On Insufficiency *

Espen Norberg
University of Oslo.

Abstract

The concept of insufficiency, introduced and developed by Le Cam, is a measure for the loss of information one suffers when considering a subexperiment of $\mathcal{E}$ instead of the experiment $\mathcal{E}$ itself. Originally, the insufficiency was defined within a lattice theoretic setting. A main result of this paper (Theorem 6), tells that the insufficiency may be obtained from $\sigma(M, L)$-continuous non-negative linear projections. This makes it possible to express the insufficiency on a measure theoretic form when considering traditional measure experiments. We find several expressions for the insufficiency. In particular, when considering dichotomies we may express the insufficiency on a very simple form, which in many situations makes it possible to find insufficiencies by direct computation.

1 Introduction

When working within the framework of a parameterized statistical model, a statistician is usually concerned with the problem of finding the "true" parameter. In doing so, he considers a random variable Y from which he is able to construct estimators, confidence intervals etc.

The random variable Y may be laborious, and for this reason the statistician may wish to use a simpler random variable on the form $T(Y)$. This is clearly justified whenever $T(Y)$ provides the same amount of information as Y. That is, when T is sufficient (for the experiment induced by Y). However, the requirement that T should be sufficient seems too severe. Even when loosing information, the statistician may prefer $T(Y)$, provided the loss is small. Thus a natural question arises; how insufficient is $T(Y)$ compared with Y? A particular situation of interest is the case where $Y = (X_1, \ldots, X_n)$ is a vector of random variables and $T(Y) = (X_1, \ldots, X_k)$ where $1 \leq k < n$. Now, how much information is contained in the additional observations $(X_{k+1}, \ldots, X_n)$?

*The present work is the author's thesis for the degree cand.scient. at the University of Oslo. It is dedicated with gratitude and affection to Erik Torgersen who initiated the project and supervised it until his untimely death in 1994.

In order to answer these questions Le Cam (1974, 1986) suggested three quantities η, η_1, and η_2. It is common to all of them that they are intended to measure "loss of information". Following the terminology from Le Cam (1974) the quantities η and η_1 are named the *insufficiency* and the *lack of sufficiency*, respectively. The quantity η_1 measures, in a sense, how "far" T is from being sufficient with respect to the experiment induced by Y. Meanwhile the insufficiency η is based on ideas of measuring how much the experiment induced by Y should be modified in order to make T sufficient. Thus, when η and η_1 are small this also indicates that little information is lost when considering $T(Y)$ instead of Y. Since these quantities originally were defined within a lattice theoretic setting, we will postpone the precise definitions until Section 4. Le Cam (1986, p.64) showed that one always has the inequalities

$$\eta_1 \leq \eta \leq \eta_2 \leq 2\eta_1. \tag{1}$$

One of our main results tells that $\eta = \eta_2$. This result enables us to establish a general measure theoretic definition for the insufficiency. We will also observe a situation where $\eta_1 \neq \eta$.

The main references of the paper are Le Cam (1974) and Le Cam (1986). Section 2 contains definitions and technical results from the formalism of Le Cam (1964, 1974 and 1986). Section 3 considers Le Cam's notion of sufficiency within the lattice theoretic setting. We extend Le Cam's definition of sufficiency to a larger class of experiments. Section 4 contains our main results: Theorem 6 and Theorem 9. We give the formal definitions of the lack of sufficiency and the insufficiency. We also prove that η (Definition 8(i)) is the same as η_2 (Definition 8(ii)). In Section 5 we list some properties of the insufficiency. The main reference of Section 6 is Torgersen (1991). Efforts are made to show the connection between the lattice theoretic and measure theoretic definitions of insufficiency – no regularity conditions are required. Results from Section 4 are taken into measure theoretic results. In Section 7 we give some results on binary experiments (dichotomies) and examples.

2 Lattice theoretic framework

An abstract L-space (respectively, an abstract M-space) is a Banach lattice with a norm $||\cdot||$ such that $||v_1 + v_2|| = ||v_1|| + ||v_2||$ ($||v_1 \vee v_2|| = ||v_1|| \vee ||v_2||$) whenever $v_1, v_2 \geq 0$. Following the classical notation for a vector lattice V, we write $v_1 \vee v_2 = \max\{v_1, v_2\}$, $v_1 \wedge v_2 = \min\{v_1, v_2\}$, $v^+ = v \vee 0$, $v^- = (-v) \vee 0$ and $|v| = v^+ + v^-$. Two vectors v_1, v_2 are called disjoint if $|v_1| \wedge |v_2| = 0$. The order interval $[v_1, v_2]$ determined by $v_1, v_2 \in V$ is the set $\{v : v \in V, v_1 \leq v \leq v_2\}$.

A set in V is called order bounded if it is contained in some order interval. A subset W in V is called order complete if any non-empty order bounded subset of W has a supremum (infimum) in W. The vector lattice V is called order complete if it is order complete as a subset of itself. A subset W in V is called a solid if $v \in W$ whenever $|v| \leq |w|$ and $w \in W$.

A sub-vector lattice of an abstract L-space is called a band if it is an order complete solid. For any subset A of an abstract L-space, there is a unique smallest band containing A. This band is called the band generated by A, and is the intersection of all bands in the abstract L-space that contain the set A.

Throughout L will denote a certain abstract L-space. The order dual [1] of L equipped with the dual ordering and the dual norm defined by $||u|| = \sup\{u(\lambda) : \lambda \in L, ||\lambda|| \leq 1\}$ is an abstract M-space and will be noted M. Similarly, the order dual of M is an abstract L-space and will be noted M^*. The evaluation map is an isometry and lattice isomorphism from L onto a band in the abstract L-space M^*. Thus, when convenient, one may assume that L is contained in M^*. This will be done without any further warnings.

A sub-vector lattice H of M is called a uniform sublattice if H is closed for the dual norm and the unit 1 of M belongs to H. A uniform sublattice H is called complete, or a complete sublattice, if it is closed for the weak $\sigma(M, L)$ topology (the weak L-topology on M). Another interesting topology on M is the locally convex-solid topology $|\sigma|(M, L)$, i.e. the absolute weak L-topology on M. This is the topology generated by the seminorms $\{\rho_\lambda : \lambda \in L\}$ where $\rho_\lambda(u) = |\lambda|(|u|)$ for all $u \in M$. If H is a complete sublattice of M, then the following holds:

(i) H is complete relative to $|\sigma|(M, L)$;

(ii) H is order complete.

We will only sketch the proof. The reader may consult the sources for details: The topological dual of $(M, |\sigma|(M, L))$ is L (see e.g. Aliprantis/Burkinshaw 1978, Theorem 6.6, p.40). From this one gets that $|\sigma|(M, L)$-closures and $\sigma(M, L)$-closures of convex subsets of M are identical (see Schaefer 1966, Corollary 2, p.65). Therefore H is closed for the $|\sigma|(M, L)$ topology. Since M is $|\sigma|(M, L)$-complete (Aliprantis/Burkinshaw 1978, Theorem 19.7, p.128), it follows that also H is $|\sigma|(M, L)$-complete. This proves part (i). Furthermore, from (Aliprantis/Burkinshaw 1978, Theorem 19.7, p.128) it also follows that $|\sigma|(M, L)$ is a Lebesgue topology. From this it is easy to verify that H is order complete.

For an element $P \in M^*$, the value of P at $u \in M$ will be noted uP or by the natural pairing $\langle u, P\rangle$. We let uv; $u, v \in M$ denote the product for the unique multiplication from $M \times M$ into M such that $1u = u1 = u$; and $u^+v^+ \geq 0$ (see Le Cam 1964, Proposition 3). The element of M^* which has density u with respect to P will be noted $u \cdot P$, i.e. $\langle v, u \cdot P\rangle = \langle uv, P\rangle$ when $v \in M$. The same notation is used for a uniform sublattice H and its order dual H^*, e.g. $\langle u, P'\rangle$ denotes the natural pairing of H and H^* when $u \in H$, $P' \in H^*$. The restriction

[1] As shown in e.g. Torgersen 1991, p.198 the order dual of an abstract L-space (abstract M-space) coincides with the Banach space dual.

of $P \in M^*$ to a uniform sublattice H of M will be noted P^H. If L_0 is a band in M^*, then the set of restrictions of L_0 to H will be noted L_0^H.

We will let non-negative linear projections Π from M onto a complete sublattice H act on the left; the image of $u \in M$ will be denoted $u\Pi$. The adjoint of Π is an isometric lattice homomorphism from H^* into M^* and will be denoted by the same letter, now acting on its right. Thus $\langle u\Pi, P^H \rangle$, $\langle u, \Pi P^H \rangle$, and $u\Pi P^H$ denote the same number for $u \in M, P \in M^*$. Let L_0 be a band in M^* such that $L_0^H = L^H$. From Kelley/Namioka 1963 (21.1, p.199) we have that a non-negative linear projection Π from M onto H is $\sigma(M, L_0)$-continuous if and only if the adjoint of Π maps L_0^H into L_0.

Let H be a complete sublattice of M and let $\mu \in M^*$ such that $\mu \geq 0$ and $\mu^H \in L^H$. Then there is a smallest idempotent J_μ in H such that $\langle J_\mu, \mu \rangle = ||\mu||$. If ν is another positive element in M^*, $\nu^H \in L^H$, such that μ and ν are disjoint when restricted to H, then J_μ and J_ν are likewise disjoint. Let H_μ denote the vector lattice given by $H_\mu = \{vJ_\mu : v \in H\}$. From Le Cam 1986, Lemma 3, p.60, one easily verifies that there exists a unique non-negative linear projection Π from M onto H_μ such that:

(i) $1\Pi = J_\mu$;

(ii) $\langle u\Pi, \mu \rangle = \langle u, \mu \rangle$ for all $u \in M$.

This projection will be called the minimal conditional expectation operator associated with μ and H.

Le Cam (1986, Lemma 4, p.61) also has shown the existence of $\sigma(M, L)$-continuous non-negative linear projections from M onto complete sublattices. The next lemma is a variant of this result and the proof follows Le Cam's argument.

LEMMA 1 *Let H be a complete sublattice of M. If L_0 is a band in M^* such that $L_0^H = L^H$, then there is a $\sigma(M, L_0)$-continuous non-negative linear projection Π from M onto H.*

Proof: Let $\{P_\alpha : \alpha \in A\}$ be a family of elements $P_\alpha \geq 0$, $||P_\alpha|| = 1$, in the band L_0 satisfying the following two requirements: (i) the restrictions $P_\alpha^H : \alpha \in A$ are pairwise disjoint; (ii) the band generated by $\{P_\alpha^H : \alpha \in A\}$ is L_0^H. Zorn's lemma yields the existence of a family having these properties.

For each α there is a smallest idempotent $u_\alpha \in H$ such that $\langle u_\alpha, P_\alpha^H \rangle = 1$. Let Π_α be the minimal expectation operator associated with P_α and H. One easily verifies that $1\Pi_\alpha = u_\alpha$ and that $\Pi = \sum_\alpha \Pi_\alpha$ is a non-negative linear projection from M onto H. It remains to show that Π is $\sigma(M, L_0)$-continuous. Put $B = \{P^H : P^H \in L_0^H, \Pi P^H \in L_0\}$. It is readily checked that B is a band in L_0^H. Since $\Pi P_\alpha^H = P_\alpha \in L_0$ it follows that B contains the family $\{P_\alpha^H : \alpha \in A\}$. The band generated by $\{P_\alpha^H : \alpha \in A\}$ is L_0^H itself. Thus the adjoint of Π maps L_0^H into L_0. □

The next lemma may appear a bit clumsy. However, for later use we find it convenient to express it on this form.

LEMMA 2 *Let $P, Q \in M^*$ and $u \in M$, $0 \leq u \leq 1$. Then $u \cdot P = P^+$ and $u \cdot Q = Q^+$ if and only if $u \cdot (P+Q) = (P+Q)^+$ and $(P+Q)^+ = P^+ + Q^+$.*

Proof: Assume that $u \cdot (P+Q) = (P+Q)^+$ and $(P+Q)^+ = P^+ + Q^+$. Clearly $u \cdot P \leq P^+$ and $u \cdot Q \leq Q^+$. Thus $P^+ = u \cdot (P+Q) - Q^+ = u \cdot P + (u \cdot Q - Q^+) \leq u \cdot P$ and, for the same reason, we have that $Q^+ \leq u \cdot Q$. This proves the if-part.

Now, assume that $u \cdot P = P^+$ and $u \cdot Q = Q^+$. Then $u \cdot P + u \cdot Q = P^+ + Q^+ \geq (P+Q)^+ \geq u \cdot (P+Q)$. Since the left and right side are the same, equality holds throughout. Thus $u \cdot (P+Q) = (P+Q)^+$ and $P^+ + Q^+ = (P+Q)^+$. □

LEMMA 3 *Let H be a complete sublattice of M and let Π be a non-negative linear projection from M onto H. If $P \in M^*$ and $u \in H$ satisfies $u \cdot P^H = (P^H)^+$, then $u \cdot \Pi P^H = [\Pi P^H]^+$.*

Proof: Let $v \in M$ and let $u \in H$ such that $u \cdot P^H = (P^H)^+$. Then completeness of H ensures that $(uv)\Pi = (v\Pi)u$ (see Le Cam 1986, p.59, Lemma 2). From this one obtains that $\langle v, u \cdot \Pi P^H \rangle = \langle vu, \Pi P^H \rangle = \langle (uv)\Pi, P^H \rangle = \langle (v\Pi)u, P^H \rangle = \langle v\Pi, u \cdot P^H \rangle = \langle v\Pi, (P^H)^+ \rangle = \langle v, \Pi(P^H)^+ \rangle = \langle v, [\Pi P^H]^+ \rangle$. The latter equation follows since the adjoint of Π is a lattice homomorphism. It follows that $u \cdot \Pi P^H = [\Pi P^H]^+$. □

3 Sufficiency

In this section we consider Le Cam's notion of sufficiency for complete sublattices. An experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ is a map $\theta \mapsto P_\theta$ from a non-empty set Θ to an abstract L-space, say L, such that $P_\theta \geq 0$ and $||P_\theta|| = 1$. We will say that $\mathcal{E}$ is an experiment in L when $\mathcal{E}$ maps Θ into L.

DEFINITION 4 *Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment in M^* and let H be a complete sublattice of M. We call H sufficient for $\mathcal{E}$ if there is a non-negative linear projection Π from M onto H such that $\langle u\Pi, P_\theta^H \rangle = \langle u, P_\theta \rangle$ whenever $u \in M$, $\theta \in \Theta$.*

If H is sufficient we may obtain the original experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ from the sub-experiment $\mathcal{E}^H = (P_\theta^H : \theta \in \Theta)$ by a projection Π that does not depend on θ, i.e. $\mathcal{E} = \Pi\mathcal{E}^H$. The statistical interpretation of sufficiency is that there is no loss of information (on the parameter) when passing from the experiment $\mathcal{E}$ to the experiment $\mathcal{E}^H$.

A small distinction from earlier definitions of sufficiency is that H is defined to be sufficient for experiments in M^* and not only for experiments contained in L (Of course, the above definition does not exclude the possibility that $\mathcal{E}$ may be contained in L). The present definition seems to represent more adequately the statistical notion of sufficiency. In order to see this, consider an experiment on the form $\mathcal{E} = (\Pi P_\theta^H : \theta \in \Theta)$ where $P_\theta \in L$ for each $\theta \in \Theta$. It is easily seen that $\mathcal{E}$ can be obtained from its own restriction to H by the projection Π, that is, $\Pi(\Pi P_\theta^H)^H \equiv_\theta \Pi P_\theta^H$. Thus, one should expect H to be sufficient for $\mathcal{E}$. However, due to the fact that ΠP_θ^H is not necessarily contained in L one encountered technical difficulties with the earlier definitions. These difficulties are avoided by the above definition. Another interesting way to deal with these problems was suggested by Le Cam (1986) who introduced the notion *completion sufficiency.*

LEMMA 5 *Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment in M^* and let H be a complete sublattice of M. Suppose $\mathcal{E}^H$ is an experiment in L^H. Then H is sufficient for $\mathcal{E}$ if and only if for each finite sum $p = \sum a_j P_{\theta_j}$, a_j real, there is a $u \in H$, $0 \le u \le 1$, such that $u \cdot p = p^+$.*

Proof: Let $\overline{M}$ denote the dual of M^*. Then M can be identified with a $\sigma(\overline{M}, M^*)$-dense subset of $\overline{M}$. Let $\overline{H}$ be the $\sigma(\overline{M}, M^*)$-closure of H in $\overline{M}$.

The 'only if' follows upon Lemma 3. Now, assume that the latter condition holds. Then from Le Cam (1986), Property (CI), p.63, Theorem 1, p.65, and Proposition 2, p.36, it follows that there is a non-negative linear projection Π from $\overline{M}$ onto $\overline{H}$ such that $\langle u\Pi, P_\theta\rangle = \langle u, P_\theta\rangle$ whenever $u \in \overline{M}$, $\theta \in \Theta$. Let $u \in M$. Since the $|\sigma|(\overline{M}, M^*)$-closure of H coincides with $\overline{H}$ (Schaefer 1966, Corollary 2, p. 65) it follows that there is a net (u_α) in H such that $u_\alpha \longrightarrow u\Pi$ with respect to $|\sigma|(\overline{M}, M^*)$. This implies that (u_α) is a Cauchy net for the $|\sigma|(M, L)$-topology. Since H is complete relative to the $|\sigma|(M, L)$-topology there is a unique $uT \in H$ such that $u_\alpha \longrightarrow uT$ for $|\sigma|(M, L)$. Thus $u_\alpha \longrightarrow uT$ for the $\sigma(M, L)$-topology. It is readily checked that the map $u \mapsto uT$ is a non-negative linear projection from M onto H. Since $P_\theta^H \in L^H$ for each $\theta \in \Theta$ it follows that $\langle uT, P_\theta\rangle = \lim_\alpha\langle u_\alpha, P_\theta\rangle = \langle u\Pi, P_\theta\rangle = \langle u, P_\theta\rangle$ whenever $u \in M$, $\theta \in \Theta$. Thus H is sufficient for $\mathcal{E}$. □

4 Insufficiency

We start this section by establishing one main result.

THEOREM 6 *Let H be a complete sublattice of M. Let L_0 be a band in M^* such that $L_0^H = L^H$. If Π is a non-negative linear projection from M onto H, then there is a $\sigma(M, L_0)$-continuous non-negative linear projection $\overline{\Pi}$ from M onto H such that $\|Q - \overline{\Pi}P^H\| \le \|Q - \Pi P^H\|$ whenever $P, Q \in L_0$.*

Proof: Let $L_0^\perp$ denote the set of vectors in M^* which are disjoint from L_0. For each $P \in L_0$ decompose ΠP^H into $\Pi P^H = \mu_P + \nu_P$ such that $\mu_P \in L_0$ and $\nu_P \in L_0^\perp$. Put $\Theta = \{P \in L_0 : \|P\| = 1, P \ge 0\}$. Clearly H is sufficient for the experiment $(\Pi P^H : P \in \Theta)$. It is easy to see that $\nu_P^H \in L_0^H$ and hence from Lemma 1 there is a non-negative linear projection $\widehat{\Pi}$ from M onto H such that $\widehat{\Pi}\nu_P^H \in L_0$ for all $P \in \Theta$. Define the experiment $(\widehat{P} : P \in \Theta)$ by $\widehat{P} = \mu_P + \widehat{\Pi}\nu_P^H$. Let $p = \sum a_j(\Pi P_j^H)$, $\hat{p} = \sum a_j \widehat{P}_j$, $\mu = \sum a_j \mu_{P_j}$ and $\nu = \sum a_j \nu_{P_j}$ be finite sums with a_j real. Then $p = \mu + \nu$ and $\hat{p} = \mu + \widehat{\Pi}\nu^H$ and by sufficiency and Lemma 5 there is a $u \in H$, $0 \leq u \leq 1$, such that $u \cdot p = p^+$. Since μ and ν are disjoint, it follows that $(\mu + \nu)^+ = \mu^+ + \nu^+$ and hence from Lemma 2 and Lemma 3 we obtain $u \cdot \hat{p} = \hat{p}^+$. It is readily checked that $P^H = \widehat{P}^H$. Thus Lemma 5 yields that H is sufficient for $(\widehat{P} : P \in \Theta)$. From the definition of sufficiency there is a non-negative linear projection $\overline{\Pi}$ from M onto H such that $\widehat{P} = \overline{\Pi}\widehat{P}^H = \overline{\Pi}P^H$ whenever $P \in \Theta$, and it follows that the adjoint of $\overline{\Pi}$ maps L_0^H into L_0. Therefore $\overline{\Pi}$ is $\sigma(M, L_0)$-continuous. Finally, if $P, Q \in L_0$, then

$$
\begin{aligned}
||Q - \overline{\Pi}P^H|| &= ||Q - \mu_P - \widehat{\Pi}\nu_P^H|| \\
&\leq ||Q - \mu_P|| + ||\widehat{\Pi}\nu_P^H|| \\
&\leq ||Q - \mu_P|| + ||\nu_P|| && (2) \\
&= ||Q - \mu_P - \nu_P|| && (3) \\
&= ||Q - \Pi P^H||.
\end{aligned}
$$

Equation (2) holds since the adjoints of non-negative linear projections are isometric, and so $||\widehat{\Pi}\nu_P^H|| = ||\nu_P^H|| \leq ||\nu_P||$. Equation (3) holds since $(Q - \mu_P)$ and ν_P are disjoint. □

COROLLARY 7 *Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment in L and let H be a complete sublattice of M. If Π is a non-negative linear projection from M onto H, then there is a $\sigma(M, L)$-continuous non-negative linear projection $\overline{\Pi}$ from M onto H such that $||P_\theta - \overline{\Pi}P_\theta^H|| \leq ||P_\theta - \Pi P_\theta^H||$ for each $\theta \in \Theta$.*

Proof: Follows from Theorem 6 upon taking $L_0 = L$. □

The concept of insufficiency, introduced and developed by Le Cam, is a measure for loss of information. It tells, in a sense, how much one must modify the experiment in order to make H sufficient. We give two alternative definitions for the insufficiency. Le Cam (1986, p.69) put forward the problem of showing that these quantities were the same (i.e. $\eta = \eta_2$). Corollary 7 yields a solution to this problem.

DEFINITION 8 *Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment in L and let H be a complete sublattice of M. The insufficiency of H for $\mathcal{E}$ is the quantity*

$$\eta(H, \mathcal{E}) = \inf_{\Pi} \sup_{\theta} ||P_\theta - \Pi P_\theta^H||$$

obtained by allowing Π to range over the set of (i) all non-negative linear projections from M onto H; or alternatively, (ii) the set of all $\sigma(M, L)$-continuous non-negative linear projections from M onto H.

The sets of projections given in the above definition possess different topological properties. In particular, the set $\mathcal{P}$ of all non-negative linear projections from M onto a complete sublattice H is compact for the topology of pointwise convergence on $M \times L^H$. That is, the weakest topology such that $u\Pi\lambda$, regarded as a function of Π, is continuous for each $u \in M, \lambda \in L^H$.

The lack of sufficiency η_1 of a complete sublattice H for an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ in L, is the number

$$\eta_1(H, \mathcal{E}) = \inf_{\mathcal{F}} \sup_{\theta} ||P_\theta - Q_\theta||$$

where infimum ranges over all experiments $\mathcal{F} = (Q_\theta : \theta \in \Theta)$ in L such that H is sufficient for $\mathcal{F}$.

The lack of sufficiency is in general much more difficult to deal with than the insufficiency. In Section 7 we will observe a situation where the lack of sufficiency does not behave well as a measure for loss of information. For this reason we will not pay much attention to this quantity.

In order to proceed we will need a result from game theory. A two-person zero-sum game is a triple $\mathcal{G} = (A, B, \Gamma)$ where A and B are arbitrary sets and Γ is a function from $A \times B$ to $[-\infty, \infty]$. The game involves two players, player I and player II, say. For player I (player II) the set of available strategies is A (B). When player I uses the strategy $a \in A$ and player II uses the strategy $b \in B$, then player II pays player I the amount $\Gamma(a, b)$. The next theorem is well known, and is often referred to as the fundamental theorem of game theory.

The Fundamental Theorem in Game Theory: *Let $\mathcal{G} = (A, B, \Gamma)$ be a game where A and B are convex subsets of linear spaces and where Γ is a real-valued function that is concave on A and convex on B. If A and B are both compact topological spaces such that Γ is upper semicontinuous on A and lower semicontinuous on B, then*

$$\sup_a \inf_b \Gamma(a, b) = \inf_b \sup_a \Gamma(a, b).$$

For a proof the reader may consult e.g. Torgersen 1991, p.127 or general works on the subject. We will now establish the second main result of this section:

THEOREM 9 *Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment in L and let H be a complete sublattice in M. Let $P_\kappa = \sum_\theta \kappa(\theta) P_\theta$ for prior probabilities κ on Θ with countable support. Then the following quantities are the same:*

(i) $\inf_\Pi \sup_\theta ||P_\theta - \Pi P_\theta^H||$;

(ii) $\inf_\Pi \sup_\kappa ||P_\kappa - \Pi P_\kappa^H||$;

(iii) $\inf_\Pi \sup_\kappa \sum_\theta \kappa(\theta) ||P_\theta - \Pi P_\theta^H||$;

(iv) $\sup_\kappa \inf_\Pi \sum_\theta \kappa(\theta) ||P_\theta - \Pi P_\theta^H||$.

Here infimum is taken over the set of all non-negative linear projections from M onto H and in (ii)–(iv) supremum is taken over all prior probability distributions κ over Θ with a countable support.

Remark: From Corollary 7 the same quantities are obtained when restricting Π to range over the set of all $\sigma(M, L)$-continuous linear projections from M onto H.

Proof: Let $\mathcal{P}$ be the set of all non-negative linear projections from M onto H and let $\mathcal{K}$ be the set of all prior probability distributions over Θ with countable support.

Since $\sup_\theta ||P_\theta - \Pi P_\theta^H|| \geq \sum_\theta \kappa(\theta)||P_\theta - \Pi P_\theta^H||$ when $\Pi \in \mathcal{P}$ and $\kappa \in \mathcal{K}$, we have

$$\begin{aligned}\inf_\Pi \sup_\theta ||P_\theta - \Pi P_\theta^H|| &\geq \inf_\Pi \sup_\kappa \sum_\theta \kappa(\theta)||P_\theta - \Pi P_\theta^H|| \\ &\geq \inf_\Pi \sup_\kappa ||P_\kappa - \Pi P_\kappa^H|| \\ &\geq \inf_\Pi \sup_\theta ||P_\theta - \Pi P_\theta^H||.\end{aligned}$$

The latter inequality follows since $\mathcal{K}$ contains the Dirac measures (i.e. one-point probabilities) on Θ. Thus (i), (ii), and (iii) are the same number.

Consider the two-person zero-sum game $(\mathcal{K}, \mathcal{P}, \Gamma)$ where the pay-off function Γ is defined by $\Gamma(\kappa, \Pi) = \sum_\theta \kappa(\theta)||P_\theta - \Pi P_\theta^H||$ when $\kappa \in \mathcal{K}$ and $\Pi \in \mathcal{P}$. Equip $\mathcal{K}$ and $\mathcal{P}$ with the topology of pointwise convergence on, respectively, Θ and $M \times L^H$. Then both $\mathcal{K}$ and $\mathcal{P}$ are convex and compact subsets of linear spaces. It is readily checked that Γ is affine-convex in (κ, Π). Furthermore, Γ is continuous in κ and the topology on $\mathcal{P}$ is chosen so that $u\Pi P^H$ is continuous in Π for each $u \in M$ and $P \in L$. From this $||P_\theta - \Pi P_\theta^H|| = \sup_{|u| \leq 1} |uP_\theta - u\Pi P_\theta^H|$ is lower semicontinuous in Π for each $\theta \in \Theta$, and hence $\Gamma(\kappa, \Pi)$ is lower semicontinuous in Π for each $\kappa \in \mathcal{K}$. The fundamental theorem in game theory yields

$$\inf_\Pi \sup_\kappa \Gamma(\kappa, \Pi) = \sup_\kappa \inf_\Pi \Gamma(\kappa, \Pi),$$

and so (iii) and (iv) coincide. □

The restriction of $\mathcal{E}$ to a subset F of Θ will be denoted $\mathcal{E}_F = (P_\theta : \theta \in F)$. Let H be a complete sublattice of M. By an argument related to the proof above, Le Cam (see e.g. 1974 or 1986) proved that

$$\eta(H, \mathcal{E}) = \sup_F \eta(H, \mathcal{E}_F),$$

where supremum is taken to range over all finite subsets F of Θ. It is easy to verify that this result also follows as a direct consequence of Theorem 9(iv).

5 Properties

Suppose $\mathcal{E}$ is an experiment in L and that H is a complete sublattice of M. If $\eta(H,\mathcal{E}) = 0$ then it follows from Definition 8(i) that there is a net (Π_α) in $\mathcal{P}$ such that $|uP_\theta - u\Pi_\alpha P_\theta^H| \longrightarrow 0$ for each $u \in M$ and each $\theta \in \Theta$. Since $\mathcal{P}$ is compact for the topology of pointwise convergence on $M \times L^H$, it follows that there is a non-negative linear projection Π from M onto H such that $||P_\theta - \Pi P_\theta^H|| = \sup_{|u|\leq 1} |uP_\theta - u\Pi P_\theta^H| = 0$ for each $\theta \in \Theta$. Therefore H is sufficient for $\mathcal{E}$. Next to follow is a list on properties of the insufficiency. Note that the properties (i), (v), and (vi) are known from earlier works of Le Cam.

THEOREM 10 *Let $\mathcal{E}$ be an experiment in L and let H and H' be a complete sublattices of M where $H' \subseteq H$. The insufficiency satisfies the following properties:*

(i) $\eta(H,\mathcal{E}) = 0$ *if and only if H is sufficient for $\mathcal{E}$;*

(ii) $\eta(H',\mathcal{E}^H) \leq \eta(H',\mathcal{E})$;

(iii) $\eta(H,\mathcal{E}) \leq \eta(H',\mathcal{E}^H) + \eta(H',\mathcal{E})$;

(iv) $\eta(H',\mathcal{E}) \leq \eta(H',\mathcal{E}^H) + \eta(H,\mathcal{E})$;

(v) $\eta(H',\mathcal{E}) = \eta(H,\mathcal{E})$ *if H' is sufficient for $\mathcal{E}^H$;*

(vi) $\eta(H',\mathcal{E}) = \eta(H',\mathcal{E}^H)$ *if H is sufficient for $\mathcal{E}$.*

Remark: By $\eta(H',\mathcal{E}^H)$ we mean the insufficiency of H' for $\mathcal{E}^H$ where $\mathcal{E}^H$ is regarded as an experiment in the abstract L-space L^H, say.

Proof: Property (i) follows from the argument above. Let $\mathcal{P}_H^M$, $\mathcal{P}_{H'}^M$, and $\mathcal{P}_{H'}^H$ be families of projections where e.g. $\mathcal{P}_H^M$ denotes the set of all non-negative linear projections from M onto H, and so on. Note that H' is closed for the $\sigma(H, L^H)$ topology on H. Property (ii): We have that

$$\begin{aligned}
\eta(H',\mathcal{E}) &= \inf_{\Pi\in\mathcal{P}_{H'}^M} \sup_\theta ||P_\theta - \Pi P_\theta^{H'}|| \\
&\geq \inf_{\Pi\in\mathcal{P}_{H'}^M} \sup_\theta ||P_\theta^H - (\Pi P_\theta^{H'})^H|| \\
&\geq \inf_{\Pi_0\in\mathcal{P}_{H'}^H} \sup_\theta ||P_\theta^H - \Pi_0 P_\theta^{H'}|| \\
&= \eta(H',\mathcal{E}^H).
\end{aligned}$$

Thus (ii) holds. Property (vi): Assume that H is sufficient for $\mathcal{E}$. Choose $\Pi \in \mathcal{P}_H^M$ such that $P_\theta = \Pi P_\theta^H$ for each $\theta \in \Theta$. Then

$$\eta(H',\mathcal{E}^H) = \inf_{\Pi_0\in\mathcal{P}_{H'}^H} \sup_\theta ||P_\theta^H - \Pi_0 P_\theta^{H'}||$$

$$
\begin{aligned}
&= \inf_{\Pi_0 \in \mathcal{P}^H_{H'}} \sup_\theta ||\Pi(P^H_\theta - \Pi_0 P^{H'}_\theta)|| \qquad (4)\\
&= \inf_{\Pi_0 \in \mathcal{P}^H_{H'}} \sup_\theta ||P_\theta - \Pi(\Pi_0 P^{H'}_\theta)||\\
&\geq \inf_{\Pi_1 \in \mathcal{P}^M_{H'}} \sup_\theta ||P_\theta - \Pi_1 P^{H'}_\theta||\\
&= \eta(H', \mathcal{E}).
\end{aligned}
$$

Equation (4) holds since the adjoint of Π is an isometric map from H^* into M^*. Hence, from property (ii) it follows that (vi) holds. Property (iv): Using the triangle inequality we get

$$
\begin{aligned}
&\inf_{\Pi \in \mathcal{P}^M_{H'}} \sup_\theta ||P_\theta - \Pi P^{H'}_\theta||\\
\leq\ &\inf_{\Pi_0 \in \mathcal{P}^M_H} (\sup_\theta ||P_\theta - \Pi_0 P^H_\theta|| + \inf_{\Pi \in \mathcal{P}^M_{H'}} \sup_\theta ||\Pi_0 P^H_\theta - \Pi P^{H'}_\theta||).
\end{aligned}
$$

Thus there is a $\sigma(M, L)$-continuous $\Pi_0 \in \mathcal{P}^M_H$ such that

$$
\eta(H', \mathcal{E}) \leq \eta(H, \mathcal{E}) + \eta(H', \Pi_0 \mathcal{E}^H).
$$

Since H is sufficient for $\Pi_0 \mathcal{E}^H$ it follows from property (vi) that $\eta(H', \Pi_0 \mathcal{E}^H) = \eta(H', \mathcal{E}^H)$. Thus (iv) holds. The proof of (iii) is similar. Property (v) is an immediate consequence of (i), (iii), and (iv). □

It might be the case that Property (iii) may be replaced by the stronger Property (iii)' $\eta(H, \mathcal{E}) \leq \eta(H', \mathcal{E})$. However, we do not have any formal argument for this. It should be noted that the converse statements of (v) and (vi) do not hold. That is, $\eta(H', \mathcal{E}) = \eta(H, \mathcal{E})$ does not imply that H' is sufficient for $\mathcal{E}^H$; and $\eta(H', \mathcal{E}) = \eta(H', \mathcal{E}^H)$ does not imply that H is sufficient for $\mathcal{E}$. Thus, when the insufficiency assigns H' and H the same value, we cannot from this alone conclude that H' is just as informative as H.

6 Insufficiency and measure theory

Traditionally a statistical experiment is defined as a pair $((\mathcal{X}, \mathcal{A}), (P_\theta : \theta \in \Theta))$ where $(\mathcal{X}, \mathcal{A})$ is a measurable space and $(P_\theta : \theta \in \Theta)$ is a family of probability measures over $(\mathcal{X}, \mathcal{A})$ indexed by some non-empty set Θ, the parameter space.

A measure experiment $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ is called dominated if there is a non-negative σ-finite measure μ such that P_θ is absolutely continuous with respect to μ for each $\theta \in \Theta$. An experiment $\mathcal{E}$ is called coherent if for each uniformly bounded net (δ_s) of real variables there corresponds a subnet $(\delta_{s'})$ and a real variable δ such that $\int \delta_{s'} h \, dP_\theta \longrightarrow \int \delta h \, dP_\theta$ when $\int |h| \, dP_\theta < \infty$ and $\theta \in \Theta$.

The 'L-space of the experiment $\mathcal{E}$' is the space of finite (signed) measures μ over $(\mathcal{X}, \mathcal{A})$ such that μ is absolutely continuous with respect to a measure on the form $\sum_{i=1}^{\infty} 2^{-i} P_{\theta_i}$ where $\theta_1, \theta_2, \ldots$ is a countable sequence in Θ. The 'L-space of $\mathcal{E}$' will be noted $L(\mathcal{E})$. Equipped with the norm of total variation and the setwise '$\geq$'-ordering $L(\mathcal{E})$ is indeed an abstract L-space. The topological dual of $L(\mathcal{E})$ is called the M-space of $\mathcal{E}$ and will be noted $M(\mathcal{E})$.

Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment. A family of real valued variables $(v_\theta : \theta \in \Theta)$ is called coherent in $\mathcal{E}$ if there is a real-valued variable v such that $P_\theta[v_\theta \neq v] = 0$ whenever $\theta \in \Theta$. It is called consistent in $\mathcal{E}$ if to each two-point subset F of Θ there corresponds a variable v_F such that $P_\theta[v_\theta \neq v_F] = 0$ whenever $\theta \in F$. Two families $(v_\theta : \theta \in \Theta)$ and $(w_\theta : \theta \in \Theta)$ of real variables are said to be $\mathcal{E}$-equivalent if $P_\theta[v_\theta \neq w_\theta] = 0$ for each $\theta \in \Theta$.

Torgersen (1991) has shown that to each uniformly bounded and consistent family $(v_\theta : \theta \in \Theta)$ of real variables there is a unique linear functional v in $M(\mathcal{E})$ such that

$$v(\lambda) = \int v_C \, d\lambda \tag{5}$$

whenever $C = \{\theta_1, \theta_2, \ldots\} \subseteq \Theta$, $\lambda \ll \sum_i 2^{-i} P_{\theta_i}$ and $v_\theta = v_C$ a.s. P_θ when $\theta \in C$. Furthermore, any linear functional v in $M(\mathcal{E})$ may be obtained from consistent and uniformly bounded families of real variables determined by v up to equivalence.

Suppose $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$. We call the sub-σ-algebra $\mathcal{B}$ sufficient for $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ if for each $A \in \mathcal{A}$ there is a common $\mathcal{B}$-measurable version of the conditional probabilities $P_\theta[A|\mathcal{B}] : \theta \in \Theta$. We call $\mathcal{B}$ pairwise sufficient for $\mathcal{E}$ if the family $(E_\theta[v_\theta|\mathcal{B}] : \theta \in \Theta)$ is consistent in $(\mathcal{E}|\mathcal{B}) = (\mathcal{X}, \mathcal{B}; (P_\theta|\mathcal{B}) : \theta \in \Theta)$ whenever the family $(v_\theta : \theta \in \Theta)$ is uniformly bounded and consistent in $\mathcal{E}$.

Another useful result from Torgersen (1991, p. 14) is that an experiment $\mathcal{E}$ is coherent if and only if consistent families of real variables in $\mathcal{E}$ are coherent. Thus the well-known implications:

$$\begin{array}{c} (\mathcal{E}|\mathcal{B}) \text{ coherent \& pairwise sufficiency} \\ \Downarrow \\ \text{sufficiency} \\ \Downarrow \\ \text{pairwise sufficiency.} \end{array}$$

Suppose that $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$. Then the M-space $M(\mathcal{E}|\mathcal{B})$ of the experiment $(\mathcal{X}, \mathcal{B}; (P_\theta|\mathcal{B}) : \theta \in \Theta)$ may be identified as the uniform sublattice $\widehat{M}(\mathcal{E}|\mathcal{B})$ of $M(\mathcal{E})$ consisting of all linear functionals v in $M(\mathcal{E})$ that may be represented by a uniformly bounded family of $\mathcal{B}$-measurable variables which are consistent in $(\mathcal{E}|\mathcal{B})$. The following example shows that families of $\mathcal{B}$-measurable variables which are consistent in $\mathcal{E}$ are not necessarily consistent in $(\mathcal{E}|\mathcal{B})$.

EXAMPLE 1 Put $\mathcal{X} = \{1, 2, \ldots, 5\}$ and equip $\mathcal{X}$ with the σ-algebra $\mathcal{A}$ consisting of all subsets of $\mathcal{X}$. Consider the subsets $B_0 = \{1, 2, 3\}$ and $B_1 = \{3, 4, 5\}$ and let $\mathcal{B}$ be the sub-σ-algebra generated by B_0 and B_1. Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta = 0, 1)$ be the experiment where P_0, P_1 are given by the point probabilities:

	1	2	3	4	5
P_0	0	1/3	1/3	0	1/3
P_1	1/3	0	1/3	1/3	0

Put $A = \{2, 3, 4\}$. Then $P_\theta[I_{B_\theta} \neq I_A] = 0$ when $\theta = 0, 1$, and so $(I_{B_\theta} : \theta = 0, 1)$ is consistent (coherent) in $\mathcal{E}$. However, the reader may easily verify that the family is not consistent (coherent) in $(\mathcal{E}|\mathcal{B})$.

The next lemma is a slight extension of a result by Torgersen 1991, Theorem 7.3.8(iv). When proving 'only if' we follow Torgersen's argument.

LEMMA 11 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and suppose that $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ is an experiment in $L(\mathcal{E})$. Let $\mathcal{B}$ be a sub-σ-algebra in $\mathcal{A}$. Then $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$ and $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$ if and only if there is a non-negative linear projection Π from $M(\mathcal{E})$ onto $\widehat{M}(\mathcal{E}|\mathcal{B})$ such that $\langle v\Pi, P_\theta \rangle \equiv_\theta \langle v, Q_\theta \rangle$ for each $v \in M(\mathcal{E})$.*

Proof: We will first prove 'only if'. Without loss of generality it may be assumed that Q_θ is absolutely continuous with respect to P_θ for each θ in Θ. If this is not the case, replace P_θ with $\frac{1}{2}(P_\theta + Q_\theta)$. Clearly this replacement leaves $M(\mathcal{E})$ and $(\mathcal{E}|\mathcal{B})$ invariant. Let v be an element in $M(\mathcal{E})$. Then v may be represented by a family $(v_\theta : \theta \in \Theta)$ of real variables which is uniformly bounded and $\mathcal{E}$-consistent (i.e. consistent in $\mathcal{E}$). Since $Q_\theta \ll P_\theta$ for each $\theta \in \Theta$, $(v_\theta : \theta \in \Theta)$ is also $\mathcal{F}$-consistent. Put $\tilde{v}_\theta = E_{Q_\theta}[v_\theta|\mathcal{B}]$. Then, by pairwise sufficiency the family $(\tilde{v}_\theta : \theta \in \Theta)$ is consistent in $(\mathcal{E}|\mathcal{B})$ and thus represents a functional $v\Pi$ in $\widehat{M}(\mathcal{E}|\mathcal{B})$. Clearly $v\Pi$ does not depend neither on the choice of variables v_θ nor on the specification of the conditional expectations. Thus the map $v \mapsto v\Pi$ is a well defined map from $M(\mathcal{E})$ into $\widehat{M}(\mathcal{E}|\mathcal{B})$. It is easily seen that this map is a projection from $M(\mathcal{E})$ onto $\widehat{M}(\mathcal{E}|\mathcal{B})$. Furthermore, since $P_\theta|\mathcal{B} = Q_\theta|\mathcal{B}$, it follows that $\langle v\Pi, P_\theta \rangle = \int \tilde{v}_\theta \, dP_\theta = \int \tilde{v}_\theta \, d(P_\theta|\mathcal{B}) = \int \tilde{v}_\theta \, d(Q_\theta|\mathcal{B}) = \int v_\theta \, dQ_\theta = \langle v, Q_\theta \rangle$.

As for the proof of 'if', let $\dot{I}_B$ denote the functional in $M(\mathcal{E})$ that may be represented by the real variable I_B , $B \in \mathcal{B}$. Then $P_\theta(B) = \int I_B \, dP_\theta = \langle \dot{I}_B \Pi, P_\theta \rangle = \langle \dot{I}_B, Q_\theta \rangle = \int I_B \, dQ_\theta = Q_\theta(B)$ and thus $\mathcal{E}|\mathcal{B} = \mathcal{F}|\mathcal{B}$. Of the same reason as above we may, without loss of generality, assume that $Q_\theta \ll P_\theta$. Let $(v_\theta : \theta \in \Theta)$ be a family of uniformly bounded real variables which is $\mathcal{F}$-consistent. Then $(v_\theta : \theta \in \Theta)$ represents a linear functional v on $L(\mathcal{F})$. Since $L(\mathcal{F})$ is a band in $L(\mathcal{E})$, there is a linear and bounded extension $w \in M(\mathcal{E})$ of v to the abstract L-space $L(\mathcal{E})$. Therefore there is a uniformly bounded and $\mathcal{E}$-consistent family $(w_\theta : \theta \in \Theta)$ representing the linear functional w. The projection Π maps $M(\mathcal{E})$ onto $\widehat{M}(\mathcal{F}|\mathcal{B})$ and thus $w\Pi$ may be represented as a family

$(\tilde{w}_\theta : \theta \in \Theta)$ of uniformly bounded and $(\mathcal{F}|\mathcal{B})$-consistent $\mathcal{B}$-measurable variables. It follows that $\int_B \tilde{w}_\theta \, dQ_\theta = \int \tilde{w}_\theta I_B \, dQ_\theta = \int \tilde{w}_\theta I_B \, dP_\theta = \langle (w\dot{I}_B)\Pi, P_\theta \rangle = \langle w\dot{I}_B, Q_\theta \rangle = \int w_\theta I_B \, dQ_\theta = \int_B w_\theta \, dQ_\theta = \int_B v_\theta \, dQ_\theta$ for each $B \in \mathcal{B}$. The latter equality follows from Equation (5) since $Q_\theta \ll P_\theta$. Therefore $\tilde{w}_\theta$ is a version of $E_{Q_\theta}[v_\theta|\mathcal{B}]$ and hence $(E_{Q_\theta}[v_\theta|\mathcal{B}] : \theta \in \Theta)$ is consistent in $(\mathcal{F}|\mathcal{B})$. Thus $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$. □

Next comes a measure theoretic version of Le Cam's Lemma 1. In the proof we make use of *almost randomizations.* A formal definition of this notion is given prior to Corollary 17.

LEMMA 12 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and $\mathcal{B}$ a sub-σ-algebra of $\mathcal{A}$. Then there exists an experiment $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ in $L(\mathcal{E})$ such that $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$ and $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$.*

Proof: By Zorn's lemma there is a family $\{\pi_s : s \in S\} \subseteq L(\mathcal{E})$ of probability measures such that the restrictions $(\pi_s|\mathcal{B}) : s \in S$ have pairwise disjoint supports $B_s : s \in S$ in $\mathcal{B}$ and such that the band generated by $(\pi_s|\mathcal{B}) : s \in S$ is $L(\mathcal{E}|\mathcal{B})$.

For each $\theta \in \Theta$ let S_θ be a countable subset of S such that $\cup_{s \in S_\theta} B_s$ supports $(P_\theta|\mathcal{B})$. Define the almost randomization Γ_θ by $\Gamma_\theta(A|\cdot) = \sum_{s \in S_\theta} \pi_s[A|\mathcal{B}] I_{B_s}$ for each $A \in \mathcal{A}$ and put $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ where $Q_\theta \equiv_\theta (P_\theta|\mathcal{B})\Gamma_\theta$. It is readily checked that $\mathcal{F}$ is an experiment in $L(\mathcal{E})$ such that $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$. Thus it remains to show that $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$. Let $(v_\theta : \theta \in \Theta)$ be a family of uniformly bounded and $\mathcal{F}$-consistent variables. Let F be a two-point subset of Θ and choose v_F such that $Q_\theta[v_\theta \neq v_F] = 0$ when $\theta \in F$. Define Γ_F by $\Gamma_F(A|x) = \bigvee_{\theta \in F} \Gamma_\theta(A|x)$ for each $A \in \mathcal{A}$ and $x \in \mathcal{X}$. Let ϕ_F be the $\mathcal{B}$-measurable function given by $\phi_F(x) = \int v_F(y) \, \Gamma_F(dy|x)$. Then

$$\begin{aligned}
\int_B E_{Q_\theta}[v_\theta|\mathcal{B}] \, d(Q_\theta|\mathcal{B}) &= \int_B v_\theta \, dQ_\theta \\
&= \int_B v_F \, d(P_\theta|\mathcal{B})\Gamma_\theta \\
&= \int_B v_F \, d(P_\theta|\mathcal{B})\Gamma_F \\
&= \int_B \phi_F \, d(P_\theta|\mathcal{B}) \\
&= \int_B \phi_F \, d(Q_\theta|\mathcal{B}),
\end{aligned}$$

whenever $\theta \in F$ and $B \in \mathcal{B}$. Thus $(E_{Q_\theta}[v_\theta|\mathcal{B}] : \theta \in \Theta)$ is consistent in $(\mathcal{F}|\mathcal{B})$. This yields the result. □

In order to show that $\widehat{M}(\mathcal{E}|\mathcal{B})$ is a complete sublattice of $M(\mathcal{E})$, let $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ be any experiment in $L \equiv L(\mathcal{E})$ such that $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$

and $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$. Then from Lemma 11 there exists a non-negative linear projection Π from $M(\mathcal{E})$ onto $\widehat{M}(\mathcal{E}|\mathcal{B})$ such that $\langle v\Pi, P_\theta\rangle = \langle v, Q_\theta\rangle$ whenever $v \in M \equiv M(\mathcal{E})$ and $\theta \in \Theta$. Let (u_α) be a net in $\widehat{M}(\mathcal{E}|\mathcal{B})$ which converges to an element u in M for the weak $\sigma(M, L)$ topology. Then $|\langle u_\alpha\Pi - u\Pi, P_\theta\rangle| = |\langle u_\alpha - u, Q_\theta\rangle| \longrightarrow 0$. Thus $u_\alpha = u_\alpha\Pi$ converges to u as well as $u\Pi$, hence $u = u\Pi \in \widehat{M}(\mathcal{E}|\mathcal{B})$. It follows that $\widehat{M}(\mathcal{E}|\mathcal{B})$ is closed for the weak $\sigma(M, L)$ topology.

We will now define the insufficiency of a sub-σ-algebra $\mathcal{B}$. Similar to the previous sections the idea is to measure how much the experiment $\mathcal{E}$ must be modified in order to make $\mathcal{B}$ pairwise sufficient.

DEFINITION 13 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. The insufficiency of $\mathcal{B}$ for $\mathcal{E}$ is the quantity*

$$\eta(\mathcal{B}, \mathcal{E}) = \inf_{\mathcal{F}} \sup_{\theta} ||Q_\theta - P_\theta||$$

where infimum is taken over all experiments $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ such that

(i) $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$;

(ii) $(Q_\theta|\mathcal{B}) = (P_\theta|\mathcal{B})$ for each $\theta \in \Theta$.

From Lemma 12 we see that the insufficiency is well-defined. In the above definition we take infimum to range over a possible very large class of experiments. For instance, suppose $\mathcal{E}$ is dominated by a σ-finite measure μ. Then, from the definition it is clear that the experiment $\mathcal{F}$ does not need to meet the same requirement. However, the next theorem tells us that nothing will be changed if we restrict infimum to range over experiments $\mathcal{F}$ that are dominated by μ. It should also be noted that the next result is a measure theoretic version of Theorem 6.

THEOREM 14 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ be an experiment satisfying the conditions (i) and (ii) in Definition 13. Then there is another experiment $\overline{\mathcal{F}} = (\mathcal{X}, \mathcal{A}; \overline{Q}_\theta : \theta \in \Theta)$ in $L(\mathcal{E})$ with the same properties such that*

$$||P_\theta - \overline{Q}_\theta|| \leq ||P_\theta - Q_\theta||$$

for all $\theta \in \Theta$.

Proof: Define the experiment $\widehat{\mathcal{F}} = (\mathcal{X}, \mathcal{A}; \widehat{Q}_\theta : \theta \in \Theta)$ by $\widehat{Q}_\theta \equiv_\theta \frac{1}{2}(P_\theta + Q_\theta)$. Put $M = M(\widehat{\mathcal{F}})$ and $H = \widehat{M}(\widehat{\mathcal{F}}|\mathcal{B})$. From Lemma 11 there is a non-negative linear projection Π from M onto H such that $\Pi\widehat{Q}_\theta^H \equiv_\theta Q_\theta$. Theorem 6 implies that there is a non-negative linear projection $\overline{\Pi}$ from M onto H such that $\overline{\Pi}\widehat{Q}_\theta^H \in L(\mathcal{E})$ and

$$||P_\theta - \overline{\Pi}\widehat{Q}_\theta^H|| \leq ||P_\theta - \Pi\widehat{Q}_\theta^H||$$

for each $\theta \in \Theta$. Define the experiment $\overline{\mathcal{F}}$ by $\overline{Q}_\theta \equiv_\theta \overline{\Pi}\widehat{Q}_\theta^H$. Then Lemma 11 yields $\overline{\mathcal{F}}|\mathcal{B} = \mathcal{E}|\mathcal{B}$ and further that $\mathcal{B}$ is pairwise sufficient for $\overline{\mathcal{F}}$. □

Under certain regularity conditions Le Cam (1974) established, by means of Markov kernels, a connection between the lattice theoretic and measure theoretic definitions of insufficiency. Next comes a generalization of this result.

THEOREM 15 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a sub-σ-algebra in $\mathcal{A}$. Then the insufficiency of $\mathcal{B}$ for $\mathcal{E}$ coincides with the insufficiency of the complete sublattice $\widehat{M}(\mathcal{E}|\mathcal{B})$ for the corresponding abstract experiment $\widehat{\mathcal{E}} = (P_\theta : \theta \in \Theta)$ in the abstract L-space $L(\mathcal{E})$. That is,*

$$\eta(\mathcal{B}, \mathcal{E}) = \eta(\widehat{M}(\mathcal{E}|\mathcal{B}), \widehat{\mathcal{E}}).$$

Proof: Follows from Lemma 11, Theorem 14, and Definition 8(ii). □

Thus the results on the insufficiency from Section 4 also holds in the measure theoretic setting:

COROLLARY 16 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Then the following quantities coincide:*

(a) $\inf_{\mathcal{F}} \sup_\theta \|Q_\theta - P_\theta\|$;

(b) $\inf_{\mathcal{F}} \sup_\kappa \sum_\theta \kappa(\theta)\|Q_\theta - P_\theta\|$;

(c) $\sup_\kappa \inf_{\mathcal{F}} \sum_\theta \kappa(\theta)\|Q_\theta - P_\theta\|$,

where supremum in (b) and (c) is taken over all prior probabilities κ over Θ with countable support. In all the three expressions infimum is taken to range over all experiments $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ such that

(i) $\mathcal{B}$ is pairwise sufficient for $\mathcal{F}$;

(ii) $(Q_\theta|\mathcal{B}) = (P_\theta|\mathcal{B})$ for each $\theta \in \Theta$.

The quantities (a),(b), and (c) will remain unchanged if we take infimum over all experiments $\mathcal{F}$ which in addition to (i) and (ii) satisfy

(iii) $Q_\theta \in L(\mathcal{E})$ for each $\theta \in \Theta$.

Proof: Due to Theorem 9 the result is immediate. □

Now, consider the experiment $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $\mathrm{ba}(\mathcal{X}, \mathcal{A})$ denote the set of all bounded and additive set functions over $(\mathcal{X}, \mathcal{A})$. A function $\Gamma : \mathcal{A} \times \mathcal{X} \mapsto [-\infty, \infty]$ is called an almost randomization from $L(\mathcal{E}|\mathcal{B})$ into $\mathrm{ba}(\mathcal{X}, \mathcal{A})$ if $\Gamma(A|\cdot)$ is $\mathcal{B}$-measurable for each $A \in \mathcal{A}$ and satisfies:

(i) $\Gamma(\emptyset|\cdot) = 0$, $\Gamma(\mathcal{X}|\cdot) = 1$; $(P_\theta|\mathcal{B})$-a.s. when $\theta \in \Theta$;

(ii) $0 \leq \Gamma(A|\cdot) \leq 1$; $(P_\theta|\mathcal{B})$-a.s. when $A \in \mathcal{A}$ and $\theta \in \Theta$;

(iii) $\Gamma(A_1 \cup A_2 \cup \cdots|\cdot) = \Gamma(A_1|\cdot) + \Gamma(A_2|\cdot) + \cdots$; $(P_\theta|\mathcal{B})$-a.s. whenever $\theta \in \Theta$ and $A_1, A_2, \ldots$ is a sequence of disjoint events in $\mathcal{A}$.

It is readily checked that $\mathcal{B}$ is sufficient for $\mathcal{E}$ if and only if there is an almost randomization Γ from $L(\mathcal{E}|\mathcal{B})$ into $\mathrm{ba}(\mathcal{X}, \mathcal{A})$ such that

$$\int_B \Gamma(A|\cdot)\, d(P_\theta|\mathcal{B}) = P_\theta(A \cap B)$$

whenever $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $\theta \in \Theta$. Thus, when $\mathcal{B}$ is sufficient, then the original experiment $\mathcal{E}$ may be reconstructed from the sub-experiment $(\mathcal{E}|\mathcal{B})$ by a 'random mechanism' (i.e. almost randomization) that does not depend on the parameter, that is, $P_\theta \equiv_\theta (P_\theta|\mathcal{B})\Gamma$ where $[(P_\theta|\mathcal{B})\Gamma](A) \equiv \int \Gamma(A|\cdot)\, d(P_\theta|\mathcal{B})$ for each $A \in \mathcal{A}$ and $\theta \in \Theta$.

It is also easy to see that if an almost randomization Γ satisfies the additional property

(iv) $\Gamma(A \cap B|\cdot) = I_B \Gamma(A|\cdot)$; $(P_\theta|\mathcal{B})$-a.s. when $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $\theta \in \Theta$;

then $(\mathcal{X}, \mathcal{A}; (P_\theta|\mathcal{B})\Gamma : \theta \in \Theta)$ is an experiment for which $\mathcal{B}$ is sufficient.

COROLLARY 17 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and suppose that $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$ such that $(\mathcal{E}|\mathcal{B})$ is coherent. Then the insufficiency $\eta(\mathcal{B}, \mathcal{E})$ coincides with the following quantities:*

(a) $\inf_\Gamma \sup_\theta ||(P_\theta|\mathcal{B})\Gamma - P_\theta||$;

(b) $\inf_\Gamma \sup_\kappa \sum_\theta \kappa(\theta)||(P_\theta|\mathcal{B})\Gamma - P_\theta||$;

(c) $\sup_\kappa \inf_\Gamma \sum_\theta \kappa(\theta)||(P_\theta|\mathcal{B})\Gamma - P_\theta||$,

where supremum in (b) and (c) is taken over all prior probabilities κ over Θ with countable support. In (a), (b), and (c) infimum is taken over all almost randomizations Γ from $L(\mathcal{E}|\mathcal{B})$ into $\mathrm{ba}(\mathcal{X}, \mathcal{A})$ *that satisfy the requirement*

(i) $\Gamma(A \cap B|\cdot) = I_B \Gamma(A|\cdot)$; *$(P_\theta|\mathcal{B})$-a.s. when $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $\theta \in \Theta$.*

The quantities remain unchanged if we take infimum over all almost randomizations which in addition to (i) satisfy

(ii) $\Gamma(A|\cdot) = 0$; *$(P_\theta|\mathcal{B})$-a.s. for each $\theta \in \Theta$ whenever $P_\theta(A) \equiv_\theta 0$.*

Proof: Due to Corollary 16 and the previous discussions on almost randomizations and sufficiency, the result is immediate. Note that $(P_\theta|\mathcal{B})\Gamma \in L(\mathcal{E})$ if and only if Γ satisfies condition (ii). □

If $\mathcal{E}'$ is a sub-experiment of $\mathcal{E}$ in the sense that $\mathcal{E}'$ is on the form $(\mathcal{E}|\mathcal{B})$, we will sometimes find it convenient to write $\eta(\mathcal{E}', \mathcal{E})$ instead of $\eta(\mathcal{B}, \mathcal{E})$. Let $X_1, X_2, \ldots$ be a sequence of random variables. Suppose $\mathcal{E}'$ and $\mathcal{E}$ are the experiments induced by $(X_1, \ldots, X_k)$ and $(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n)$, respectively. Then $\mathcal{E}'$ can be identified with a sub-experiment of $\mathcal{E}$. We will call $\eta(\mathcal{E}', \mathcal{E})$ for the insufficiency of $(X_1, \ldots, X_k)$ associated with the additional observation $(X_{k+1}, \ldots, X_n)$.

Suppose $X_1, X_2, \ldots$ are iid copies of a random variable X. In order to find an upper bound for the insufficiency of $(X_1, \ldots, X_n)$ with respect to the additional observation X_{n+1}, we may construct a suitable almost randomization – for instance, by means of an estimator $\hat{\theta}_n$. This idea is related to the constructions of Helgeland (1982) and Mammen (1986).

EXAMPLE 2 Let $\mathcal{E} = (\Re, \mathcal{R}; P_\theta : -\infty < \theta < \infty)$, where $\mathcal{R}$ is the family of Borel sets on the real line $\Re$ and $P_\theta = N(\theta, 1)$. Put $\mathcal{B}_n = \mathcal{R}^n \times \{\emptyset, \Re\}$. Now, for each rectangle $A = A_1 \times \cdots \times A_{n+1}$ in $\mathcal{R}^{n+1}$ define

$$\Gamma(A|\mathbf{x}) = I_{A_1}(x_1) \cdots I_{A_n}(x_n) P_{\hat{\theta}_n(\mathbf{x})}(A_{n+1})$$

where $\mathbf{x} = (x_1, \ldots, x_{n+1}) \in \Re^{n+1}$ and $\hat{\theta}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n x_i$. Using π–λ system arguments one easily verifies that Γ extends to a unique almost randomization from $L(\mathcal{E}^{n+1}|\mathcal{B}_n)$ to $\mathrm{ba}(\Re^{n+1}, \mathcal{R}^{n+1})$. We denote the extension of Γ by the same letter. Let $Z \sim N(0, 1)$. Then we have that

$$\begin{aligned}
\eta(\mathcal{E}^n, \mathcal{E}^{n+1}) &\leq \sup_\theta \|(P_\theta^{n+1}|\mathcal{B}_n)\Gamma - P_\theta^{n+1}\| \\
&= \|(P_0^{n+1}|\mathcal{B}_n)\Gamma - P_0^{n+1}\| \\
&= \int \mathrm{Prob}\{|Z| \leq \frac{1}{2n}|\sum_{i=1}^n x_i|\}\, P_0^n(d(x_1, \ldots, x_n)) \\
&= \int \mathrm{Prob}\{|Z| \leq \frac{|y|}{2\sqrt{n}}\}\, P_0(dy) \\
&= \frac{2}{\pi}\tan^{-1}(\frac{1}{2\sqrt{n}}). \qquad (6)
\end{aligned}$$

This upper bound is interesting in view of the lower bound

$$\frac{1}{\pi\sqrt{n}}\exp\{-\frac{1}{4n}\}$$

which was found by Le Cam (1974). Note that the upper bound in (6) always is less than $\frac{1}{\pi\sqrt{n}}$. Clearly these bounds are very sharp.

COROLLARY 18 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment dominated by a σ-finite measure μ and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Put $f_\theta \equiv_\theta dP_\theta/d\mu$. Let*

$h_0 \in L^1(\mu)$ be a function that does not depend on θ and let $(g_\theta : \theta \in \Theta)$ be a family of $\mathcal{B}$-measurable functions such that $h_0 g_\theta$ is a p.d.f. with respect to μ for each $\theta \in \Theta$ and $\int_B h_0 g_\theta \, d\mu \equiv_\theta P_\theta(B)$ whenever $B \in \mathcal{B}$. Then the insufficiency $\eta(\mathcal{B}, \mathcal{E})$ is the same as the following quantities:

(a) $\inf_h \sup_\theta \int |hg_\theta - f_\theta| \, d\mu$;

(b) $\inf_h \sup_\kappa \sum_\theta \kappa(\theta) \int |hg_\theta - f_\theta| \, d\mu$;

(c) $\sup_\kappa \inf_h \sum_\theta \kappa(\theta) \int |hg_\theta - f_\theta| \, d\mu$,

where supremum in (b) and (c) is taken over all prior probabilities κ over Θ with countable support. In all the quantities infimum is taken over functions h in $L^1(\mu)$ such that

(i) h does not depend on θ;

(ii) hg_θ is a p.d.f. for all $\theta \in \Theta$;

(iii) $\int_B hg_\theta \, d\mu \equiv_\theta P_\theta(B)$ whenever $B \in \mathcal{B}$.

Remark: The existence of a function h_0 and a family $(g_\theta : \theta \in \Theta)$ satisfying the requirements in the text follows by the factorization criterion and Lemma 12.

Proof: Let $h_0, g_\theta : \theta \in \Theta$, and μ be as specified above. Let $\mathcal{F} = (\mathcal{X}, \mathcal{A}; Q_\theta : \theta \in \Theta)$ be an experiment in $L(\mathcal{E})$ such that $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$ and $\mathcal{B}$ is sufficient for $\mathcal{F}$. We need only to show that the Radon-Nikodym derivatives $dQ_\theta/d\mu$ may be specified to be on the form $\hat{h}g_\theta$ where $\hat{h} \in L^1(\mu)$ does not depend on θ. From the factorization criterion we may specify $dQ_\theta/d\mu$ as the product $h'g'_\theta$ when $\theta \in \Theta$ where $g'_\theta : \theta \in \Theta$ are $\mathcal{B}$-measurable and $h' \in L^1(\mu)$ does not depend on θ. Let π be a probability measure such that $\pi \ll \mu$ and $\mu \ll \pi$. Put $\phi = E_\pi[h_0(d\mu/d\pi)|\mathcal{B}]$ and $\phi' = E_\pi[h'(d\mu/d\pi)|\mathcal{B}]$. Then $\mathcal{F}|\mathcal{B} = \mathcal{E}|\mathcal{B}$ implies that

$$E_\pi[g_\theta h_0(d\mu/d\pi)|\mathcal{B}] = E_\pi[g'_\theta h'(d\mu/d\pi)|\mathcal{B}]$$

π-almost surely. Put

$$\tilde{g}_\theta = g_\theta(\phi/\phi')I_{[\phi'>0]}$$

and let

$$\hat{h} = h'(\phi/\phi')I_{[\phi'>0]}.$$

It is readily checked that $\hat{h} \in L^1(\mu)$. Furthermore we have that $dQ_\theta/d\mu = h'g'_\theta = h'\tilde{g}_\theta = \hat{h}g_\theta$ μ-a.e. The desired result now follows by Corollary 16. $\square$

7 Dichotomies

Dichotomies are experiments with a two-point parameter space. In this section we will see that the insufficiency can be expressed on a very simple form when considering dichotomies. When proving our next result we will make use of the fact that

$$\min\{a_0, a_1\}||v_0 - v_1|| = \inf_{w \in V}[a_0||w - v_0|| + a_1||w - v_1||] \tag{7}$$

when V is a normed real linear space and $v_0, v_1 \in V$, $a_0, a_1 > 0$. The proof is simple: '$\leq$' follows from the triangle inequality and '$\geq$' follows by taking $w = v_0$ or $w = v_1$.

THEOREM 19 *Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta = 0, 1)$ be a dichotomy and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let μ be a σ-finite measure that dominates $\mathcal{E}$ and let $f_\theta = dP_\theta/d\mu \; : \theta = 0, 1$. Suppose the product $h_0 g_\theta$ is a p.d.f. with respect to μ such that $\int_B h_0 g_\theta \, d\mu \equiv_\theta P_\theta(B)$ whenever $B \in \mathcal{B}$, where $g_\theta : \theta = 0, 1$ are $\mathcal{B}$-measurable and $h_0 \in L^1(\mu)$ does not depend on θ. Then*

$$\eta(\mathcal{B}, \mathcal{E}) = \sup_\lambda \inf_B \int \lambda |f_1 \frac{g_0}{g_1} - f_0| I_B + (1 - \lambda)|f_0 \frac{g_1}{g_0} - f_1| I_{\mathcal{X} \setminus B} \, d\mu \,,$$

where supremum is taken over all $\lambda \in [0, 1]$ and infimum is taken over all events B in $\mathcal{B}$.

Proof: It follows from Corollary 18 that $\eta(\mathcal{E}, \mathcal{B})$ may be written

$$\eta(\mathcal{B}, \mathcal{E}) = \sup_\lambda \inf_h \int \lambda |hg_0 - f_0| + (1 - \lambda)|hg_1 - f_1| \, d\mu \,, \tag{8}$$

where $\lambda \in [0, 1]$ and infimum is taken over all $h \in L^1(\mu)$ that do not depend on θ such that the product hg_θ is a p.d.f. with respect to μ and $\int_B hg_\theta \, d\mu = P_\theta(B)$ for each $B \in \mathcal{B}$ and $\theta = 0, 1$.

Let us first assume that $\mathcal{B}$ is finite, i.e. generated by a finite disjoint partition $B_1, B_2, \ldots, B_n \in \mathcal{A}$. Then for fixed $\lambda \in [0, 1]$, the h minimizing the integral in Equation (8) also minimizes

$$\int_{B_i} \lambda |hg_0 - f_0| + (1 - \lambda)|hg_1 - f_1| \, d\mu$$

for each $i = 1, 2, \ldots, n$. Let a_0, a_1 denote the values of, respectively, λg_0 and $(1 - \lambda) g_1$ on B_i. Without loss of generality we may assume that both g_0 and g_1 are greater than 0 on B_i. From Equation (7) we get

$$\inf_h \int_{B_i} \lambda |hg_0 - f_0| + (1 - \lambda)|hg_1 - f_1| \, d\mu$$

$$
\begin{aligned}
&= \inf_h \int_{B_i} a_0|h - \frac{f_0}{g_0}| + a_1|h - \frac{f_1}{g_1}|\,d\mu \\
&= \min\{a_0, a_1\} \int_{B_i} |\frac{f_0}{g_0} - \frac{f_1}{g_1}|\,d\mu \\
&= \int_{B_i} \lambda|\frac{g_0}{g_1} f_1 - f_0|\,d\mu \wedge \int_{B_i} (1-\lambda)|\frac{g_1}{g_0} f_0 - f_1|\,d\mu.
\end{aligned}
$$

This gives the result when $\mathcal{B}$ is finite.

Suppose $\mathcal{B}$ is infinite. We may without loss of generality assume that $g_\theta \geq 0$, $\theta = 0,1$. For each natural number n define the finite σ-algebra $\mathcal{B}_n$ generated by the sets on the form

$$[(k-1)2^{-n} \leq g_\theta < k2^{-n}],$$

where $\theta = 0,1$ and $k = 1,2,\ldots,n2^n$. Let $\varepsilon > 0$ and choose $h \in L^1(\mu)$ to be a function that does not depend on θ such that the product hg_θ is a p.d.f. and $\int_B hg_\theta\,d\mu \equiv_\theta P_\theta(B)$ whenever $B \in \mathcal{B}$ and

$$\int |hg_\theta - f_\theta|\,d\mu < \eta(\mathcal{B},\mathcal{E}) + \varepsilon,$$

when $\theta = 0,1$. Let $\nu = h\,d\mu$ and define $g_{\theta,n} = d(P_\theta|\mathcal{B}_n)/d(\nu|\mathcal{B}_n)$, $\theta = 0,1$. Clearly $\int_B hg_{\theta,n}\,d\mu \equiv_\theta P_\theta(B)$ whenever $B \in \mathcal{B}_n$ and it is readily checked that the product $hg_{\theta,n}$ is a p.d.f. when $\theta = 0,1$. Note that

$$\lim_n \int |hg_{\theta,n} - f_\theta|\,d\mu < \lim_n \int |hg_{\theta,n} - hg_\theta|\,d\mu + \eta(\mathcal{B},\mathcal{E}) + \varepsilon\,,$$

and hence from Lebesgue's dominated convergence theorem we have

$$\lim_n \int |hg_{\theta,n} - f_\theta|\,d\mu < \eta(\mathcal{B},\mathcal{E}) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary it follows that $\liminf \eta(\mathcal{B}_n,\mathcal{E}) \leq \eta(\mathcal{B},\mathcal{E})$. For each n, choose $B_n \in \mathcal{B}_n$ such that $B = B_n$ minimizes the expression

$$\int \lambda|\frac{g_{0,n}}{g_{1,n}} f_1 - f_0|I_B + (1-\lambda)|\frac{g_{1,n}}{g_{0,n}} f_0 - f_1|I_{\mathcal{X}\setminus B}\,d\mu$$

when allowing B to range over all events in $\mathcal{B}_n$. Put

$$\phi_{\lambda,n} = \lambda|\frac{g_{0,n}}{g_{1,n}} f_1 - f_0|I_{B_n} + (1-\lambda)|\frac{g_{1,n}}{g_{0,n}} f_0 - f_1|I_{\mathcal{X}\setminus B_n}\,.$$

Let C be an event in $\mathcal{B}$ such that $\liminf B_n \subseteq C \subseteq \limsup B_n$ and define

$$\phi_\lambda = \lambda|\frac{g_0}{g_1} f_1 - f_0|I_C + (1-\lambda)|\frac{g_1}{g_0} f_0 - f_1|I_{\mathcal{X}\setminus C}\,.$$

It is easy to verify that $\phi_{\lambda,n}$ converges pointwise μ-a.e. to ϕ_λ. Thus, by Fatou's lemma we get

$$\begin{aligned}
\eta(\mathcal{B},\mathcal{E}) &\geq \liminf \eta(\mathcal{B}_n,\mathcal{E}) \\
&= \liminf(\sup_\lambda \int \phi_{\lambda,n}\, d\mu) \\
&\geq \sup_\lambda(\liminf \int \phi_{\lambda,n}\, d\mu) \\
&\geq \sup_\lambda \int (\liminf \phi_{\lambda,n}) d\mu \\
&= \sup_\lambda \int \phi_\lambda\, d\mu \\
&\geq \sup_\lambda \inf_B \int \lambda |\frac{g_0}{g_1} f_1 - f_0| I_B + (1-\lambda)|\frac{g_1}{g_0} f_0 - f_1| I_{\mathcal{X}\setminus B}\, d\mu\,.
\end{aligned}$$

In the latter expression infimum is taken over all events B in $\mathcal{B}$. The converse inequality follows from Corollary 18. This finishes the proof. □

An experiment $\mathcal{E} = (\mathcal{X},\mathcal{A};P_\theta : \theta \in \Theta)$ is called totally non-informative if $P_\theta \equiv_\theta P$ for some probability measure P over $(\mathcal{X},\mathcal{A})$. Similarly, we say a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is totally non-informative for the experiment $\mathcal{E}$ if the sub-experiment $(\mathcal{E}|\mathcal{B})$ is totally non-informative.

COROLLARY 20 *If the sub-σ-algebra $\mathcal{B}$ is totally non-informative for the dichotomy $\mathcal{E} = (\mathcal{X},\mathcal{A};P_\theta : \theta = 0,1)$, then $\eta(\mathcal{B},\mathcal{E}) = \frac{1}{2}\|P_0 - P_1\|$.*

Proof: Put $f_\theta = dP_\theta/d\pi$ and $g_\theta = d(P_\theta|\mathcal{B})/d(\pi|\mathcal{B})$ where $\pi = \frac{1}{2}(P_0 + P_1)$. Then from Theorem 19 we have that

$$\begin{aligned}
\eta(\mathcal{B},\mathcal{E}) &= \sup_\lambda \inf_B \int \lambda |f_1 \frac{g_0}{g_1} - f_0| I_B + (1-\lambda)|f_0 \frac{g_1}{g_0} - f_1| I_{\mathcal{X}\setminus B}\, d\pi \\
&= \sup_\lambda \inf_B \int (\lambda I_B + (1-\lambda) I_{\mathcal{X}\setminus B})|f_0 - f_1|\, d\pi \\
&= \sup_\lambda \int (\lambda \wedge (1-\lambda))|f_0 - f_1|\, d\pi \\
&= \frac{1}{2}\int |f_0 - f_1|\, d\pi\,.
\end{aligned}$$

This yields the result. □

Let $\mathcal{E} = (\mathcal{X},\mathcal{A};P_\theta : \theta \in \Theta)$ be an experiment and suppose Y is a random variable (i.e. a measurable map from $(\mathcal{X},\mathcal{A})$ into some other measurable space $(\mathcal{Y},\mathcal{C})$, say). Then Y induces a new experiment $\mathcal{E}Y^{-1} = (\mathcal{Y},\mathcal{C};P_\theta Y^{-1} : \theta \in \Theta)$.

Here $P_\theta Y^{-1}$ denotes the distribution of Y when the parameter θ prevails. Two random variables X, Y are said to be independent with respect to the underlying experiment $\mathcal{E}$ if X, Y are independent with respect to the underlying probability space $(\mathcal{X}, \mathcal{A}, P_\theta)$, for each $\theta \in \Theta$. Note that $\mathcal{E}X^{-1}$ may be regarded as an sub-experiment of $\mathcal{E}(X,Y)^{-1}$ when X, Y are independent random variables with respect to the underlying experiment $\mathcal{E}$.

THEOREM 21 *Suppose X, Y are independent random variables with respect to some underlying dichotomy $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta = 0, 1)$. Let $Z = (X, Y)$. Then the insufficiency $\eta(\mathcal{E}X^{-1}, \mathcal{E}Z^{-1})$ is the number*

$$\sup_\lambda ||\lambda(P_0X^{-1}) \wedge (1-\lambda)(P_1X^{-1})|| \, ||(P_0Y^{-1}) - (P_1Y^{-1})||$$

where $\lambda \in [0, 1]$. Furthermore, an upper bound for the insufficiency is given by the number

$$||(P_0X^{-1}) \wedge (P_1X^{-1})|| \, ||(P_0Y^{-1}) - (P_1Y^{-1})||.$$

Proof: Let $\mu = (P_0X^{-1} + P_1X^{-1})$ and $\nu = (P_0Y^{-1} + P_1Y^{-1})$. Define $f_\theta = dP_\theta Z^{-1}/d(\mu \times \nu)$, $g_\theta = dP_\theta X^{-1}/d\mu$ and $h_\theta = dP_\theta Y^{-1}/d\nu$ when $\theta = 0, 1$. Note that $f_\theta(x, y) = g_\theta(x)h_\theta(y)$ almost everywhere $[\mu \times \nu]$. Thus $|f_0\frac{g_1}{g_0} - f_1| = |h_0 - h_1|g_1$ and $|f_1\frac{g_0}{g_1} - f_0| = |h_0 - h_1|g_0$. Then from Theorem 19 and the Tonelli-Fubini theorem the insufficiency $\eta(\mathcal{E}X^{-1}, \mathcal{E}Z^{-1})$ may be written

$$\begin{aligned} & \sup_\lambda \int |h_0 - h_1|(\lambda g_0 \wedge (1-\lambda)g_1)\, d(\mu \times \nu) \\ = \; & \sup_\lambda (\int \lambda g_0 \wedge (1-\lambda)g_1 \, d\mu)(\int |h_0 - h_1| \, d\nu). \end{aligned}$$

Writing this out we obtain the first expression. As for the upper bound we note that $||\lambda(P_0X^{-1}) \wedge (1-\lambda)(P_1X^{-1})||$ is the minimum Bayes risk for the prior probability λ of the event '$\theta = 0$' and the experiment $\mathcal{E}X^{-1}$. The minimum Bayes risk $||\lambda(P_0X^{-1}) \wedge (1-\lambda)(P_1X^{-1})||$ regarded as a function $b(\,\cdot\,|\mathcal{E}X^{-1})$ of λ, is known as the dual Neyman-Pearson (N-P) function of $\mathcal{E}X^{-1}$. The dual (N-P) function satisfies the inequality

$$2b(\frac{1}{2}|\mathcal{E}X^{-1}) \geq b(\lambda|\mathcal{E}X^{-1})$$

for all $\lambda \in [0, 1]$, see Torgersen 1991, p.46. This gives the upper bound. □

EXAMPLE 3 Let $X_1, X_2, \ldots$ be a sequence of independent copies of $X \sim N(\theta, 1)$, where the expectation $\theta \in \{\theta_0, \theta_1\}$ is unknown. Let $\mathcal{E} = (\Re, \mathcal{R}; P_\theta : \theta = \theta_0, \theta_1)$, where $\mathcal{R}$ is the family of Borel sets on the real line $\Re$ and $P_\theta = N(\theta, 1)$.

Let $\mathcal{E}^n$ denote the n-th order product of $\mathcal{E}$. Clearly $\mathcal{E}^n$ is the experiment induced by $(X_1, \ldots, X_n)$. Simple calculus yields

$$\frac{1}{2}||P_{\theta_0} - P_{\theta_1}|| = \text{Prob}\{|Z| \leq \frac{|\theta_0 - \theta_1|}{2}\}$$

and

$$||P^n_{\theta_0} \wedge P^n_{\theta_1}|| = \text{Prob}\{|Z| > \frac{\sqrt{n}|\theta_0 - \theta_1|}{2}\}$$

where $Z \sim N(0,1)$. It is also easy to verify that $\sup_\lambda ||\lambda P^n_{\theta_0} \wedge (1-\lambda)P^n_{\theta_1}|| = \frac{1}{2}||P^n_{\theta_0} \wedge P^n_{\theta_1}||$. Thus, from Theorem 21 we have

$$\eta(\mathcal{E}^n, \mathcal{E}^{n+1}) = \text{Prob}\{|Z| \leq \frac{|\theta_0 - \theta_1|}{2}\}\text{Prob}\{|Z| > \frac{\sqrt{n}|\theta_0 - \theta_1|}{2}\}. \qquad (9)$$

From Feller (1968, Lemma 2, p.175) it follows that the latter probability in (9) converges exponentially to 0 when $\theta_0 \neq \theta_1$.

We remind the reader that the insufficiency $\eta(\mathcal{E}^n, \mathcal{E}^{n+1})$ may be interpreted as the loss of information one suffers by not taking the additional X_{n+1} when already observing $X_1, \ldots, X_n$. In other words; the information contained in the additional observation X_{n+1}.

EXAMPLE 4 Let $X_1, X_2, \ldots$ be a sequence of independent copies of a uniformly distributed $X \sim \mathcal{U}[0,\theta]$ where $\theta \in \{\theta_0, \theta_1\}$, $0 < \theta_0 \leq \theta_1$. For each natural number n, let $\mathcal{E}^n$ denote the experiment realized by $X_1, \ldots, X_n$. Then simple calculus yields

$$\sup_\lambda \int |\lambda(\frac{1}{\theta_1})^n I_{[0,\theta_1]^n}(\mathbf{x}) \wedge (1-\lambda)(\frac{1}{\theta_0})^n I_{[0,\theta_0]^n}(\mathbf{x})|\, d\mathbf{x} = \frac{(\frac{1}{\theta_1})^n}{(\frac{1}{\theta_0})^n + (\frac{1}{\theta_1})^n},$$

and furthermore

$$\int |\frac{1}{\theta_0} I_{[0,\theta_0]}(x) - \frac{1}{\theta_1} I_{[0,\theta_1]}(x)|\, dx = 2(1 - \frac{\theta_0}{\theta_1}).$$

Thus we have

$$\eta(\mathcal{E}^n, \mathcal{E}^{n+1}) = 2(1 - \frac{\theta_0}{\theta_1})(\frac{\theta_0}{\theta_1})^n \frac{1}{1 + (\frac{\theta_0}{\theta_1})^n}$$

which converges exponentially to 0 when $\theta_0 < \theta_1$.

We will now return to the problems of the inequality (1). We proved in Section 4 that one always have that $\eta = \eta_2$. The next example shows a situation where $\eta \neq \eta_1$.

EXAMPLE 5 Let $\mathcal{X} = \{1,2,3,4\}$ and let $\mathcal{A}$ be the family of all subsets of $\mathcal{X}$. For each $\alpha \in [0,1]$ define the experiment $\mathcal{E}_\alpha$ by

$\mathcal{E}_\alpha$	1	2	3	4
P_0	0	α /2	(2-α)/4	(2-α)/4
P_1	(2-α)/4	(2-α)/4	α/2	0

Consider the sub-σ-algebra $\mathcal{B}_0 = \{\emptyset, \mathcal{X}\}$ and let $\mathcal{B}_1$ be the σ-algebra generated by the subset $\{1,2\}$. From Theorem 19 and Corollary 20 we get

$$\begin{aligned}
\eta(\mathcal{B}_0, \mathcal{E}_\alpha) &= \max\{\alpha/2, (1-\alpha)\}\,; \\
\eta(\mathcal{B}_1, \mathcal{E}_\alpha) &= \alpha/2\,; \\
\eta(\mathcal{B}_0, (\mathcal{E}_\alpha|\mathcal{B}_1)) &= 1-\alpha.
\end{aligned}$$

In particular, when $\alpha = 2/3$ we have that

$$\eta(\mathcal{B}_0, \mathcal{E}_\alpha) = \eta(\mathcal{B}_1, \mathcal{E}_\alpha) = \eta(\mathcal{B}_0, (\mathcal{E}_\alpha|\mathcal{B}_1)) = 1/3.$$

Note that $\mathcal{B}_0$ is not sufficient for $(\mathcal{E}_\alpha|\mathcal{B}_1)$ and that $\mathcal{B}_1$ is not sufficient for $\mathcal{E}_\alpha$. Thus we see that the converse statements of Theorem 10(v) & (vi) do not hold (See the discussion in the end of Section 5).

Now, consider the case where $\alpha = 1$. Then both $\mathcal{B}_0$ and $\mathcal{B}_1$ are totally non-informative for the experiment $\mathcal{E}_1$. According to Corollary 20 we have that

$$\eta(\mathcal{B}_1, \mathcal{E}_1) = \frac{1}{2} = \eta(\mathcal{B}_0, \mathcal{E}_1).$$

However, the lack of sufficiency behaves differently. In order to see this, define the experiment $\mathcal{F}$ by

$\mathcal{F}$	1	2	3	4
Q_0	1/6	1/2	1/4	1/12
Q_1	1/12	1/4	1/2	1/6

It is readily seen that $\mathcal{B}_1$ is sufficient for $\mathcal{F}$. Thus we have

$$\eta_1(\mathcal{B}_1, \mathcal{E}_1) \le \sup_\theta ||Q_\theta - P_\theta|| = \frac{1}{3} < \frac{1}{2} = \eta_1(\mathcal{B}_0, \mathcal{E}_1).$$

We see that the lack of sufficiency fails to satisfy the property that corresponds to Theorem 10(v). Thus we have a situation where $\eta \neq \eta_1$.

By using the properties of the insufficiency we may obtain an interesting bound for the statistical distance of direct products.

COROLLARY 22 *Let P, Q be probability measures over the same measurable space $(\mathcal{X}, \mathcal{A})$. Then*

$$||P^n - Q^n|| \le ||P-Q||(1 + 2\sum_{k=1}^{n-1} \sup_{\lambda_k} ||\lambda_k P^k \wedge (1-\lambda_k) Q^k||)$$

where $\lambda_k \in [0,1]$, $k = 1, 2, \cdots, n-1$ and where $\|\cdot\|$ denotes the respective norms of total variation.

Proof: Put $\mathcal{B}_0 = \{\emptyset, \mathcal{X}^n\}$ and let $\mathcal{B}_k : k = 1, 2, \ldots, n$ be the sub-σ-algebras of $\mathcal{A}^n$ where $\mathcal{B}_k$ is the σ-algebra generated by the sets on the form

$$A_1 \times \cdots \times A_k \times \mathcal{X} \times \cdots \times \mathcal{X}$$

where $A_i \in \mathcal{A}$, $i = 1, 2, \ldots, k$. Let $\mathcal{E}$ be the direct product of order n of the binary experiment $(\mathcal{X}, \mathcal{A}; (P, Q))$. From Theorem 10(iv) we have

$$\eta(\mathcal{B}_0, \mathcal{E}) \leq \sum_{k=0}^{n-1} \eta(\mathcal{B}_k, (\mathcal{E}|\mathcal{B}_{k+1})).$$

Writing this out, using Corollary 20 and Theorem 21 we get the desired result. □

References

Aliprantis, C.D./Burkinshaw, O. 1978. Locally solid Riesz spaces, *Academic Press*, New York.

Feller, W. 1968. An Introduction to Probability Theory and Its Applications, vol. I, ed. 3, *Wiley*, New York.

Helgeland, J. 1982. Additional observations and statistical information in the case of 1-parameter exponential distributions. *Z. Wahrsch. Verw. Gebiete.* **59**, pp. 77-100.

Kelley, J.L./Namioka, I. 1963. Linear Topological Spaces. *Van Nostrand*, Princeton, New Jersey.

Le Cam, L. 1964. Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **2**, pp. 1419-55.

Le Cam, L. 1974. On the information contained in additional observations. *Ann. Statist.* **2**, pp. 630-49.

Le Cam, L. 1986. Asymptotic Methods in Statistical Decision Theory. *Springer Verlag*, New York.

Mammen, E. 1986. The statistical information contained in additional observations. *Ann. Statist.* **14**, pp. 665-78.

Schaefer, H.H. 1966. Topological vector spaces. *The Macmillan Company.* New York.

Torgersen, E.N. 1991. Comparison of Statistical Experiments. *Cambridge University Press.*