A simplified approach to the Hajek-LeCam bound

F. Liese, I. Steinke
University of Rostock

Summary
Let E_n be a sequence of experiments weakly converging to a limit experiment E. One of the basic objectives of asymptotic decision theory is to derive asymptotically "best" decisions in E_n from optimal decisions in the limit experiment E. A central statement in this context is the Hajek-LeCam bound which represents a lower bound for the maximum risk of a sequence of decisions.
It is given a simplified proof for the existence of accumulation points of a sequence of generalized decision functions using the ε-Blackwell sufficiency of sufficient statistics defined on finite experiments. That is the main step for deriving the Hajek-LeCam bound for weakly convergent sequences of experiments.

AMS Classification: 62F05, 62F12

Key Words: Hajek-LeCam bound, weakly convergent experiments, generalized decision function

1 Introduction

The lower Hajek-LeCam bound is a central statement of asymptotic decision theory. The traditional way (see Strasser [4], LeCam [1]) to establish this statement is carried out in the following way. Using the concept of ε-deficiency, a metric is introduced which describes the so-called strong convergence of statistical experiments. A first step to the Hajek-LeCam bound is the relation between randomisation and deficiency. A next step concerns the fact that for finite experiments the strong and weak convergence coincide. The combination of these two results leads to the existence of accumulation points of sequences of decisions which belong to a weakly convergent sequence of experiments. This result is the key to prove the Hajek-LeCam bound for the maximum risk of a sequence of decisions.

This way to get the Hajek-LeCam bound is well developed in the monographs by Strasser and LeCam. This approach is going parallel with a systematic development of the whole asymptotic decision theory. The results provide a deep insight into the structure of convergent sequences of experiments. But for lectures in mathematical statistics this way is connected with some disadvantages. A lot of sophisticated results from analysis, topology and other fields are necessary for these considerations although several interesting "by-products" are obtained. Therefore a direct and simplified approach to the Hajek-LeCam bound is desirable.

We do not prove a general randomisation criterion in this note. We only deal with finite experiments and prove that a sufficient statistic T for $(P_\vartheta)_{\vartheta\in\mathcal{J}}$, $|\mathcal{J}|<\infty$, is ε-Blackwell sufficient, i.e. there is a kernel K_ε so that $\|P_\vartheta - K_\varepsilon P_\vartheta\circ T^{-1}\|_v < \varepsilon$ for every $\vartheta\in\mathcal{J}$. Consequently, every finite experiment is up to some ε equivalent to the standard experiment. Hence it suffices to deal with standard experiments whose standard measures σ_n converge weakly to the standard measure σ of the limit experiment. We approximate the projection onto the coordinate in the unit simplex by step functions measurable w.r.t. the algebra of σ-continuity sets in order to prove that for any weakly convergent sequence of standard experiments the distributions of the sequence are approximately randomisations by kernels of the corresponding distributions in the limit experiment. This and the ε-Blackwell sufficiency of any sufficient statistic take us in the position to prove the existence of accumulation points of sequences of decisions being the main step towards the Hajek-LeCam bound.

2 Notations and results

Let $(\Omega,\mathfrak{F})$ be a measurable space and $\Theta\neq\emptyset$ an arbitrary set. A **statistical experiment** with parameter set Θ is defined to be a triple $E=(\Omega,\mathfrak{F},P_\vartheta,\vartheta\in\Theta)$, where $(\Omega,\mathfrak{F})$ is the sample space and $\mathcal{P}=(P_\vartheta)_{\vartheta\in\Theta}$ is a family of probability measures on $(\Omega,\mathfrak{F})$. The experiment E is said to be **finite** if $|\Theta|<\infty$.

Let $A(\Theta)$ be the system of all finite subsets of Θ. For $\mathcal{J}\in A(\Theta)$, we denote by $\Sigma_\mathcal{J}$ the set of all mappings $x:\mathcal{J}\to I\!\!R_1$ so that $0\le x(\vartheta)\le 1$, $\sum\limits_{\vartheta\in\mathcal{J}} x(\vartheta)=1$. Then the unit simplex $\Sigma_\mathcal{J}$ becomes a Polish space with respect to the Euclidean distance $d_\mathcal{J}$; its σ-algebra of Borel sets is denoted by $\mathfrak{B}_\mathcal{J}$. Let $\nu_\mathcal{J}=\sum\limits_{\vartheta\in\mathcal{J}} P_\vartheta$ and introduce the $\mathfrak{F}$-$\mathfrak{B}_\mathcal{J}$ measurable mapping $T_\mathcal{J}$ by

$$\Big(T_\mathcal{J}(\omega)\Big)_\vartheta = \left(\frac{dP_\vartheta}{d\nu_\mathcal{J}}\right)(\omega) \quad , \quad \vartheta\in\mathcal{J}\ . \tag{1}$$

It holds

$$I\!\!E_{\nu_\mathcal{J}}\Big(X\,|\,\sigma(T_\mathcal{J})\Big) = I\!\!E_{P_\vartheta}\Big(X\,|\,\sigma(T_\mathcal{J})\Big) \quad \nu_\mathcal{J}-a.s.$$

for every $\vartheta\in\mathcal{J}$ and every bounded r.v. X on $(\Omega,\mathfrak{F})$. This means that $T_\mathcal{J}$ is a sufficient statistic for $(P_\vartheta)_{\vartheta\in\mathcal{J}}$.

It is well-known that only under additional assumptions on the measurable space $(\Omega,\mathfrak{F})$ the statistic $T_\mathcal{J}$ is Blackwell sufficient, i.e. there exists a stochastic kernel $K:(\Sigma_\mathcal{J},\mathfrak{B}_\mathcal{J})\Rightarrow(\Omega,\mathfrak{F})$ so that $K\,P_\vartheta\circ T_\mathcal{J}^{-1}=P_\vartheta$ for every $\vartheta\in\mathcal{J}$. Here and in the following $K\mu$ denotes the application of the kernel K to the measure μ. Let, furthermore,

$$\|\mu-\nu\|_v = 2\sup_{A\in\mathfrak{F}}|\mu(A)-\nu(A)|$$

be the **variational distance** of the two probability measures μ and ν on $(\Omega,\mathfrak{F})$.

In general, $T_{\mathcal{J}}$ is not Blackwell sufficient because of the non-existence of regular conditional distributions. But $T_{\mathcal{J}}$ is approximately Blackwell-sufficient in the following sense.

Proposition 2.1 *Let $(\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \mathcal{J})$ be a finite experiment, $(R, \mathfrak{R})$ a measurable space and $T : (\Omega, \mathfrak{F}) \to (R, \mathfrak{R})$ a sufficient statistic for $(P_\vartheta)_{\vartheta \in \Theta}$. Then for every $\varepsilon > 0$ there exists a stochastic kernel $L_\varepsilon : (R, \mathfrak{R}) \Rightarrow (\Omega, \mathfrak{F})$ so that*

$$\|P_\vartheta - L_\varepsilon \, P_\vartheta \circ T^{-1}\|_v < \varepsilon$$

for every $\vartheta \in \mathcal{J}$.

For the proof see section 3.
Given the experiment $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$ and $\mathcal{J} \in A(\Theta)$, the measure $\sigma_{\mathcal{J}} = \nu_{\mathcal{J}} \circ T_{\mathcal{J}}^{-1}$ is called **standard measure**.

Introduce distributions Q_ϑ on $(\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}})$ by $dQ_\vartheta = \pi_\vartheta \, d\sigma_{\mathcal{J}}$, where π_ϑ is the projection onto the ϑ-th coordinate of $I\!R^{\mathcal{J}}$, $\vartheta \in \mathcal{J}$. Then it is clear that $Q_\vartheta = P_\vartheta \circ T_{\mathcal{J}}^{-1}$. Because of the sufficiency, the finite experiments $E^{\mathcal{J}} = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \mathcal{J})$ and $E_{\mathcal{J}} = (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}}, Q_\vartheta, \vartheta \in \mathcal{J})$ are equivalent from the statistical point of view.

The notion of weak convergence of experiments is basic in asymptotic decision theory. Let $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$, $E_n = (\Omega_n, \mathfrak{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$ be experiments. E_n is called **weakly convergent** to E, written $E_n \Rightarrow E$, if for every finite subset $\mathcal{J} \subseteq \Theta$

$$\sigma_{n,\mathcal{J}} \Longrightarrow \sigma_{\mathcal{J}}$$

as $n \to \infty$, where $\Rightarrow$ is the symbol for weak convergence of finite measures defined on $(\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}})$.

Note that for $E_n \Rightarrow E$ the corresponding finite subexperiments $E_{n,\mathcal{J}}$ and $E_{\mathcal{J}}$ have a special structure. Let $\mathfrak{C}_{\mathcal{J}}$ denote the algebra of $\sigma_{\mathcal{J}}$-continuity sets, i.e. the family of all sets $A \in \mathfrak{B}_{\mathcal{J}}$ with $\sigma_{\mathcal{J}}(\partial A) = 0$. As we will show later, the projections π_ϑ being the densites in all experiments $E_{n,\mathcal{J}}$ and $E_{\mathcal{J}}$ may be approximated uniformly by step functions over sets from $\mathfrak{C}_{\mathcal{J}}$. On the other hand, the sequence of measures $\sigma_{n,\mathcal{J}}$ converges pointwise to $\sigma_{\mathcal{J}}$ on $\mathfrak{C}_{\mathcal{J}}$, i.e. $\lim_{n\to\infty} \sigma_{n,\mathcal{J}}(A) = \sigma_{\mathcal{J}}(A)$ for every $A \in \mathfrak{C}_{\mathcal{J}}$. For experiments of this type the pointwise convergence of the standard measures $\sigma_{n,\mathcal{J}}$ implies approximate randomisation of the corresponding distributions of the experiments by stochastic kernels.

Given an algebra $\mathfrak{A}$ of subsets of $\mathcal{X}$, we denote by $I\!B(\mathfrak{A})$ the set of all bounded, real-valued, $\mathfrak{A}$-measurable functions and by $I\!B_0(\mathfrak{A}) \subseteq I\!B(\mathfrak{A})$ the set of all those f assuming only a finite number of different values. Let $\|f\|_u$ be the sup-norm.

Proposition 2.2 *Let $(\mathcal{X}, \mathfrak{B})$ be a measurable space and $\mathfrak{A} \subseteq \mathfrak{B}$ a subalgebra with $\mathfrak{B} = \sigma(\mathfrak{A})$. Suppose that $\sigma, \sigma_1, \sigma_2, \ldots$ are finite measures on $(\mathcal{X}, \mathfrak{B})$ with $0 < \sigma(\mathcal{X}) = \sigma_n(\mathcal{X})$, $n = 1, 2, \ldots,$ and*

$$\lim_{n\to\infty} \sigma_n(A) = \sigma(A)$$

for every $A \in \mathfrak{A}$. Assume that $\{f_\vartheta \in I\!B(\mathfrak{B}), \vartheta \in \mathcal{J}\}$ is a finite family of non-negative functions with

$$\inf_{g \in \mathbb{B}_0(\mathfrak{A})} \|f_\vartheta - g\|_u = 0$$

and $\int f_\vartheta d\sigma_n = \int f_\vartheta d\sigma = 1$*. Let the distributions* Q_ϑ *and* $Q_{n,\vartheta}$ *be defined by* $dQ_\vartheta = f_\vartheta d\sigma$ *and* $dQ_{n,\vartheta} = f_\vartheta d\sigma_n$*, respectively, on* $(\mathcal{X}, \mathfrak{B})$*. Then for every* $\varepsilon > 0$ *there exist* n_ε *and stochastic kernels* $K_{n,\varepsilon} : (\mathcal{X}, \mathfrak{B}) \Rightarrow (\mathcal{X}, \mathfrak{B})$ *so that*

$$\|Q_{n,\vartheta} - K_{n,\varepsilon} Q_\vartheta\|_v < \varepsilon$$

for every $\vartheta \in \mathcal{J}$*,* $n \geq n_\varepsilon$*.*

For the proof see section 3.
Now we are able to establish the randomisation criterion for weakly convergent experiments.

Theorem 2.3 *Let* $E_n = (\Omega_n, \mathfrak{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$, $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$ *be experiments so that* E_n *converges weakly to* E*. Then for every* $\mathcal{J} \in A(\Theta)$ *and every* $\varepsilon > 0$ *there exists a stochastic kernel* $K_{n,\varepsilon,\mathcal{J}} : (\Omega, \mathfrak{F}) \Rightarrow (\Omega_n, \mathfrak{F}_n)$ *so that*

$$\|P_{n,\vartheta} - K_{n,\varepsilon,\mathcal{J}} P_\vartheta\|_v < \varepsilon$$

for every $\vartheta \in \mathcal{J}$ *and every sufficiently large* n*.*

Proof: The proof is devided into several steps.
1) Let $\mathcal{J} \in A(\Theta)$ be fixed and $\varepsilon > 0$. For every $a \in \Sigma_\mathcal{J}$ and $r > 0$ we denote by $K_r(a) \subseteq \Sigma_\mathcal{J}$ the open sphere with radius r and centre a. Let σ be a finite measure on $(\Sigma_\mathcal{J}, \mathfrak{B}_\mathcal{J})$. For different $r_i > 0$ the boundaries $\partial K_{r_i}(a)$ of the spheres $K_{r_i}(a)$, $i = 1, 2$, are disjoint. As σ is finite, the set $\{r : \sigma(\partial K_r(a)) > 0\}$ is at most countable. If δ is any positive real number, then the system

$$\mathfrak{K}_\delta = \{K_r(a) \,|\, r \leq \delta, a \in \Sigma_\mathcal{J} \text{ with } \sigma(\partial K_r(a)) = 0\}$$

covers the compact set $\Sigma_\mathcal{J}$. Hence there exists a finite subcover $\mathfrak{M}_\delta$. Denote by $\mathfrak{B}_m$ the algebra generated by $\bigcup_{j=1}^{m} \mathfrak{M}_{\frac{1}{j}}$. By construction, the atoms of $\mathfrak{B}_m$ have a diameter not exceeding $\frac{1}{m}$.
2) Consider the standard experiments

$$\begin{aligned} E_{n,\mathcal{J}} &= (\Sigma_\mathcal{J}, \mathfrak{B}_\mathcal{J}, Q_{n,\vartheta}, \vartheta \in \mathcal{J}) \;, \\ E_\mathcal{J} &= (\Sigma_\mathcal{J}, \mathfrak{B}_\mathcal{J}, Q_\vartheta, \vartheta \in \mathcal{J}) \end{aligned}$$

which belong to E_n and E, respectively. Let $T_{n,\mathcal{J}}$ and $T_\mathcal{J}$ be defined as in (1). Then $Q_{n,\vartheta} = P_{n,\vartheta} \circ T_{n,\mathcal{J}}^{-1}$, $Q_\vartheta = P_\vartheta \circ T_\mathcal{J}^{-1}$, $\sigma_{n,\mathcal{J}} = (\sum_{\vartheta \in \mathcal{J}} P_{n,\vartheta}) \circ T_{n,\mathcal{J}}^{-1}$, $\sigma_\mathcal{J} = (\sum_{\vartheta \in \mathcal{J}} P_\vartheta) \circ T_\mathcal{J}^{-1}$, and it holds

$$\frac{dQ_\vartheta}{d\sigma_\mathcal{J}} = \pi_\vartheta \qquad \sigma_\mathcal{J} - a.s. \quad , \quad \frac{dQ_{n,\vartheta}}{d\sigma_{n,\mathcal{J}}} = \pi_\vartheta \qquad \sigma_{n,\mathcal{J}} - a.s. \;,$$

where π_ϑ is the projection onto the ϑ-th coordinate of $\mathbb{R}^\mathcal{J}$, $\vartheta \in \mathcal{J}$.

Since $\Sigma_{\mathcal{J}}$ is compact, the continuous function π_ϑ is even uniformly continuous. Hence for every $\varepsilon > 0$ there exist some $\delta_\varepsilon > 0$ so that

$$|\pi_\vartheta(x) - \pi_\vartheta(y)| < \varepsilon$$

for every $x, y \in \Sigma_{\mathcal{J}}$ with $d_{\mathcal{J}}(x,y) < \delta_\varepsilon$ and every $\vartheta \in \mathcal{J}$. If $m > 1/\delta_\varepsilon$, then the atoms of $\mathfrak{B}_m$ have a diameter not exceeding δ_ε. Hence there are step-functions g_ϑ being $\mathfrak{B}_m$-measurable so that $\|\pi_\vartheta - g_\vartheta\|_u < \varepsilon$, $\vartheta \in \mathcal{J}$. Introduce the algebra $\mathfrak{A}$ by $\mathfrak{A} = \bigcup_{m=1}^{\infty} \mathfrak{B}_m$. By construction, for every $a \in \Sigma_{\mathcal{J}}$ and every $\varepsilon > 0$ there exists an open sphere $K_{\varepsilon'}(a')$, $\varepsilon' \leq \varepsilon$, $a \in K_{\varepsilon'}(a') \subseteq K_\varepsilon(a)$, which belongs to $\mathfrak{A}$. As $\Sigma_{\mathcal{J}}$ is separable, $\sigma(\mathfrak{A})$ contains all open subsets of $\Sigma_{\mathcal{J}}$. Hence $\mathfrak{B}_{\mathcal{J}} = \sigma(\mathfrak{A})$. Every $\mathfrak{B}_m$ is generated by sets with boundary of $\sigma_{\mathcal{J}}$-measure zero. Therefore $\mathfrak{A}$ is a subalgebra of the algebra of all Borel sets $B \subseteq \Sigma_{\mathcal{J}}$ with $\sigma_{\mathcal{J}}(\partial B) = 0$.

Now take $\sigma = \sigma_{\mathcal{J}}$. Then all assumptions of Proposition 2.2 are fulfilled. Consequently, for every $\varepsilon > 0$ there exist stochastic kernels $K_{n,\varepsilon} : (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}}) \Rightarrow (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}})$ so that

$$\|Q_{n,\vartheta} - K_{n,\varepsilon} Q_\vartheta\|_v < \frac{\varepsilon}{2}$$

for every $\vartheta \in \mathcal{J}$ and every sufficiently large n.

3) Apply Proposition 2.1 to $T = T_{n,\mathcal{J}}$, $(R, \mathfrak{R}) = (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}})$ and $Q_{n,\vartheta} = P_{n,\vartheta} \circ T_{n,\mathcal{J}}^{-1}$: There exist stochastic kernels $L_{n,\varepsilon} : (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}}) \Rightarrow (\Omega_n, \mathfrak{F}_n)$ so that

$$\|P_{n,\vartheta} - L_{n,\varepsilon} Q_{n,\vartheta}\|_v < \frac{\varepsilon}{2}$$

for every sufficiently large n and every $\vartheta \in \mathcal{J}$. Let $\delta_{T_{\mathcal{J}}} : (\Omega, \mathfrak{F}) \Rightarrow (\Sigma_{\mathcal{J}}, \mathfrak{B}_{\mathcal{J}})$ be the kernel induced by the $\mathfrak{F}$-$\mathfrak{B}_{\mathcal{J}}$ measurable mapping $T_{\mathcal{J}}$ and put

$$K_{n,\varepsilon,\mathcal{J}} = L_{n,\varepsilon} K_{n,\varepsilon} \delta_{T_{\mathcal{J}}} \ .$$

Then $\|P_{n,\vartheta} - K_{n,\varepsilon,\mathcal{J}} P_\vartheta\|_v < \varepsilon$ for every $\vartheta \in \mathcal{J}$ and sufficiently large n which completes the proof. $\square$

Let $(\Omega, \mathfrak{F})$ be a measurable space and $ca(\Omega, \mathfrak{F})$ the family of all finite, signed measures defined on $(\Omega, \mathfrak{F})$. If $\mu = \mu^+ - \mu^-$ is the Hahn-decomposition of $\mu \in ca(\Omega, \mathfrak{F})$ and $|\mu| = \mu^+ + \mu^-$, then $\|\mu\| = |\mu|(\Omega)$ is said to be the **total variation** of μ. $ca(\Omega, \mathfrak{F})$ equipped with the total variation as norm is a Banach space. Note that $\|P - Q\| \leq \|P - Q\|_v$ if P and Q are probability measures on $\mathfrak{F}$. Given an experiment $E = (\Omega, \mathfrak{F}, \mathcal{P})$, we fix a linear subspace $\mathfrak{M} \subseteq ca(\Omega, \mathfrak{F})$ so that

(M1) $\quad \mathcal{P} \subseteq \mathfrak{M}$,

(M2) $\quad$ if $\mu \in \mathfrak{M}$ and $\nu \ll \mu$, then $\nu \in \mathfrak{M}$.

$\mathcal{M}(E)$ is defined to be the family of all linear subspaces of $ca(\Omega, \mathfrak{F})$ satisfying **(M1)** and **(M2)**. Let $\mu = \mu^+ - \mu^-$ be the Hahn-decomposition of $\mu \in ca(\Omega, \mathfrak{F})$. As $\mu^+ \ll \mu$ and $\mu^- \ll \mu$, it follows from **(M2)** that $\mu \in \mathfrak{M}$ implies $|\mu| \in \mathfrak{M}$.

Commonly, the L-space of the experiment $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$ is defined by

$$L(E) = \{\mu \,:\, \mu \in ca(\Omega, \mathfrak{F}) \,,\, \mathcal{P}^\perp \subseteq \mu^\perp\} \ ,$$

where $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ and $\mathcal{P}^\perp = \{\nu : \nu \perp \varrho$ for every $\varrho \in \mathcal{P}\}$. Since $\mathcal{P} \subseteq L(E)$ and $\nu \ll \mu$ implies $\mu^\perp \subseteq \nu^\perp$, it is easy to see that the L-space of E fulfils **(M1)** and **(M2)**.

Assume $I\!D$ is a given decision space equipped with a σ-algebra $\mathfrak{D}$. Let $\mathcal{L}$ be a linear subspace of $I\!B(\mathfrak{D})$, the space of all bounded, $\mathfrak{D}$-measurable, real-valued functions, and $\mathfrak{M} \in \mathcal{M}(E)$. By $\mathcal{B}(\mathcal{L}, \mathfrak{M})$ we shall denote the set of all **generalized decision functions**, i.e. the set of bilinear functions $\beta : \mathcal{L} \times \mathfrak{M} \to I\!R_1$ which are supposed to have the following properties:

(D1) $$|\beta(f, \mu)| \leq \|f\|_u \|\mu\| \ ,$$

(D2) $$\beta(f, \mu) \geq 0 \quad \text{if } f \geq 0 \ , \ \mu \geq 0 \ ,$$

(D3) $$\beta(1, \mu) = \mu(\Omega) \ .$$

Obviously, every stochastic kernel $K : (\Omega, \mathfrak{F}) \Rightarrow (I\!D, \mathfrak{D})$ defines a generalized decision function by

$$\beta_K(f, \mu) = \iint f(x)\, K(\omega, dx)\, \mu(d\omega)$$

for any $\mathfrak{M}$ and $\mathcal{L}$, even if $\mathfrak{M} = ca(\Omega, \mathfrak{F})$ and $\mathcal{L} = I\!B(\mathfrak{D})$.

Let $\mathcal{L}_0 \subseteq \mathcal{L}$, $\mathfrak{M}_0 \subseteq \mathfrak{M}$ be subspaces. On $\mathcal{B}(\mathcal{L}, \mathfrak{M})$ we introduce the **topology** $\tau(\mathcal{L}_0, \mathfrak{M}_0)$ as the weakest topology for which all mappings $\beta \to \beta(f, \mu)$, $f \in \mathcal{L}_0$, $\mu \in \mathfrak{M}_0$, are continuous.

Proposition 2.4 *$\mathcal{B}(\mathcal{L}, \mathfrak{M})$ is convex and compact w.r.t. the topology $\tau(\mathcal{L}_0, \mathfrak{M}_0)$.*

If $\mathcal{L}_0 = \mathcal{L}$ and $\mathfrak{M}_0 = \mathfrak{M}$, then the statement of Proposition 2.4 follows with the same arguments as in the proof of Strasser [4], Theorem 42.3. To establish Proposition 2.4 we have only to remark that $\mathcal{B}(\mathcal{L}, \mathfrak{M})$ being compact w.r.t. to $\tau(\mathcal{L}, \mathfrak{M})$ is again compact w.r.t. the weaker topology $\tau(\mathcal{L}_0, \mathfrak{M}_0)$.

It turns out that every $\beta \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$ which is defined only for $f \in \mathcal{L}$, $\mu \in \mathfrak{M}$, may be extended to a bilinear functional $\bar{\beta}$ which is defined for every $\mu \in ca(\Omega, \mathfrak{F})$ where the properties **(D1)**, **(D2)** and **(D3)** are preserved.

Proposition 2.5 *Assume that $\mathfrak{M} \in \mathcal{M}(E)$. Then for any $\beta \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$ there exists a $\bar{\beta} \in \mathcal{B}(\mathcal{L}, ca(\Omega, \mathcal{F}))$ such that*

$$\bar{\beta}(f, \mu) = \beta(f, \mu) \quad \forall\, f \in \mathcal{L}\ \forall\, \mu \in \mathfrak{M} \ .$$

For a proof of this proposition see section 3.

Suppose now that $E_n = (\Omega_n, \mathfrak{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$ is a sequence of experiments which weakly converges to the experiment $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$. Let $\mathfrak{M} \in \mathcal{M}(E)$, $\mathfrak{M}_n \in \mathcal{M}(E_n)$, and let $\beta_n \in \mathcal{B}(\mathcal{L}, \mathfrak{M}_n)$ be any sequence of generalized decision functions. $\beta \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$ is called an **accumulation point in distribution** of $\{\beta_n\}$, if for every $\mathcal{J} \in A(\Theta)$ and every finite subset $G \subseteq \mathcal{L}$ there exists a sequence $n'(\mathcal{J}, G)$ so that

$$\lim_{n' \to \infty} \beta_{n'}(g, P_{n',\vartheta}) = \beta(g, P_\vartheta)$$

for every $g \in G$, $\vartheta \in \mathcal{J}$.

The existence of accumulation points plays a key role in the proof of the Hajek-LeCam bound for weakly convergent sequences of experiments.

Proposition 2.6 *Assume that $E_n \Rightarrow E$, $\mathfrak{M} \in \mathcal{M}(E)$, and $\mathfrak{M}_n \in \mathcal{M}(E_n)$. Then every sequence $\{\beta_n\}$, $\beta_n \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$, $n \in \mathbb{N}$, has an accumulation point.*

Proof: (based on Strasser [4], Theorem 62.3) Let $\delta > 0$. According to Theorem 2.3, for any $\mathcal{J} \in A(\Theta)$ there exists a stochastic kernel $K_{n,\delta,\mathcal{J}} : (\Omega, \mathfrak{F}) \Rightarrow (\Omega_n, \mathfrak{F}_n)$ so that

$$\|P_{n,\vartheta} - K_{n,\delta,\mathcal{J}} P_\vartheta\|_v < \delta \quad , \ \forall \vartheta \in \mathcal{J}, \tag{2}$$

for every sufficiently large n.

Denote by $\bar{\beta}_n \in \mathcal{B}(\mathcal{L}, ca(\Omega_n, \mathfrak{F}_n))$ an extension of $\beta_n \in \mathcal{B}(\mathcal{L}, \mathfrak{M}_n)$ according to Proposition 2.5 and introduce $\beta_{n,\mathcal{J}} \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$ by

$$\beta_{n,\mathcal{J}}(f, \mu) = \bar{\beta}_n(f, K_{n,\delta,\mathcal{J}}\mu) \quad , \ f \in \mathcal{L} \ , \ \mu \in \mathfrak{M} \ .$$

Let $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathfrak{M}_0 \subseteq \mathfrak{M}$ be any subspaces. Then the compactness of $\mathcal{B}(\mathcal{L}, \mathfrak{M})$ w.r.t. $\tau(\mathcal{L}_0, \mathfrak{M}_0)$ implies that for any net $(\beta_{n,\mathcal{J}})_{n \in \mathbb{N}}$ there is an accumulation point $\beta_\mathcal{J} \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$ and for the net $(\beta_\mathcal{J})_{\mathcal{J} \in A(\Theta)}$ there is again an accumulation point $\beta \in \mathcal{B}(\mathcal{L}, \mathfrak{M})$. Here $A(\Theta)$ is supposed to be directed by inclusion.

It is to be shown that β is an accumulation point of $(\beta_n)_{n \in \mathbb{N}}$ in distribution. Let $\mathcal{J}_0 \in A(\Theta)$, $G \subseteq \mathcal{L}$ a finite subset and $\varepsilon > 0$. The proof is complete if we show that for any $n_0 \in \mathbb{N}$ there is a $n_\varepsilon \in \mathbb{N}$ so that $n_\varepsilon \geq n_0$ and

$$|\beta_{n_\varepsilon}(f, P_{n,\vartheta}) - \beta(f, P_\vartheta)| < \varepsilon$$

for every $f \in G$ and every $\vartheta \in \mathcal{J}$.

Since β is an accumulation point of $\beta_\mathcal{J}$, $\mathcal{J} \in A(\Theta)$, there exists a $\mathcal{J}_\varepsilon \in A(\Theta)$, $\mathcal{J}_0 \subseteq \mathcal{J}_\varepsilon$, so that

$$|\beta_{\mathcal{J}_\varepsilon}(f, P_\vartheta) - \beta(f, P_\vartheta)| < \frac{\varepsilon}{3} \tag{3}$$

for every $f \in G$ and every $\vartheta \in \mathcal{J}_0$. Take $a = \max\{\|f\|_u | f \in G\}$ and $\delta = \varepsilon/(3a)$. Choose n_1 sufficiently large so that the kernel $K_{n,\delta,\mathcal{J}_\varepsilon}$ from (2) fulfils

$$\|P_{n,\vartheta} - K_{n,\delta,\mathcal{J}_\varepsilon} P_\vartheta\|_v < \frac{\varepsilon}{3a} \tag{4}$$

for every $\vartheta \in \mathcal{J}_0$ and $n \geq n_1$. For $\beta_{\mathcal{J}_\varepsilon}$ is an accumulation point of $\beta_{n,\mathcal{J}_\varepsilon}$, there exists a $n_\varepsilon \geq \max\{n_0, n_1\}$ such that

$$|\beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f, P_\vartheta) - \beta_{\mathcal{J}_\varepsilon}(f, P_\vartheta)| < \frac{\varepsilon}{3} \tag{5}$$

for every $f \in G$ and every $\vartheta \in \mathcal{J}_0$. Consequently,

$$\begin{aligned} |\beta_{n_\varepsilon}(f, P_{n_\varepsilon,\vartheta}) - \beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f, P_\vartheta)| &= |\beta_{n_\varepsilon}(f, P_{n_\varepsilon,\vartheta}) - \beta_{n_\varepsilon}(f, K_{n_\varepsilon,\delta,\mathcal{J}_\varepsilon} P_\vartheta)| \\ &\leq \|f\|_u \|P_{n_\varepsilon,\vartheta} - K_{n_\varepsilon,\delta,\mathcal{J}_\varepsilon} P_\vartheta\|_v < \delta\, a = \frac{\varepsilon}{3} \ . \end{aligned} \tag{6}$$

The triangle inequality in conjunction with (3), (5), and (6) implies

$$\begin{aligned} |\beta(f,P_\vartheta)-\beta_{n_\varepsilon}(f,P_{n_\varepsilon,\vartheta})| &\le |\beta(f,P_\vartheta)-\beta_{\mathcal{J}_\varepsilon}(f,P_\vartheta)|+|\beta_{\mathcal{J}_\varepsilon}(f,P_\vartheta)-\beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f,P_\vartheta)| \\ &+ |\beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f,P_\vartheta)-\beta_{n_\varepsilon}(f,P_{n_\varepsilon,\vartheta})| < \varepsilon \end{aligned}$$

for every $\vartheta \in \mathcal{J}_0$ and every $f \in G$ which completes the proof. □

Assume that we are given a decision space $(I\!D, \mathfrak{D})$ and a family $\{W_\vartheta, \vartheta \in \Theta\} \subseteq \mathcal{L}$ of loss functions. Let $E_n = (\Omega_n, \mathfrak{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$, $E = (\Omega, \mathfrak{F}, P_\vartheta, \vartheta \in \Theta)$ be experiments and fix $\mathcal{L}_0 \subseteq \mathcal{L}$, $\mathfrak{M}_0 \subseteq \mathfrak{M}$ in order to introduce the topology $\tau_0 = \tau(\mathcal{L}_0, \mathfrak{M}_0)$ for $\mathcal{B}(\mathcal{L}, \mathfrak{M})$.

Theorem 2.7 *(Hajek-LeCam bound) Suppose that $E_n \Rightarrow E$ and assume that $\beta_n \in \mathcal{B}(\mathcal{L}, \mathfrak{M}_n)$ is a sequence of generalized decision functions. If $\mathfrak{M}_n \in \mathcal{M}(E_n)$, $\mathfrak{M} \in \mathcal{M}(E)$, $\Theta_0 \subseteq \Theta$, and $\mathcal{H} = \mathcal{H}(\{\beta_n, n \in I\!N\})$ is the closed convex hull of the accumulation points of $\{\beta_n\}$ w.r.t. τ_0, then*

$$\liminf_{n\to\infty} \sup_{\vartheta\in\Theta_0} \beta_n(W_\vartheta, P_{n,\vartheta}) \ge \inf_{\beta\in\mathcal{H}} \sup_{\vartheta\in\Theta_0} \beta(W_\vartheta, P_\vartheta) \ .$$

The proof of Theorem 2.7 may be carried out in the same way as in Strasser [4], Theorem 62.5. The essential part of this proof is the existence of accumulation points which was established in Proposition 2.6. It is clear that the right-hand side of the Hajek-LeCam bound depends on the choice of $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathfrak{M}_0 \subseteq \mathfrak{M}$.

If $\mathcal{L}$ is large enough, the definition of generalized decision functions may be extended to measurable functions bounded from below, in the same way as was done in Strasser [4], Definition 43.1. Then the result of Theorem 2.7 is valid for W_ϑ bounded from below.

3 Proofs of the auxiliary results

Proof of Proposition 2.1: Let $|\mathcal{J}|$ be the number of elements of $\mathcal{J}$ and put

$$\mu = \frac{1}{|\mathcal{J}|} \sum_{\vartheta\in\mathcal{J}} P_\vartheta \ .$$

As $\mathcal{P} = (P_\vartheta)_{\vartheta\in\mathcal{J}}$ and μ are equivalent, we obtain from the Halmos-Savage-Theorem that there exist versions g_ϑ of the densities $\frac{dP_\vartheta}{d\mu}$ which are $\sigma(T)$-measurable. Consequently, there exist functions $h_\vartheta : R \to I\!R_1$ being $\mathfrak{R}$-$\mathfrak{B}_1$ measurable so that $g_\vartheta = h_\vartheta(T)$. The σ-algebra $\mathfrak{R}_h = \sigma(h_\vartheta, \vartheta \in \mathcal{J})$ is countably generated. Hence there are finite algebras $\mathfrak{R}_1 \subseteq \mathfrak{R}_2 \subseteq \cdots$ with $\mathfrak{R}_h = \sigma(\bigcup_{n=1}^{\infty} \mathfrak{R}_n)$. But then $\mathfrak{F}_n = T^{-1}(\mathfrak{R}_n)$ is an increasing sequence of algebras with

$$\sigma(\bigcup_{n=1}^{\infty} \mathfrak{F}_n) = T^{-1}(\mathfrak{R}_h) \ .$$

Put $g_{n,\vartheta} = I\!E_\mu(g_\vartheta|\mathfrak{F}_n)$. Then Levy's martingale convergence theorem implies

$$\lim_{n\to\infty} I\!E_\mu|\, g_{n,\vartheta} - g_\vartheta \,| = 0$$

for every $\vartheta \in \mathcal{J}$, i.e. for any $\varepsilon > 0$ there is a $n_\varepsilon \in I\!N$ such that

$$\int |g_{n,\vartheta}(x) - g_\vartheta(x)|\, \mu(dx) < \frac{\varepsilon}{4} \tag{7}$$

for every $n > n_\varepsilon$ and every $\vartheta \in \mathcal{J}$. Let I_B denote the indicator function of the set B. Denote the atoms of $\mathfrak{R}_n$ by $B_{n,i}$, $i = 1, \ldots m_n$, put $A_{n,i} = T^{-1}(B_{n,i}) \in \mathfrak{F}_n$, and introduce the stochastic kernel $L_\varepsilon : (R, \mathfrak{R}) \Rightarrow (\Omega, \mathfrak{F})$ by

$$L_\varepsilon(x, A) = \sum_{i:\mu(A_{n,i})>0} \frac{\mu(A \cap A_{n,i})}{\mu(A_{n,i})}\, I_{B_{n,i}}(x) + \tilde{\mu}(A)\, I_{N_n}(x)\ , \tag{8}$$

where $N_n = \bigcup\{B_{n,i} \,:\, \mu(A_{n,i}) = 0\}$ and $\tilde{\mu}$ is any fixed probability measure on $(\Omega, \mathfrak{F})$. Then for any $A \in \mathfrak{F}$ holds

$$\begin{aligned}
&|\int L_\varepsilon(x, A)\, (P_\vartheta \circ T^{-1})(dx) - P_\vartheta(A)| \\
&= \quad |\int L_\varepsilon(T(\omega), A)\, g_\vartheta(\omega)\, \mu(d\omega) - \int_A g_\vartheta(\omega)\, \mu(d\omega)| \\
&\leq \quad |\int L_\varepsilon(T(\omega), A)\, (g_\vartheta(\omega) - g_{n,\vartheta}(\omega))\, \mu(d\omega)| + |\int_A (g_{n,\vartheta}(\omega) - g_\vartheta(\omega))\, \mu(d\omega)| \\
&\quad + \quad |\int L_\varepsilon(T(\omega), A) g_{n,\vartheta}(\omega)\, \mu(d\omega) - \int_A g_{n,\vartheta}(\omega)\, \mu(d\omega)|\ ,
\end{aligned}$$

and by (7) and (8),

$$< 2\,\frac{\varepsilon}{4} + |\sum_{i:\mu(A_{n,i})>0} \frac{\mu(A \cap A_{n,i})}{\mu(A_{n,i})} \int_{A_{n,i}} g_{n,\vartheta}(\omega)\mu(d\omega) - \int_A g_{n,\vartheta}(\omega)\mu(d\omega)|\ .$$

As $g_{n,\vartheta}$ is $\mathfrak{F}_n$-measurable, this function is constant on every $A_{n,i}$ which shows that the second term on the right side is zero. Hence $\|L_\varepsilon\, P_\vartheta \circ T^{-1} - P_\vartheta\|_v < \varepsilon$, $\vartheta \in \mathcal{J}$. □

Proof of Proposition 2.2: Let $\varepsilon > 0$ be given and choose $g_\vartheta \in I\!B_0(\mathfrak{A})$ with

$$\|f_\vartheta - g_\vartheta\|_u < \frac{\varepsilon}{6\sigma(\mathcal{X})} \tag{9}$$

for every $\vartheta \in \mathcal{J}$. Denote by $\mathfrak{A}_\mathcal{J} \subseteq \mathfrak{A}$ the smallest algebra w.r.t. which all g_ϑ, $\vartheta \in \mathcal{J}$, are measurable. As $g_\vartheta \in I\!B_0(\mathfrak{A})$ takes on only a finite number of different values, the algebra $\mathfrak{A}_\mathcal{J}$ is finite. The atoms of $\mathfrak{A}_\mathcal{J}$ are denoted by $A_1, \ldots, A_N$. Now we choose n_ε so that

$$\sum_{i=1}^{N} |\sigma_n(A_i) - \sigma(A_i)| < \frac{\varepsilon/6}{\max\limits_{\vartheta\in\mathcal{J}} \|f_\vartheta\|_u + \frac{\varepsilon}{6\sigma(\mathcal{X})}} \tag{10}$$

for every $n \geq n_\varepsilon$. The stochastic kernel $K_{n,\varepsilon} : (\mathcal{X}, \mathfrak{B}) \Rightarrow (\mathcal{X}, \mathfrak{B})$ is introduced by

$$K_{n,\varepsilon}(x, B) = \sum_{i:\sigma_n(A_i)>0} \frac{\sigma_n(B \cap A_i)}{\sigma_n(A_i)} I_{A_i}(x) + \tilde{\mu}(B)\, I_{N_n}(x)\ ,$$

where $N_n = \bigcup\{A_i : \sigma_n(A_i) = 0\}$ and $\tilde{\mu}$ is any fixed probability measure on $(\mathcal{X}, \mathfrak{B})$. Note that $K_{n,\varepsilon}\, \sigma_n = \sigma_n$. Since g_ϑ is $\mathfrak{A}_{\mathcal{J}}$-measurable, it can be written in the form $g_\vartheta = \sum_{l=1}^{N} c_{l,\vartheta} I_{A_l}$. It follows that

$$
\begin{aligned}
&|(K_{n,\varepsilon} Q_\vartheta)(B) - Q_{n,\vartheta}(B)| \\
&\quad = \; |\int K_{n,\varepsilon}(x,B)\, f_\vartheta(x)\, \sigma(dx) - \int_B f_\vartheta(x)\, \sigma_n(dx)| \\
&\quad \leq \; \int |f_\vartheta(x) - g_\vartheta(x)|\, K_{n,\varepsilon}(x,B) \sigma(dx) + \int_B |f_\vartheta(x) - g_\vartheta(x)|\, \sigma_n(dx) \\
&\qquad + \Big| \int \sum_{i:\sigma_n(A_i)>0} \sum_{l=1}^{N} \frac{\sigma_n(B \cap A_i)}{\sigma_n(A_i)} I_{A_i}(x) I_{A_l}(x)\, c_{l,\vartheta}\, \sigma(dx) - \sum_{l=1}^{N} c_{l,\vartheta}\, \sigma_n(A_l \cap B) \Big| \\
&\quad \leq \; \|f_\vartheta - g_\vartheta\|_u \sigma_n(\mathcal{X}) + \|f_\vartheta - g_\vartheta\|_u \sigma(\mathcal{X}) + \|g_\vartheta\|_u \sum_{i=1}^{N} |\sigma_n(A_i) - \sigma(A_i)| < \frac{\varepsilon}{2}
\end{aligned}
$$

by (9) and (10) for every $n \geq n_\varepsilon$. $\square$

Proof of Proposition 2.5: Let $\mathcal{K}$ be a finite subset of $\mathfrak{M}$. Then by assumption **(M2)**

$$\mu_{\mathcal{K}} = \sum_{\mu \in \mathcal{K}} |\mu| \; \in \mathfrak{M} \; .$$

Any $\mu \in ca(\Omega, \mathfrak{F})$ can be decomposed into $\mu^a_{\mathcal{K}}$ being absolutely continuous w.r.t. $\mu_{\mathcal{K}}$ and $\mu^s_{\mathcal{K}}$ being singular w.r.t. $\mu_{\mathcal{K}}$:

$$\mu = \mu^a_{\mathcal{K}} + \mu^s_{\mathcal{K}} \; .$$

Let $\nu_0 \in ca(\Omega, \mathfrak{F})$ be any fixed probability measure. For any $\mu \in ca(\Omega, \mathfrak{F})$ we define by

$$\beta_{\mathcal{K}}(f, \mu) = \beta(f, \mu^a_{\mathcal{J}}) + \mu^s_{\mathcal{K}}(\Omega) \beta(f, \nu_0)$$

a bilinear functional on $\mathcal{L} \times ca(\Omega, \mathfrak{F})$. It is easy to show that $\beta_{\mathcal{K}} \in \mathcal{B}(\mathcal{L}, ca(\Omega, \mathfrak{F}))$. Moreover, if $\mu \in \mathcal{K}$, then follows $\mu^a_{\mathcal{K}} = \mu$ and $\mu^s_{\mathcal{K}} = 0$. Hence

$$\beta_{\mathcal{K}}(f, \mu) = \beta(f, \mu) \quad \forall\, f \in \mathcal{L} \; , \; \forall \mu \in \mathcal{K} \; .$$

Since $\mathcal{B}(\mathcal{J}, ca(\Omega, \mathfrak{F}))$ is compact w.r.t. the weak topology $\tau(\mathcal{L}_0, \mathfrak{M}_0)$ for any fixed subspaces $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathfrak{M}_0 \subseteq \mathfrak{M}$, there exists an accumulation point $\bar{\beta} \in \mathcal{B}(\mathcal{L}, ca(\Omega, \mathfrak{F}))$ of the net $(\beta_{\mathcal{K}})_{\mathcal{K} \in A(ca(\Omega, \mathfrak{F}))}$, where $A(ca(\Omega, \mathfrak{F}))$ is the family of all finite subsets of $ca(\Omega, \mathfrak{F})$, which is directed by inclusion. By defintion, for any $\varepsilon > 0$, any $f \in \mathcal{L}$, and any $\mu \in \mathfrak{M}$ there exists a set $\mathcal{K}_\varepsilon \in A(ca(\Omega, \mathfrak{F}))$ that contains μ so that

$$|\bar{\beta}(f, \mu) - \beta(f, \mu)| = |\bar{\beta}(f, \mu) - \beta_{\mathcal{K}}(f, \mu)| < \varepsilon$$

for every $K \supseteq K_\varepsilon$. That completes the proof. $\square$

References

[1] LeCam, L. (1986): *Asymptotic Methods in Statistical Decision Theory.* Springer, Berlin, Heidelberg, New York.

[2] Neuhaus, G. (1989): *Einige Kapitel der finiten und asymptotischen Entscheidungstheorie von LeCam.* Hamburg.

[3] Rüschendorf, L. (1987): *Asymptotische Statistik.* Skripten zur Mathematischen Statistik Nr. 13, Münster.

[4] Strasser, H. (1985): *Mathematical Theory of Statistics.* de Gruyter, Berlin.