimal quadratic risk among all equivariant estimators, but the convolution property does neither hold for this optimal estimator nor for the canonical estimator.

2.3 The structure of the proof of the convolution theorem

It is illuminating to study the structure of the proof of the convolution theorem in the finite-dimensional case. The proof of the convolution theorem is based on two basic facts.

The first basic fact is concerned with the structure of strictly equivariant estimators.

Previously, we have seen that for full shift experiments non-randomized strictly equivariant estimators have a very simple structure. Below, we will see that also for randomized strictly equivariant estimators a similar simple structure can be established. To be explicit, it can be shown that for full shift experiments randomized but strictly equivariant estimators are convolutions with canonical estimators. This fact is the algebraic background of the convolution theorem.

However, this algebraic structure can be established by more or less direct computation only for strictly equivariant estimators. In order to cover arbitrary equivariant estimators, too, we have to consider a second basic fact. It is concerned with the problem of replacing equivariant estimators by strictly equivariant estimators.

The following theorem contains a version of this second basic fact in a form which is needed subsequently.

(2.10) THEOREM *Let $(P_h)_{h\in H}$ be a dominated shift experiment as defined in section 2.1 and let ρ be an equivariant estimator.*

(1) There exists an estimator ρ^* such that

$$\int \rho(x,B)\, P_h(dx) = \int \rho^*(x,B)\, P_h(dx) \quad \text{for all } B \in \mathcal{B} \text{ and } h \in H$$

and such that

$$\rho^*(.,B) = \rho^*(.+\tau(h), B + f(h)) \quad P_0\text{-a.e.},$$

where the exceptional set depends on $h \in H$ but not on $B \in \mathcal{B}$.

(2) For every finite-dimensional subspace $L \subseteq H$ the estimator ρ^* can be chosen in such a way it is even strictly equivariant for $h \in L$.

For completeness we give a proof of Theorem (2.10) in the appendix.

The counterpart of Theorem (2.10) in testing theory is the Theorem of Hunt and Stein. The Theorem of Hunt and Stein deals with testing problems which are invariant under the operation of a transformation group on the sample space. If this transformation group has an invariant probability measure then the proof of the theorem is easy. In general, the proof is based on the Markov-Kakutani fixpoint theorem. This method of proving the Theorem of Hunt and Stein is due to LeCam and it has been used by LeCam, [3], to prove a general version of the convolution theorem.

Next, we show how the finite-dimensional convolution theorem for full shift experiments can be proved along the lines described above.

(2.11) COROLLARY *Assume that H is of finite dimension and let $E = (P_h)_{h\in H}$ be a dominated full shift experiment. Then the canonical estimators have the convolution property.*

Proof: Let ρ be an arbitrary equivariant estimator of f. By Theorem (2.10), part 2, we may assume that ρ is even strictly equivariant. Defining the canonical estimator by $T_0 := f \circ \tau^{-1}$, we obtain

$$\rho(x,B) = \rho(x - \tau(\tau^{-1}(x)), B - f(\tau^{-1}(x)) = \rho(0, B - T_0).$$

Denoting $\mu(B) := \rho(0,B)$ it follows that

$$\int \rho(x,B)\, P_h(dx) = \Big(\mu * \mathcal{L}(T_0|P_h)\Big)(B).$$

$\square$

It should be noted that an infinite-dimensional counterpart of this result makes no sense since dominated full shift experiments do not exist for infinite dimensional Hilbert spaces. The only sensible question can be whether the infinite-dimensional convolution theorem holds for dominated shift experiments where $\tau(H)$ is dense in Ω. But for such cases the method of the preceding proof cannot be applied. Any strictly equivariant estimator ρ would be uniquely defined on $\tau(H)$, but it may happen - and in fact is typical - that $\tau(H)$ has P_0-measure zero. Then the distribution of ρ is completely undetermined. A continuity argument as in the definition of canonical estimators is not possible since even if ρ is continuous, the continuity cannot be maintained when Theorem (2.10) is applied.

2.4 Examples and counterexamples

We will present three examples showing that for infinite-dimensional shift experiments very different situations may appear. In particular, it will turn out that the convolution theorem is not valid, in general.

Let us make the following global assumptions:

Let $H = L^2([0,1])$ and $\Omega = \mathcal{C}_0([0,1])$. Define the embedding by (1). Assume that under P_0 the coordinate process (X_t) has continuous paths and let $P_h := P_0 * \delta_{\tau(h)}$, $h \in H$. We are going to consider the estimation problem for the linear function

$$f : h \mapsto \int_0^1 h(s)\, ds, \quad h \in H.$$

Our first example is concerned with a Gaussian shift situation of Example (2.6) where it is well-known that the convolution theorem is true. We state and prove the assertion for completeness.

(2.12) THEOREM *Suppose that the coordinate process (X_t) is the standard Wiener process under P_0. Then the canonical estimator $T_0 = X_1$ has the convolution property.*

Proof: Let ρ be an arbitrary equivariant estimator of f. By Theorem (2.10) we may assume that ρ is even strictly equivariant on the orthogonal complement of $\ker(f)$.

By strict equivariance of ρ it follows that

$$\rho(x, B) = \rho(x + x(1)\mathrm{Id}_{\mathbb{R}} - x(1)\mathrm{Id}_{\mathbb{R}}, B) = \rho(x - x(1)\mathrm{Id}_{\mathbb{R}}, B - x(1)),$$

since for $h \equiv x(1)$ the equations $f(h) = x(1)$ and $\tau(h) = x(1)\mathrm{Id}_{\mathbb{R}}$ are true. This implies

$$\begin{aligned}\int \rho(x, B)\, P_h(dx) &= \int \rho(x + \tau(h), B)\, P_0(dx) \\ &= \int \rho(x, B - f(h))\, P_0(dx) \\ &= \int \rho(x - x(1)\mathrm{Id}_{\mathbb{R}}, B - x(1) - f(h))\, P_0(dx)\end{aligned}$$

Since (X_t) is a Wiener process under P_0, the process $(X_t - X_1\mathrm{Id}_{\mathbb{R}})$ and the random variable X_1 are stochastically independent. If we define

$$R(B) := \int \rho(x - x(1)\mathrm{Id}_{\mathbb{R}}, B)\, P_0(dx),$$

then we obtain

$$\begin{aligned}\int \rho(x, B)\, P_h(dx) &= \int R(B - x(1) - f(h))\, P_0(dx) \\ &= \int R(B - x(1))\, P_h(dx).\end{aligned}$$

Hence, the distribution of $\rho\, P_h$ is for all $h \in H$ a convolution of the distribution of X_1 with R. $\square$

In this particular case the canonical estimator T_0 is also unbiased and has minimal quadratic risk among all equivariant estimators.

Next we turn to the first counterexample.

Assume that under P_0 the coordinate process (X_t) has the distribution of $(\sqrt{Z}W_t)$, where (W_t) is a Wiener process and Z is a nonnegative random variable which is stochastically independent of (W_t). In this case the quadratic variation process $(\langle X, X\rangle_t)$ of (X_t) comes into the game. The quadratic variation process has the following properties: The distribution of $(\langle X, X\rangle_t)$ under each P_h, $h \in H$, coincides with the distribution of (tZ), under P_0, and therefore does not depend on $h \in H$.

It follows that the function $T_1 := X_1 - \langle X, X\rangle_1$ is an equivariant estimator which is not equivalent to the canonical estimator $T_0 = X_1$. Our first counterexample will show that there are distributions of Z such that the distribution of T_1 cannot be written as a convolution with the distribution of T_0. Nevertheless the canonical estimator T_0 is unbiased and has minimal quadratic risk among all equivariant estimators.

(2.13) THEOREM *Assume that under P_0 the coordinate process (X_t) has the distribution of $(\sqrt{Z}W_t)$, where (W_t) is a Wiener process and Z is a nonnegative and P_0-integrable random variable which is stochastically independent of (W_t). Then the following assertions are true:*

1. The canonical estimator T_0 is unbiased and has minimal quadratic risk among all equivariant estimators. .

2. There are distributions of Z, such that T_0 does not have the convolution property.

Proof: Let us start with the proof of 1.

Let ρ by any equivariant estimator. By Theorem (2.10) we may assume that ρ is even strictly equivariant on the orthogonal complement of $\ker(f)$.

Let $P_0^z := P_0(.|\langle X, X\rangle_1 = z)$, $z \in \mathbb{R}$, be a regular version of the conditional probability and define $P_h^z := P_0^z * \delta_{\tau(h)}$, $h \in H$. Since the distribution of $(\langle X, X\rangle_t)$ under P_h does not depend on $h \in H$, it follows that $P_h^{\langle X,X\rangle_1} = P_h(.|\langle X, X\rangle_1)$ P_0-a.e., for all $h \in H$.

Under P_0^z the distribution of the coordinate process (X_t) is that of $(\sqrt{z}W_t)$, where (W_t) denotes a standard Wiener process. Hence it follows by the same arguments as in the proof of Theorem (2.12) that for each experiment $(P_h^z)_{h\in H}$, $z \geq 0$, the canonical estimator $T_0 = X_1$ has the convolution property and has thus smaller quadratic risk than ρ.

By integration with respect to the distribution of $\langle X, X\rangle_1$ the risk inequality extends to the experiment $(P_h)_{h\in H}$.

In order to prove 2., we show that in general the distribution of T_1 cannot be written as a convolution with the distribution of T_0.

Define the distribution of Z to be $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Then we have

$$T_0 \sim \sqrt{Z}W_1 \sim \frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0, 1),$$

and

$$T_1 \sim \sqrt{Z}W_1 - Z \sim \frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(-1, 1).$$

Assume that μ is a probability measure satisfying

$$\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(-1, 1) = \Big(\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0, 1)\Big) * \mu. \tag{4}$$

Then the singular parts must coincide, i.e.

$$\Big(\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(-1, 1)\Big)_s = \Big(\Big(\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0, 1)\Big) * \mu\Big)_s,$$

whence

$$\frac{1}{2}\delta_0 = \big(\frac{1}{2}\delta_0 * \mu\big)_s = \frac{1}{2}\delta_0 * \mu_s.$$

This implies $\mu_s = \delta_0$ and $\mu = \delta_0$ which is a contradiction to (4). $\square$

Our second counterexample shows that things can even be more complicated.

(2.14) THEOREM *Assume that under P_0 the coordinate process (X_t) has the distribution of $(Zt + \sqrt{Z}W_t)$, where (W_t) is a Wiener process and Z is a nonnegative, integrable and nonconstant random variable which is stochastically independent of (W_t). Then the following assertions are true:*

1. The estimator $T_1 = X_1 - \langle X, X\rangle_1$ is equivariant and has minimal quadratic risk among all equivariant estimators, but it is not equivalent to the canonical estimator T_0.

2. There exist random variables Z such that neither T_1 nor T_0 have the convolution property.

Proof: In order to prove 1., we proceed as in the proof of Theorem (2.13).

Under each P_0^z the coordinate process (X_t) has the distribution of $(zt + \sqrt{z}W_t)$, where (W_t) is a Wiener prozess. Therefore it follows for the same reasons as in Theorem (2.12) that for each experiment $(P_h^z)_{h\in H}$, $z \geq 0$, the canonical estimator $T_0 = X_1$ has the convolution property. However, T_0 is not unbiased. For $(P_h^z)_{h\in H}$, $z \geq 0$, the canonical unbiased estimators are rather the functions $T_{1,z} := T_0 - z = X_1 - z$, which are equivalent to T_0, and which have minimal quadratic risk among all equivariant estimators.

Now, let ρ be any equivariant estimator which is strictly equivariant on the orthogonal complement of ker(f). Then we have, for every $z \geq 0$,

$$\int (\xi - f(h))^2 \rho(., d\xi)\, dP_h^z \geq \int (T_{1,z} - f(h))^2\, dP_h^z.$$

This implies

$$\begin{aligned}&\int (\xi - f(h))^2 \rho(., d\xi)\, dP_h(.|\langle X, X\rangle_1)\\ &\quad \geq \int (X_1 - \langle X, X\rangle_1 - f(h))^2\, dP_h(.|\langle X, X\rangle_1)\ P_h\text{-a.e.}\end{aligned}$$

By integration with respect to the distribution of $\langle X, X\rangle_1$ the assertion follows.

In order to prove assertion 2, we consider the estimators T_0 and T_1 separately. It is trivial that the canonical estimator T_0 cannot have the convolution property, since T_1 has a smaller variance than T_0. But also T_1 in general does not have the convolution property since for some distribution of Z (e.g. $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$) the distribution of T_0 cannot be written as a convolution with the distribution of T_1. This follows from the proof of Theorem (2.13). □

3 The counterexamples for cylindrical measures

In this section we present assertion 2 of Theorem (2.14) in terms of cylindrical experiments.

(3.1) THEOREM *There is a cylindrical measure $P = P_0$ on the separable Hilbert space $H = L^2[0,1]$ such that*

(1) the full shift $(P_h)_{h\in H}$ forms a dominated experiment.

In fact, the abstract L-space generated by $(P_h)_{h\in H}$ can be described in an explicit way: There is an injective continuous map $\tau : H \to C_0[0,1]$ such that $\tau(P_h) = (\tau(P))_{\tau(h)}$ are Borel-measures on $C_0[0,1]$ and are dominated by a Borel probability measure μ (e.g. $\mu = \tau(P_0)$) on $C_0[0,1]$ equivalent to $(\tau(P_h))_{h\in H}$. We identify the family $(P_h)_{h\in H}$ with elements of $L^1(\mu) := L^1(C_0[0,1], \mu)$.

(2) There is a continuous linear projection $A : H \to \mathbb{R}$ which may be factored as $A = \phi_A \circ \tau$, where ϕ_A is a continuous projection from $C_0[0,1]$ to $\mathbb{R}$, and a continuous positive contraction $\gamma : L^1(\mu) \to L^1(\mathbb{R}, \lambda)$, where λ denotes the Lebesque-measure on $\mathbb{R}$, such that

(i) $\gamma(P_h) = \gamma(P_0) * \delta_{Ah}$ for $h \in H$,

(ii) $\gamma(P_0)$ is not the convolution of $A(P_0)$ with any probability measure.

Define $U \subseteq H$ to be the one-dimensional subspace formed by the constant functions and W to be its orthogonal complement. We write the elements $h \in H$ as $h = (u, w)$ and denote by π_U and π_W the respective orthogonal projections.

To define the cylindrical measure $P = P_0$ on H we first fix an auxiliary random variable Z uniformly distributed on [0,1]. We have to define the law of P on each finite-dimensional subspace V of H; we may and do suppose w.l.g. that V contains U so that we may again write the elements $v \in V$ as $v = (u, w)$ where $u \in U$ and $w \in V \cap W$. Given $\{Z = s\}$ for some $s \in [0,1]$ we define the law of $P[\,\cdot\,|Z = s]$ on V as the normal distribution with mean $(s, 0)$ and covariance matrix equal to $s \cdot \mathrm{Id}$. It is straightforward to check that this welldefines a cylindrical measure on H, which does not extend to a sigma-additive Borel-measure on H.

The experiment E now is defined as the full shift $E = (P_h)_{h\in H}$.

To obtain a concrete respresentation of the abstract L-space generated by E define

$$\tau : L^2[0,1] \to C_0[0,1] \tag{5}$$

$$f(\,\cdot\,) \mapsto g(\,\cdot\,) = \int_0^{\cdot} f(t)dt \tag{6}$$

A basic fact of the Wiener process going back to the original paper by N. Wiener, [8], implies that τ maps P to a sigma-additive Borel measure $\mu = \tau(P)$ on $C_0[0,1]$, which can be described explicitly: Given $\{Z = s\}$ the measure $P(\,\cdot\,|Z = s)$ equals Wiener measure on $C_0[0,1]$ with variance and drift both equal to s. Denoting by B_s the Borel subset of $C_0[0,1]$

$$B_s = \{g \in C_0[0,1] : \langle g, g\rangle_t = st, \text{ for all } t \in [0,1]\} \tag{7}$$

where $\langle g, g\rangle_t$ denotes the quadratic variation function, the set B_s has full measure under the law $P(\,\cdot\,|Z = s)$. We observe that the sets B_s are invariant under shifts by elements in the image $\tau(H)$ of H, a fact which will turn out to be of crucial importance. Also note that by Cameron-Martin-Girsanov the measures $\tau(P_h)$ are all equivalent to $\tau(P_0)$, for any $h \in H$. In particular $\tau(P_h)$ all are elements of $L^1(\mu)$ which shows in particular that the experiment E is dominated (in the abstract sense). Letting $B = \bigcup_{s\in[0,1]} B_s$ we exhibit a Borel subset of $C_0[0,1]$ of full μ-measure.

We now define $A : H \mapsto \mathbb{R}$ as the orthogonal projection $A = \pi_U$. The mapping A may be represented as $A = \phi_A \circ \tau$ where $\phi_A : C_0[0,1] \mapsto \mathbb{R}$ is given by $\phi_A(g) = g(1)$. The measure $A(P_0)$ now has the following law: Given $\{Z = s\}$ it is $\mathcal{N}(s,s)$-distributed.

To define the map $\gamma : L^1(\mu) \mapsto L^1(\mathbb{R}, \lambda)$ we first define a map $p_U : B \mapsto \mathbb{R}$ by letting $p_U(g) = \phi_A(g) - s$, if $g \in B_s$. We may write p_U explicitly by

$$p_U(g) = g(1) - \langle g, g\rangle_1 \text{ for } g \in B. \tag{8}$$

Note that p_U is a Borel-measurable function on B. We therefore may extend p_U to a stochastic kernel

$$\begin{aligned} \rho : B &\rightarrow \mathcal{M}(\mathbb{R}) && (9)\\ g &\mapsto \rho(g,.) := \delta_{p_U(g)} && (10) \end{aligned}$$

where $\delta_{p_U(g)}$ denotes Dirac-measure at $p_U(g)$. The kernel ρ induces a positive contraction $\gamma : L^1(\mu) \rightarrow L^1(\mathbb{R}, \lambda)$. By construction and the invariance of B_s under shifts by $\tau(h)$, for $h \in H$, we have

$$\gamma \circ S_{\tau(h)} = S_{Ah} \circ \gamma, \qquad \text{for } h \in H, \tag{11}$$

where $S_f : g \mapsto f + g$ denotes translation by g in $C_0[0,1]$ and $S_x : y \mapsto x + y$ the translation in $\mathbb{R}$. Whence by identifying $(P_h)_{h\in H}$ with the elements $\tau(P_h)$ of $L^1(\mu)$ we get the desired invariance property

$$\gamma(P_h) = \gamma(P_0) * \delta_{Ah}, \qquad \text{for } h \in H. \tag{12}$$

Finally note that $\gamma(P)(\,\cdot\,|Z = s)$ is normally distributed on $\mathbb{R}$ with mean equal to zero and variance equal to s, i.e., $\gamma(P)$ is a mixture of $\mathcal{N}(0,s)$-distributions. As $A(P_0)$ is a mixture of $\mathcal{N}(s,s)$-distributions one easily verifies for the variances that $V(A(P_0)) > V(\gamma(P_0))$ and in particular $\gamma(P_0)$ is not the convolution of $A(P_0)$ with any probability measure.

4 Appendix: Proof of Theorem (2.10)

Proof: Let ρ be an arbitrary equivariant estimator, i.e., a Markov kernel satisfying

$$\int\int g(\xi)\rho(x+\tau(h), d\xi + f(h))\, P_0(dx) = \int\int g(\xi)\rho(x, d\xi)\, P_0(dx)$$

for all $h \in H$ and $g \in \mathcal{C}_b := \mathcal{C}_b(\mathbb{R})$.

Let ν be a probability measure which is equivalent to the family $(P_h)_{h\in H}$ and denote $L^1 := L^1(\Omega, \mathcal{A}, \nu)$. By $\mathcal{B}$ we denote the set of all bilinear functions β on $\mathcal{C}_b \times L^1$ which are continuous and satisfy

$$\begin{aligned} &|\beta(g,k)| \le ||g||_u ||k||_1 && \text{if } g \in \mathcal{C}_b,\ k \in L^1,\\ &\beta(g,k) \ge 0 && \text{if } g \ge 0,\ k \ge 0,\\ &\beta(1,k) = \textstyle\int k\, d\nu && \text{if } k \in L^1. \end{aligned}$$

We endow $\mathcal{B}$ with the topology of pointwise convergence on $\mathcal{C}_b \times L^1$. With respect to this topology $\mathcal{B}$ is a compact set. It is easy to see that each element of $\mathcal{B}$ can be represented by a substochastic kernel (cf. Strasser, [7], Lemma 42.6).

For every $h \in H$, let β_h be the bilinear function

$$\beta_h : (g,k) \mapsto \int\int g(\xi)\rho(x+\tau(h), d\xi + f(h))\, k(x)\, \nu(dx),$$

where $g \in \mathcal{C}_b,\ k \in L^1$. Define

$$K := \overline{\text{co}}\{\beta_h : h \in H\}.$$

Then K is a convex and compact subset of $\mathcal{B}$.

For every $h \in H$ let $T_h : \mathcal{B} \to \mathcal{B}$ be the linear transformation defined by

$$T_h\beta : (g,k) \mapsto \beta(g(. - \tau(h)), k(. - f(h))), \quad g \in \mathcal{C}_b,\ k \in L^1.$$

Then $(T_h)_{h\in H}$ is an Abelian group of continuous transformations, mapping K into K. Thus, by the Markov-Kakutani fixpoint theorem there is a fixpoint in K, i.e. a bilinear function $\beta^* \in K$ such that $T_h\beta^* = \beta^*$ for all $h \in H$.

By equivariance of ρ we have $\beta_h(g, dP_0/d\nu) = \beta_0(g, dP_0/d\nu)$ for all $g \in \mathcal{C}_b$ and $h \in H$. This property extends to all elements of K and therefore it follows that

$$\beta^*(g, dP_0/d\nu) = \beta_0(g, dP_0/d\nu), \quad \text{for all } g \in \mathcal{C}_b \text{ and } h \in H.$$

It follows that β^* can be represented by a stochastic kernel ρ^* having the same distribution under P_0 as ρ.

Since β^* is a fixpoint we have

$$\int\int g(\xi)\rho^*(x+\tau(h), d\xi + f(h))\, k(x)\nu(dx) = \int\int g(\xi)\rho^*(x, d\xi)\, k(x)\nu(dx) \quad (13)$$

for all $h \in H$, $g \in \mathcal{C}_b$ and $k \in L^1$. It follows that for all $h \in H$ and $B \in \mathcal{B}$ we have

$$\rho^*(., B) = \rho^*(. + \tau(h), B + f(h)) \quad P_0\text{-a.e.},$$

where the exceptional set depends on $h \in H$ but not on $B \in \mathcal{B}$.

In order to finish the proof of Theorem 2.10, we need a lifting argument. One may apply Ionescu-Tulcea's lifting theorem (see LeCam, [4], section 8.3, Theorem 3) or a direct argument like that of Strasser, [7], Theorem 48.9. □

References

[1] C. Boll. *Comparison of experiments in the infinite case and the use of invariance in establishing sufficiency*. PhD thesis, Stanford, (1955).

[2] J. Hajek. A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitsth. verw. Geb.*, **14** (1970), 323–330.

[3] L. LeCam. Limits of experiments In: L. LeCam, editor, *Proc. 6th Berkeley Symp.*, (1972), pp. 245–261.

[4] L. LeCam. *Asymptotic Methods in Statistical Decision Theory*. Springer, (1986).

[5] L. LeCam. An infinite dimensional convolution theorem. In S.S. Gupta and J.O. Berger, editors, *Statistical Theory and related Topics V*, Springer, New York, (1994), pp. 401–411.

[6] M.W. Moussatat. *On the asymptotic theory of statistical experiments and some of its applications*. PhD thesis, Berkeley, (1976).

[7] H. Strasser. *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*, volume 7 of *De Gruyter Studies in Mathematics*. de Gruyter, (1985).

[8] N. Wiener. Differential space. *J. Math. Phys.*, **2**, 132–174.

[9] A. van der Vaart. An asymptotic representation theorem. *Internat. Statist. Review*, **58** (1991), 97–121.

[10] G. Yang. A conversation with Lucien LeCam. *Statistical Science*, 14 (1999), 223–241.

Klaus Pötzelberger
Department of Statistics
Vienna University of Economics and Business Administration

Walter Schachermayer
Department of Statistics, Probability Theory and Actuarial Mathematics
Technical University of Vienna

Helmut Strasser
Department of Statistics
Vienna University of Economics and Business Administration