Tatra Mt. Math. Publ. **26** (2003), 255–267

**TATRA MOUNTAINS**
*Mathematical Publications*

# ON THE STRUCTURE OF UNDOMINATED STATISTICAL EXPERIMENTS: SUFFICIENCY, INVARIANCE, AND OPTIMALITY IN UNBIASED ESTIMATION

(extended abstract)

Immanuel M. Bomze

ABSTRACT. There are three different (although equivalent) ways to define the $L$-space $L(\mathcal{E})$ of a statistical experiment $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{P}$ is an (arbitrarily complex) set of probability measures on a measurable space $(\Omega, \mathcal{F})$. This $L(\mathcal{E})$ is an ordered Banach lattice substitute of $L^1(Q)$ if $Q$ is a probability measure on $(\Omega, \mathcal{F})$ equivalent to all considered probabilities $P \in \mathcal{P}$ of the experiment in case that $\mathcal{E}$ (i.e., $\mathcal{P}$) is dominated by a $\sigma$-finite measure. The (topological) dual $M(\mathcal{E}) = \left[L(\mathcal{E})\right]^*$ is called $M$-space of $\mathcal{E}$, and generalizes the space of all bounded random variables in the experiment. We discuss relations between measure-theoretic properties of $\mathcal{P}$ like (weak) domination, and functional-analytic ones of the $L$- and $M$-spaces as (order complete) Banach lattices. It will turn out, that many structural properties which facilitate analysis of sufficiency, invariance and optimality in unbiased estimation carry over from the dominated to the undominated case. This way, following *LeCam's* decision theoretic approach [L. LeCam: *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York, 1986], functional analytic means are employed to avoid involved measure theoretic arguments.

## Introduction, motivation, and basic theory

The purpose of this extended abstract is to illuminate the basic structure underlying statistical decision theory under the perspective of geometry rather than measure theory, in the otherwise hard case of undominated experiments. As a substitute for the space of all densities in question (in the undominated case, no such densities exist) we have to construct the L-space of the experiment.

---

2000 Mathematics Subject Classification: 62B05, 62B15, 62C07 62C15, 62G05.
Keywords: (order complete) Banach lattice, solid subsets of lattices, band, abstract $L$-space, abstract $M$-space, statistical decision theory, undominated experiment, coherent experiment, Randon measures; finite subset property; Radon-Nikodym theorem.

It will turn out, that many structural properties facilitating analysis carry over from the dominated to the undominated case, at the cost of generalizing the notion of a random variable: generalized observables now will be linear functionals taking the underlying probability distributions and related measures to real numbers, in a way corresponding to integration in the classical context. Thus, following L e C a m 's decision theoretic approach [8], functional analytic means are employed to avoid involved measure theoretic arguments. The geometric notions used emerge quite naturally if we want to keep the key linear and order structures which are so helpful in classical analysis of statistical reasoning. This approach allows for general results on sufficiency, invariance, and optimality in unbiased estimation.

The presentation is organized as follows: after introducing the general setup, we introduce and discuss several equivalent ways to define L- and M-spaces of the experiment. Thereafter, the notions of sufficient reduction and subalgebras are presented, followed by a treatment of admissibility and optimality in the context of generalized unbiased estimation. Finally, we cover the case of invariance, to indicate how sufficient projections can be calculated exactly or approximately.

In a situation where several or even an infinity of observations can be obtained simultaneously, it is useful to consider random vectors, or random elements in infinite-dimensional linear spaces, respectively. It is possible to transfer the obtained results to generalized random elements which are defined analogously by functional analytic means. In this way, one can cope with invariance, admissiblity, and optimality in the estimation of multivariate or even infinite dimensional parameters, too. However, for conciseness of presentation, this generalization is not included here. For details see [1], from where most of the results presented below are taken.

The most frequently used model in context of statistical decision theory [3, 8, 11, 12, 14] is that of a *statistical experiment* $\mathcal{E}$. Formally, this is $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{P}$ is an arbitrarily complex set of probability measures on measurable space $(\Omega, \mathcal{F})$. A statistical decision problem is to determine which of the distributions in $\mathcal{P}$ are most likely to generate the observations (or data) collected.

To each experiment $\mathcal{E}$ there is associated the system of null sets

$$\mathcal{N}(\mathcal{E}) := \left\{A \in \mathcal{F} \colon P(A) = 0 \text{ for all } P \in \mathcal{P}\right\}.$$

If there exists a $\sigma$-finite measure $\mu$ on $(\Omega, \mathcal{F})$ such that $\mathcal{N}(\mathcal{E}) \supseteq \left\{F \in \mathcal{F} \colon \mu(F) = 0\right\}$, then there also exists a probability measure $Q \in \mathrm{ca}(\Omega, \mathcal{F})$ with $\mathcal{N}(\mathcal{E}) = \left\{F \in \mathcal{F} \colon Q(F) = 0\right\}$. Then $\mathcal{E}$ is called *dominated* and we write $\mathcal{P} \cong Q$. In the dominated case, analysis is relatively easy: by the Radon-Nikodym Theorem, all distributions $P \in \mathcal{P}$ have densities $\frac{dP}{dQ} \in L^1(Q)$ with respect to a fixed probability $Q$. Then most of statistical theory can be analyzed within the space $L^1(Q)$.

If, however, the distributions in question $P \in \mathcal{P}$ are too diverse, no such $Q$ exists. This is the *undominated* case. In this case, we need a substitute for $L^1(Q)$. One of the key properties of $L^1(Q)$ is that it is an *ordered vector space*, which is closed under (pointwise) maximization/minimization. Hence observables can be linearly combined and (order-) compared with each other. If we want to keep these features, we have to seek a *Banach lattice* containing all observables. In a first step, we "linearize" the set of distributions, i.e., extend from $P_i \in \mathcal{P}$ to (limits of) linear combinations of $P_i$, with the aim to employ (convex, linear, and order) analysis.

Now, the space $\mathrm{ca}(\Omega, \mathcal{F})$ of *all* signed finite measures on $(\Omega, \mathcal{F})$ would allow for this, but this space has no modelling power: it is too large, because it ignores structure of $\mathcal{P}$. Nevertheless, the natural order inherited from $\mathrm{ca}(\Omega, \mathcal{F})$ will prove to be useful: for any two $\mu, \nu \in \mathrm{ca}(\Omega, \mathcal{F})$,

$$\mu \leq \nu \quad \text{means} \quad \mu(F) \leq \nu(F) \qquad \text{for all } F \in \mathcal{F}.$$

With this order, $\mathrm{ca}(\Omega, \mathcal{F})$ is a vector lattice. In particular, for any two $\mu, \nu \in \mathrm{ca}(\Omega, \mathcal{F})$, there are always $\sup(\mu, \nu) \in \mathrm{ca}(\Omega, \mathcal{F})$ and $\inf(\mu, \nu) \in \mathrm{ca}(\Omega, \mathcal{F})$. Denote by $|\mu| := \sup(\mu, -\mu) \in \mathrm{ca}(\Omega, \mathcal{F})$. Then the *variational norm* can be written as $\|\mu\| := |\mu|(\Omega)$. Under this norm, $\mathrm{ca}(\Omega, \mathcal{F})$ is an *order complete* Banach lattice, more precisely, an *abstract L-space* [10], which means that the norm $\|\cdot\|$ is additive on the positive cone $\mathrm{ca}(\Omega, \mathcal{F})_+ = \{\mu \in \mathrm{ca}(\Omega, \mathcal{F}) \colon \mu \geq o\}$ of $\mathrm{ca}(\Omega, \mathcal{F})$:

$$\mu \geq o, \ \nu \geq o \implies \|\mu + \nu\| = \|\mu\| + \|\nu\|.$$

As indicated above, we consider a subspace of $\mathrm{ca}(\Omega, \mathcal{F})$, which is related to $\mathcal{E}$ and which respects both linear and order structure of measures: this will be called the *L-space* $L(\mathcal{E})$ of the experiment introduced in [7]. This L-space is an order-complete Banach sub-lattice [10] of $\mathrm{ca}(\Omega, \mathcal{F})$, and will be defined in three different ways below [1]. But before, we need some more terminology: a linear subspace $\mathcal{D} \subseteq \mathrm{ca}(\Omega, \mathcal{F})$ is called *solid*, if $\mathcal{D}$ contains all intervals with $\mathcal{D}$-endpoints:

$$\mu \leq \nu \leq \sigma \text{ and } \mu, \sigma \in \mathcal{D} \implies \nu \in \mathcal{D}.$$

A solid linear sub-space $\mathcal{D} \subseteq \mathrm{ca}(\Omega, \mathcal{F})$ is called a *band* [10], if $\sup_{i \in I} \mu_i \in \mathcal{D}$ whenever $\mu_i \in \mathcal{D}$ satisfy $\mu_i \leq \mu \in \mathrm{ca}(\Omega, \mathcal{F})$ for all $i \in I$. Finally, we use $\mu \perp \nu$ as an abbreviation for $\inf(|\mu|, |\nu|) = o$.

**Theorem 1 and Definition.** *If $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{P})$ is a statistical experiment,*

(a) *let $L_1(\mathcal{E})$ be the smallest band (with respect to $\subseteq$) in $\mathrm{ca}(\Omega, \mathcal{F})$ containing $\mathcal{P}$;*

(b) *let $L_2(\mathcal{E})$ be the $\|\cdot\|$-closure of $L'(\mathcal{E})$, where $L'(\mathcal{E}) := \Big\{\mu \in \mathrm{ca}(\Omega, \mathcal{F}) \colon |\mu| \leq \sum_{i=1}^{n} \alpha_i P_i$ for some $P_i \in \mathcal{P}$, $\alpha_i \geq 0$, $1 \leq i \leq n$; $n \in \mathbb{N}\Big\}$; and*

(c) *let* $L_3(\mathcal{E}) := \{\mu \in \mathrm{ca}(\Omega, \mathcal{F}) \colon \mu \perp \sigma$ *for all* $\sigma \perp \mathcal{P}\}$.

*Then* $L_1(\mathcal{E}) = L_2(\mathcal{E}) = L_3(\mathcal{E})$. *This space is called the L-space of* $\mathcal{E}$, *denoted by* $L(\mathcal{E})$.

Proof. See [1]; cf. [12]. □

The three representations above can be interpreted as follows:

(a) the representation $L_1(\mathcal{E})$ ensures that linear and (even infinite) order operations within $L(\mathcal{E})$ make sense;

(b) the representation $L_2(\mathcal{E})$ means that we can "almost" use densities with respect to (variable) $Q = \sum_{i=1}^{n} \alpha_i P_i$ within $L(\mathcal{E})$;

(c) the representation $L_3(\mathcal{E})$ guarantees that we can ignore distributions which are irrelevant to $\mathcal{P}$, i.e., those measures $\mu$ which are concentrated on an $\mathcal{E}$-null set $F \in \mathcal{N}(\mathcal{E})$: indeed, $\mu \perp \nu$ means that $\mu$ is concentrated on a $\nu$-null set and vice versa.

In the dominated case $\mathcal{P} \cong Q$, the L-space $L(\mathcal{E})$ is identical to the set of all measures in $\mathrm{ca}(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $Q$. If $\mathcal{P} \cong Q$, then $L(\mathcal{E})$ is, as a *Banach* lattice, isomorphic to $L^1(Q)$ (see, e.g., [12, p. 227f.], where also the inclusion $L_1(\mathcal{E}) \subseteq L_3(\mathcal{E})$ in the above theorem is shown for general $\mathcal{E}$; cf. [11]).

The situation for undominated experiments is different. As an abstract L-space, $L(\mathcal{E})$ is always isomorphic to $L^1(m)$, with $m$ a *Radon measure* on a locally compact topological space [4]. But in general $m$ is not even *semi-finite* and then there is no representation of the topological dual $M(\mathcal{E}) := L(\mathcal{E})^*$ as $L^\infty(m)$. Recall that $m$ is said to be semi-finite [9] on a $\sigma$-field $\mathcal{A}$, if $m(A) = \sup\{m(B) \colon B \in \mathcal{A},\ B \subseteq A,\ m(B) < +\infty\}$ for all $A \in \mathcal{A}$. This is called *finite subset property* in [15].

Although $M(\mathcal{E})$ may not coincide with $L^\infty(m)$, it is an order complete abstract M-space, if furnished by the induced order [10, p. 121]. This means that whenever $f, g \in M(\mathcal{E})$ with $f \geq o$ and $g \geq o$, then $\|\sup(f,g)\| = \max\{\|f\|, \|g\|\}$. Thus, according to [5] and [6], $M(\mathcal{E})$ is, as a Banach lattice, isometrically isomorphic to the space $C(Z)$ of all continuous, real-valued functions topological space $Z$ which is compact and *extremally disconnected* (i.e., the closure of every open set $U \subseteq Z$ is open). Despite this strange topological property of $Z$, the isometric isomorphism between $M(\mathcal{E})$ and $C(Z)$ will be helpful in working with products within $M(\mathcal{E})$ in the following section, because at first sight it may not be obvious how to multiply linear functionals such that the product is again linear.

$M(\mathcal{E})$ is called the *M-space of the experiment* $\mathcal{E}$ and generalizes the space of equivalence classes of bounded random variables in the following sense. Let $X$

denote the set of all real-valued functions defined on $\Omega$ that are $\mathcal{F}$-measurable and bounded. For any $\varphi \in X$, denote by $\dot\varphi$ the map assigning $\int_\Omega \varphi \, d\mu$ to every $\mu \in L(\mathcal{E})$. Furthermore, for any pair $\{\varphi, \psi\} \subseteq X$, the relation $\varphi \sim \psi$ means that $\varphi = \psi$ holds $P$-almost surely for all $P \in \mathcal{P}$. Then $X/_\sim$ is isomorphic to $\dot X := \{\dot\varphi \colon \varphi \in X\} \subseteq M(\mathcal{E})$, and one can show that $M(\mathcal{E})$ coincides with the $\sigma\big(M(\mathcal{E}), L(\mathcal{E})\big)$-closure of $\dot X$:

**Theorem 2.** *$M(\mathcal{E})$ coincides with the $\sigma\big(M(\mathcal{E}), L(\mathcal{E})\big)$-closure of $\dot X$.*

*To be more specific: for every $f \in M(\mathcal{E})$ there are $\mathcal{F}$-measurable, bounded $\varphi_i \colon \Omega \to \mathbb{R}$, which in addition satisfy $|\varphi_i(\omega)| \le \|f\|$ for all $i$, all $\omega \in \Omega$, such that, as $i$ gets large,*

$$\int_\Omega \varphi_i \, d\mu \to f(\mu) \qquad \text{for all } \mu \in L(\mathcal{E}).$$

Proof. See [1]; cf. [7, 12]. □

To formulate an alternative representation of $M(\mathcal{E})$ due to Torgersen [13], we denote for any family of classical bounded random variables $\tilde f = (f_P)_{P\in\mathcal{P}} \in \bigotimes_{P\in\mathcal{P}} L^\infty(P)$ by

$$\|\tilde f\|_{\mathcal{P}} := \sup_{P\in\mathcal{P}} \|f_P\|_{\infty,P} \le +\infty .$$

**Theorem 3.** *The M-space $M(\mathcal{E})$ of $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{P})$ is, as a Banach lattice, isometrically isomorphic to*

$$\Big\{\tilde f = (f_P)_{P\in\mathcal{P}} \in \bigotimes_{P\in\mathcal{P}} L^\infty(P) \colon \|\tilde f\|_{\mathcal{P}} < \infty \text{ and } \bigcap_{P\in\mathcal{Q}} f_P \ne \emptyset \colon \text{ all finite } \mathcal{Q} \subset \mathcal{P}\Big\}$$

*if $\tilde M(\mathcal{E})$ is equipped by the above norm and furnished by the natural ordering:*

$$\tilde f \le \tilde g \qquad \text{means} \qquad f_P \le g_P \quad \text{for all } P \in \mathcal{P}.$$

Proof. See [1]; cf. [13]. □

The idea behind Torgersen's representation can be read as follows. Instead of one observable $\varphi$, pick for any distribution $P \in \mathcal{P}$ an *individual* version $f_P$ and form strings $(f_P)_{P\in\mathcal{P}}$ across *all* $P \in \mathcal{P}$, in a way compatible with any *finite* collection $\mathcal{Q}$ of distributions: $\varphi \in \bigcap_{P\in\mathcal{Q}} f_P$.

In the dominated case $\mathcal{P} \cong Q$, all $f_P = \varphi_Q$, and $X/_\sim$ is the same as $L^\infty(Q)$. On the other hand, here $M(\mathcal{E})$ clearly coincides with $\dot X$. Hence the fact that $X/_\sim$ and $\dot{X}$ are isomorphic reduces in the dominated case to the familiar isomorphism between $\left[L^1(Q)\right]^*$ and $L^\infty(Q)$:

$$\left[L^1(Q)\right]^* \cong \left[L(\mathcal{E})\right]^* = M(\mathcal{E}) = \dot{X} \cong X/_\sim \cong \tilde{M}(\mathcal{E}) \cong L^\infty(Q)\,.$$

We proceed from dominated experiments to the "next simple" case: an experiment $\mathcal{E}$ is called *coherent*, if $M(\mathcal{E}) = \dot{X}$. By the immediately preceding arguments, every dominated experiment is also coherent, the reverse implication being false in general (for even larger classes of statistical experiments see, e.g., [9]). But every coherent experiment is *weakly dominated* (and vice versa) in the following sense [11]: there exists a semi-finite (not $\sigma$-finite, in general) and *localizable* measure $\mu$ on $(\Omega, \mathcal{F})$, such that $\mathcal{N} = \mu^{-1}(\{0\})$. This result is an alternative interpretation of the fact that $L^\infty(\mu)$ is isomorphic to $\left[L^1(\mu)\right]^*$ if and only if $\mu$ is semi-finite and localizable [15]. As usual, $\mu$ is called localizable on $\mathcal{F}$ if and only if $L^\infty(\mu)$ is order complete or, equivalently, if for any family $\{F_i \colon i \in I\} \subseteq \mathcal{F}$, there is a set $F \in \mathcal{F}$ such that $\mu(F) = \sup\limits_{i \in I} \mu(F_i)$.

The experiment $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{P})$ with $\Omega = [0,1]$, $\mathcal{F}$ the Borel field, and $\mathcal{P} = \left\{\delta_x \colon x \in [0,1]\right\}$ is not coherent since the counting measure $\mu$ is not localizable on $\mathcal{F}$, because $\mathcal{F}$ is countably generated but $\mu$ is not $\sigma$-finite [9] (this argument needs the assumption that each uncountable metric space contains a non-Borel set). An alternative (still non-constructive) proof of this is based upon Theorem 3:

One can show [11] that $\mathcal{E}$ is coherent in the sense that $M(\mathcal{E}) = \dot{X}$ if and only if for all $\tilde{f} = (f_P)_{P \in \mathcal{P}} \in \tilde{M}(\mathcal{E})$ there is a function $\psi \in X$ fulfilling $\psi \in \bigcap\limits_{P \in \mathcal{P}} f_P$. The latter condition has been used in [2] to define coherence originally. Now consider again the experiment $\mathcal{E}$ with $\Omega = [0,1]$, $\mathcal{F}$ the Borel $\sigma$-field, and $\mathcal{P} = \left\{\delta_x \colon x \in [0,1]\right\}$. Pick a function $h \colon \Omega \to [0,1]$ which is not $\mathcal{F}$-measurable, and consider $\tilde{f} = (f_x)_{x \in [0,1]}$ with $f_x(\omega) = h(x)$, a constant, for all $x \in [0,1]$ (evidently $f_x$ is short for the more tedious notation $f_{\delta_x}$). Then clearly $\|\tilde{f}\| \leq 1$. Furthermore, for all finite subsets $\mathcal{Q} \subset [0,1]$ there is a $\varphi \in X$ which solves the problem $\varphi(x) = h(x)$ for all $x \in \mathcal{Q}$, e.g., the *Lagrange* interpolation polynomial. Hence $\bigcap\limits_{x \in \mathcal{Q}} f_x \neq \emptyset$ for all finite subsets $\mathcal{Q}$ of $[0,1]$. Thus $\tilde{f} \in \tilde{M}(\mathcal{E})$. But any $\psi \in \bigcap\limits_{P \in \mathcal{P}} f_P$ necessarily must coincide with $h$, and hence cannot belong to $X$, so that $\mathcal{E}$ cannot be coherent.

## Sufficiency

A problem reduction, provided one knows that the relevant information is contained in a sub-$\sigma$-field $\mathcal{C}$ of $\mathcal{F}$, is, in the classical measure theoretic context, accomplished by means of the conditional expectation $\mathbb{E}_P(\varphi|\mathcal{C})$ with respect to $P$ of a random variable $\varphi$, given $\mathcal{C}$. The counterparts of sub-$\sigma$-fields $\mathcal{C}$ of $\mathcal{F}$ are in the undominated case *sub-algebras* $\mathcal{A}$ of the M-space $M(\mathcal{E})$. Indeed, looking at Torgersen's representation of $M(\mathcal{E})$, we see that one can define a bilinear multiplication taking a pair $(f,g) \in M(\mathcal{E}) \times M(\mathcal{E})$ into the product $f \cdot g \in M(\mathcal{E})$, which respects order and which is jointly continuous in the norm topology. To visualize this multiplication, we also may employ the representation of $M(\mathcal{E})$ as the space $C(Z)$ of real-valued continuous functions defined on a compact space $Z$, using pointwise multiplication of functions within $C(Z)$. This way, one can even show that multiplication is separately weakly continuous [1]. Given our multiplication, a linear subspace $\mathcal{A}$ of $M(\mathcal{E})$ is called a sub-algebra if it is closed under multiplication: $a, b \in \mathcal{A}$ implies $a \cdot b \in \mathcal{A}$.

Also, to generalize the transition from $\varphi$ to $\mathbb{E}_P(\varphi|\mathcal{C})$, a projection operator $\Pi_P\colon M(\mathcal{E}) \to \mathcal{A}$ is introduced, again based upon the isometry of $M(\mathcal{E})$ to $C(Z)$. Indeed, one can show that to any weakly closed subalgebra $\mathcal{A}$ there exists a sub-$\sigma$-field $\hat{\mathcal{A}}$ of Baire sets in $Z$ and vice versa, and then we simply put $\Pi_P f$ that element in $\mathcal{A}$ corresponding to a suitable continuous representative of $\mathbb{E}_{\hat{P}}(\hat{f}|\hat{\mathcal{A}})$. For details, we refer again to [1], but let us mention that the ideas underlying the following two results more or less go back to the seminal article by Le Cam [7].

**Theorem 4.** *If $\mathcal{A} \subseteq M(\mathcal{E})$ is a $\sigma\big(M(\mathcal{E}), L(\mathcal{E})\big)$-closed sub-algebra, then there is a linear, positive, idempotent projection $\Pi_P\colon M(\mathcal{E}) \to \mathcal{A}$ which satisfies $\Pi_P g \leq g$ for all $g \in \mathcal{A}$ with $g \geq o$.*

Proof. See [1]; cf. [7]. □

We now concentrate on the case where the reduction mentioned above is possible without loss of information: in this situation, the sub-algebra $\mathcal{A}$ is termed *sufficient* in analogy to the classical notion of sufficient sub-$\sigma$-fields $\mathcal{C}$. Recall that in the classical context, a sub-$\sigma$-field $\mathcal{C}$ is called sufficient if there is a version of $\mathbb{E}_P(\varphi|\mathcal{C})$ which does not depend on $P$:

**Theorem 5 and Definition.** *Let $\mathcal{A} \subseteq M(\mathcal{E})$ be a $\|\cdot\|$-closed sub-algebra containing the functional $e\colon \mu \mapsto \mu(\Omega)$. If there is a linear, positive, idempotent projection $\Pi\colon M(\mathcal{E}) \to \mathcal{A}$ satisfying $\Pi g \leq g$ for all $g \in \mathcal{A}$ with $g \geq o$, which does not depend on $P$, then $\mathcal{A}$ is called sufficient. If $\mathcal{A}$ is sufficient, this sufficient projection $\Pi$ satisfies $\Pi g = g$ for all $g \in \mathcal{A}$ and also*

$$(\Pi f)(P) = f(P) \qquad \text{for all } P \in \mathcal{P},\ f \in M(\mathcal{E}).$$

*Moreover, the projection is the adjoint $\Pi = \Lambda^*$ of a positive linear operator $\Lambda\colon L(\mathcal{E}) \to L(\mathcal{E})$; in particular, $\Pi$ is weakly continuous and $\mathcal{A}$ is $\sigma\big(M(\mathcal{E}), L(\mathcal{E})\big)$-closed.*

Proof. See [1]; cf. [7]. □

The main result in this context is that the property of sufficiency is preserved under arbitrarily many intersections. Remember that this result does not hold true in the classical context for undominated experiments. As a consequence, in undominated experiments the minimal sufficient sub-$\sigma$-field need *not* exist in the classical sense.

**THEOREM 6 AND DEFINITION.** *Any intersection of sufficient sub-algebras is again sufficient. Hence there is always a minimal sufficient sub-algebra $\mathcal{A}_m$, which coincides with the intersection of all sufficient sub-algebras.*

P r o o f. See [1]; cf. [7]. □

**COROLLARY.** *If $\mathcal{E}$ is coherent, there exists always a minimal sufficient sub-$\sigma$-field in the classical sense.*

P r o o f. See [1]. □

Recall that this result is incorrect without any assumptions on $\mathcal{E}$: see [2, 11].

## Admissibility and optimal unbiased estimation

This chapter is devoted to optimality and admissibility in unbiased estimation. A parameter $\vartheta$ of interest is (unbiasedly) estimable, if there is a generalized random variable $f \in M(\mathcal{E})$ fulfilling

$$f(P) = \vartheta(P) \quad \text{for all } P \in \mathcal{P}.$$

In this case we call $f$ a *generalized unbiased estimator* of $\vartheta$. The risk of $f$ with respect to a given convex loss function can be defined in the present context as follows:

**DEFINITION.** Let $L = (L_P)_{P\in\mathcal{P}}$ be a collection of convex loss functions over $\mathbb{R}$, for example, the quadratic loss $L_P(t) = (t-\theta_P)^2$. The *L-risk* of a $f \in M(\mathcal{E})$ under $P \in \mathcal{P}$ is defined as

$$L_P * f(P) := \sup\left\{ \sum_{i=1}^{n} \mu_i(\Omega) L_P\left[f(\bar{\mu}_i)\right] \right\},$$

where $\bar{\mu}_i := \frac{1}{\mu_i(\Omega)}\mu_i$ and the sup ranges over all $\mu_i \in L(\mathcal{E})$ with $\mu_i > o$ and $\sum_i \mu_i = P$.

The idea behind this concept is a sort of coarse integration dominated by $P$, accomplished by decomposition of $P$ into $\mu_i$ as above. Then the usual pointwise integration of classical loss $(L_P \circ \varphi)$ w.r.t. $P$ is replaced with the expression for

$L_P * f(P)$. A probably more transparent way to define $L_P * f$ uses again the isomorphism of $M(\mathcal{E})$ with $C(Z)$, observing that $L_P \circ \hat{f} \in C(Z)$ if $\hat{f} \in C(Z)$. In an intrinsic, purely vector-lattice way, one may define $L_P * f = \sup\{\alpha_n f + \beta_n e \colon n \in \mathbb{N}\} \in M(\mathcal{E})$, where $e(\mu) = \mu(\Omega)$ and $\alpha_n$, $\beta_n$ are real-valued sequences satisfying $L_P(t) = \sup_n[\alpha_n t + \beta_n]$ for all $t \in \mathbb{R}$ (recall that every convex function can be written as the upper envelope of countably many affine functions). The latter representation also yields the formula for $L_P * f(P)$ in the definition of the risk above. Remark that we may regain multiplication intrinsically via $f \cdot g = \frac{1}{2}\big[S * (f+g) - S * f - S * g\big]$ if $S(t) = t^2$.

In short, $L_P * f(P)$ generalizes the classical risk $\int\limits_\Omega L_P(\varphi)\,dP$; indeed, $L_P * \dot{\varphi}(P) = \int\limits_\Omega L_P(\varphi)\,dP$, and more precisely, $(L_P \circ \varphi)^{\cdot} = L_P * \dot{\varphi}$ for any $\varphi \in X$.

Now, if $f \in M$ minimizes the $L$-risk among all generalized unbiased estimators of $\vartheta$, then it is obvious to call $f$ *optimal* with respect to this loss function $L$; if there is no alternative generalized unbiased estimator with lower risk, then $f$ is termed *admissible*; note that alternative generalized unbiased estimators are of the form $g = f + v$ with

$$v \in V := \{v \in M(\mathcal{E}) \colon v(P) = 0 \text{ for all } P \in \mathcal{P}\}\,.$$

**Definition.** An estimator $f \in M(\mathcal{E})$ is called:

(a) *$L$-admissible* if for all $v \in V$,

$$L_P * (f+v)(P) \le L_P * f(P) \text{ for all } P \in \mathcal{P}$$

implies $v = o$;

(b) *$L$-optimal* if for all $v \in V$ we have

$$L_P * f(P) \le L_P * (f+v)(P) \text{ for all } P \in \mathcal{P}\,.$$

(c) *optimal* if $f$ is $L$-optimal for all convex $L$.

Also, there is a generalization of Jensen's inequality, relating (sufficient) projection with the risk:

**Theorem 7.** *Let $\mathcal{A}$ be a sufficient sub-algebra and $\Pi$ the sufficient projection onto $\mathcal{A}$. For any convex loss function $L$, we have*

$$L_P * (\Pi f) \le \Pi(L_P * f) \qquad \text{for all } P \in \mathcal{P}$$

*and thus*

$$L_P * (\Pi f)(P) \le L_P * f(P) \qquad \text{for all } P \in \mathcal{P}\,.$$

Proof. See [1]. □

By consequence, risk is reduced by sufficient projections: recall that $\Pi f(P) = f(P)$ for all $P \in \mathcal{P}$ if $\Pi f$ is a sufficient projection, i.e., $v = \Pi f - f \in V$ for all estimators $f \in M(\mathcal{E})$, so $\Pi f$ is an improving unbiased estimator. To summarize, we see that Rao-Blackwellization works also in the generalized context. Similarly, we also can generalize Rao's covariance principle which states that an estimator is optimal if and only if its covariance with any unbiased estimator of zero vanishes (note that we are in the context of bounded random variables, so that infinite variances are no issue here):

**THEOREM 8 AND DEFINITION.** *$f \in M(\mathcal{E})$ is optimal if and only if $f \in \mathcal{A}_0$ where*

$$\mathcal{A}_0 := \{f \in M(\mathcal{E}) \colon f \cdot v \in V \text{ for all } v \in V\}\,.$$

*$\mathcal{A}_0$ is a closed sub-algebra, the so-called optimal algebra.*

Proof. See [1]. □

Further generalizations of classical results follow:

**THEOREM 9.** (a) *If $L$ is strictly convex, then every $L$-optimal $f$ is $L$-admissible and this determines $f$ uniquely among unbiased alternatives;*

(b) *Every $L$-admissible $f$ belongs to the minimal sufficient algebra $\mathcal{A}_m$, and therefore $\mathcal{A}_0 \subseteq \mathcal{A}_m$.*

Proof. See [1]. □

Similarly we can generalize existence criteria of non-trivial optimal unbiased estimators in terms of the structure of $V$: call $V$ *reducible* if there are non-trivial, weakly closed linear subspaces $V_1$ and $V_2$ of $V$ such that $V_1 \oplus V_2 = V$ and $V_1 \cdot V_2 = \{o\}$. Otherwise, $V$ is called *irreducible.* Then

**THEOREM 10.** *$V$ is irreducible if and only if $\mathcal{A}_0$ is trivial, i.e., if $\mathcal{A}_0 = \mathbb{R}e$ holds.*

Proof. See [1]. □

We close with the counterpart to the Lehmann-Scheffé Theorem:

**THEOREM 11.** *The following assertions are equivalent:*

(a) *$\mathcal{A}_0$ is sufficient;*

(b) *$\mathcal{A}_0 = \mathcal{A}_m$;*

(c) *$\mathcal{A}_m$ is (boundedly) complete, i.e., $\mathcal{A}_m \cap V = \{o\}$;*

(d) *there exists a boundedly complete and sufficient sub-algebra;*

(e) *every unbiased estimable parameter $\theta_P$ has an optimal estimator;*

(f) *every unbiased estimable parameter $\theta_P$ has a UMVUE.*

Proof. See [1]. □

## Invariance

If we know that the models $(\Omega, \mathcal{F}, P)$, $P \in \mathcal{P}$, are left unchanged under some transformations, then this knowledge is a gain of information which is closely related to sufficiency, and therefore also connected with optimality and admissibility. This is the central aspect in the treatment of invariance with respect to groups $G$ of transformations which act on the experiment. Again, we obtain results on generalized random variables that in the classical context can be derived only under additional assumptions, e.g., about the structure of the group $G$. However, exploiting further information about group structure, or about the way $G$ acts on the experiment, allows to calculate explicitly the sufficient projection $\Pi_G$, which in turn enables us to transform generalized estimators into invariant ones without losing relevant information.

Consider a transformation group of transitions $\{T_g : g \in G\}$ acting on the L-space, i.e., linear, positive, isometric operators $T_g : L(\mathcal{E}) \to L(\mathcal{E})$ w.r.t. variational norm $\|\cdot\|$, such that $T_g P = P$ for all $P \in \mathcal{P}$. This reflects invariance of $\mathcal{E}$ under the group $G$.

EXAMPLE. Let $g : \Omega \to \Omega$ be *model preserving*, i.e., invertible with both $g$ and $g^{-1}$ measurable such that the image measure given by $(T_g P)(F) := P\big(g^{-1}(F)\big)$, all $F \in \mathcal{F}$, coincides with $P$, i.e., $T_g P = P$ for all $P \in \mathcal{P}$. Then $T_g$ satisfies the above conditions.

**THEOREM 12.** *The adjoint $T_g^* : M(\mathcal{E}) \to M(\mathcal{E})$ is an isometric Banach lattice isomorphism; furthermore, the sub-algebra*

$$\mathcal{A}_g := \big\{ f \in M(\mathcal{E}) : T_g^* f = f \big\}$$

*of invariant generalized observables is sufficient, and its projection is given by the Césaro ergodic limit*

$$\Pi_g f := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \big[T_g^*\big]^k f .$$

P r o o f. See [1]. □

As an immediate consequence we obtain an important relationship between invariance and sufficiency:

**THEOREM 13 AND DEFINITION.** *Let*

$$\mathcal{A}_G := \big\{ f \in M(\mathcal{E}) : T_g^* f = f \ \text{ for all } g \in G \big\} .$$

*Then $\mathcal{A}_G = \bigcap_{g\in G} \mathcal{A}_g$ is sufficient, hence $\mathcal{A}_m \subseteq \mathcal{A}_G$. $\mathcal{A}_G$ is called the invariant sub-algebra.*

P r o o f. See [1]. □

To answer the natural question whether $\mathcal{A}_G$ is minimal-sufficient, we introduce, for any group $G$ of model preserving transitions,

$$V(G) := \left\{ \sum_{i=1}^{n} \left[ f_i - T^*_{g_i} f_i \right] : f_i \in M(\mathcal{E}),\ g_i \in G \right\}.$$

Then $V(G) \subseteq V$, i.e., $v(P) = 0$ for all $P \in \mathcal{P}$ and all $v \in V(G)$.

**THEOREM 14.** *$\mathcal{A}_G = \mathcal{A}_m = \mathcal{A}_0$ if and only if $V(G)$ is $\sigma\big(M(\mathcal{E}), L(\mathcal{E})\big)$-dense in $V$.*

Proof. See [1]. □

Now consider the *full model* where $\mathcal{P}$ coincides with *all* invariant distributions under $G$:

$$\mathcal{P} = \left\{ P \in \mathrm{ca}(\Omega, \mathcal{F}) : P \geq o,\ P(\Omega) = 1,\ T_g P = P \text{ for all } g \in G \right\}.$$

Again, this situation has been studied in the classical context. However, in that context, the following results hold only under additional assumptions.

**THEOREM 15.** *In the full model where $\mathcal{P}$ coincides with all invariant distributions under $G$, we always have $\mathcal{A}_G = \mathcal{A}_m = \mathcal{A}_0$. Hence every unbiasedly estimable parameter has a generalized* UMVUE, *which is invariant.*

Proof. See [1]. □

We finalize by indicating how to calculate of $\Pi_G$ explicitly for proofs, see again [1]:

- for finite groups $G$ we have
$$\Pi_G = \frac{1}{n} \sum_{g \in G} T^*_g ;$$
- for commutative groups, $\Pi_G$ is approximated by the iterates $\Pi_{g_1} \circ \cdots \circ \Pi_{g_n}$ (note that the ordering of $g_1, \ldots, g_n$ is irrelevant due to commutativity);
- for amenable groups under appropriate technical assumptions, we have
$$\Pi_G f(\mu) = \lim_{i \in I} \frac{1}{\eta(K_i)} \int_{K_i} f(T_g \mu)\, \eta(dg) \qquad \text{for all } (f, \mu) \in M(\mathcal{E}) \times L(\mathcal{E}).$$

  The assumptions are as follows: $\eta$ is a (not necessarily $\sigma$-finite) measure on the measurable space $(G, \mathcal{G})$; group multiplication is $(\mathcal{G} \otimes \mathcal{G})$-$\mathcal{G}$-measurable; $g \mapsto f(T_g \mu)$ is $\mathcal{G}$-measurable for all $(f, \mu) \in M(\mathcal{E}) \times L(\mathcal{E})$; and finally, there exists a summing net $(K_i)_{i \in I}$, i.e., $\eta(K_i) < \infty$ for all $i \in I$ and $\lim_{i \in I} \int_{K_i} [\varphi(g) - \varphi(gh)] \eta(dg) = 0$ for all bounded, $\mathcal{G}$-measurable $\varphi \colon G \to \mathbb{R}$ and all $h \in G$.

## REFERENCES


[1] BOMZE, I. M.: *A Functional Analytic Approach to Statistical Experiments*, Longman, Harlow, 1990.

[2] HASEGAWA, M.—PERLMAN, M. D.: *On the existence of a minimal sufficient subfield*, Ann. Statist. **2** (1974), 1049–1055; *Correction*, Ann. Statist. **3** (1975), 1371–1372.

[3] HEYER, H.: *Theory of Statistical Experiments*, Springer-Verlag, Berlin, 1982.

[4] KAKUTANI, S.: *Concrete representation of abstract L-spaces and the mean ergodic theorem*, Ann. Math. **42** (1941), 523–537.

[5] KAKUTANI, S.: *Concrete representation of abstract M-spaces*, Ann. Math. **42** (1941), 994–1024.

[6] KREIN, M.—KREIN, S.: *On an inner characteristic of the set of all continuous functions defined on a bicompact Hausdorff space*, Dokl. Akad. Nauk SSSR **27** (1940), 427–430. (Russian)

[7] LECAM, L.: *Sufficiency and approximate sufficiency*, Ann. Math. Statist. **35** (1964), 1419–1455.

[8] LECAM, L.: *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York, 1986.

[9] LUSCHGY, H.—MUSSMANN, D.: *Products of majorized experiments*, Statist. Decision **4** (1986), 321–335.

[10] SCHAEFER, H. H.: *Banach Lattices and Positive Operators*, Springer-Verlag, Berlin, 1974.

[11] SIEBERT, E.: *Pairwise sufficiency*, Z. Wahrscheinlichkeitstheor. Verw. Geb. **46** (1979), 237–246.

[12] STRASSER, H.: *Mathematical Theory of Statistics*, de Gruyter, Berlin, 1985.

[13] TORGERSEN, E. N.: *On complete sufficient statistics and uniformly minimum variance unbiased estimators*, in: Teoria statistica delle decisioni. Sympos. Math. **25**, 137–153 (1980).

[14] TORGERSEN, E. N.: *Comparison of Statistical Experiments*, Cambridge Univ. Press, Cambridge, 1991.

[15] ZAANEN, A. C.: *Integration*, North Holland, Amsterdam, 1967.


Received April 9, 2002

*Institut für Statistik*
*Universität Wien*
*Universitätsstrasse 5*
*A–1010 Wien*
*AUSTRIA*
*E-mail*: immanuel.bomze@univie.ac.at