KYBERNETIKA — VOLUME 33 (1997), NUMBER 5, PAGES 477–487

A NOTE ON THE HÁJEK–LECAM BOUND

FRIEDRICH LIESE AND INGO STEINKE

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 Hájek–LeCam bound which is an asymptotic lower bound for the maximum risk of a sequence of decisions. To give a simplified proof for the Hájek-LeCam bound we use the concept of approximate Blackwell–sufficiency.

1. INTRODUCTION

The lower Hájek–LeCam bound is a central statement of asymptotic decision theory. The traditional way (see Strasser [7], LeCam [2]) 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 Hájek–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 Hájek–LeCam bound for the maximum risk of a sequence of decisions.

This way to get the Hájek–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 Hájek–LeCam bound is desirable.

Such simplified and direct approach is due to Millar [4] and LeCam [2] provided the infimum is taken over all decisions in the Hájek–LeCam bound. Sometimes the optimal decision for the limit experiment may be found only in restricted classes of decisions. The approach given by Strasser also provides lower bounds in this situation. Millar pointed out that his simplified approach is not applicable to the more general situation of restricted classes of decisions.

The aim of the paper is to simplify Strasser's approach to the Hájek–LeCam bound and to avoid topological assumptions on the space of decisions.

2. APPROXIMATIONS OF FINITE EXPERIMENTS

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

Let $\|\mu - \nu\| = 2 \sup_{A \in F} |\mu(A) - \nu(A)|$ be the variational distance of two probability measures μ, ν on $(\Omega, \mathcal{F})$. Given a further measurable space $(R, \mathcal{R})$ and a stochastic kernel $K : (\Omega, \mathcal{F}) \Rightarrow (R, \mathcal{R})$, we denote by Kf the application of K to a bounded $\mathcal{R}$-measurable function f and by $K\mu$ the application to a distribution μ on $(\Omega, \mathcal{F})$. Note that Kf is a bounded function on $(\Omega, \mathcal{F})$ and $K\mu$ is a distribution function on $(R, \mathcal{R})$.

Denote by

$$H(P, Q) = \left[\int (p^{\frac{1}{2}} - q^{\frac{1}{2}})^2 \, \mathrm{d}\mu \right]^{\frac{1}{2}}$$

the Hellinger distance of two distributions where μ is a dominating σ-finite measure and p, q are the corresponding densities. The following inequality is well-known (see Strasser [7])

$$H(P, Q) \leq \|P - Q\|^{\frac{1}{2}} . \tag{1}$$

Furthermore by Csiszár's inequality

$$H(KP, KQ) \leq H(P, Q) \tag{2}$$

for every stochastic kernel where equality holds iff the kernel K is sufficient for $\{P, Q\}$. For details we refer to Csiszár [1] and Liese, Vajda [3].

If $T : (\Omega, \mathcal{F}) \to (R, \mathcal{R})$ is a sufficient statistic for $(P_\vartheta)_{\vartheta \in \Theta}$ and $(R, \mathcal{R})$ is a Borel space then there is a stochastic kernel $L : (R, \mathcal{R}) \Rightarrow (\Omega, \mathcal{F})$ such that $P_\vartheta = L(P_\vartheta \circ T^{-1})$, i.e. T is Blackwell-sufficient.

Without any conditions on the measurable space $(R, \mathcal{R})$ regular conditional distributions do not exist so that not every sufficient statistic is also Blackwell-sufficient. This leads to the following definition.

Definition 2.1. Let $(\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$ be an experiment and $T : (\Omega, \mathcal{F}) \to (R, \mathcal{R})$ a statistic. T is called *approximate Blackwell-sufficient* if for every $\varepsilon > 0$ there is a stochastic kernel $L_\varepsilon : (R, \mathcal{R}) \Rightarrow (\Omega, \mathcal{F})$ such that

$$\|P_\vartheta - L_\varepsilon (P_\vartheta \circ T^{-1})\| < \varepsilon \quad \text{for every} \quad \vartheta \in \Theta .$$

Let $A(\Theta)$ be the system of all finite subsets of Θ.

Theorem 2.2. A statistic $T : (\Omega, \mathcal{F}) \to (R, \mathcal{R})$ is sufficient for the experiment $(\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$ iff T is approximate Blackwell-sufficient for every finite subexperiment $(\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \mathcal{J})$, $\mathcal{J} \in A(\Theta)$.

Proof. Assume T is approximate Blackwell-sufficient for every pure subexperiment, fix $\vartheta_1, \vartheta_2 \in \Theta$ and set $\mathcal{J} = \{\vartheta_1, \vartheta_2\}$. Then by Csiszár's inequality (2)

$$H(P_{\vartheta_1} \circ T^{-1}, P_{\vartheta_2} \circ T^{-1}) \leq H(P_{\vartheta_1}, P_{\vartheta_2}). \tag{3}$$

For every $\varepsilon > 0$ there is a stochastic kernel $L_\varepsilon : (R, \mathcal{R}) \Rightarrow (\Omega, \mathcal{F})$ such that

$$\|P_{\vartheta_i} - L_\varepsilon(P_{\vartheta_i} \circ T^{-1})\| < \varepsilon, \quad i = 1, 2.$$

Consequently by (1) and (2)

$$\begin{aligned} H(P_{\vartheta_1}, P_{\vartheta_2}) &\leq H(P_{\vartheta_1}, L_\varepsilon(P_{\vartheta_1} \circ T^{-1})) + H(L_\varepsilon(P_{\vartheta_1} \circ T^{-1}), L_\varepsilon(P_{\vartheta_2} \circ T^{-1})) \\ &\quad + H(L_\varepsilon(P_{\vartheta_2} \circ T^{-1}), P_{\vartheta_2})) \\ &\leq 2\sqrt{\varepsilon} + H(P_{\vartheta_1} \circ T^{-1}, P_{\vartheta_2} \circ T^{-1}). \end{aligned}$$

As $\varepsilon > 0$ was arbitrary we obtain in conjunction with (3)

$$H(P_{\vartheta_1}, P_{\vartheta_2}) = H(P_{\vartheta_1} \circ T^{-1}, P_{\vartheta_2} \circ T^{-1}).$$

Consequently T is sufficient for $\{P_{\vartheta_1}, P_{\vartheta_2}\}$. As $\{\vartheta_1, \vartheta_2\}$ was arbitrarily chosen we see that T is sufficient for $(P_\vartheta)_{\vartheta \in \Theta}$.

Assume now that T is sufficient and fix $\mathcal{J} \in A(\Theta)$. 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 $\frac{\mathrm{d}P_\vartheta}{\mathrm{d}\mu}$ which are $\sigma(T)$-measurable. Consequently, there exist functions $h_\vartheta : R \to \mathbb{R}_1$ being $\mathcal{R}$-$\mathcal{B}_1$ measurable so that $g_\vartheta = h_\vartheta(T)$. The σ-algebra $\mathcal{R}_h = \sigma(h_\vartheta, \vartheta \in \mathcal{J})$ is countably generated. Hence there are finite algebras $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \cdots$ with $\mathcal{R}_h = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{R}_n\right)$. But then $\mathcal{F}_n = T^{-1}(\mathcal{R}_n)$ is an increasing sequence of algebras with

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

Put $g_{n,\vartheta} = \mathbb{E}_\mu(g_\vartheta | \mathcal{F}_n)$. Then Levy's martingale convergence theorem implies

$$\lim_{n \to \infty} \mathbb{E}_\mu | g_{n,\vartheta} - g_\vartheta | = 0$$

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

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

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 $\mathcal{R}_n$ by $B_{n,i}$, $i = 1, \dots m_n$, put $A_{n,i} = T^{-1}(B_{n,i}) \in \mathcal{F}_n$, and introduce the stochastic kernel $L_\varepsilon : (R, \mathcal{R}) \Rightarrow (\Omega, \mathcal{F})$ by

$$L_\varepsilon(x, A) = \sum_{i:\mu(A_{n,i})>0} \frac{\mu(A \cap A_{n,i})}{\mu(A_{n,i})} I_{B_{n,i}}(x) + \widetilde{\mu}(A) \, I_{N_n}(x) , \tag{5}$$

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

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

and by (4) and (5),

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

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

Let $\mathcal{X}$ be a compact metric space and $\mathcal{B}$ be the σ-algebra of Borel sets. Suppose $f_\vartheta : \mathcal{X} \to [0, \infty)$, $\vartheta \in \mathcal{J}$, is a family of bounded and continuous functions. Denote by $\mathcal{S}$ the set of all probability measures σ on $(\mathcal{X}, \mathcal{B})$ such that

$$\int f_\vartheta \mathrm{d}\sigma = 1 \quad \text{for every } \vartheta \in \mathcal{J} .$$

For $\sigma \in \mathcal{S}$ we set

$$Q_{\sigma,\vartheta}(B) = \int_B f_\vartheta \, \mathrm{d}\sigma .$$

The symbol $\Rightarrow$ will be used for weak convergence of distributions.

Theorem 2.3. Assume $|\mathcal{J}| < \infty$ and $\sigma, \sigma_n \in \mathcal{S}$. If $\sigma_n \Rightarrow \sigma$ as $n \to \infty$, then for every $\varepsilon > 0$ there exist n_ε and stochastic kernels $K_{n,\varepsilon} : (\mathcal{X}, \mathcal{B}) \Rightarrow (\mathcal{X}, \mathcal{B})$ such that

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

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

Proof. Let ϱ be the metric in $\mathcal{X}$ and denote by $S_r(x) = \{y : \varrho(y,x) < r\}$ the open sphere with centre $x \in \mathcal{X}$ and radius r. For every fixed $x \in \mathcal{X}$ the set

$$\{r : \sigma(\{y \,:\, \varrho(y,x) = r\}) > 0\}$$

is at most countable. Consequently for every $\delta > 0$ any compact set K may be covered by a finite number of spheres with diameter not exceeding δ and being σ-continuity sets. As the system of σ-continuity sets is an algebra we see that the compact set $\mathcal{X}$ may be covered by a finite number $A_1, \ldots, A_N$ of disjoint σ-continuity sets with diameter not exceeding δ. As the f_ϑ are uniformly continuous on $\mathcal{X}$ for every $\varepsilon > 0$ we find an $\delta_\varepsilon > 0$ such that $\varrho(x,y) < \delta_\varepsilon$ implies $|f_\vartheta(x) - f_\vartheta(y)| < \varepsilon$.

Choose $x_i \in A_i$ and set

$$g_\vartheta(x) = \sum_{i=1}^{N} f_\vartheta(x_i)\, I_{A_i}(x)\,.$$

The algebra $\mathcal{A}$ generated by $A_1, \ldots, A_N$ is finite. Consequently there exist stochastic kernels $K_{n,\varepsilon}, K_\varepsilon$: $(\mathcal{X}, \mathcal{A}) \Rightarrow (\mathcal{X}, \mathcal{B})$ such that

$$\begin{aligned} \int_A K_{n,\varepsilon}(x,B)\,\sigma_n(\mathrm{d}x) &= \sigma_n(A \cap B) \\ \int_A K_\varepsilon(x,B)\,\sigma(\mathrm{d}x) &= \sigma(A \cap B) \end{aligned}$$

for every $A \in \mathcal{A}$, $B \in \mathcal{B}$. Then $K_{n,\varepsilon} g = g \quad \sigma_n$-a.s. and $K_\varepsilon g = g \quad \sigma$-a.s. for every $\mathcal{A}$-measurable g. Hence

$$\begin{aligned} |(K_{n,\varepsilon}\, Q_{\sigma,\vartheta})(B) - Q_{\sigma_n,\vartheta}(B)| &= \left| \int_B (K_{n,\varepsilon} f_\vartheta)(x)\,\sigma(\mathrm{d}x) - \int_B f_\vartheta(x)\,\sigma_n(\mathrm{d}x) \right| \\ \leq\;\; 2\varepsilon + \left| \int_B (K_{n,\varepsilon} g_\vartheta)(x)\,\sigma(\mathrm{d}x) - \int_B (K_{n,\varepsilon} g_\vartheta)(x)\,\sigma_n(\mathrm{d}x) \right| \\ \leq\;\; 2\varepsilon + 2 \sup_{x \in \mathcal{X}} |g_\vartheta(x)| \sup_{A \in \mathcal{A}} |\sigma(A) - \sigma_n(A)| \end{aligned}$$

as $K_{n,\varepsilon} g_\vartheta$ is $\mathcal{A}$-measurable. Since $\mathcal{A}$ is finite and $\sigma(\partial A) = 0$ for every $A \in \mathcal{A}$ we obtain $\sup_{A \in \mathcal{A}} |\sigma(A) - \sigma_n(A)| \underset{n \to \infty}{\longrightarrow} 0$ which proves the statement for compact $\mathcal{X}$. □

3. THE HÁJEK–LE CAM BOUND

Given the experiment $E = (\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$ and $\mathcal{J} \in A(\Theta)$ we set $\nu_\mathcal{J} = \frac{1}{|\mathcal{J}|} \sum_{\vartheta \in \mathcal{J}} P_\vartheta$. Denote by $\Sigma_\mathcal{J}$ the set of all mappings $x_\vartheta : \mathcal{J} \to [0, \infty)$, $\mathcal{J}$ with $\sum_{\vartheta \in \mathcal{J}} x_\vartheta \leq |\mathcal{J}|$. $\Sigma_\mathcal{J}$ is a compact metric space, its σ-algebra of Borel sets is denoted by $\mathcal{B}_\mathcal{J}$. Introduce the $\mathcal{F}$-$\mathcal{B}_\mathcal{J}$-measurable mapping $T_\mathcal{J}$ by

$$(T_\mathcal{J}(\omega))_\vartheta = \frac{\mathrm{d}P_\vartheta}{\mathrm{d}\nu_\mathcal{J}}(\omega), \quad \vartheta \in \mathcal{J}. \tag{6}$$

The statistic $T_\mathcal{J}$ is sufficient for P_ϑ, $\vartheta \in \mathcal{J}$.

The notion of weak convergence of experiments is basic in asymptotic decision theory. Let $E = (\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$, $E_n = (\Omega_n, \mathcal{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$ 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 $\sigma_{n,\mathcal{J}} = \nu_{n,\mathcal{J}} \circ T_{n,\mathcal{J}}^{-1}$, $\sigma_\mathcal{J} = \nu_\mathcal{J} \circ T_\mathcal{J}^{-1}$ are the corresponding standard measures.

Now we are able to establish the randomisation criterion for weakly convergent experiments.

Theorem 3.1. Let $E_n = (\Omega_n, \mathcal{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$, $E = (\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$ 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 $K_{n,\varepsilon,\mathcal{J}} : (\Omega, \mathcal{F}) \Rightarrow (\Omega_n, \mathcal{F}_n)$ so that

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

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

Proof. The proof is divided into several steps.
Let $T_{n,\mathcal{J}}$ and $T_\mathcal{J}$ be defined as in (6).

1. Denote by $\pi_{\mathcal{J},\vartheta}$ the projection of $\Sigma_\mathcal{J}$ into the ϑth coordinate. $\pi_{\mathcal{J},\vartheta}$ is a bounded continuous function and we have

$$\frac{\mathrm{d}(P_{n,\vartheta} \circ T_{n,\mathcal{J}}^{-1})}{\mathrm{d}\sigma_{n,\mathcal{J}}} = \pi_{\mathcal{J},\vartheta}, \quad \frac{\mathrm{d}(P_\vartheta \circ T_\mathcal{J}^{-1})}{\mathrm{d}\sigma_\mathcal{J}} = \pi_{\mathcal{J},\vartheta}.$$

Hence with $f_\vartheta = \pi_{\mathcal{J},\vartheta}$ and the notations in Theorem 2.3:

$$P_{n,\vartheta} \circ T_{n,\mathcal{J}}^{-1} = Q_{\sigma_{n,\mathcal{J}},\vartheta}, \quad P_\vartheta \circ T_\mathcal{J}^{-1} = Q_{\sigma_\mathcal{J},\vartheta}.$$

From Theorem 2.3 we obtain the existence of a stochastic kernel $K_{n,\varepsilon} : (\Sigma_\mathcal{J}, \mathcal{B}_\mathcal{J}) \Rightarrow (\Sigma_\mathcal{J}, \mathcal{B}_\mathcal{J})$ such that

$$||Q_{\sigma_{n,\mathcal{J}},\vartheta} - K_{n,\varepsilon} Q_{\sigma_\mathcal{J},\vartheta}|| < \frac{\varepsilon}{2}$$

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

2. Apply Theorem 2.2 to $T = T_{n,\mathcal{J}}$, $(R, \mathcal{R}) = (\Sigma_{\mathcal{J}}, \mathcal{B}_{\mathcal{J}})$ and $Q_{\sigma_n,\mathcal{J},\vartheta} = P_{n,\vartheta} \circ T_{n,\mathcal{J}}^{-1}$: There exist stochastic kernels $L_{n,\varepsilon} : (\Sigma_{\mathcal{J}}, \mathcal{B}_{\mathcal{J}}) \Rightarrow (\Omega_n, \mathcal{F}_n)$ so that

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

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

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

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

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

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

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

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

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

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

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 $\nu \ll \mu$ imply $\mu^\perp \subseteq \nu^\perp$, it is easy to see that the L-space of E fulfils (M1) and (M2).

Assume $\mathbb{D}$ is a given decision space equipped with a σ-algebra $\mathcal{D}$. Let $\mathcal{L}$ be a linear subspace of $\mathbb{B}(\mathcal{D})$, the space of all bounded, $\mathcal{D}$-measurable, real-valued functions, and $\mathcal{M} \in \mathcal{M}(E)$. Let $\|f\|$ denote the supremum norm of f. By $\mathcal{B}(\mathcal{L}, \mathcal{M})$ we shall denote the set of all *generalized decision functions*, i.e. the set of bilinear functions $\beta : \mathcal{L} \times \mathcal{M} \to \mathbb{R}_1$ which are supposed to have the following properties:

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

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

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

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

$$\beta_K(f, \mu) = \int \int f(x)\, K(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega)$$

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

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

Proposition 3.2. $\mathcal{B}(\mathcal{L}, \mathcal{M})$ is convex and compact w.r.t. the topology $\tau(\mathcal{L}_0, \mathcal{M}_0)$.

If $\mathcal{L}_0 = \mathcal{L}$ and $\mathcal{M}_0 = \mathcal{M}$, then the statement of Proposition 3.2 follows with the same arguments as in the proof of Strasser [7], Theorem 42.3. To establish Proposition 3.2 we have only to remark that $\mathcal{B}(\mathcal{L}, \mathcal{M})$ being compact w.r.t. to $\tau(\mathcal{L}, \mathcal{M})$ is again compact w.r.t. the weaker topology $\tau(\mathcal{L}_0, \mathcal{M}_0)$.

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

Proposition 3.3. Assume that $\mathcal{M} \in \mathcal{M}(E)$. Then for any $\beta \in \mathcal{B}(\mathcal{L}, \mathcal{M})$ there exists a $\bar{\beta} \in \mathcal{B}(\mathcal{L}, ca(\Omega, \mathcal{F}))$ such that

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

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

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

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

$$\mu = \mu_{\mathcal{K}}^a + \mu_{\mathcal{K}}^s .$$

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

$$\beta_{\mathcal{K}}(f, \mu) = \beta(f, \mu_{\mathcal{J}}^a) + \mu_{\mathcal{K}}^s(\Omega)\beta(f, \nu_0)$$

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

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

Since $\mathcal{B}(\mathcal{L}, ca(\Omega, \mathcal{F}))$ is compact w.r.t. the weak topology $\tau(\mathcal{L}_0, \mathcal{M}_0)$ for any fixed subspaces $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathcal{M}_0 \subseteq \mathcal{M}$, there exists an accumulation point $\bar{\beta} \in \mathcal{B}(\mathcal{L}, ca(\Omega, \mathcal{F}))$ of the net $(\beta_{\mathcal{K}})_{\mathcal{K} \in A(ca(\Omega, \mathcal{F}))}$, where $A(ca(\Omega, \mathcal{F}))$ is the family of all finite subsets of

$ca(\Omega, \mathcal{F})$, which is directed by inclusion. By definition, for any $\varepsilon > 0$, any $f \in \mathcal{L}$, and any $\mu \in \mathcal{M}$ there exists a set $\mathcal{K}_\varepsilon \in A(ca(\Omega, \mathcal{F}))$ 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. □

Suppose now that $E_n = (\Omega_n, \mathcal{F}_n, P_{n,\vartheta}, \vartheta \in \Theta)$ is a sequence of experiments which weakly converges to the experiment $E = (\Omega, \mathcal{F}, P_\vartheta, \vartheta \in \Theta)$. Let $\mathcal{M} \in \mathcal{M}(E)$, $\mathcal{M}_n \in \mathcal{M}(E_n)$, and let $\beta_n \in \mathcal{B}(\mathcal{L}, \mathcal{M}_n)$ be any sequence of generalized decision functions. $\beta \in \mathcal{B}(\mathcal{L}, \mathcal{M})$ 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 Hájek–LeCam bound for weakly convergent sequences of experiments.

Proposition 3.4. Assume that $E_n \Rightarrow E$, $\mathcal{M} \in \mathcal{M}(E)$, and $\mathcal{M}_n \in \mathcal{M}(E_n)$. Then every sequence $\{\beta_n\}$, $\beta_n \in \mathcal{B}(\mathcal{L}, \mathcal{M})$, $n \in I\!N$, has an accumulation point.

Proof. (based on Strasser [7], Theorem 62.3) Let $\delta > 0$. According to Theorem 3.1, for any $\mathcal{J} \in A(\Theta)$ there exists a stochastic kernel $K_{n,\delta,\mathcal{J}} : (\Omega, \mathcal{F}) \Rightarrow (\Omega_n, \mathcal{F}_n)$ so that

$$\|P_{n,\vartheta} - K_{n,\delta,\mathcal{J}} P_\vartheta\| < \delta, \quad \forall \vartheta \in \mathcal{J}, \tag{7}$$

for every sufficiently large n.

Denote by $\bar{\beta}_n \in \mathcal{B}(\mathcal{L}, ca(\Omega_n, \mathcal{F}_n))$ an extension of $\beta_n \in \mathcal{B}(\mathcal{L}, \mathcal{M}_n)$ according to Proposition 3.3 and introduce $\beta_{n,\mathcal{J}} \in \mathcal{B}(\mathcal{L}, \mathcal{M})$ by

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

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

It is to be shown that β is an accumulation point of $(\beta_n)_{n \in I\!N}$ 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 $n_0 \in I\!N$ there is a $n_\varepsilon \in I\!N$ so that $n_\varepsilon \geq n_0$ and

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

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

$$|\beta_{\mathcal{J}_\varepsilon}(f, P_\vartheta) - \beta(f, P_\vartheta)| < \frac{\varepsilon}{3} \tag{8}$$

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 (7) fulfils

$$\|P_{n,\vartheta} - K_{n,\delta,\mathcal{J}_\varepsilon} P_\vartheta\| < \frac{\varepsilon}{3a} \tag{9}$$

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

$$|\beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f, P_\vartheta) - \beta_{\mathcal{J}_\varepsilon}(f, P_\vartheta)| < \frac{\varepsilon}{3} \tag{10}$$

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

$$\begin{aligned} |\beta_{n_\varepsilon}(f, P_{n_\varepsilon,\vartheta}) - \beta_{n_\varepsilon,\mathcal{J}_\varepsilon}(f, P_\vartheta)| &= |\beta_{n_\varepsilon}(f, P_{n_\varepsilon,\vartheta}) - \beta_{n_\varepsilon}(f, K_{n_\varepsilon,\delta,\mathcal{J}_\varepsilon} P_\vartheta)| \\ &\leq \|f\|_u \|P_{n_\varepsilon,\vartheta} - K_{n_\varepsilon,\delta,\mathcal{J}_\varepsilon} P_\vartheta\| < \delta\, a = \frac{\varepsilon}{3}\,. \end{aligned} \tag{11}$$

The triangle inequality in conjunction with (8), (10), and (11) implies

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

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

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

Theorem 3.5. (Hájek–LeCam bound) Suppose that $E_n \Rightarrow E$ and assume that $\beta_n \in \mathcal{B}(\mathcal{L}, \mathcal{M}_n)$ is a sequence of generalized decision functions. If $\mathcal{M}_n \in \mathcal{M}(E_n)$, $\mathcal{M} \in \mathcal{M}(E)$, $\Theta_0 \subseteq \Theta$, and $\mathcal{H} = \mathcal{H}(\{\beta_n, n \in \mathbb{N}\})$ is the closed convex hull of the accumulation points of $\{\beta_n\}$ w.r.t. τ_0, then

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

The proof of Theorem 3.5 may be carried out in the same way as in Strasser [7], Theorem 62.5. The essential part of this proof is the existence of accumulation points which was established in Proposition 3.4. It is clear that the right-hand side of the Hájek–LeCam bound depends on the choice of $\mathcal{L}_0 \subseteq \mathcal{L}$ and $\mathcal{M}_0 \subseteq \mathcal{M}$ because $\mathcal{L}_0$ and $\mathcal{M}_0$ have influence on the structure of the topology $\tau(\mathcal{L}_0, \mathcal{M}_0)$ which determines the set $\mathcal{H}$.

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 [7], Definition 43.1. Then the result of Theorem 3.5 is valid for W_ϑ bounded from below.

(Received April 26, 1996.)

REFERENCES

[1] I. Csiszár: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität Markoffscher Ketten. Publ. Math. Inst. Hungar. Acad. Sci., Ser. A, *8* (1963), 85–108.
[2] L. LeCam: Asymptotic Methods in Statistical Decision Theory. Springer–Verlag, Berlin – Heidelberg – New York 1986.
[3] F. Liese and I. Vajda: Convex Statistical Distances. Teubner Verlag, Leipzig 1987.
[4] P. W. Millar: The minimax principle in asymptotic statistical theory. In: Lecture Notes in Mathematics 974, pp. 75–265. Springer–Verlag, Berlin – Heidelberg – New York 1983.
[5] G. Neuhaus: Einige Kapitel der finiten und asymptotischen Entscheidungstheorie von LeCam. Hamburg 1989.
[6] L. Rüschendorf: Asymptotische Statistik. Skripten zur Mathematischen Statistik Nr. 13, Münster 1987.
[7] H. Strasser: Mathematical Theory of Statistics. de Gruyter, Berlin 1985.

Prof. Dr. Fiedrich Liese and Ingo Steinke, University of Rostock, Department of Mathematics, Universitätsplatz 1, D-18 051 Rostock. Germany.