PROHOROV AND CONTINUOUS MAPPING THEOREMS IN THE HOFFMANN-JØRGENSEN WEAK CONVERGENCE THEORY, WITH APPLICATIONS TO CONVOLUTION AND ASYMPTOTIC MINIMAX THEOREMS

by

Aad W. van der Vaart [1] and Jon A. Wellner[2]

Free University, Amsterdam, and University of Washington, Seattle

January, 1989

Revised: December, 1989

AMS 1980 Subject Classifications: Primary 28A33, 60B10, 60F05, 62G05, 62G20; Secondary 60B12, 60F17, 62G30, 46G05.

Key Words and Phrases: almost uniform convergence, asymptotic minimax theorem, continuous mapping theorems, convergence in outer probability, convolution theorem in function spaces, Dudley's version of the Skorokhod theorem, Markov chain, missing data, net convergence, Prohorov's theorem, regular estimators, stationary distribution, weak convergence.

[1] **Research supported by:** the Netherlands Organization for the Advancement of Pure Research (Z.W.O.)

[2] **Research supported by:** National Science Foundation Grant DMS-8708083 and the John Simon Guggenheim Foundation

PROHOROV AND CONTINUOUS MAPPING THEOREMS IN THE HOFFMANN-JØRGENSEN WEAK CONVERGENCE THEORY, WITH APPLICATIONS TO CONVOLUTION AND ASYMPTOTIC MINIMAX THEOREMS

by

Aad W. van der Vaart and Jon A. Wellner
Free University, Amsterdam and University of Washington

ABSTRACT

First, we review and further develop the general weak convergence theory of Hoffmann-Jørgensen and Dudley for nonmeasurable functions. Our new results include: (i) a generalized Prohorov theorem; (ii) an extension of Le Cam's third lemma; (iii) some new uniformity results.

The general weak convergence theory is then used to formulate several convolution and asymptotic minimax theorems for nonmeasurable "estimators"; these theorems are of statistical interest and importance since they provide lower bounds for estimation. The reformulation using the general weak convergence theory permits a significant weakening of measurability hypotheses. Examples considered include estimation of an unknown measure based on i.i.d. observations, estimation of a bivariate distribution P with missing marginal data, and estimation of the stationary distribution of a Markov chain.

PROHOROV AND CONTINUOUS MAPPING THEOREMS IN THE HOFFMANN-JØRGENSEN WEAK CONVERGENCE THEORY, WITH APPLICATIONS TO CONVOLUTION AND ASYMPTOTIC MINIMAX THEOREMS

by

Aad W. van der Vaart and Jon A. Wellner

Outline

0. Introduction

One of the key goals of asymptotic statistical estimation theory concerns asymptotic efficiency:

Question A: What (asymptotic) "lower bounds" for estimation are available (in a given problem)?

Question B: Can we construct estimators which achieve the bounds?

Since the key papers by Hájek (1970, 1972) and Le Cam (1972), many statisticians have often answered the first question in terms of either "convolution theorems" or "local asymptotic minimax theorems." The first theorems of this type were given by Hájek (1970), (1972) and Le Cam (1972) with earlier important contributions by Le Cam (1953) and Chernoff (1956). Hájek (1972), page 177, gives an interesting discussion of the early history of these theorems. Since then, these theorems have been extended, generalized, refined, and applied to both "nonparametric "and "semiparametric models" (see e.g. Koshevnik and Levit (1976), Levit (1978), Beran (1977), Millar (1979), (1983), and Begun, Hall, Huang, and Wellner (1983)).

More recently, convolution theorems have been developed for estimation of functions in more abstract spaces. Millar (1985) treats convolution theorems for estimators with values in a separable Banach space, while van der Vaart (1988a), (1988b) proves convolution and asymptotic minimax theorems for estimators with values in a general vector space under various measurability assumptions.

Simultaneous with these developments in statistics, the theory of general empirical processes has developed rapidly since the key paper by Dudley (1978) (see e.g. Dudley (1984), Giné and Zinn (1984), Dudley and Philipp (1983), Pollard (1984)), and has spurred the formulation of a more general theory of weak convergence which neatly handles measurability difficulties which have encumbered earlier work. This new weak convergence theory is due to Hoffmann-Jørgensen (1984) and Dudley (1985) in a development from earlier work by Dudley (1966) and Wichura (1968), (1970). It generalizes both the weak convergence theory for separable metric spaces as given in Billingsley (1968) and the weak convergence of measures defined on the sigma - field generated by closed balls due to Dudley (1966) as expounded in Gaenssler (1983) and Pollard (1984).

Typically the problem of estimation of abstract functions requires the use of nonseparable metric spaces. This is already apparent in the simple problem of estimating a cumulative distribution or hazard function, which are most naturally viewed as elements of the Skorokhod space D equipped with the supremum metric. Furthermore, for estimating a distribution on a higher dimensional Euclidean space, one usually views this as indexed by a collection of Borel sets such as balls or half spaces, in which case it can be seen as an element of the l^∞ space employed in empirical process theory. In view of this it is necessary to derive lower bound theorems for estimators with values in nonseparable spaces. Since for such spaces the usual 'estimators' will often not be Borel measurable, the new theory of weak convergence provides the right language for the formulation of such results.

Our goal here is to combine these two developments by presenting new convolution and asymptotic minimax theorems (answers to question A) formulated in terms of the general weak convergence theory of Hoffmann-Jørgensen (1984) and Dudley (1985).

We begin in section 1 with a review and synopsis of the the new weak convergence theory, with some new developments. Substantial parts of this section are due to Hoffmann-Jørgensen (1984) and Dudley (1985). Several results, which are crucial for the development in sections 2 and 3, are new, however. The generalizations of Prohorov's theorem given in theorems 1.1 and 1.3 are both new and are important tools in the proofs of the convolution and asymptotic minimax theorems given in sections 2 and 3. The extension of Le Cam's "third lemma" to this new weak convergence theory, which we give in lemma 1.6, is also apparently new. Some new results on the uniformity of convergence (in the new theory) over subclasses are presented in section 1.6. Section 1 concludes with net versions of some of the results in sections 1.1 - 1.6; these are stated in section 1.7. In particular, we state the net version of the generalization of Prohorov's theorem; this is a crucial tool in the proof of the local asymptotic minimax theorem presented in section 3. The proofs of the results for nets are sketched in section 5 and complete proofs are given in an appendix.

A convolution theorem formulated in terms of the new general weak convergence theory is given in section 2. The theorem covers the basic (and, for applications, most important) case of models which satisfy a local asymptotic normality (LAN) condition. Similarly, an asymptotic minimax theorem formulated in terms of the new general weak convergence theory is given in section 3. Four Banach spaces of particular interest for statistical applications are singled out in corollaries. In section 4 we discuss three examples: estimating a measure P based on i.i.d. observations, estimating a bivariate measure with missing marginal data, and estimating the stationary distribution of a Markov chain. The examples raise two interesting problems concerning "question B" which we leave as open: i.e. can the bounds be achieved? Finally, proofs for most of the results of section 1 are given in section 5.

1. Weak Convergence

1.1. Definitions and Basic Results

Often the random functions with which we work are, unfortunately, *not* measurable with respect to the Borel sigma - field **M** of the metric space M, and hence do not induce probability measures thereon. This typically occurs when M is nonseparable; for example $D[0,1]$ with the supremum (or uniform) metric $\|\cdot\|_\infty$, or $l^\infty(\mathbf{F})$ with the supremum metric $\|\cdot\|_{\mathbf{F}}$. The theory we outline below, due to Hoffmann-Jørgensen (1984) and Dudley (1985) following an evolution from Dudley (1966), gives up the goal of inducing distributions on M equipped with some sigma - field of subsets. It gives a theory of "weak convergence of laws without laws being defined" -- except asymptotically.

Much of the material presented in this section is known or almost known; we give references to original sources, as best known to us, throughout the section. The main new results in this section are: the extension of Prohorov's theorem theorem 1.1); the continuous mapping theorems (propositions 1.1, 1.4, and 1.5); and the extended version of Le Cam's third lemma (lemma 1.6). We have not seen the main reference given above, the manuscript of Hoffmann-Jørgensen (1984), during the writing of this paper, and are, of this date (January, 1990) unaware of its exact contents.

Suppose that $(\Omega, \mathbf{A}, P)$ is a probability space and $f : \Omega \to R$ is a (completely arbitrary) function.

Definition 1.1. f^* denotes any measurable function from $(\Omega, \mathbf{A})$ to $(\bar{R}, \mathbf{B})$ such that

(i) $f^* \geq f$ a.s.

(ii) If $h \geq f$ and h is measurable, then $h \geq f^*$ a.s.

We will show below that f^* exists and is unique. Define

(1) $$f_* \equiv -((-f)^*).$$

Then $f_* : (\Omega, \mathbf{A}) \to \bar{R}$ is measurable with

(iii) $f_* \leq f$ a.s.

(iv) If $h \leq f$ and h is measurable, then $h \leq f^*$ a.s..

We summarize (i) and (ii), and (iii) and (iv), respectively, by the notation

(2) $$f^* \equiv \text{ess.inf}\{ h : h \geq f \,,\ h \text{ measurable} \}$$
$$f_* \equiv \text{ess.sup}\{ h : h \leq f \,,\ h \text{ measurable} \}.$$

Let

(3) $$E_P^* f \equiv \inf\{ E_P h : f \leq h \text{ and } h \text{ is measurable} \},$$
$$E_{*P} f \equiv \sup\{ E_P h : f \geq h \text{ and } h \text{ is measurable} \}.$$

It is shown in lemma 1.2 below that

(4) $\quad E_P^* f = E_P f^* \quad$ and $\quad E_{*P} f = E_P f_*$.

We also define, as usual,

$$P^*(A) \equiv \inf\{P(B) : B \supset A\ ,\ B \in \mathbf{A}\},$$
$$P_*(A) \equiv \sup\{P(B) : B \subset A\ ,\ B \in \mathbf{A}\}.$$

Our first two lemmas summarize the basic properties of f^* and f_* . All the proofs for this section are given in section 5. The following lemma 1.1 and (i), (ii), and (xi) of lemma 1.2 are also given in Dudley and Philipp (1983), while (xiii) of lemma 1.2 is in Dudley (1984) lemma 3.1.6.

Lemma 1.1. f^* exists; moreover we can choose $f^* \geq f$ everywhere.

Lemma 1.2. (Some facts about f^* and f_*)

(i) $(f+g)^* \leq f^* + g^*$ a.s.

(ii) $(f-g)^* \geq f^* - g^*$ a.s. whenever both sides are defined a.s.

(iii) $|f^* - g^*| \leq |f - g|^*$ a.s. whenever both sides are defined a.s.

(iv) $(f+g)_* \geq f_* + g_*$ a.s.

(v) $(f-g)_* \leq f_* - g_*$ a.s. whenever both sides are defined a.s.

(vi) $|f_* - g_*| \leq |f - g|^*$ a.s. whenever both sides are defined a.s.

(vii) If $g : \Omega \to R$ is measurable, then $(f\,g)^* = f^* g\, 1_{[g \geq 0]} + f_* g\, 1_{[g<0]}$ a.s.

(viii) $(1+f)^* = 1 + f^*$ a.s., $(1+f)_* = 1 + f_*$ a.s.

(ix) For any $A \subset \Omega$ there is a measurable set $A_0 \in \mathbf{A}$ such that $A \subset A_0$ and $1_A^* = 1_{A_0}$ a.s.

(x) $1_A^* + 1_{A^c *} = 1$ a.s.

(xi) $E_P^* f = E_P f^*$ and $E_{*P} f = E_P f_*$.

(xii) $P^*(A) = E(1_A)^*$.

(xiii) $1_{[f>\varepsilon]}^* = 1_{[f^*>\varepsilon]}$ a.s. and $P^*(f>\varepsilon) = P(f^*>\varepsilon)$.

Suppose that (M, d) is a metric space (nonseparable in general), $\{(\mathrm{X}_n, \mathrm{A}_n, P_n)\}_{n \geq 0}$ is a sequence of probability spaces, and

(5) $\quad \mathbb{X}_n : \mathrm{X}_n \to M\ , \quad$ for $n = 0, 1, 2, \cdots$

are arbitrary maps. Let $C_b(M)$ be the collection of bounded, continuous functions h from M to R .

Definition 1.2. We say that $\mathbb{X}_n$ *converges weakly to a random element* $\mathbb{X}_0$ in $(M, \mathbf{M})$, and write $\mathbb{X}_n \Rightarrow \mathbb{X}_0$, if for every $h \in \mathrm{C}_b(M)$,

(6) $$E^* h(\mathbf{X}_n) \to E\, h(\mathbf{X}_0)\,, \quad \text{as} \quad n \to \infty\,.$$

Call $\mathbf{X}_0$ *separable* if there exists a separable Borel set $M_0 \subset M$ with $P_0(\mathbf{X}_0 \in M_0) = 1$. With the possible exception of set-theoretic pathological cases, it is no loss of generality to assume that a random element in the Borel sigma-field is separable. See the discussion in Dudley (1985), pages 148 - 149, and Dudley (1976), theorem 5.5.

Since (6) holds for $-h$, it follows from definition 1.2 that, for $h \in C_b(M)$, also

(7) $$E_* h(\mathbf{X}_n) \to E\, h(\mathbf{X}_0) \quad \text{as} \quad n \to \infty\,.$$

Note that to show $\mathbf{X}_n \Rightarrow \mathbf{X}_0$, it suffices to prove

(8) $$\limsup_{n \to \infty} E^* h(\mathbf{X}_n) \le Eh(\mathbf{X}_0)$$

for all $h \in C_b(M)$: (8) implies that

$$\limsup_{n \to \infty} E^*(-h(\mathbf{X}_n)) \le E(-h(\mathbf{X}_0))\,,$$

or

$$\liminf_{n \to \infty} E_* h(\mathbf{X}_n) \ge Eh(\mathbf{X}_0)\,,$$

and hence

(9) $$Eh(\mathbf{X}_0) \le \liminf E_* h(\mathbf{X}_n) \le \liminf E^* h(\mathbf{X}_n) \le \limsup E^* h(\mathbf{X}_n) \le Eh(\mathbf{X}_0)\,,$$

so that (6) holds.

An appropriate generalization of the tightness definition for this theory is as follows:

Definition 1.3. The sequence $\{\mathbf{X}_n\}_{n \ge 1}$ is *tight* if and only if for every $\varepsilon > 0$ there exists a compact set $K \subset M$ such that for all $\delta > 0$

(10) $$\liminf_n P_{n*}(\mathbf{X}_n \in K^\delta) \ge 1 - \varepsilon\,;$$

where $K^\delta \equiv \{x \in M : d(x, K) < \delta\}$, or, equivalently

(11) $$\limsup_n P_n^*(\mathbf{X}_n \notin K^\delta) < \varepsilon\,.$$

This definition is due to Hoffmann-Jørgensen (1984) in a development from Dudley (1966) and (1976, page 23.3); see Andersen and Dobrić (1987), page 167.

Call a random element $\mathbf{X}_0$ in (M, M) *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset M$ such that $P_0(\mathbf{X}_0 \in K) \ge 1 - \varepsilon$. We note that, by theorem 3.2 of Parthasarathy (1967) or theorem 1.4 of Billingsley (1968), any separable $\mathbf{X}_0$ in a *complete* metric space (M, d) is tight.

With these definitions, most of the usual results in the theory of weak convergence of measures on metric spaces as outlined in Billingsley (1968), and elsewhere, have analogues or extensions to the nonmeasurable HJ - Dudley theory. One notable exception, however, is the direct part of Prohorov's theorem (theorem 6.1, page 37, Billingsley (1968)), as will be seen in example 1.1 below. We begin with analogues of several results from the standard theory. The following theorem is stated by Andersen and Dobrić (1987) (see their remark 2.13, page 168), and has been also used by Giné and Zinn (1986), page 61; we do not claim it as new.

Lemma 1.3. (Portmanteau theorem). The following statements are equivalent:

(i) $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ as $n \to \infty$.

(ii) $\liminf_{n\to\infty} P_{n*}(\mathbb{X}_n \in G) \geq P_0(\mathbb{X}_0 \in G)$ for every open set $G \subset M$.

(iii) $\limsup_{n\to\infty} P_n^*(\mathbb{X}_n \in F) \leq P_0(\mathbb{X}_0 \in F)$ for every closed set $F \subset M$.

(iv) $\liminf_{n\to\infty} E_* h(\mathbb{X}_n) \geq Eh(\mathbb{X}_0)$ for every bounded, lower semi-continuous function h.

(v) $\limsup_{n\to\infty} E^* h(\mathbb{X}_n) \leq Eh(\mathbb{X}_0)$ for every bounded, upper semi-continuous function h.

(vi) $\lim_{n\to\infty} P_n^*(\mathbb{X}_n \in A) = P(\mathbb{X}_0 \in A)$ for every Borel set A with $P_0(\mathbb{X}_0 \in \partial A) = 0$.

Lemma 1.4. (Weak convergence to a tight limit implies tightness). Suppose that $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ where $\mathbb{X}_0$ is tight. Then $\{\mathbb{X}_n\}$ is tight.

For $n = 1, 2, \cdots$, let $\mathbb{X}_n : \mathrm{X}_n \to M_1$ and $\mathbb{Y}_n : \mathrm{X}_n \to M_2$ be maps into metric spaces M_1 and M_2. Equip $M_1 \times M_2$ with the metric

$$d((x_1,x_2),(y_1,y_2)) = d_1(x_1,y_1) \vee d_2(x_2,y_2).$$

Lemma 1.5. (Marginal tightness implies joint tightness). If both $\{\mathbb{X}_n\}$ and $\{\mathbb{Y}_n\}$ are tight, then $\{(\mathbb{X}_n, \mathbb{Y}_n)\}$ is tight too.

Proposition 1.1. (Continuous mapping theorem; CM). Suppose that:

(i) $g : M \to M'$ is continuous on a Borel set $M_0 \subset M$.

(ii) $\mathbb{X}_0$ is Borel measurable and $P_0(\mathbb{X}_0 \in M_0) = 1$.

Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ implies that $g(\mathbb{X}_n) \Rightarrow g(\mathbb{X}_0)$.

1.2. Le Cam's third lemma

For $n = 1, 2, \cdots$ let P_n and Q_n be probability measures on measurable spaces $(\mathrm{X}_n, \mathrm{A}_n)$. Let Λ_n be the log-likelihood ratio of Q_n with respect to P_n : i.e. given densities q_n and p_n with respect to a σ − finite dominating measure μ_n (e.g. $\mu_n = P_n + Q_n$), let

$$\Lambda_n = \log(\frac{q_n}{p_n}),$$

where $\log \frac{a}{b} = -\infty$ if $a = 0 < b$, $+\infty$ if $b = 0 < a$, and 0 if $a = b$.

Let $\mathbb{X}_n : \mathrm{X}_n \to M$ as before. Equip $M \times \bar{R}$ with the metric

$$d((x,r),(y,s)) = d(x,y) \vee \arctan |r - s| .$$

Lemma 1.6. (An extension of Le Cam's third lemma). Let P_n and Q_n be contiguous, and suppose that

(12) $\qquad (\mathbb{X}_n, \Lambda_n) \Rightarrow (\mathbb{X}, \Lambda) \qquad$ under P_n

where $(\mathbb{X}, \Lambda) : \mathrm{X}_0 \to M \times R$ is Borel measurable. Then

(13) $\qquad \mathbb{X}_n \Rightarrow \mathbb{Z} \qquad$ under Q_n

where $\mathbb{Z} : \mathrm{X}_0 \to M$ is Borel measurable and

(14) $\qquad P(\mathbb{Z} \in B) = E\, 1_B(\mathbb{X}) e^{\Lambda}$.

Furthermore, if $\mathbb{X}$ is separable (or tight), then $\mathbb{Z}$ may be taken to be separable (or tight) too.

Remark. If $\mathbb{X}$ is Borel measurable in M and has separable range, then $(\mathbb{X}, \Lambda)$ is Borel measurable in $(M, \bar{R})$. Proof: Let M_0 be the range of $\mathbb{X}$. Trivially $(\mathbb{X}, \Lambda)$ is measurable as a map in $M_0 \times \bar{R}$ with the product $\sigma-$field of the Borel $\sigma-$fields of M_0 and $\bar{R}$. The latter $\sigma-$field equals the Borel $\sigma-$field of $M_0 \times \bar{R}$, by separability of M_0 and R. Now for any Borel set B in $M \times \bar{R}$, $\{(\mathbb{X}, \Lambda) \in B\} = \{(\mathbb{X}, \Lambda) \in B \cap (M_0 \times \bar{R})\} \in \mathrm{A}_0$ since $B \cap (M_0 \times \bar{R})$ is a Borel set in $M_0 \times \bar{R}$.

1.3. Prohorov's theorem

Here is an example to show that Prohorov's theorem (tightness implies relative compactness) fails for $\Rightarrow$ without additional hypotheses.

Example 1.1. Let $(\mathrm{X}_n, \mathrm{A}_n, P_n) = (\mathrm{X}, \mathrm{A}, P) = ([0,1], \{\emptyset, [0,1]\}, \lambda)$ for all $n \geq 1$ where λ denotes Lebesgue measure. Let $M = [0,1]$, and define $\mathbb{X}(\omega) = \omega$ for $\omega \in \mathrm{X} \equiv [0,1]$. Consider the sequence $\{\mathbb{X}_n\}$ defined by $\mathbb{X}_n \equiv \mathbb{X}$, $n \geq 1$. Then $P_{n*}(\mathbb{X}_n \in [0,1]) = \lambda_*(\mathbb{X} \in [0,1]) = 1$ for all $n = 1, 2, \cdots$, so $\{\mathbb{X}_n\}$ is tight. But $\mathbb{X}^* = 1$, $\mathbb{X}_* = 0$; and hence for the bounded, continuous function $h(x) = x$ on M it follows that

$$E^* h(\mathbb{X}_n) = E\, \mathbb{X}_n^* = 1, \quad \text{and}$$

$$E_* h(\mathbb{X}_n) = E\, \mathbb{X}_{n*} = 0 \qquad \text{for all } n = 1, 2, \cdots .$$

Thus $\{\mathbb{X}_n\}$ has no subsequence for which $\Rightarrow$ holds.

Of course the key difficulty here is that $\Rightarrow$ implies some "measurability in the limit" since (6) and (7) imply that

(15) $$E^* h(\mathbf{X}_n) - E_* h(\mathbf{X}_n) \to 0 \quad \text{as } n \to \infty$$

for every $h \in C_b(M)$. In the example there is no measurability to start with, so there is no measurability in the limit, and weak convergence in the sense of definition 1.2 fails; thus it is not true in general that $\mathbf{X} \Rightarrow \mathbf{X}$. In fact $\mathbf{X} \Rightarrow \mathbf{X}$ is true if and only if $\mathbf{X}$ is Borel measurable.

Here is a modification of Prohorov's theorem appropriate for the present (Hoffmann-Jørgensen and Dudley) definition of weak convergence. We first need a definition:

Definition 1.4. We say that a subset $\mathbf{H} \subset C_b(M)$ *approximates the unit ball of* $C_b(M)$ *on compacts* if and only if for every $h \in C_b(M)$ with $\|h\|_\infty \le 1$, $\eta > 0$ and compact set $K \subset M$ there is an $h_0 \in \mathbf{H}$ with $\|h_0\|_\infty \le 2$ and $\sup_{y \in K} |h(y) - h_0(y)| \le \eta$.

Theorem 1.1. (Extension of Prohorov's theorem). Suppose that:

(i) $\{\mathbf{X}_n\}$ is tight.

There exists a subset $\mathbf{H} \subset C_b(M)$ such that:

(ii) $\mathbf{H}$ aproximates the unit ball of $C_b(M)$ on compacts.

(iii) $E^* h(\mathbf{X}_n) - E_* h(\mathbf{X}_n) \to 0$ as $n \to \infty$ for all $h \in \mathbf{H}$.

Then there exists a subsequence $\{\mathbf{X}_{n'}\} \subset \{\mathbf{X}_n\}$ such that $\mathbf{X}_{n'} \Rightarrow$ some tight $\mathbf{X}_0$.

Theorem 1.1 has a number of useful corollaries:

Corollary 1.1. For $n = 1, 2, \cdots$ suppose that $\mathbf{X}_n : \mathrm{X}_n \to M_1$ and $\mathbf{Y}_n : \mathrm{X}_n \to M_2$ are maps with

$$\mathbf{X}_n \Rightarrow \mathbf{X}_0 \quad \text{and} \quad \mathbf{Y}_n \Rightarrow \mathbf{Y}_0,$$

where $\mathbf{X}_0$ and $\mathbf{Y}_0$ are tight Borel measurable maps into M_1 and M_2 respectively. Then there exists a subsequence $\{n'\} \subset \{n\}$ such that

$$(\mathbf{X}_{n'}, \mathbf{Y}_{n'}) \Rightarrow (\mathbf{X}_0, \mathbf{Y}_0) \quad \text{as } n' \to \infty$$

for some tight joint law $\mathbf{L}(\mathbf{X}_0, \mathbf{Y}_0)$ on the product space $(M_1 \times M_2, \mathbf{M}_1 \times \mathbf{M}_2)$.

As in classical weak convergence theory for separable metric spaces, joint convergence for the full (joint) sequence in corollary 1.1 fails in general: consider $\{(\mathbf{X}_n, \mathbf{Y}_n);\ n \ge 1\}$ in $[0,1]^2$ with $(\mathbf{X}_{2n}, \mathbf{Y}_{2n})$ uniformly distributed on the line $y = x$ and $(\mathbf{X}_{2n+1}, \mathbf{Y}_{2n+1})$ uniformly distributed on the line $y = 1 - x$. Then $\mathbf{X}_n \sim$ Uniform(0,1) and $\mathbf{Y}_n \sim$ Uniform(0,1) for all n so that $\mathbf{X}_n \Rightarrow \mathbf{X}_0$ and

$\mathbb{Y}_n \Rightarrow \mathbb{Y}_0$ with $\mathbb{X}_0$ and $\mathbb{Y}_0 \sim$ Uniform(0,1). But the full sequence $\{(\mathbb{X}_n, \mathbb{Y}_n)\,;\; n \geq 1\}$ does not converge weakly. Joint convergence for the full sequence can hold under additional hypotheses; see e.g. Billingsley (1968), page 27, theorems 4.4 and 4.5.

For an arbitrary collection $\mathbf{F}$, let $l^\infty(\mathbf{F})$ denote the collection of all bounded real functions on $\mathbf{F}$. It will be equipped with the uniform norm $\|z\|_\infty = \sup_{f \in \mathbf{F}} |z(f)|$.

Corollary 1.2. Suppose that $M = l^\infty(\mathbf{F})$ and that:

(i) $\{\mathbb{X}_n\}$ is tight.

(ii) $\{\mathbb{X}_n(f)\}$ is $\mathbf{A}_n - \mathbf{B}$ measurable for every $f \in \mathbf{F}$ and $n = 1, 2, \cdots$.

Then there exist $\{n'\} \subset \{n\}$ such that $\mathbb{X}_{n'} \Rightarrow$ some tight $\mathbb{X}_0$.

When $M = l^\infty(\mathbf{F})$, (asymptotic) tightness of $\mathbb{X}_n$ in the sense of definition 1.3 is equivalent to (asymptotic) ρ − equicontinuity of $\mathbb{X}_n$ for some pseudo - metric ρ on $\mathbf{F}$ together with (asymptotic) boundedness; see e.g. Andersen and Dobrić (1987), page 167. This parallels the classical results in Billingsley (1968), theorem 8.2. Also, a tight law $\mathbf{L}(\mathbb{X}_0)$ on $l^\infty(\mathbf{F})$ is completely determined by its finite - dimensional laws $\mathbf{L}(\mathbb{X}_0(f_1), \cdots, \mathbb{X}_0(f_m))$. Hence if (i) and (ii) of corollary 1.2 hold and, moreover,

$$\mathbf{L}(\mathbb{X}_n(f_1), \cdots, \mathbb{X}_n(f_m)) \rightarrow \mathbf{L}(\mathbb{X}_0(f_1), \cdots, \mathbb{X}_0(f_m)),$$

then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$.

Corollary 1.3. Let $\mathbf{A}$ be a sigma - field of subsets of M such that for every pair $x_1 \neq x_2$ in M there exists an $\mathbf{A}$ − measurable $h \in C_b(M)$ with $h(x_1) \neq h(x_2)$. Suppose that:

(i) $\{\mathbb{X}_n\}$ is tight.

(ii) $\{\mathbb{X}_n\}$ is $\mathbf{A}_n - \mathbf{A}$ measurable for every $n = 1, 2, \cdots$.

Then there exists $\{n'\} \subset \{n\}$ such that $\mathbb{X}_{n'} \Rightarrow$ some tight $\mathbb{X}_0$.

An example is obtained by letting $\mathbf{A}$ be the sigma - field generated by the closed balls (cf. Gaenssler (1983), Dudley (1966)). Then $h(x) = (1 - p\, d(x, x_1))^+$ satisfies $h(x_1) = 1$ and $h(x_2) = 0$ for sufficiently large p. Actually, corollary 2 is a special case of corollary 3, too.

1.4. Measurability in the Ball σ − field

An alternative theory of weak convergence due to Dudley (1966) (see also Gaenssler (1983) and Gaenssler and Schneemeier (1988)) is applicable when the $\mathbb{X}_n$'s are measurable in the sigma - field $\mathbf{M}_b$ of subsets of M generated by the closed balls. In this theory, $\mathbb{X}_n \Rightarrow_{Dudley} \mathbb{X}_0$ means that

(16) $\qquad Eh(\mathbb{X}_n) \rightarrow Eh(\mathbb{X}_0) \quad$ for all $\; h \in C_b(M, \mathbf{M}_b)$

where $C_b(M, \mathbf{M}_b)$ is the collection of all bounded continuous functions defined on M which are $\mathbf{M}_b$ – measurable. In view of results of Dudley (1966) (cf. Gaenssler (1983), theorem 28, page 47), (16) is equivalent to

$$\text{(17)} \qquad \int^* h\, dP_n \circ \mathbf{X}_n^{-1} \to \int h\, dP_0 \circ \mathbf{X}_0^{-1} \qquad \text{for all } h \in C_b(M)$$

provided $\mathbf{X}_0$ has separable range.

When $\mathbf{X}_n$ is measurable in the ball sigma - field $\mathbf{M}_b$ and $\mathbf{X}_0$ has separable range the following proposition says that $\Rightarrow_{Dudley}$ is equivalent to $\Rightarrow$.

Proposition 1.2. (Equivalence of $\Rightarrow_{Dudley}$ and $\Rightarrow$ under $\mathbf{M}_b$ – measurability). Suppose that $\mathbf{X}_n : \mathrm{X}_n \to M$ is $\mathbf{M}_b$ – measurable for $n = 0, 1, \cdots$ and $P_0 \circ \mathbf{X}_0^{-1}$ is concentrated on a separable subset M_0 of M. Then $\mathbf{X}_n \Rightarrow_{Dudley} \mathbf{X}_0$ if and only if $\mathbf{X}_n \Rightarrow \mathbf{X}_0$.

On the other hand, Dudley (1985) gives an example to show that $\mathbf{X}_n \Rightarrow \mathbf{X}_0$ is *not* equivalent to

$$\int^* h\, dP_n \circ \mathbf{X}_n^{-1} \to Eh(\mathbf{X}_0) \qquad \text{for all } h \in C_b(M),$$

(where each $P_n \circ \mathbf{X}_n^{-1}$ is defined on $\{B \subset M : \mathbf{X}_n^{-1}(B) \in \mathbf{A}_n\}$), which would be a natural extension of (16) in the case that the $\mathbf{X}_n$'s are not $\mathbf{M}_b$ – measurable.

1.5. Convergence in outer probability, convergence almost uniformly, and continuous mapping theorems

Now suppose that every $\mathbf{X}_n$ is defined on the same probability space $(\mathbf{X}, \mathbf{A}, P)$, $n = 0, 1, \cdots$. We say that $\mathbf{X}_n$ converges *almost uniformly* to $\mathbf{X}_0$ if

$$d(\mathbf{X}_n, \mathbf{X}_0)^* \to_{a.s.} 0 \qquad \text{as } n \to \infty.$$

We say that $\mathbf{X}_n$ *converges in outer probability* to $\mathbf{X}_0$ if

$$d(\mathbf{X}_n, \mathbf{X}_0)^* \to_p 0 \qquad \text{as } n \to \infty.$$

It is clear that convergence almost uniformly implies convergence in outer probability. By lemma 1.2.xiii, the latter is equivalent to, for every $\varepsilon > 0$

$$\text{(18)} \qquad P^*(d(\mathbf{X}_n, \mathbf{X}_0) > \varepsilon) \to 0 \qquad \text{as } n \to \infty.$$

The definition of almost uniform convergence has a similar translation:

Proposition 1.3. (Dudley). Let $(\mathbf{X}, \mathbf{A}, P)$ be a probability space, (M, d) a metric space, and $\mathbb{X}_n$ functions from $\mathbf{X}$ into M, $n = 0, 1, \cdots$. Then the following are equivalent:

A. $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_{a.s.} 0$;

B. For every $\varepsilon > 0$, $P^*(\sup_{m \geq n} d(\mathbb{X}_m, \mathbb{X}_0) > \varepsilon) \to 0$ as $n \to \infty$;

C. For each $\delta > 0$, there is some $B \in \mathbf{A}$ with $P(B) > 1 - \delta$ such that $\mathbb{X}_n \to \mathbb{X}_0$ uniformly on B;

Proof. See proposition 1.1 of Dudley (1985). □

Both convergence in outer probability and almost uniformly are preserved by a continuous map if the limit is Borel measurable.

Proposition 1.4. (Continuous mapping theorem; CM continued). Suppose that:

(i) $g : M \to M'$ is continuous on a Borel set $M_0 \subset M$.

(ii) $\mathbb{X}_0$ is Borel measurable and $P_0(\mathbb{X}_0 \in M_0) = 1$.

Then:

A. $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_p 0$ implies that $d'(g(\mathbb{X}_n), g(\mathbb{X}_0))^* \to_p 0$.

B. $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_{a.s.} 0$ implies that $d'(g(\mathbb{X}_n), g(\mathbb{X}_0))^* \to_{a.s.} 0$.

The continuous mapping (or Mann - Wald) theorems given in propositions 1.1 and 1.4 have a further useful extension to a sequence of functions g_n. The following proposition generalizes Billingsley (1968), theorem 5.5, pages 33 - 34, and Gaenssler (1983), theorem 5, page 56. The basic idea is apparently due to H. Rubin; see Topsøe (1967) for a discussion and a refined version of the theorem in the standard (separable M) weak convergence theory. It is a very useful result in connection with Hadamard - differentiable functions and the delta method; see Wellner (1989) and Sheehy and Wellner (1988).

Proposition 1.5. (Extended continuous mapping theorem; CM_n). Suppose that g, $g_n : M \to M'$ are functions which satisfy:

(i) $\mathbb{X}_0$ is Borel measurable and $P_0(\mathbb{X}_0 \in M_0) = 1$ for some Borel subset M_0 of M.

(ii) For every $x \in M_0$ and every sequence $\{x_n\}$ with $x_n \to x$, $g_n(x_n) \to g(x)$.

Then the restriction of g to M_0 is continuous and:

A. $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ implies $g_n(\mathbb{X}_n) \Rightarrow g(\mathbb{X}_0)$.

B. $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_p 0$ implies $d'(g_n(\mathbb{X}_n), g(\mathbb{X}_0))^* \to_p 0$.

C. $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_{a.s.} 0$ implies $d'(g_n(\mathbb{X}_n), g(\mathbb{X}_0))^* \to_{a.s.} 0$.

Convergence in outer probability is stronger than convergence $\Rightarrow$:

Lemma 1.7. Let every $\mathbb{X}_n$ be defined on the same probability space $(\mathbf{X}, \mathbf{A}, P)$ and assume $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_p 0$ where $\mathbb{X}_0$ is Borel measurable. Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$.

This lemma is actually a special case of the following lemma:

Lemma 1.8. For $n = 0, 1, \cdots$ let $\mathbb{X}_n : \mathbf{X}_n \to M$ and $\mathbb{Y}_n : \mathbf{X}_n \to M$ be arbitrary maps. Suppose that $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ where $\mathbb{X}_0$ is Borel measurable and $d(\mathbb{X}_n, \mathbb{Y}_n)^* \to_p 0$. Then $\mathbb{Y}_n \Rightarrow \mathbb{X}_0$.

The almost surely convergent construction of Skorokhod (1956) (see e.g. Billingsley (1971) has a counterpart in the Hoffmann-Jørgensen theory of weak convergence; the counterpart is due to Dudley (1985). To state Dudley's theorem, we first introduce the notion of a perfect function.

Definition 1.5. Let (Ω, Σ, Q) be a probability space, let $(\mathbf{X}, \mathbf{A})$ be a measurable space, and let $\phi : \Omega \to \mathbf{X}$ be measurable. Let $P = Q \circ \phi^{-1}$. Then ϕ is *perfect* if and only if for any bounded real - valued function g on $\mathbf{X}$

$$P^*(g) = Q^*(g \circ \phi); \quad \text{i.e.} \quad E_P^* g = E_Q^* g \circ \phi .$$

As shown by Dudley (1985), theorem 2, page 146, this is equivalent to $(g \circ \phi)^* = g^* \circ \phi$ a.s. Q and to $P^* = Q^* \circ \phi^{-1}$.

The key property of perfect functions that makes them useful and important is the following: suppose that $\mathbb{X} : (\mathbf{X}, \mathbf{A}, P) \to (M, d)$ where (M, d) is a metric space. Suppose there exists a perfect measurable function $\phi : (\tilde{\Omega}, \tilde{\Sigma}, \tilde{Q}) \to (\mathbf{X}, \mathbf{A}, P)$ with $P = \tilde{Q} \circ \phi^{-1}$. Define $\tilde{\mathbb{X}} : (\tilde{\mathbf{X}}, \tilde{\Sigma}) \to (M, d)$ by $\tilde{\mathbb{X}} = \mathbb{X} \circ \phi$. Then for *any* set $B \subset M$

$$(19) \qquad \tilde{Q}^*(\tilde{\mathbb{X}} \in B) = P^*(\mathbb{X} \in B);$$

this follows from the definition since, with $g(\omega) \equiv 1_{[\mathbb{X}(\omega) \in B]}$ for $\omega \in \mathbf{X}$,

$$P^*(\mathbb{X} \in B) = \tilde{Q}^*(\mathbb{X} \circ \phi \in B) = \tilde{Q}^*(\tilde{\mathbb{X}} \in B).$$

In this sense we can say that "$\tilde{\mathbb{X}} =_d \mathbb{X}$".

Here is Dudley's (1985) fourth-generation Skorokhod theorem.

Theorem 1.2. (Skorokhod - Dudley[2] - Wichura). Let (M, d) be any metric space, $(\mathbf{X}_n, \mathbf{A}_n, P_n)$ any probability spaces, and $\mathbb{X}_n$ a function from $\mathbf{X}_n$ into M for each $n = 0, 1, \cdots$. Suppose that $\mathbb{X}_0$ has separable range $M_0 \subset M$ and is measurable. Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ if and only if there exists a probability space $(\tilde{\mathbf{X}}, \tilde{\Sigma}, Q)$ and perfect measurable functions ϕ_n from $(\tilde{\mathbf{X}}, \tilde{\Sigma})$ to $(\mathbf{X}_n, \mathbf{A}_n)$ for each $n = 0, 1, \cdots$ such that:

(i) $Q \circ \phi_n^{-1} = P_n$ on $\mathbf{A}_n$ for each $n = 0, 1, \cdots$,

(ii) $\tilde{\mathbb{X}}_n \equiv \mathbb{X}_n \circ \phi_n \rightarrow \mathbb{X}_0 \circ \phi_0 \equiv \tilde{\mathbb{X}}_0$ almost uniformly; i.e.

$$d(\tilde{\mathbb{X}}_n, \tilde{\mathbb{X}}_0)^* \rightarrow_{a.s.} 0 \quad \text{as } n \rightarrow \infty \tag{20}$$

For a complete discussion of theorem 1.2 and the notion of perfect functions, see Dudley (1985). For another discussion and exposition, see Pollard (1988).

1.6. Uniformity over subclasses $\mathbf{H} \subset C_b(M)$

Now we give several results connected with uniformity of the convergence in (6) over certain collections $\mathbf{H} \subset C_b(M)$. All of the following results depend on the separability of $\mathbb{X}_0$.

Lemma 1.9. Let $\mathbf{H} \subset C_b(M)$ be equi-continuous and bounded. If $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ and $\mathbb{X}_0$ is Borel measurable and separable, then

$$\sup_{h \in \mathbf{H}} |E^* h(\mathbb{X}_n) - Eh(\mathbb{X}_0)| \rightarrow 0 . \tag{21}$$

Now let S be a countable, dense subset of the separable subset $M_0 \subset M$. For $s \in S$, $p \in Q^+$, $q \in Q$, set

$$h_{p,q,s}(x) = q\,(1 - p\,d(x, s))^+ ,$$

and let

$$\mathbf{H}_S \equiv \{ h_{p,q,s} : p \in Q^+, q \in Q, s \in S \},$$

$$\mathbf{H}_1 \equiv \{ \min_{1 \le i \le n} h_i : h_i \in \mathbf{H}_S, i = 1, \cdots, n, n \ge 1 \} .$$

Lemma 1.10. Suppose that $\mathbb{X}_0$ is Borel measurable with separable range M_0. Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ if and only if

$$E^* h(\mathbb{X}_n) \rightarrow Eh(\mathbb{X}_0) \qquad \text{for all } h \in \mathbf{H}_1 . \tag{22}$$

Corollary 1.4. Suppose $\mathbb{X}_0$ is Borel measurable and separable. Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ if and only if

$$E^* h(\mathbb{X}_n) \rightarrow Eh(\mathbb{X}_0) \tag{23}$$

for all uniformly continuous $h \in C_b(M)$.

Corollary 1.5. Suppose $\mathbb{X}_0$ is Borel measurable and separable. Then $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ if and only if

$$d_{BL_1^*}(\mathbf{L}^*(\mathbb{X}_n), \mathbf{L}(\mathbb{X}_0)) \equiv \sup_{h \in BL_1} |E^* h(\mathbb{X}_n) - Eh(\mathbb{X}_0)| \rightarrow 0 \tag{24}$$

where

$$BL_1 \equiv \{h \in C_b(M): |h(x) - h(y)| \le d(x,y) \text{ for all } x, y \in M \text{ and } \|h\|_\infty \le 1\}.$$

Corollary 1.6. Suppose that $\mathbf{X}_0$ is Borel measurable with separable range M_0. Then there exists a countable subset $\mathbf{H}$ of $C_b(M)$ such that $\mathbf{X}_n \Rightarrow \mathbf{X}_0$ if and only if

(25) $$\sup_{h \in \mathbf{H}} |E^* h(\mathbf{X}_n) - Eh(\mathbf{X}_0)| \to 0.$$

1.7. Net Versions of Hoffmann - Jørgensen Convergence Theory

We will have occasion to use the net versions of several of the theorems developed in sections 1.1 - 1.5. Before proceeding, we briefly review some of the key definitions and properties of nets. A *directed set* A is a partially ordered set such that given α_1 and α_2 in A, there exists an α_3 in A with $\alpha_1 \le \alpha_3$ and $\alpha_2 \le \alpha_3$. A *net* $\{x_\alpha : \alpha \in A\}$ is a subset of a set X indexed by a directed set A. A net $\{x_\alpha\}$ in a topological space *converges* to a point x, if for every open neighborhood G of x there is an α_0 such that $x_\alpha \in G$ for every $\alpha \ge \alpha_0$. The *liminf* of a net of real numbers $\{x_\alpha\}$ is defined by $\text{liminf}_\alpha x_\alpha = \lim_\alpha \inf_{\beta \ge \alpha} x_\beta$. A *subnet* $\{x_{\alpha(\beta)} : \beta \in B\}$ is a subset of $\{x_\alpha : \alpha \in A\}$ which is a net itself, of course, and moreover has the property that for every $\alpha_0 \in A$ there is a $\beta_0 \in B$ with $\alpha(\beta) \ge \alpha_0$ for every $\beta \ge \beta_0$ (i.e. $\alpha(\beta)$ is 'eventually past' α_0 for every α_0). A sequence is clearly a net and a subsequence a subnet, but a subnet of a sequence is not necessarily a subsequence. The following two important properties hold:

- A set F is closed if and only if it contains the limits of every convergent net $\{x_\alpha : \alpha \in A\} \subset F$.
- A set K is compact if and only if every net $\{x_\alpha : \alpha \in A\} \subset K$ has a converging subnet with limit in K.

Proofs of these properties can be found in Kelley (1955). If the topological space does not satisfy the 'first axiom of countability' (there is not for every point a countable base for the neighborhoods of that point), then the assertions can be false if 'net' is replaced by sequence. An example of such a space is one which we will encounter in the proof of theorem 1.3 below: the product of uncountably many metric spaces equipped with the product topology is not first countable.

As in section 1.1 suppose that (M, d) is a metric space (nonseparable in general), $\{(\mathbf{X}_\alpha, \mathbf{A}_\alpha, P_\alpha)\}_{\alpha \in A}$ is a net of probability spaces, and

(26) $$\mathbf{X}_\alpha : \mathrm{X}_\alpha \to M, \quad \text{for } \alpha \in A$$

are arbitrary maps. As before, $C_b(M)$ is the collection of bounded, continuous functions h from M to R.

Definition 1.6. We say that $\mathbb{X}_\alpha$ *converges weakly to a random element* $\mathbb{X}_0$ in $(M, \mathbf{M})$, and write $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$, if for every $h \in C_b(M)$,

$$(27) \qquad \lim_\alpha E^* h(\mathbb{X}_\alpha) = E\, h(\mathbb{X}_0).$$

Definition 1.7. The net $\{\mathbb{X}_\alpha\}_{\alpha \in A}$ is *tight* if and only if for every $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon \subset M$ such that for all $\delta > 0$

$$(28) \qquad \liminf_\alpha P_{\alpha *}(\mathbb{X}_\alpha \in K^\delta) \geq 1 - \varepsilon;$$

where $K^\delta \equiv \{x \in M : d(x, K) < \delta\}$, or, equivalently

$$(29) \qquad \limsup_\alpha P^*_\alpha(\mathbb{X}_\alpha \notin K^\delta) < \varepsilon.$$

It is easily checked that lemmas 1.3 - 1.8 and proposition 1.1 go through for nets with only notational changes needed in the proofs (see the appendix for complete statements and proofs). The situation is somewhat different for Prohorov's theorem. It is true that a tight net $\{\mathbb{X}_\alpha\}$ has a converging subnet. However, this is a different result from theorem 1.1 which says that a tight sequence has a converging subquence. The proof of the net version is therefore somewhat different from the proof of theorem 1.1 too. In fact, the proof of the net version is easier because of the second key characterization of compact sets in terms of nets and subnets discussed above.

Theorem 1.3. (Extension of Prohorov's theorem for nets). Suppose that:

(i) $\{\mathbb{X}_\alpha\}_{\alpha \in A}$ is tight.

There exists a subset $\mathbf{H} \subset C_b(M)$ such that:

(ii) $\mathbf{H}$ aproximates the unit ball of $C_b(M)$ on compacts.

(iii) $\lim_\alpha \{E^* h(\mathbb{X}_\alpha) - E_* h(\mathbb{X}_\alpha)\} = 0$ for all $h \in \mathbf{H}$.

Then there exists a subnet $\{\mathbb{X}_{\alpha'}\}_{\alpha' \in A'} \subset \{\mathbb{X}_\alpha\}_{\alpha \in A}$ such that $\mathbb{X}_{\alpha'} \Rightarrow$ some tight $\mathbb{X}_0$.

Corollary 1.7. For $\alpha \in A$ suppose that $\mathbb{X}_\alpha : \mathrm{X}_\alpha \to M_1$ and $\mathbb{Y}_\alpha : \mathrm{X}_\alpha \to M_2$ are maps with

$$\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0 \quad \text{and} \quad \mathbb{Y}_\alpha \Rightarrow \mathbb{Y}_0,$$

where $\mathbb{X}_0$ and $\mathbb{Y}_0$ are tight Borel measurable maps into M_1 and M_2 respectively. Then there exists a subnet $\{(\mathbb{X}_{\alpha'}, \mathbb{Y}_{\alpha'})\}_{\alpha' \in A'} \subset \{(\mathbb{X}_\alpha, \mathbb{Y}_\alpha)\}_{\alpha \in A}$ such that

(30) $$(\boldsymbol{X}_{\alpha'}, \boldsymbol{Y}_{\alpha'}) \Rightarrow (\boldsymbol{X}_0, \boldsymbol{Y}_0)$$

for some tight joint law $\mathbf{L}(\mathbb{X}_0, \mathbb{Y}_0)$ on the product space $(M_1 \times M_2, \mathbf{M}_1 \times \mathbf{M}_2)$.

2. A Convolution Theorem

Now our goal is to use the results of section 1 to establish a convolution theorem for regular "estimators"; our definition of regularity will be formulated in terms of $\Rightarrow$. This is the key difference between theorem 2.1 presented below and earlier results of this type due to Millar (1985) and van der Vaart (1988a, 1988b). The present theorem removes measurability conditions, allows for dependent data (as in Hájek (1970)), and extends the applicability of those results substantially.

Let $\mathbf{H}$ be a linear subspace of some Hilbert space. Write $<\cdot,\cdot>$ for the inner product and $\|\cdot\|$ for the norm.

For $n = 1, 2, \cdots$ and $h \in \mathbf{H}$ let $P_{n,h}$ be a probability measure defined on a measurable space $(\mathbf{X}_n, \mathbf{A}_n)$. Assume that

$$(1) \qquad \log \frac{dP_{n,h}}{dP_{n,0}} = \Delta_{n,h} - \frac{1}{2}\|h\|^2 + o_{P_{n,0}}(1), \qquad \text{for every } h \in \mathbf{H},$$

where $\Delta_{n,h} : \mathbf{X}_n \to R$ are measurable maps with

$$(2) \qquad \mathrm{L}_0(\Delta_{n,h_1}, \cdots, \Delta_{n,h_d}) \to N_d(0, <h_i, h_j>)$$

for every finite subset $h_1, \cdots, h_d \in \mathbf{H}$. This is a very natural generalization of the local asymptotic normality (LAN) assumption of Hájek (1972).

Let $\mathbf{B}$ be a Banach space and $\nu_n(P_{n,h})$ be $\mathbf{B}$ – valued "parameters" such that

$$(3) \qquad R_n(\nu_n(P_{n,h}) - \nu_n(P_{n,0})) \to \dot{\nu}(h), \qquad \text{for every } h \in \mathbf{H}$$

for some sequence of linear maps $R_n : \mathbf{B} \to \mathbf{B}$ with $\|R_n\| \to \infty$ and a continuous linear map $\dot{\nu} : \mathbf{H} \to \mathbf{B}$.

A sequence of maps $T_n : \mathbf{X}_n \to \mathbf{B}$ is said to be *locally regular for* ν_n if, under $P_{n,h}$,

$$(4) \qquad R_n(T_n - \nu_n(P_{n,h})) \Rightarrow \mathbb{Z} \qquad \text{as } n \to \infty,$$

for every $h \in \mathbf{H}$, where $\mathbb{Z}$ is a Borel measurable tight random element in $\mathbf{B}$ which does not depend on $h \in \mathbf{H}$. Note that no measurability hypotheses are made about the mappings T_n.

Theorem 2.1 (Convolution theorem). Suppose (1) - (3) hold and T_n is regular. Then there exist tight Borel measurable elements $\mathbb{Z}_0$ and $\mathbb{W}$ in $\mathbf{B}$ with

A. $P(\mathbb{Z}_0 \in \overline{\dot{\nu}(\mathbf{H})}) = 1$.

B. $\mathrm{L}(\mathbb{Z}) = \mathrm{L}(\mathbb{Z}_0 + \mathbb{W})$.

C. $\mathbb{Z}_0$ and $\mathbb{W}$ are independent.

D. $\mathrm{L}(b^*\mathbb{Z}_0) = N(0, \|\dot{\nu}^T b^*\|^2)$ for every $b^* \in \mathbf{B}^*$.

Remark 2.1. Convolution of Borel measures on a nonseparable Banach space is not well-defined in general. However the sum of two Borel measurable random elements in $\mathbf{B}$ with separable range $\mathbf{B}_0$ (as $\mathbb{Z}_0$ and $\mathbb{W}$ can be taken to have) is Borel measurable; recall the discussion following definition 1.2.

Remark 2.2. Tightness of $\mathbb{Z}_0$ and part D of the conclusion completely determine $\mathbf{L}(\mathbb{Z}_0)$. An equivalent expression of D is that $\{b^*(\mathbb{Z}_0) : b^* \in \mathbf{B}^*\}$ is a mean - zero Gaussian process with covariance function

(5) $$Cov(b_1^*\mathbb{Z}_0, b_2^*\mathbb{Z}_0) = <\dot{\nu}^T b_1^*, \dot{\nu}^T b_2^*> = <\dot{\nu}\dot{\nu}^T b_1^*, b_2^*>$$

where $\dot{\nu}^T : \mathbf{B}^* \to \overline{\mathbf{H}}$ is the adjoint of $\dot{\nu}$ defined by $<h, \dot{\nu}^T b^*> = b^*\dot{\nu}(h)$ for every $h \in \mathbf{H}$ and $b^* \in \mathbf{B}^*$.

Remark 2.3. The set $\overline{\dot{\nu}(\mathbf{H})}$ in A is, in fact, the support of $\mathbf{L}(\mathbb{Z}_0)$. Our map $\dot{\nu}^T$ plays the role of U in the slightly more general setting of theorem 6 of Jain (1977), and $\dot{\nu}\dot{\nu}^T$ is the covariance operator S of Kuelbs (1976) and U^*U of Jain (1977).

Remark 2.4. Regular estimators $\{T_n\}$ which have limit $\mathbb{Z}_0$ (more precisely, limit law $\mathbf{L}(\mathbb{Z}_0)$ in (4)) so that $\mathbb{W} = 0$ in B, have an optimality property. We sometimes call such estimators *Hájek optimal*. This optimality claim can be made more precise in several ways as follows

Define the Levy concentration function of $\mathbf{L}(\mathbb{Z})$ by

$$K_{\mathbb{Z}}(t) = \sup_{x \in \mathbf{B}} P(\|\mathbb{Z} - x\| \le t).$$

$K_{\mathbb{Z}}$ measures the maximum probability assignable by $\mathbf{L}(\mathbb{Z})$ to a ball of radius t in $\mathbf{B}$. Then B - D imply that

(6) $$K_{\mathbb{Z}}(t) \le K_{\mathbb{Z}_0}(t) = P(\|\mathbb{Z}_0\| \le t).$$

To see that the inequality in (6) holds, let $B_t = \{b \in \mathbf{B} : \|b\| \le t\}$. Then

$$K_{\mathbb{Z}}(t) = \sup_{x \in \mathbf{B}} P(\mathbb{Z} \in B_t + x)$$
$$= \sup_{x \in \mathbf{B}} \int P(\mathbb{Z}_0 \in B_t + x - y) dP_{\mathbb{W}}(y),$$

which yields the inequality in (6). The equality in (6) follows from Anderson's (1955) inequality; see (8) below.

Another way that $\mathbb{Z}_0$ (or $\mathbf{L}(\mathbb{Z}_0)$) is optimal is via risks. Let $l : \mathbf{B} \to R$ be convex and symmetric. Then it follows from B and C that

(7) $$E\, l(\mathbb{Z}) \ge E\, l(\mathbb{Z}_0).$$

To see (7), note that by the properties of l,

$$2l(\mathbb{Z}_0) \le l(\mathbb{Z}_0 + \mathbb{W}) + l(\mathbb{Z}_0 - \mathbb{W}) = l(\mathbb{Z}_0 + \mathbb{W}) + l(-\mathbb{Z}_0 + \mathbb{W}),$$

and this yields (7) since $\mathbf{L}(\mathbb{Z}_0) = \mathbf{L}(-\mathbb{Z}_0)$, and hence $\mathbf{L}(-\mathbb{Z}_0 + \mathbb{W}) = \mathbf{L}(\mathbb{Z}_0 + \mathbb{W}) = \mathbf{L}(\mathbb{Z})$.

Inequality (7) also holds for the class of subconvex loss functions l as defined in (3.3) of the next section, and in particular for $l(b) = l_t(b) \equiv 1_{[\|b\| > t]}$, $t \in R$. This follows from B and C and Anderson's (1955) lemma (see Pfanzagl (1985), page 454, and van der Vaart (1988a, theorem 3.20, page 72). Thus, by (7) with $l = l_t$,

(8) $\qquad \|\mathbb{Z}_0\|$ is stochastically smaller than $\|\mathbb{Z}\|$

(i.e. $P(\|\mathbb{Z}_0\| \le t) \ge P(\|\mathbb{Z}\| \le t)$ for all $t > 0$). Note that this also implies the equality on the right side of (6).

Theorem 2.1 has a variety of special cases and corollaries. In particular, the case of independent and identically distributed observations is worth singling out as a corollary.

Let $\mathbf{P}$ be a collection of probability measures on a measurable space $(\mathbf{X}, \mathbf{A})$, and let $(X_1, \cdots, X_n)$ be iid $P \in \mathbf{P}$. Let $(\mathbf{B}, \|\cdot\|_{\mathbf{B}})$ be a Banach space, and suppose $\nu : \mathbf{P} \to \mathbf{B}$. An "estimator" of $\nu(P)$ is a map $T_n = t_n(X_1, \cdots, X_n)$ from $\mathbf{X}^n$ to $\mathbf{B}$; here we use quotes around "estimator" because no measurability assumptions will be imposed on t_n.

Now the differentiability of ν in (3) may be described as follows: Fix $P \in \mathbf{P}$, and let $\mathbf{P}(P)$ be a collection of sequences $\{P_n\}$ in $\mathbf{P}$ such that

$$(8) \qquad \int \left\{ \sqrt{n}\, (dP_n^{1/2} - dP^{1/2}) - \frac{1}{2} h\, dP^{1/2} \right\}^2 \to 0 \qquad \text{as} \quad n \to \infty$$

for some $h \in L_2(P)$. The set $\dot{\mathbf{P}}^0 \equiv \dot{\mathbf{P}}^0(P)$ of all such h's obtained in this way is called the *tangent set* (at P). We assume here that $\dot{\mathbf{P}}^0$ is linear. It is easily seen that $\int h\, dP = 0$ for all $h \in \dot{\mathbf{P}}^0$.

Definition 2.1. $\nu : \mathbf{P} \to \mathbf{B}$ is said to be *pathwise differentiable* at $P \in \mathbf{P}$ if there is a continuous linear map $\dot{\nu} \equiv \dot{\nu}_P$ from $\dot{\mathbf{P}}^0$ to $\mathbf{B}$ such that

$$(9) \qquad \sqrt{n}\, (\nu(P_n) - \nu(P)) \to \dot{\nu}(h) \qquad \text{as} \quad n \to \infty$$

for every $\{P_n\} \subset \mathbf{P}(P)$ satisfying (8).

Now for regularity of $\{T_n\}$ as in (4) becomes:

Definition 2.2. The "estimator" sequence $\{T_n\}$ of $\nu(P)$ is said to be *locally regular* at $P \in \mathbf{P}$ if, under P_n (so $(\mathbf{X}_n, \mathbf{A}_n, P_n)$ of earlier in this section 1 is now $(\mathbf{X}^n, \mathbf{A}^n, P_n^n)$),

$$(10) \qquad \sqrt{n}\, (T_n - \nu(P_n)) \Rightarrow \mathbb{Z} \qquad \text{as} \quad n \to \infty$$

for every $\{P_n\} \in \mathbf{P}(P)$ and a tight Borel measurable random element $\mathbb{Z}$ which does not depend on which $\{P_n\} \in \mathbf{P}(P)$ is chosen.

Corollary 2.1. (Convolution theorem; iid data from $P \in \mathbf{P}$). Suppose that:

(i) ν is pathwise differentiable at $P \in \mathbf{P}$ with derivative $\dot{\nu}$.

(ii) $\{T_n\}$ is $\Rightarrow$ − regular with limit $\mathbb{Z}$.

(iii) $\dot{\mathbf{P}}^0$ is linear.

Then there exist tight Borel measurable random elements $\mathbb{Z}_0$ and $\mathbb{W}$ in $\mathbf{B}$ such that:

A. $P(\mathbb{Z}_0 \in \overline{\dot{\nu}(\dot{\mathbf{P}}^0)}) = 1$.

B. $\mathbf{L}(\mathbb{Z}) = \mathbf{L}(\mathbb{Z}_0 + \mathbb{W})$.

C. $\mathbb{Z}_0$ and $\mathbb{W}$ are independent.

D. $\mathbb{Z}_0$ is mean 0 Gaussian with

$$Cov[b_1^*\mathbb{Z}_0, b_2^*\mathbb{Z}_0] = <\dot{\nu}^T b_1^*, \dot{\nu}^T b_2^*>_P = <\dot{\nu}\dot{\nu}^T b_1^*, b_2^*>.$$

Proof of theorem 2.1. Set

(a) $$\mathbb{Z}_{n,h} = R_n(T_n - \nu_n(P_{n,h})).$$

Let $g \equiv (g_1, \cdots, g_d)^T$ be an orthonormal subset of $\mathbf{H}$, and for every $a \in R^d$ let $\{P_{n,h}\}$ correspond to $h = a^T g \in \mathbf{H}$. Now by (1) the log likelihood ratio $\Lambda_n(P_{n,h}, P_{n,0})$ satisfies

(b) $$\Lambda_n(P_{n,h}, P_{n,0}) = \Delta_{n,h} - \frac{1}{2}\|h\|^2 + o_{P_{n,0}}(1).$$
$$= \Delta_{n,h} - \frac{1}{2}|a|^2 + o_{P_{n,0}}(1).$$

By (2)

$$\mathbf{L}_0(\Delta_{n,g_1}, \cdots, \Delta_{n,g_d}) \to N_d(0, <g_i, g_j>) = N_d(0, I) \equiv \mathbf{L}(S).$$

By corollary 1.1 there exists a subsequence $\{n'\}$ such that, under $P_{n',0}$,

(c) $$(\mathbb{Z}_{n',0}, \Delta_{n',g_1}, \cdots, \Delta_{n',g_m}) \Rightarrow (\mathbb{Z}, S) \quad \text{in} \quad (\mathbf{B} \times R^d, \|\cdot\|_0)$$

where $\|(b, r)\|_0 \equiv \|b\| \vee |r|$. Here $(\mathbb{Z}, S)$ can be taken to be a Borel measurable random element in $\mathbf{B} \times R^d$ with its Borel sigma - field and with values in a set $\mathbf{B}_0 \times R^d = \cup_{m=1}^{\infty} K_{1,m} \times R^d$, where $K_{1,1} \subset K_{1,2} \subset \cdots$ are compact subsets of $\mathbf{B}$.

Again by (2)

$$\mathbf{L}_0(\Delta_{n,g_1}, \cdots, \Delta_{n,g_d}, \Delta_{n,h}) \to \mathbf{L}(S, \Delta_h)$$

where (S, Δ_h) is $(d+1)$ − dimensional Gaussian with covariance determined by (2). From $h = a^T g$ it follows that

$$E(a^T S - \Delta_h)^2 = 0 .$$

But then

$$\sum_{i=1}^{d} a_i \Delta_{n,g_i} - \Delta_{n,h} \to_{P_{n,0}} 0 .$$

Combining this with (b), (c), and (3) we see that under $P_{n',0}$, for every $h = a^T g$, $a \in R^d$,

(d) $$(\mathbb{Z}_{n',h}, \Lambda_{n'}(P_{n',h}, P_{n',0})) \Rightarrow (\mathbb{Z} - \dot{\nu}(h), a^T S - \frac{1}{2}|a|^2)$$
$$= (\mathbb{Z} - a^T \dot{\nu}(g), a^T S - \frac{1}{2}|a|^2) .$$

Hence under $P_{n,h}$, it follows from lemma 1.5 that

(e) $$\mathbb{Z}_{n,h} \Rightarrow \mathbb{Z}_a \quad \text{on} \quad (\mathbf{B}, \|\cdot\|_{\mathbf{B}})$$

where

(f) $$P(\mathbb{Z}_a \in A) = E\, 1_A(\mathbb{Z} - a^T \dot{\nu}(g)) \exp(a^T S - \frac{1}{2}|a|^2) .$$

By regularity of $\{T_n\}$,

(g) $$\mathbf{L}(\mathbb{Z}_a) = \mathbf{L}(\mathbb{Z}) \quad \text{for all} \quad a \in R^d .$$

The remainder of the proof is the same as in van der Vaart (1988b); for completeness we repeat it here.

Let $\mathbf{B}_1$ be a closed separable subspace of $\mathbf{B}$ that contains $\cup_{m=1}^{\infty} K_{1,m}$ and $a^T \dot{\nu}(g)$ for every a. The trace of $\boldsymbol{B}_{Borel}$ on $\mathbf{B}_1$ is the Borel sigma - field of $\mathbf{B}_1$, and $P(\mathbb{Z}_a \in \mathbf{B}_1) = 1$ for every a. Thus, for every $b^* \in \mathbf{B}^*$ and every $a \in R^d$, it follows from (f) and (g) that

(h) $$E\, e^{ib^* \mathbb{Z}} = E \exp(ib^*(\mathbb{Z} - a^T \dot{\nu}(g)) + a^T S - \frac{1}{2}|a|^2) .$$

The left side of (h) is constant in $a \in R^d$ The right side is defined for $a \in C^d$ and is analytic; by uniqueness of analytic continuation, it must be constant on all of C^d. Let $a = -ib^* \dot{\nu}(g)$ to obtain

(i) $$E\, e^{ib^* \mathbb{Z}} = E\, e^{ib^*(\mathbb{Z} - S^T \dot{\nu}(g))} \exp(-\frac{1}{2}|b^* \dot{\nu}(g)|^2)$$
$$= E\, e^{ib^* \mathbb{W}_g} E\, e^{ib^* \mathbb{Z}_g}$$

for every $b^* \in \mathbf{B}^*$ where

$$\mathbb{W}_g \equiv \mathbb{Z} - S^T \dot{\nu}(g) \quad \text{and} \quad \mathbb{Z}_g \equiv S^T \dot{\nu}(g)$$

are random elements in $\mathbf{B}_1$. (The sum of random elements in a separable normed space is a random element.)

Then, given $c^* \in \mathbf{B}^*$, let $a = i(c^* - b^*)\dot{v}(g)$ in (h) to obtain

(j) $$E\, e^{ib^*\mathbb{Z}} = E\, e^{ib^*(\mathbb{Z} - S^T\dot{v}(g)) + ic^* S^T\dot{v}(g)} \exp(-\frac{1}{2}|b^*\dot{v}(g)|^2)\exp(\frac{1}{2}|c^*\dot{v}(g)|^2)\,.$$

$$= E(e^{ib^*\mathbb{W}_g + ic^*\mathbb{Z}_g})\frac{E\, e^{ib^*\mathbb{Z}_g}}{E\, e^{ic^*\mathbb{Z}_g}}\,.$$

By (i) and (j)

(k) $$E\, e^{ib^*\mathbb{W}_g + ic^*\mathbb{Z}_g} = E\, e^{ib^*\mathbb{W}_g} E\, e^{ic^*\mathbb{Z}_g}$$

for every $b^*, c^* \in \mathbf{B}^*$. It follows that $\mathbb{W}_g$ and $\mathbb{Z}_g$ are independent random elements in $\mathbf{B}_1$ and $\mathbf{L}(\mathbb{Z}) = \mathbf{L}(\mathbb{Z}_g + \mathbb{W}_g)$.

Fix $\varepsilon > 0$. For sufficiently large m

(l) $$1 - \varepsilon < P(\mathbb{Z} \in K_{1,m}) = \int P(\mathbb{Z}_g + x \in K_{1,m})\, dP_{\mathbb{W}_g}(x)\,.$$

Hence there exists $x_0 \in \mathbf{B}_1$ such that

(m) $$P(\mathbb{Z}_g \in K_{1,m} - x_0) > 1 - \varepsilon\,.$$

By symmetry of the normal distribution, (m) implies that

(n) $$P(\mathbb{Z}_g \in K_{2,m}) > 1 - 2\varepsilon$$

with $K_{2,m} \equiv \frac{1}{2}(K_{1,m} - K_{1,m})$. By by (l) and (n)

$$1 - \varepsilon < \int_{K_{2,m}} P(y + \mathbb{W}_g \in K_{1,m})\, dP_{\mathbb{Z}_g}(y) + 2\varepsilon\,,$$

so there exists $y_0 \in K_{2,m}$ such that

(o) $$P(\mathbb{W}_g \in K_{1,m} - y_0) > 1 - 3\varepsilon\,.$$

Since $K_{3,m} \equiv K_{1,m} - K_{2,m} \supset K_{1,m} - y_0$, (o) implies

$$P(\mathbb{W}_g \in K_{3,m}) > 1 - 3\varepsilon\,.$$

Let $\mathbf{B}_2$ be a closed separable subspace of $\mathbf{B}$ that contains the union of the $K_{1,m}$, the $K_{2,m}$, and the $K_{3,m}$. For every finite - dimensional subspace of $\mathbf{H}$, take an orthonormal basis g, and carry out the above construction. This yields a collection of random elements $(\mathbb{Z}_g, \mathbb{W}_g)$ in $\mathbf{B}_2 \times \mathbf{B}_2$ with its Borel sigma - field such that:

(p) $\mathbb{Z}_g$ takes values in $\bigcup_m K_{2,m}$,

(q) $\mathbb{W}_g$ takes values in $\bigcup_m K_{3,m}$,

(r) $\mathbb{Z}_g$ and $\mathbb{W}_g$ are independent,

(s) $\mathbf{L}(\mathbb{Z}) = \mathbf{L}(\mathbb{Z}_g + \mathbb{W}_g)$,

(t) $\mathbf{L}(b^* \mathbb{Z}_g)) = N(0, |b^* \dot{\nu}(g)|^2)$,

(u) $P(\mathbb{Z}_g \in \dot{\nu}(\mathbf{H})) = 1$.

Moreover, the set of laws $\{\mathbf{L}(\mathbb{Z}_g, \mathbb{W}_g) : g\}$ where g ranges over the chosen finite orthonormal subsets of $\mathbf{H}$ is tight. Let $[g] \equiv$ linear span of g, and partially order the g's by inclusion of their linear spans: $g_1 \le g_2$ if $[g_1] \subset [g_2]$. By Prohorov's theorem 1.3 the net $\{\mathbf{L}(\mathbb{Z}_g, \mathbb{W}_g) : g\}$ has a convergent subnet, say $\{\mathbf{L}(\mathbb{Z}_g, \mathbb{W}_g) : g = g(\alpha),\ \alpha \in A\}$, which converges weakly to the law of a random element $(\mathbb{Z}_0, \mathbb{W})$ in $\mathbf{B}_2 \times \mathbf{B}_2$. It is clear from (r) that $\mathbb{Z}_0$ and $\mathbb{W}$ are independent and, from (s), that $\mathbf{L}(\mathbb{Z}) = \mathbf{L}(\mathbb{Z}_0 + \mathbb{W})$; thus B and C hold. Furthermore, by (u),

$$P(\mathbb{Z}_0 \in \overline{\dot{\nu}(\mathbf{H})}) \ge \liminf_{\alpha} P(\mathbb{Z}_{g(\alpha)} \in \overline{\dot{\nu}(\mathbf{H})}) = 1,$$

which proves A.

Finally, with (orthogonal) projection in $L_2(P)$ onto $[g]$ denoted by $\Pi(\cdot \mid [g])$,

$$\Pi(\dot{\nu}^T b^* \mid [g]) = a^T g$$

with

$$a \equiv \langle \dot{\nu}^T b^*, g \rangle = b^* \dot{\nu}(g),$$

and hence

$$\|\Pi(\dot{\nu}^T b^* \mid [g])\|^2 = |a|^2 = |b^* \dot{\nu}(g)|^2.$$

It follows that

(v) $|b^* \dot{\nu}(g)|^2 \underset{g}{\rightarrow} \|\dot{\nu}^T b^*\|^2$.

Therefore, by (t) and (u),

(w) $\mathbf{L}(b^* \mathbb{Z}_g) \underset{g}{\rightarrow} N(0, \|\dot{\nu}^T b^*\|^2)$

while

(x) $\mathbf{L}(b^* \mathbb{Z}_{g(\alpha)}) \underset{\alpha}{\rightarrow} \mathbf{L}(b^* \mathbb{Z}_0)$

since $\mathbb{Z}_{g(\alpha)} \Rightarrow \mathbb{Z}_0$. Thus

(y) $\mathbf{L}(b^* \mathbb{Z}_0) = N(0, \|\dot{\nu}^T b^*\|^2)$,

so D holds. □

To see (5) of remark 2.2, use the standard (polarization) identity $\langle x, y\rangle = \frac{1}{2}(\langle x+y, x+y\rangle - \langle x, x\rangle - \langle y, y\rangle)$ as follows: $b_1^* \mathbb{Z}_0$, $b_2^* \mathbb{Z}_0$ are jointly Gaussian and

$$Cov[b_1^* \mathbb{Z}_0, b_2^* \mathbb{Z}_0] = \frac{1}{2}\{Var[(b_1^* + b_2^*)\mathbb{Z}_0] - Var[b_1^* \mathbb{Z}_0] - Var[b_2^* \mathbb{Z}_0]\}$$

$$= \frac{1}{2}\{\|\dot{v}^T(b_1^* + b_2^*)\|^2 - \|\dot{v}^T b_1^*\|^2 - \|\dot{v}^T b_2^*\|^2 \qquad \text{by D}$$

$$= <\dot{v}^T b_1^* , \dot{v}^T b_2^* >$$

$$= <\dot{v}\dot{v}^T b_1^* , b_2^*> \qquad \text{by definition of } \dot{v}^T .$$

3. Asymptotic Minimax Theorems for Gaussian Limits

In this section we obtain an asymptotic minimax theorem for the situation considered in section 2. It generalizes earlier results by Hájek (1972), Levit (1978), and Millar (1983). Let $\mathbf{H}$, $\mathbf{B}$, $P_{n,h}$, and $\nu_n(P_{n,h})$ be as in section 2, and suppose that (2.1) - (2.3) hold. In our main theorem we again do not require measurability of the "estimators," though we impose the following asymptotic measurability condition. Given a linear subspace $\mathbf{B}'$ of $\mathbf{B}^*$, let $T_n : \mathbf{X}_n \to \mathbf{B}$ be maps ("estimators") such that

(1) $$E_0^* f(R_n(T_n - \nu_n(P_{n,0})) - E_{0*} f(R_n(T_n - \nu_n(P_{n,0})) \to 0$$

for every f of the form $f(x) = g(b_1'x, \cdots, b_m'x)$ with $b_1', \cdots, b_m' \in \mathbf{B}'$, $g \in C_b(R^m)$.

Assume existence of a tight Borel measurable random element $\mathbb{Z}_0$ in $\mathbf{B}$ such that

(2) $$\mathbf{L}(b^* \mathbb{Z}_0) = N(0, \|\dot{\nu}^T b^*\|^2) \quad \text{for every} \quad b^* \in \mathbf{B}^* .$$

Call a function $l : \mathbf{B} \to R$ *subconvex* with respect to a topology τ on $\mathbf{B}$ if

$$l(0) = 0 \le l(x) \quad \text{for every} \quad x \in \mathbf{B}$$

(3) $$l(x) = l(-x)$$

$$\{x : l(x) \le c\} \quad \text{is convex and } \tau-\text{closed for every} \quad c \in R .$$

Let $\tau(\mathbf{B}')$ be the weakest topology on $\mathbf{B}$ that makes $b' : \mathbf{B} \to R$ continuous for all $b' \in \mathbf{B}'$.

Remark 3.1. A convex subset of $\mathbf{B}$ is $\tau(\mathbf{B}^*)-$closed if and only if it is $\|\cdot\|-$closed. Hence $\tau(\mathbf{B}^*)-$subconvex is identical to $\|\cdot\|-$subconvex.

Theorem 3.1. (Local asymptotic minimax theorem). Assume that (2.1) - (2.3) and (1) - (2) hold. Then, for every $\tau(\mathbf{B}')-$subconvex l

(4) $$\sup_{I \subset \mathbf{H}} \liminf_{n \to \infty} \sup_{h \in I} E_{h,*}\, l(R_n(T_n - \nu_n(P_{n,h}))) \ge El(\mathbb{Z}_0) .$$

Here the first supremum is taken over all finite subsets I in $\mathbf{H}$.

Remark 3.2. A sequence of estimators $\{T_n\}$ is said to be *local asymptotic minimax (LAM)* for the loss function l if it achieves equality in (4): i.e. if

(5) $$\limsup_{n \to \infty} \sup_{h \in I} E_{h,*}\, l(R_n(T_n - \nu_n(P_{n,h}))) = El(\mathbb{Z}_0)$$

for every finite set $I \subset \mathbf{H}$. Note that if $\{T_n\}$ is locally regular and Hájek optimal, then it is automatically LAM for bounded continuous loss functions. Proof: Local regularity and Hájek optimality imply that, under $P_{n,h}$,

(6) $$R_n(T_n - \nu_n(P_{n,h})) \Rightarrow \mathbb{Z}_0 \quad \text{for} \quad h \in \mathbf{H},$$

and since the inner supremum in (5) is just over a finite set, (6) implies that (5) holds; recall (1.7).

Here are some corollaries of the local asymptotic minimax (LAM) theorem 3.1; we assume throughout that (2) and (3) hold, and call a map T_n in a subset of $R^{\mathbf{F}}$ *projection measurable* if every coordinate $T_n(f)$ is a measurable map in R.

Corollary 3.1. Let $\mathbf{B}$ be separable. Then the LAM theorem holds for every Borel - measurable T_n and $\|\cdot\|$ – subconvex loss.

Corollary 3.2. Let $\mathbf{B} = D[a, b]$ with $\|\cdot\|_\infty$. Then the LAM theorem holds for every projection measurable T_n and $\|\cdot\|_\infty$ – subconvex loss.

Proof. Projection measurable is identical to $D[a, b]^*$ – measurable; cf. lemma 3.26 of van der Vaart (1988a). Hence we can take $\mathbf{B}' = D[a, b]^*$ and have (1) hold. □

Corollary 3.3. Let $\mathbf{B} = l^\infty(\mathbf{F})$. Then the LAM theorem holds for every projection-measurable T_n and $\tau(\Pi)$ – subconvex loss, where $\Pi = lin\{\pi_f : f \in \mathbf{F}\}$. An example of such loss is $l(x) = l_0(\|x\|_\infty)$ where $l_0 : [0, \infty) \to [0, \infty)$ is increasing and lower-semicontinuous.

Proof. The first assertion is trivial. The second follows from

$$\{x : l(x) \le c\} = \{x : \|x\|_\infty \le d\} = \bigcap_{f \in \mathbf{F}} \{x : |x(f)| \le d\}$$

since l_0 is increasing and lower-semicontinuous. □

Corollary 3.4. Let $\mathbf{B} = l^\infty(\mathbf{F})$ with $\|\cdot\|_\infty$. Then the LAM theorem holds for every projection measurable T_n for which $\{R_n(T_n - \nu_n(P_{n,0}))\}$ is tight and $\|\cdot\|_\infty$ – subconvex loss function.

Proof. Tightness and projection measurability imply that (1) is satisfied for $\mathbf{B}' = l^\infty(\mathbf{F})^*$; cf. part (2) of the proof of Prohorov's theorem 1.1. □

Proof of theorem 3.1. (i). Assume first that

(a) $$l(x) = \alpha \sum_{i=1}^{r} 1_{K_i^c}(\tau_i x),$$

where $\alpha \ge 0$ and, for every $i = 1, \cdots, r$,

$$\tau_i x = (b'_{i,1}x, \cdots, b'_{i,p_i}x)^T, \quad \text{some } b'_{i,1}, \cdots, b'_{i,p_i} \in \mathbf{B}'$$

$$K_i \subset R^{p_i} \quad \text{is convex, symmetric, and compact.}$$

It is clear that it suffices to consider the case $\alpha = 1$.

Call the expression on the left side of (4) "Risk(l)." Assume without loss of generality that it is finite. Partially order the finite subsets I of $\mathbf{H}$ by inclusion: $I_1 \le I_2$ if and only if $I_1 \subset I_2$. There exists a subnet $\{n_I : I \subset \mathbf{H}, \text{ finite}\}$ such that

$$Risk(l) = \lim_{I} \sup_{h \in I} E_{h,*} l(R_{n_I}(T_{n_I} - \nu_{n_I}(P_{n_I,h}))) .$$

For instance, take n_I to be any natural number with $n_I \geq |I|$ = the cardinality of I, and

$$|r(n_I, I) - \lim_{n \to \infty} r(n, I)| < |I|^{-1} ,$$

where

$$r(n, I) \equiv \sup_{h \in I} E_{h,*} l(R_n(T_n - \nu_n(P_{n,h}))) .$$

Set

$$\mathbb{Z}_{n,h} = R_n(T_n - \nu_n(P_{n,h})) .$$

For every $i = 1, \cdots, r$ consider $\tau_i \mathbb{Z}_{n,h}$ as a map from $\mathbf{X}_n$ into the one-point compactification D_i of R^{p_i}. Trivially $\{\tau_i \mathbb{Z}_{n,0}\}$ is tight under $P_{n,0}$. Furthermore, by (1)

$$E_0^* f(\tau_i \mathbb{Z}_{n,0}) - E_{0*} f(\tau_i \mathbb{Z}_{n,0}) \to 0 , \quad \text{as } n \to \infty , \qquad \text{for every } f \in C_b(D_i) .$$

By the Prohorov theorem 1.3, the subnet $\{n_I\}$ has a further subnet such that for every $i = 1, \cdots, r$

$$\tau_i \mathbb{Z}_{n',0} \Rightarrow W_i \qquad \text{on } D_i .$$

Here, as in the sequel of the this proof, we abuse notation in writing $\{n'\}$ for every new subnet that will be extracted.

Fix an orthonormal subset $\{g_1, \cdots, g_m\}$ of $\mathbf{H}$. As in the proof of the convolution theorem (cf. (a) - (e)) there exists a further subnet such that, under $P_{n',h}$,

$$\tau_i \mathbb{Z}_{n',h} \Rightarrow W_{a,i} \qquad \text{on } D_i$$

for every $h = a^T g$ and $a \in R^m$, and $i = 1, \cdots, r$. Here (cf. (f))

(c) $$P(W_{a,i} \in A) = E 1_A(W_{a,i} - \tau_i a^T \dot{\nu}(g)) e^{a^T S - \frac{1}{2}|a|^2}$$

(where $\infty + y = \infty$ for every $y \in R^{p_i}$).

Next we use

(d) $$e^{a^T s - \frac{1}{2}|a|^2} dN_d(0, \lambda^{-1} I)(a) = c(s) dN_d(e(s), (1+\lambda)^{-1} I)(a)$$

where

$$e(s) = (1+\lambda)^{-1} s$$

$$c(s) = (1+\lambda^{-1})^{d/2} \exp(-\frac{1}{2}(1+\lambda)^{-1} |s|^2) .$$

By (c) - (d)

$$\int P(W_{a,i} \in A)\, dN_d(0,\lambda^{-1}I)(a)$$
$$= \int E\, 1_A(W_i - \Sigma_i e(S) - \Sigma_i a)\, c(S)\, dN_m(0,(1+\lambda)^{-1}I)(a)$$
$$= N_{p_i}(0,(1+\lambda)^{-1}\Sigma_i\Sigma_i^T) * M_{i,\lambda}(A),$$

where Σ_i is the $p_i \times d$ matrix with $j-$th column $\tau_i \dot{\nu}(g_j)$ and $M_{i,\lambda}$ is the probability measure on D_i given by

$$M_{i,\lambda}(A) = E\, 1_A(W_i - \Sigma_i e(S))\, c(S).$$

Now, for any $h = a^T g$,

$$Risk(l) \geq \lim_{n'} E_{h*}\, l(\mathbb{Z}_{n',h})$$
$$\geq \sum_{i=1}^{r} \liminf_{n'} P_{n',h*}(\tau_i \mathbb{Z}_{n',h} \in K_i^c)$$
$$\geq \sum_{i=1}^{r} P(W_{a,i} \in K_i^c).$$

Thus for every $\lambda > 0$

$$Risk(l) \geq \sum_{i=1}^{r} \int P(W_{a,i} \in K_i^c)\, dN_d(0,\lambda^{-1}I)(a)$$
$$= \sum_{i=1}^{r} N_{p_i}(0,\Sigma_i\Sigma_i^T(1+\lambda)^{-1}) * M_{i,\lambda}(K_i^c)$$
$$\geq \sum_{i=1}^{r} N_{p_i}(0,\Sigma_i\Sigma_i^T(1+\lambda)^{-1})(K_i^c),$$

where the last step follows by Anderson's (1955) lemma; see e.g. Pfanzagl (1985), page 454.

Let $\lambda \downarrow 0$ to obtain that

$$Risk(l) \geq \sum_{i=1}^{r} N_{p_i}(0,\Sigma_i\Sigma_i^T)(K_i^c)$$
$$= \sum_{i=1}^{r} P(\tau_i \mathbb{Z}_g \in K_i^c),$$

where $\mathbb{Z}_g$ is as in (i) of the proof of the convolution theorem.

This is true for every finite, orthonormal subset g of $\mathbf{H}$. Partially order these subsets as in the proof of the convolution theorem. Since every τ_i depends on only finitely many elements from $\mathbf{B}' \subset \mathbf{B}^*$, we have by (w) and (y) of the proof of the convolution theorem that

$$\mathbf{L}(\tau_i \mathbb{Z}_g) \underset{g}{\rightarrow} \mathbf{L}(\tau_i \mathbb{Z}_0).$$

Hence

$$Risk(l) \geq \liminf_g \sum_{i=1}^{r} P(\tau_i \mathbb{Z}_g \in K_i^c)$$

$$\geq \sum_{i=1}^{r} P(\tau_i \mathbb{Z}_0 \in K_i^c) = E\, l(\mathbb{Z}_0).$$

(ii). Suppose that (4) holds for every member of a sequence $\{l_r\}$ such that

$$\text{(e)} \qquad \begin{array}{l} 0 \leq l_r \leq l \\ l_r \uparrow l \quad \text{a.e.} \quad \mathbf{L}(\mathbb{Z}_0). \end{array}$$

Then

$$R(l) \geq \limsup_{r \to \infty} R(l_r) \geq \limsup_{r \to \infty} El_r(\mathbb{Z}_0) = El(\mathbb{Z}_0).$$

Hence (4) holds for l, too.

(iii). Let $C \subset \mathbf{B}$ be $\tau(\mathbf{B}')$ – closed, convex, and symmetric. Then there exists a sequence of compact subsets $K_p \subset R^p$ and maps $\tau_p(x) = (b'_1 x, \cdots, b'_p x)$ (for some $b'_1, \cdots, b'_p \in \mathbf{B}'$) such that

$$\text{(f)} \qquad \begin{array}{l} 0 \leq 1_{K_p^c}(\tau_p \cdot) \leq 1_{C^c} \\ 1_{K_p^c}(\tau_p \cdot) \uparrow 1_{C^c} \quad \text{a.e.} \quad \mathbf{L}(\mathbb{Z}_0) \end{array}$$

Proof of (iii). By the Hahn-Banach theorem applied to the locally convex, topological vector space $(\mathbf{B}, \tau(\mathbf{B}'))$,

$$C = \bigcap_{b' \in \mathbf{B}'} \{x : |b'x| \leq c_{b'}\},$$

for constants $c_{b'} \in \bar{R}$. Let S be a separable Borel set with $P(\mathbb{Z}_0 \in S) = 1$. We have

$$C^c \cap S = \bigcup_{b' \in \mathbf{B}'} \{x \in S : |b'x| > c_{b'}\}.$$

Every set $\{x \in S : |b'x| > c_{b'}\}$ is relatively $\tau(\mathbf{B}')$ – open in S. Since S is separable with respect to the norm topology, it is Lindelöf with respect to the norm topology, hence certainly with respect to the $\tau(\mathbf{B}')$ – topology. Thus there is a countable subset $\{b'_i\}$ such that

$$C^c \cap S = \bigcup_{i=1}^{\infty} \{x \in S : |b'_i x| > c_{b'_i}\}.$$

Now (f) is satisfied with

$$K_p = \{y \in R^p : |y_i| \leq c_{b'_i},\ i = 1, \cdots, p\}.$$

(iv). Finally, let l be an arbitrary $\tau(\mathbf{B}')-$ subconvex function. Set

$$l_r(x) = r^{-1}\sum_{i=1}^{r^2} 1_{C_{r,i}}(x),$$

where $C_{r,i} = \{x : l(x) > i/r\}$. Then $0 \le l_r \uparrow l$ everywhere. Moreover, every $C_{r,i}$ is $\tau(\mathbf{B}')-$ closed, convex and symmetric. Hence by (iii), every l_r can be approximated from below in the sense of (e) by functions of type (a). By (i) - (ii), (4) holds for every l_r. Apply (ii) again to see that (4) then also holds for l. □

4. Examples

In this section we give several examples to illustrate the convolution and asymptotic minimax theorems of sections 2 and 3.

Example 4.1. (Estimation of a probability measure P from iid sampling.) Suppose that $(\mathbf{X}, \mathbf{A})$ is a measurable space, and let $X_1, \cdots, X_n$ be independent and identically distributed random elements in $(\mathbf{X}, \mathbf{A})$ with unknown distribution P. Suppose that $\mathbf{F} \subset L_2(P) \equiv L_2(\mathbf{X}, \mathbf{A}, P)$, and let $\mathbf{B} = l^\infty(\mathbf{F})$. We consider estimation of $\nu(P) \in l^\infty(\mathbf{F})$ defined by

(1) $$\nu(P)(f) = \int f \, dP \equiv P(f), \quad f \in \mathbf{F}.$$

Let $\mathbf{H}_0 \equiv \{$all bounded $L_2(P)$ functions $h : \int h \, dP = 0\}$; thus $\overline{\mathbf{H}_0} = \mathbf{H} = \{ h \in L_2(P) : \int h \, dP = 0\} \subset L_2(P)$ with the usual inner product $<f, g> = P(fg) = E f(X) g(X)$. For $h \in \mathbf{H}_0$, define P_n by

(2) $$\frac{dP_n}{dP} = 1 + n^{-1/2} h$$

for n sufficiently large. Then (2.6) holds, so that $\dot{\mathbf{P}}^0 = \mathbf{H}_0$ is linear. Of course (2.6) entails that (2.1) holds with $\Delta_{n,h} = n^{-1/2} \sum_{i=1}^n h(X_i)$.

Furthermore, (2.7) holds with $\dot{\nu}(h)$ given by

(3) $$\dot{\nu}(h)(f) = \int h f \, dP = \int h (f - P(f)) dP$$
$$= <h, f - P(f)> = <h, \dot{\nu}^T \pi_f>, \quad f \in \mathbf{F}$$

since $<1, h> = 0$; here π_f is the element of $\mathbf{B}^* = l^\infty(\mathbf{F})^*$ given by $\pi_f(b) = b(f)$ for $b \in \mathbf{B} = l^\infty(\mathbf{F})$. Hence the limit process $\mathbb{Z}_0$ of theorems 2.1 and 3.1 has covariance function

(4) $$Cov(\mathbb{Z}_0(f), \mathbb{Z}_0(g)) = <f - P(f), g - P(g)> = P(fg) - P(f)P(g).$$

It is known that the empirical measure $\mathbb{P}_n \equiv n^{-1} \sum_{i=1}^n \delta_{X_i}$ achieves this bound in the sense of theorems 2.1 and 3.1 for certain classes $\mathbf{F}$ ($P-$Donsker classes in the terminology of Dudley (1984), (1985)) or Gin´e and Zinn (1984)); i.e. for a $P-$Donsker class $\mathbf{F}$

(5) $$\sqrt{n}\,(\nu(\mathbb{P}_n) - \nu(P)) \Rightarrow \mathbb{Z}_0 \quad \text{in } l^\infty(\mathbf{F}).$$

Regularity of the empirical measure estimator $\mathbb{P}_n$ is established for classes $\mathbf{F}$ with (uniformly) square integrable envelope function F in Sheehy and Wellner (1988).

It is not always possible to compute $\dot{\nu}$ and $\dot{\nu}^T$ explicitly, as is illustrated by our next example.

Example 4.2. (Bivariate three - sample model). Suppose that (X, Y) has joint distribution P on the product measure space $(\mathbb{X} \times \mathbb{Y}, \mathbf{A} \times \mathbf{B})$. Let P_X and P_Y

denote the marginal distributions of X and Y. Suppose that we observe

$$
(6) \qquad \begin{array}{ll}
(X_{11},Y_{11}), \cdots ,(X_{1n_1},Y_{1n_1}) & \text{iid } P \\
X_{21}, \cdots ,X_{2n_2} & \text{iid } P_X \\
Y_{31}, \cdots ,Y_{3n_3} & \text{iid } P_Y
\end{array}
$$

where the three samples are independent. We write $\underline{Z} \equiv ((X_{11},Y_{11}), \cdots ,(X_{1n_1},Y_{1n_1}))$, $\underline{X}_2 \equiv (X_{21}, \cdots ,X_{2n_2})$, and $\underline{Y}_3 \equiv (Y_{31}, \cdots ,Y_{3n_3})$. This can, of course, be viewed as a missing data model: the Y's are missing in the second sample, while the X's are missing in the third sample. Let $n \equiv n_1+n_2+n_3$, and set $\lambda_{ni} \equiv n_i/n$, $i = 1,2,3$. We assume that $\lambda_{ni} \to \lambda_i \in (0,1)$, $i = 1,2,3$.

Let $\mathbf{G}_0 \equiv \{$all bounded $L_2(P)$ functions $g : \int g\, dP = 0\}$. Let $L_2^0 \equiv \{g \in L_2(P) : \int g\, dP = 0\}$. For $g \in \mathbf{G}_0$, define P_n by

$$
(7) \qquad \frac{dP_n}{dP} = 1 + n^{-1/2} g .
$$

Now set

$$
(8) \qquad g_1(x) = E(g(X,Y) \mid X = x),
$$

$$
(9) \qquad g_2(y) = E(g(X,Y) \mid Y = y),
$$

and note that

$$
\frac{dP_{nX}}{dP_X} = 1 + n^{-1/2} g_1, \qquad \frac{dP_{nY}}{dP_Y} = 1 + n^{-1/2} g_2 .
$$

Thus we take $\mathbf{H} = \overline{\mathbf{H}_0}$ where $\mathbf{H}_0$ is the subspace

$$
\mathbf{H}_0 \equiv \{(g, g_1, g_2) : g \in \mathbf{G}_0\},
$$

(where g_1 and g_2 are defined by (8) and (9)) of the Hilbert space $L_2^0(P) \times L_2^0(P_X) \times L_2^0(P_Y)$ with inner product

$$
<(g,a,b),(h,c,d)> = \lambda_1 E(gh) + \lambda_2 E(ac) + \lambda_3 E(bd) .
$$

It follows that in the present model

P_{n0} is the law of $(\underline{Z}, \underline{X}_2, \underline{Y}_3)$ under P, and

P_{nh} is the law of $(\underline{Z}, \underline{X}_2, \underline{Y}_3)$ under P_n.

Let E, E_n denote expectations under P, P_n respectively.

Note that (2.1) and (2.2) hold with

$$
\Delta_{n,h} = n^{-1/2}\{\sum_{i=1}^{n_1} g(X_{1i}, Y_{1i}) + \sum_{i=1}^{n_2} g_1(X_{2i}) + \sum_{i=1}^{n_3} g_2(Y_{3i})\}
$$

$$
(10) \qquad = \lambda_1^{1/2} n_1^{-1/2} \sum_{i=1}^{n_1} g(X_{1i}, Y_{1i}) + \lambda_2^{1/2} n_2^{-1/2} \sum_{i=1}^{n_2} g_1(X_{2i})
$$

$$+ \lambda_3^{1/2} n_3^{-1/2} \sum_{i=1}^{n_3} g_2(Y_{3i}) + o_P(1)$$

and

$$\|h\|^2 = \lambda_1 E g^2(X,Y) + \lambda_2 E g_1^2(X) + \lambda_3 E g_2^2(Y)$$

where again $h \equiv (g, g_1, g_2) \in \mathbf{H}_0$.

As in example 4.1, suppose that $\mathbf{F} \subset L_2(P) = L_2(\mathbb{X} \times \mathbb{Y}, \mathbf{A} \times \mathbf{B}, P)$, and let $\mathbf{B} = l^\infty(\mathbf{F})$. Consider estimation of the measure P indexed by $\mathbf{F}$: i.e. $\nu_n(P_{n,0}) \in l^\infty(\mathbf{F})$ defined by

$$\nu_n(P_{n,0})(f) = \frac{1}{n_1} \sum_{i=1}^{n_1} Ef(X_{1i}, Y_{1i}) = \int f \, dP = P(f).$$

Thus

$$\nu_n(P_{n,h})(f) = P_n(f)$$

and

$$\sqrt{n}\,(\nu_n(P_{n,h}) - \nu_n(P_{n,0}))(f) = \int g\,(f - Pf)\,dP,$$

so (2.3) holds trivially with $R_n = \sqrt{n}$ and, for $h = (g, g_1, g_2)$,

$$\dot{\nu}(h)(f) = \langle \frac{1}{\lambda_1}(f - Pf, 0, 0), (g, g_1, g_2) \rangle$$

$$= \langle \frac{1}{\lambda_1}(f - Pf - a - b, \frac{\lambda_1}{\lambda_2} a, \frac{\lambda_1}{\lambda_3} b), (g, g_1, g_2) \rangle$$

(11) $$\equiv \langle \dot{\nu}^T \pi_f, (g, g_1, g_2) \rangle$$

where π_f is, as in example 4.1, the evalution map, and $a = a(x)$ and $b(y)$ satisfy, with $f^0 \equiv f - Pf$,

$$E(f^0(X,Y) - a(X) - b(Y) \mid X = x) = \frac{\lambda_1}{\lambda_2} a(x)$$

and

$$E(f^0(X,Y) - a(X) - b(Y) \mid Y = y) = \frac{\lambda_1}{\lambda_3} b(y),$$

or equivalently

(12) $$E(\lambda_2 f^0(X,Y) - (\lambda_1 + \lambda_2) a(X) - \lambda_2 b(Y) \mid X = x) = 0$$

and

(13) $$E(\lambda_3 f^0(X,Y) - \lambda_3 a(X) - (\lambda_1 + \lambda_3) b(Y) \mid Y = y) = 0.$$

Thus

$$Cov(\mathbb{Z}_0(f), \mathbb{Z}_0(g)) = <\dot{v}^T \pi_f, \dot{v}^T \pi_g>$$

$$= \frac{1}{\lambda_1} E(f^0 - a_f - b_f)(g^0 - a_g - b_g)$$

$$+ \frac{1}{\lambda_2} E(a_f a_g) + \frac{1}{\lambda_3} E(b_f b_g)$$

$$= \frac{1}{\lambda_1} E(f^0 - a_f - b_f) g^0 . \tag{14}$$

Note that

$$\|\dot{v}^T \pi_f\|^2 \le \|\frac{1}{\lambda_1}(f^0, 0, 0)\|^2 = \frac{1}{\lambda_1}\int (f - Pf)^2 dP ,$$

where the right hand side is the variance of the estimator of $v(P)(f) = P(f)$ based on using the $\underline{Z}$'s only. Strict inequality holds if $(a + b, \lambda_1 a/\lambda_2, \lambda_1 b/\lambda_3) \neq 0$ in $L_2^0(P) \times L_2^0(P_X) \times L_2^0(P_Y)$.

The pair of equations (12), (13) reduce to the "ACE equations" when $\lambda_1 = 0$ and $\lambda_2 + \lambda_3 = 1$ with $\lambda_2, \lambda_3 > 0$. These equations arise in connection with projections on sum spaces; see e.g. Bickel, Ritov, and Wellner (1989) for these equations in the closely related problem in which the P_X and P_Y margins are known, corresponding to $n_2 = n_3 = \infty$ and $\lambda_1 = 0$, $\lambda_2, \lambda_3 > 0$.

Example 4.3. (Estimation of the stationary distribution of a Markov chain.) Let $X_0, X_1, X_2, \cdots$ be a stationary Markov chain with state space $(\mathbf{X}, \mathbf{A})$ and unknown transition kernel (Markov kernel) $Q(x, A)$. We assume that the chain has a unique stationary distribution π, and consider estimation of π seen as an element of $l^\infty(\mathbf{C})$, based on observation of $X_0, X_1, \cdots, X_n$ of the chain. Here $\mathbf{C}$ is a subset of $\mathbf{A}$.

Let $\overline{\mathbf{H}}$ be the set of all measurable functions $h(x, y)$ with

$$E_Q h^2(X_0, X_1) < \infty , \tag{15}$$

$$E_Q(h(X_0, X_1) | X_0) = 0 . \tag{16}$$

$\overline{\mathbf{H}}$ is a Hilbert space with respect to the inner product

$$<h_1, h_2> = E_Q h_1(X_0, X_1) h_2(X_0, X_1) . \tag{17}$$

Let $\mathbf{H}$ be the subspace of uniformly bounded $h \in \overline{\mathbf{H}}$.

For $h \in \mathbf{H}$ and sufficiently large n

$$Q_{n,h}(x, A) = \int_A (1 + n^{-1/2} h(x, y)) Q(x, dy) \tag{18}$$

is a transition kernel by (16).

Now suppose that the chain is uniformly ergodic (cf. Nummelin (1984), 6.6), i.e.

$$(19) \qquad \sup_{x,A} |Q^m(x,A) - \pi(A)| \to 0 \quad \text{as} \quad m \to \infty$$

where Q^m is the mth - step transition kernel. Under (19), every $Q_{n,h}$ admits a unique stationary distribution $\pi_{n,h}$ too, and $\pi_{n,h}$ asymptotically depends smoothly on h. The following lemma is essentially due to Penev (1988).

Lemma 4.1. Suppose that (19) holds. Then, for sufficiently large n, $Q_{n,h}$ given by (18) admits a unique stationary distribution $\pi_{n,h}$ and

$$(20) \qquad \sqrt{n}\,(\pi_{n,h} - \pi) \to \dot{\nu}(h) \quad \text{in } l^\infty(\mathbf{C})$$

where $\dot{\nu}(h)(A) = \langle h, \tilde{\nu}_A \rangle$ with

$$(21) \qquad \tilde{\nu}_A(x,y) = \sum_{m=0}^{\infty} (Q^m(y,A) - Q^{m+1}(x,A)).$$

Moreover $\frac{d\pi_{n,h}}{d\pi}(X_0) \to 1$ in π – probability.

Let $P_{n,h}$ be the distribution of the stationary chain $X_0, X_1, \cdots, X_n$ under $Q_{n,h}$. Then, under (19)

$$(22) \qquad \log \frac{dP_{n,h}}{dP_{n,0}}(X_0, \cdots, X_n) = \Delta_{n,h} - \frac{1}{2}\|h\|^2 + o_{P_{n,0}}(1)$$

where

$$(23) \qquad \Delta_{n,h} = n^{-1/2} \sum_{i=1}^{n} h(X_{i-1}, X_i)$$

satisfies (2.2). To see this, note first that (19) implies that the chain with transition kernel Q is mixing, and hence ergodic. Next, (22) and (2.2) follow either by a short direct argument or as in Roussas (1972).

Relations (20) - (22) imply that the conditions for the convolution theorem are satisfied. The same is true for the asymptotic minimax theorem, provided the optimal limit process $\mathbb{Z}_0$ exists. The latter requires a Vapnik - Chervonenkis type of condition on the extensiveness of the class $\mathbf{C}$. If $\mathbb{Z}_0$ exists as a tight random element in $l^\infty(\mathbf{C})$, then it should have uniformly bounded sample paths and have zero - mean Gaussian marginals with

$$(24) \qquad Cov(\mathbb{Z}_0(A), \mathbb{Z}_0(B)) = \langle \tilde{\nu}_A, \tilde{\nu}_B \rangle.$$

(One could view the stationary distribution π as an element of some other space too, of course. For instance, if $\mathbf{X} = R$, then it is natural to identify π with its distribution function, which can be seen as an element of the Skorokhod space $D[-\infty, \infty]$ with supremum norm. This set-up rules out certain estimators. On the other hand, one obtains asymptotic minimax theorems for a different set of loss functions.)

Before proving lemma 4.1, note that the process corresponding to the empirical measure $I\!P_n$ converges marginally in distribution to $\mathbb{Z}_0$. Specifically, set

$$(25) \qquad \mathbb{Z}_{n,h}(A) \equiv n^{1/2}(I\!P_n(A) - \pi_{n,h}(A)) = n^{-1/2}\sum_{i=1}^{n}(1_A(X_i) - \pi_{n,h}(A)) \quad \text{for} \quad A \in \mathbf{C}.$$

Then

$$(26) \qquad \mathbf{L}_{Q_{n,h}}(\mathbb{Z}_{n,h}(A_1), \cdots, \mathbb{Z}_{n,h}(A_k)) \to \mathbf{L}(\mathbb{Z}_0(A_1), \cdots, \mathbb{Z}_0(A_k))$$

for every $A_1, \cdots, A_k$ in $\mathbf{A}$. This implies that the empirical measure $I\!P_n$ is efficient for the estimation of π provided some tightness conditions are satisfied. Efficiency of $I\!P_n$ has been established for the case $\mathbf{X} = R$ and $\mathbf{C} = \{(-\infty, t] : t \in R\}$; see e.g. Billingsley (1968), section 22, and Penev (1988).

In fact, it holds that

$$(27) \qquad n^{-1/2}\sum_{i=1}^{n}(1_A(X_i) - \pi(A)) = n^{-1/2}\sum_{i=1}^{n}\bar{v}_A(X_{i-1}, X_i) + o_{P_{n,0}}(1),$$

which implies (26) for $h = 0$, and hence for general h via lemma 1.6 and (20). To see the validity of (27), let

$$\phi_A(x) = \sum_{m=0}^{\infty}(Q^m(x, A) - \pi(A)).$$

By (19), $\phi_A(X_0)$ is square integrable. Then (27) follows from

$$(28) \qquad n^{-1/2}\sum_{i=1}^{n}(\bar{v}_A(X_{i-1}, X_i) - \bar{v}_A(X_i, X_i)) = n^{-1/2}\sum_{i=1}^{n}(\phi_A(X_{i-1}) - \phi_A(X_i)) = n^{-1/2}(\phi_A(X_0) - \phi_A(X_n)) = o_{P_{n,0}}(1),$$

and, by (21) and (19),

$$(29) \qquad \bar{v}_A(X_i, X_i) = 1_A(X_i) - \pi(A)$$

since $Q^0(x, A) = 1_A(x)$.

The asymptotic covariance function of $\mathbb{Z}_{n,0}$ is easily computed (cf. e.g. Billingsley (1968), page 197). Combination with (26) yields

$$(30) \qquad E\,\mathbb{Z}_0(A)\mathbb{Z}_0(B) = Cov(1_A(X_0), 1_B(X_0)) + \sum_{k=1}^{\infty} Cov(1_A(X_0), 1_B(X_k)) + \sum_{k=1}^{\infty} Cov(1_A(X_k), 1_B(X_0)).$$

This formula can also be derived directly from (24) and (21), but this is somewhat tedious.

Proof of lemma 1. Let $\mathbf{M}$ be the set of finite signed measures on $(\mathbf{X},\mathbf{A})$. Equip $\mathbf{M}$ with the total variation norm $\|\cdot\|$. View Q as the operator $Q:\mathbf{M}\to\mathbf{M}$ given by $Q\mu(A)=\int Q(x,A)\,d\mu(x)$, and define $\Pi:\mathbf{M}\to\mathbf{M}$ by $\Pi\mu=\mu(\mathbf{X})\pi$. By definition $Q\pi=\pi$. It is easily checked that $\Pi Q=Q\Pi=\Pi^2=\Pi$ and

(31) $\qquad (Q-\Pi)^m = Q^m - \Pi \quad \text{for } m=1,2,\cdots .$

By (19) and (31)

$$\|(Q-\Pi)^m\| \to 0 .$$

By a standard argument one sees that this convergence occurs at a geometric rate so that

(32) $\qquad \sum_{m=0}^{\infty} \|(Q-\Pi)^m\| < \infty .$

Therefore, the operator $R\equiv I-Q+\Pi$ has a bounded inverse $R^{-1}=\sum_{m=0}^{\infty}(Q-\Pi)^m$. For $h\in\mathbf{H}$ define the operator Δ_h by

$$\Delta_n\mu(A) \equiv \iint_A h(x,y)\,Q(x,dy)\,d\mu(x) .$$

It is easily seen that $\|\Delta_n\| \le 2\|h\|_\infty$. Hence for $\|R^{-1}\|\,2\,\|h\|_\infty < \sqrt{n}$,

$$(R-n^{-1/2}\Delta_n)^{-1} : \mathbf{M} \to \mathbf{M}$$

exists as a bounded operator. Set

$$\pi_{n,h} = (R-n^{-1/2}\Delta_h)^{-1}\pi .$$

Since $(R-n^{-1/2}\Delta_h)\mu(\mathbf{X})=\mu(\mathbf{X})$ for every $\mu\in\mathbf{M}$, we have $\pi_{n,h}(\mathbf{X})=\pi(\mathbf{X})=1$. Next,

$$(I-Q-n^{-1/2}\Delta_h)\pi_{n,h} = (R-n^{-1/2}\Delta_h)\pi_{n,h} - \Pi\pi_{n,h} = 0 ,$$

so that $\pi_{n,h}$ is a stationary distribution for the transition kernel $Q_{n,h}$ given by (18). Reversing the argument shows that $\pi_{n,h}$ is unique.

Now

$$\begin{aligned}\sqrt{n}\,(\pi_{n,h}-\pi) &= \sqrt{n}\,\{(R-n^{-1/2}\Delta_h)^{-1}\pi - R^{-1}\pi\} \\ &= (R-n^{-1/2}\Delta_h)^{-1}\sqrt{n}\,\{I-(R-n^{-1/2}\Delta_h)R^{-1}\}\pi \\ &\to R^{-1}\Delta_h R^{-1}\pi = R^{-1}\Delta_h\pi \equiv \dot{\nu}(h) .\end{aligned}$$

Use (31) to see that

$$\dot{\nu}(h)(A) = \Delta_h\pi(A) + \sum_{m=1}^{\infty}\iint Q^m(y,A)\,h(x,y)\,Q(x,dy)\,d\pi(x)$$

$$= \langle h , \dot{\nu}_A \rangle ,$$

where

$$\dot{\nu}_A(x, y) = \sum_{m=0}^{\infty} (Q^m(y, A) - \pi(A)) .$$

Here $\dot{\nu}_A$ is square integrable (in fact uniformly bounded by (32)), but is not contained in $\overline{\mathbf{H}}$ in general. Its orthogonal projection on $\overline{\mathbf{H}}$ is

$$\tilde{\nu}_A(x,y) = \dot{\nu}_A(x,y) - E(\dot{\nu}_A(X_0, X_1) | X_0 = x) ,$$

which equals the right side of (21). The last assertion of the lemma follows from

$$\pi\left(\left|\frac{d\pi_{n,h}}{d\pi}(X_0) - 1\right| > \varepsilon\right) \le \varepsilon^{-1} \int \left|\frac{d\pi_{n,h}}{d\pi} - 1\right| d\pi$$

$$\le \varepsilon^{-1} \|\pi_{n,h} - \pi\| \to 0 . \quad \square$$

5. Proofs for Section 1

Proof of lemma 1.1. Set $L = \inf\{ \int arctan\, h\, dP : h \geq f\ ,\ h$ measurable $\}$. Choose $h_m \geq f$ measurable with $\int arctan\, h_m\, dP \downarrow L$. Set $f_m \equiv \inf_{k \leq m} h_k$. Define $f^*(\omega) \equiv \lim_{m \to \infty} f_m(\omega)$ (finite or $-\infty$). Then $f^*(\omega) \geq f(\omega)$ everywhere, f^* is measurable, and $\int arctan\, f^*\, dP = L$. Moreover, for every $h \geq f$ that is measurable, it follows that

(a) $$\int arctan\,(h \wedge f^*)\, dP = \lim_{m \to \infty} \int arctan\,(h \wedge f_m)\, dP \geq L = \int arctan\, f^*\, dP\ .$$

Since $arctan(f^*) - arctan(f^* \wedge h) \geq 0$, and

(b) $$\int (arctan\, f^* - arctan\,(f^* \wedge h)\, dP \leq 0\ , \quad \text{by (a)}\ ,$$

it follows that

(c) $$arctan\, f^* = arctan(f^* \wedge h) \quad \text{a.s.}\ ,$$

implying that $f^* = f^* \wedge h$ a.s., or $f^* \leq h$ a.s.. □

Proof of lemma 1.2. (i) is trivial. (ii) follows from (i) and

$$f^* = (f - g + g)^* \leq (f - g)^* + g^* \quad \text{a.s.}\ .$$

To prove (iii), note that $f^* - g^* \leq (f - g)^*$ a.s. by (ii) if both sides are well defined a.s., which is obviously smaller than $|f - g|^*$.

(iv) - (vi) follow from (i) - (iii) by symmetry.

The proof of (vii) proceeds in steps: First, we claim that if A is measurable then

(a) $$(f\, 1_A)^* = f^*\, 1_A\ .$$

To see (a), note that $(f\, 1_A)^* \leq f^*\, 1_A$ is clearly true. But if $h \geq f\, 1_A$, then $f \leq h\, 1_A + f^*\, 1_{A^c}$. Hence $f^* \leq h\, 1_A + f^*\, 1_{A^c}$, so that $f^*\, 1_A \leq h$. Thus (a) holds. Now

(b) $$\begin{aligned}(f\, g)^* &= (f\, g)^*\, 1_{[g > 0]} + (f\, g)^*\, 1_{[g < 0]} + (f\, g)^*\, 1_{[g = 0]} \\ &= (f\, g\, 1_{[g > 0]})^* + (f\, g\, 1_{[g < 0]})^* + (f\, g\, 1_{[g = 0]})^* \quad \text{by (a)}\ .\end{aligned}$$

Here

(c) $$(f\, g\, 1_{[g = 0]})^* = 0^* = 0\ ,$$

and

(d) $$(f\, g\, 1_{[g > 0]})^* = f^*\, g\, 1_{[g > 0]}\ .$$

To see (d), first note that $\leq$ clearly holds. But if $h \geq f\, g\, 1_{[g > 0]}$, then $h\, 1_{[g \leq 0]} \geq 0$ and $h\, 1_{[g > 0]} \geq f\, g\, 1_{[g > 0]}$. Hence $(h/g)\, 1_{[g > 0]} \geq f\, 1_{[g > 0]}$. Thus $f^*\, 1_{[g \geq 0]} = (f\, 1_{[g > 0]})^* \leq (h/g)\, 1_{[g > 0]}$, and we conclude that

$$f^*\, g\, 1_{[g > 0]} \leq h\, 1_{[g > 0]} + 0 \leq h\, 1_{[g > 0]} + h\, 1_{[g \leq 0]} = h\ ,$$

so (d) holds. Similarly,

(e) $\qquad (f\, g\, 1_{[g<0]})^* = f_*\, g\, 1_{[g<0]}\,.$

Combining (c)-(e) with (b) yields the claim.

(viii) is trivial.

To prove (ix), let $h = 1_A^*$. Then $A_0 = [h = 1]$ satisfies $1_A \le 1_{A_0}$ and $1_{A_0} \le h$.

(x) follows from (viii): $(1_{A^c})_* = (1 - 1_A)_* = 1 + (-1_A)_* = 1 - 1_A^*$ where (viii) is used to get the second equality.

To prove (xi): First, $E_P^* f \equiv \int^* f\, dP \le \int f^*\, dP \equiv E_P f^*$ is trivially true. To prove the reverse inequality, by definition of $\int^* f\, dP$ there exists a sequence $h_m \ge f$, h_m measurable such that

$$\int^* f\, dP + \frac{1}{2^m} \ge \int h_m\, dP$$

for each $m = 1, 2, \cdots$. But any such sequence $\{h_m\}$ also satisfies $h_m \ge f^*$ by definition of f^*, and hence we have

$$h_m \ge g_m \equiv \inf_{1 \le k \le m} h_k \ge f^*$$

where $g_m \downarrow$. It follows that

$$\begin{aligned} \int^* f\, dP &\ge \liminf_{m \to \infty} \int h_m\, dP \ge \liminf_{m \to \infty} \int g_m\, dP \ge \liminf_{m \to \infty} \int (g_m \quad M)\, dP \\ &\ge \int \liminf_{m \to \infty} g_m \quad M\, dP \\ &= \int f^* \quad M\, dP \qquad \text{for every } M, \end{aligned}$$

and the right side converges to $\int f^*\, dP$ as $M \to \infty$ whenever the latter is finite.

Proof of (xii):

$$\begin{aligned} P^*(A) &\equiv \inf\{ P(B) : B \supset A,\ B \text{ measurable} \} \\ &\ge \inf\{ E\, h : h \ge 1_A,\ h \text{ measurable} \} \\ &= E(1_A)^* \qquad \text{by (xi)} \\ &= E\, 1_{A_0} \qquad \text{by (ix)} \\ &= P(A_0) \ge P^*(A). \end{aligned}$$

Proof of (xiii): We show that $A \supset \{f > \varepsilon\}$ with A measurable implies that $A \supset \{f^* > \varepsilon\}$. Thus $P(f^* > \varepsilon) \le P^*(f > \varepsilon)$.

Define $g \equiv f^* 1_A + (f^* \wedge \varepsilon) 1_{A^c}$. Then $g \geq f$ a.s. and is measurable, and hence $g \geq f^*$. But this implies $f^* \wedge \varepsilon \geq f^*$ on A^c, or $A^c \subset \{f^* \leq \varepsilon\}$, and hence $A \supset \{f^* > \varepsilon\}$.

Note that this implies the first assertion: $1^*_{[f > \varepsilon]} \leq 1_{[f^* > \varepsilon]}$ easily since $f^* \geq f$; moreover, by (xii), $E\, 1^*_{[f > \varepsilon]} = P^*(f > \varepsilon) = P(f^* > \varepsilon) = E\, 1_{[f^* > \varepsilon]}$, which implies the assertion. □

Proof of lemma 1.3. The equivalence of (ii) and (iii) is trivial; similarly, the equivalence of (iv) and (v) is also trivial. That (iv) and (v) together imply (i) is easy since a continuous function is both upper and lower semicontinuous.

Now we show that (i) implies (ii). First, there exists a sequence $\{h_m\} \subset C_b(M)$ with $0 \leq h_m \leq 1_G$ and $h_m \uparrow 1_G$. Now for every $m = 1, 2, \cdots$

$$\liminf_{n \to \infty} P_{n*}(\boldsymbol{X}_n \in G) \geq \liminf_{n \to \infty} E_* h_m(\boldsymbol{X}_n) \geq E_* h_m(\boldsymbol{X}_0).$$

By monotone convergence $E h_m(\boldsymbol{X}_0) \uparrow P(\boldsymbol{X}_0 \in G)$ as $m \to \infty$. To show that (iii) implies (v), assume without loss of generality that $0 \leq f \leq 1$. Fix r. For $p = 0, \cdots, r$, set $F_p = \{x : f(x) \geq p/r\}$ (a closed set by upper semicontinuity of f) and $f_r(x) = r^{-1} \sum_{p=0}^{r} 1_{F_p}(x)$. Then $f_r \geq f$ and $\|f_r - f\|_\infty \leq 1/r$. Now

$$\limsup E^* f(\boldsymbol{X}_n) \leq \limsup E^* f_r(\boldsymbol{X}_n) \leq \frac{1}{r} \sum_{p=0}^{r} \limsup_n P_n^*(\boldsymbol{X}_n \in F_r)$$

$$\leq \frac{1}{r} \sum_{p=0}^{r} P_0(\boldsymbol{X}_0 \in F_p) \quad \text{by (iii)}$$

$$= E f_r(\boldsymbol{X}_0).$$

Let $r \to \infty$ to get the conclusion.

Now we show that (vi) implies (iii). We have $\partial F^\varepsilon = \{x : d(x, F) = \varepsilon\}$. Thus we have $\partial F^{\varepsilon_1} \cap \partial F^{\varepsilon_2} \neq \emptyset$ if $\varepsilon_1 \neq \varepsilon_2$, so that $P_0(\boldsymbol{X}_0 \in \partial F^\varepsilon)$ can be nonzero for at most countably many $\varepsilon > 0$. Choose $\varepsilon_m \downarrow 0$ such that $P_0(\boldsymbol{X}_0 \in \partial F^{\varepsilon_m}) = 0$, $m = 1, 2, \cdots$. Then

$$\limsup_n P_n^*(\boldsymbol{X}_n \in F) \leq \limsup_n P_n^*(\boldsymbol{X}_n \in F^{\varepsilon_m}) = P(\boldsymbol{X}_0 \in F^{\varepsilon_m}).$$

Letting $m \to \infty$ yields (iii).

To complete the proof, we show that (ii) and (iii) together imply (vi). For $A \subset M$ with $P_0(\boldsymbol{X}_0 \in \partial A) = 0$,

$$P_0(\boldsymbol{X}_0 \in \operatorname{Int} A) \leq \liminf_n P_{n*}(\boldsymbol{X}_n \in \operatorname{Int} A)$$

$$\leq \liminf P_n^*(\boldsymbol{X}_n \in A)$$

$$\leq \limsup P_n^*(\mathbf{X}_n \in A)$$

$$\leq \limsup P_n^*(\mathbf{X}_n \in \bar{A}) \leq P_0(\mathbf{X}_0 \in \bar{A}).$$

But the left and right sides are equal since $P_0(\mathbf{X}_0 \in \partial A) = 0$, and hence (vi) follows. □

Proof of lemma 1.4. For every $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon$ with $P_0(\mathbf{X}_0 \in K) \geq 1 - \varepsilon$. By lemma 1

$$\liminf_{n \to \infty} P_{n*}(\mathbf{X}_n \in K^\delta) \geq P_0(\mathbf{X}_0 \in K^\delta) \geq P_0(\mathbf{X}_0 \in K) \geq 1 - \varepsilon. \quad \square$$

Proof of lemma 1.5. Let $\varepsilon > 0$, and let K_1 and K_2 be compact sets in M_1 and M_2 respectively with $\liminf_{n \to \infty} P_{n*}(\mathbf{X}_n \in K_1^\delta) \geq 1 - \varepsilon$ and $\liminf_{n \to \infty} P_{n*}(\mathbf{Y}_n \in K_2^\delta) \geq 1 - \varepsilon$ for every $\delta > 0$. Then $(K_1 \times K_2)^\delta = K_1^\delta \times K_2^\delta$ and

$$\limsup_{n \to \infty} P_n^*((\mathbf{X}_n, \mathbf{Y}_n) \in (K_1^\delta \times K_2^\delta)^c)$$

$$\leq \limsup_{n \to \infty} P_n^*((\mathbf{X}_n, \mathbf{Y}_n) \in (K_1^\delta)^c \times M_2 \cup M_1 \times (K_2^\delta)^c)$$

$$\leq \limsup_{n \to \infty} \{ P_n^*((\mathbf{X}_n, \mathbf{Y}_n) \in (K_1^\delta)^c \times M_2) + P_n^*((\mathbf{X}_n, \mathbf{Y}_n) \in M_1 \times (K_2^\delta)^c) \}$$

$$\leq 2\varepsilon. \quad \square$$

Proof of Proposition 1.1. This is almost as in Billingsley (1968): Let $F \subset M'$ be closed and let $C(g) \subset M$ denote the continuity set of g. Then, since $\overline{g^{-1}(F)} \subset C(g)^c \cup g^{-1}(F)$ and $P_0(\mathbf{X}_0 \in C^c(g)) = 0$,

$$\limsup_n P_n^*(g(\mathbf{X}_n) \in F) = \limsup_n P_n^*(\mathbf{X}_n \in g^{-1}(F))$$

$$\leq \limsup_n P_n^*(\mathbf{X}_n \in \overline{g^{-1}(F)})$$

$$\leq P_0^*(\mathbf{X}_0 \in \overline{g^{-1}(F)}) \quad \text{by (i) implies (iii) of lemma 1.3,}$$

$$\leq P_0^*(\mathbf{X}_0 \in C(g)^c \cup g^{-1}(F))$$

$$= P_0(\mathbf{X}_0 \in g^{-1}(F))$$

$$= P_0(g(\mathbf{X}_0) \in F)$$

which yields the conclusion C by (iii) implies (i) of lemma 1.3. □

Proof of lemma 1.6. By contiguity, $E\,e^{\Lambda} = 1$. Furthermore, there exists a compact $K_\varepsilon \subset R$ with

$$Q_n(\Lambda_n \in (K_\varepsilon)^c) \le \varepsilon, \qquad n = 1, 2, \cdots .$$

Here $(K_\varepsilon)^c = M - K_\varepsilon$, the complement in M of K_ε. Since $1 = E\exp(\Lambda) < \infty$, K_ε can be chosen such that

$$E\, 1_{K_\varepsilon^c}(\Lambda)e^{\Lambda} \le \varepsilon .$$

Let g be continuous, with compact support such that $0 \le 1_{K_\varepsilon} \le g \le 1$. Hence $|g - 1| \le 1_{K_\varepsilon^c}$. Let $h \in C_b(M)$. Since $h(\mathbf{X}_n)^* \le \|h\|_\infty$,

$$|E^*_{Q_n} h(\mathbf{X}_n) - Eh(\mathbf{X})e^{\Lambda}|$$

$$\le |E_{Q_n} h(\mathbf{X}_n)^*(1 - g(\Lambda_n))| + |E_{Q_n} h(\mathbf{X}_n)^* g(\Lambda_n) - Eh(\mathbf{X})g(\Lambda)e^{\Lambda}|$$

$$+ |Eh(\mathbf{X})(g(\Lambda) - 1)e^{\Lambda}|$$

$$\le \|h\|_\infty E_{Q_n} |1 - g(\Lambda_n)| + |E_{Q_n} h(\mathbf{X}_n)^* g(\Lambda_n) - Eh(\mathbf{X})g(\Lambda)e^{\Lambda}|$$

$$+ \|h\|_\infty E|g(\Lambda) - 1|e^{\Lambda}$$

(a) $$\le 2\|h\|_\infty \varepsilon + |E_{Q_n} h(\mathbf{X}_n)^* g(\Lambda_n) - Eh(\mathbf{X})g(\Lambda)e^{\Lambda}| .$$

But

$$E_{Q_n} h(\mathbf{X}_n)^* g(\Lambda_n) = E_{P_n} h(\mathbf{X}_n)^* g(\Lambda_n)e^{\Lambda_n},$$

$$= E_{P_n}(h(\mathbf{X}_n)g(\Lambda_n)e^{\Lambda_n})^* \qquad \text{by lemma 1.2.vii}$$

(b) $$\to Eh(\mathbf{X})g(\Lambda)e^{\Lambda},$$

since $(x, \lambda) \to h(x)g(\lambda)e^{\lambda} \in C_b(M \times R)$. Combining (a) and (b) completes the proof of (13).

The second assertion of the lemma is obvious. □

Proof of theorem 1.1. For $m = 1, 2, \cdots$, let K_m be a compact such that $\liminf_n P_{n*}(\mathbf{X}_n \in K_m^\delta) \ge 1 - 1/m$ for every $\delta > 0$.

(1). There exists $\{n'\}$ such that $\lim_{n' \to \infty} E^* h(\mathbf{X}_{n'})$ exists for every $h \in C_b(M)$; say

(a) $$\lim_{n' \to \infty} E^* h(\mathbf{X}_{n'}) = T(h) .$$

Proof of (1). Since K_m is compact, $C(K_m)$ is separable. The same is true for $\{h \in C(K_m) : \|h\|_\infty \le 1\}$. By the Tietze extension theorem (Jameson (1974), theorem 12.4, page 113), every $h \in C(K_m)$ with $\sup_{y \in K_m} |h(y)| \le 1$ has an extension to an $h \in C_b(M)$ with $\|h\|_\infty \le 1$. Thus, there exists a countable subset of $\{h \in C_b(M) : \|h\|_\infty \le 1\}$, of which the restrictions to K_m are dense in

$\{h \in C(K_m) : \|h\|_\infty \le 1\}$.

Let $\{h_j\}_{j=1}^{\infty} \subset \{h \in C_b(M) : \|h\|_\infty \le 1\}$ have this property for every $m = 1, 2, \cdots$. By a diagonalization argument one can find $\{n'\} \subset \{n\}$ such that

$$E^* h_j(\boldsymbol{X}_{n'}) \to T(h_j), \qquad \text{as } n \to \infty$$

for every $j = 1, 2, \cdots$ and numbers $T(h_j) \in [-1, 1]$.

Fix $h \in C_b(M)$ with $\|h\|_\infty \le 1$. Given $\varepsilon > 0$ and m there exists h_j with

$$\sup_{y \in K_m} |h(y) - h_j(y)| < \varepsilon .$$

As a consequence, there exists $\delta > 0$ such that

$$\sup_{y \in K_m^\delta} |h(y) - h_j(y)| < 2\varepsilon .$$

(Indeed, suppose that there exists $y_p \in K_m^{1/p}$ with $|h(y_p) - h_j(y_p)| \ge \varepsilon$ for every $p = 1, 2, \cdots$. Then $d(y_p, K_m) \le 1/p \to 0$ as $p \to \infty$, so that there exists $x_p \in K_m$ with $d(y_p, x_p) \to 0$. Extract a subsequence $\{p'\}$ with $x_{p'} \to x \in K_m$. Then $y_{p'} \to x$ too, and by continuity of $h - h_j$ it would follow that $|h(x) - h_j(x)| \ge \varepsilon$.)

Now since $1_{K_m^\delta}(\boldsymbol{X}_n)_* + 1_{(K_m^\delta)^c}(\boldsymbol{X}_n)^* \equiv 1$ by lemma 1.2.x,

$$\begin{aligned}
&|E^* h(\boldsymbol{X}_n) - E^* h_j(\boldsymbol{X}_n)| \\
&\quad \le |E(h(\boldsymbol{X}_n)^* - h_j(\boldsymbol{X}_n)^*) 1_{K_m^\delta}(\boldsymbol{X}_n)_*| \\
&\qquad + |E\, h(\boldsymbol{X}_n)^* 1_{(K_m^\delta)^c}(\boldsymbol{X}_n)^*| + |E\, h_j(\boldsymbol{X}_n)^* 1_{(K_m^\delta)^c}(\boldsymbol{X}_n)^*| \\
&\quad \le E\, |h(\boldsymbol{X}_n) - h_j(\boldsymbol{X}_n)|^* 1_{A_n} + 2P_n^*(\boldsymbol{X}_n \in (K_m^\delta)^c),
\end{aligned}$$

for some measurable $A_n \in \mathbf{A}_n$ with $A_n \subset [\mathbb{X}_n \in K_m^\delta]$; this follows from lemma 1.2.x and 1.2.ix. The last expression is smaller than $2\varepsilon + (2/m)$ for sufficiently large n.

We conclude that $\{E^* h(\mathbb{X}_{n'})\}$ has the property that for every $\eta > 0$ there is a converging sequence of numbers that is eventually within distance η. Thus $E^* h(\mathbb{X}_{n'})$ converges, and (1) is proved.

(2). For every $h \in C_b(M)$

$$E^* h(\boldsymbol{X}_n) - E_* h(\boldsymbol{X}_n) \to 0 \qquad \text{as } n \to \infty .$$

Proof of (2). Fix $h \in C_b(M)$ with $\|h\|_\infty \le 1$, m, and $\varepsilon > 0$. By (ii) and an argument as above, there exists a $h_0 \in \mathbf{H}$ with $\|h_0\|_\infty \le 2$ and $\delta > 0$ such that

$$\sup_{y \in K_m^\delta} |h(y) - h_0(y)| \le 2\varepsilon .$$

Then, by lemma 1.2.x and 1.2.vi,

$$|E_* h(\mathbf{X}_n) - E_* h_0(\mathbf{X}_n)|$$

$$\le E\, |h(\mathbf{X}_n)_* - h_0(\mathbf{X}_n)_*|\, 1_{K_m^\delta}(\mathbf{X}_n)_*$$

$$+ E\, |h(\mathbf{X}_n)_*\, 1_{(K_m^\delta)^c}(\mathbf{X}_n)^*|$$

$$+ E\, |h_0(\mathbf{X}_n)_*\, 1_{(K_m^\delta)^c}(\mathbf{X}_n)^*|$$

$$\le E\, |h(\mathbf{X}_n) - h_0(\mathbf{X}_n)|^*\, 1_{K_m^\delta}(\mathbf{X}_n)_* + 3P_n^*(\mathbf{X}_n \in (K_m^\delta)^c)\,.$$

(a) $$\le 2\varepsilon + \frac{3}{m} \qquad \text{for } n \ge \text{some } N_m\,.$$

Similarly, by lemma 1.2.x and 1.2.iii,

$$|E^* h(\mathbf{X}_n) - E^* h_0(\mathbf{X}_n)|$$

(b) $$\le E\, |h(\mathbf{X}_n) - h_0(\mathbf{X}_n)|^*\, 1_{K_m^\delta}(\mathbf{X}_n)_* + 3P^*(\mathbf{X}_n \in (K_m^\delta)^c)\,.$$

For $n \to \infty$ the latter expression is less than $2\varepsilon + (3/m)$. Since

(c) $$E^* h(\mathbf{X}_n) - E_* h(\mathbf{X}_n)$$

$$= E^* h_0(\mathbf{X}_n) - E_* h_0(\mathbf{X}_n) + E^* h(\mathbf{X}_n) - E^* h_0(\mathbf{X}_n)$$

$$+ E_* h_0(\mathbf{X}_n) - E_* h(\mathbf{X}_n)\,,$$

(2) follows from (a), (b), and (iii).

(3). The map $T : C_b(M) \to R$ is an abstract integral; i.e. $T(h)$ is linear, positive, and continuous at 0 on $C_b(M)$.

Proof of (3). A. First linearity: For $h_1,\ h_2 \in C_b(M)$

$$T(h_1 + h_2) = \lim_{n' \to \infty} E^*(h_2 + h_2)(\mathbf{X}_{n'}) \le \lim_{n' \to \infty} (E^* h_1(\mathbf{X}_{n'}) + E^* h_2(\mathbf{X}_{n'}))$$

$$= T(h_1) + T(h_2) = \lim_{n' \to \infty} (E_* h_1(\mathbf{X}_{n'}) + E_* h_2(\mathbf{X}_{n'}))$$

$$\le \lim_{n' \to \infty} E_*(h_1(\mathbf{X}_{n'}) + h_2(\mathbf{X}_{n'})) = T(h_1 + h_2)$$

by (1) and (2). By a similar argument $T(ch) = c\,T(h)$ for $c \in R$ and $h \in C_b(M)$.

B. T is positive. This is trivial.

C. If $\{h_p\} \subset C_b(M)$ with $h_p \downarrow 0$ pointwise, then $T(h_p) \downarrow 0$. Indeed, by Dini's theorem $h_p \downarrow 0$ uniformly on compacts. Fix $\varepsilon > 0$ and m. For sufficiently large p,

$$\sup_{y \in K_m} |h_p(y)| < \varepsilon\,.$$

Then, as above, there exists $\delta \equiv \delta_p > 0$ such that

$$\sup_{y \in K_m^\delta} |h_p(y)| < 2\varepsilon .$$

Thus for sufficiently large p

$$T(h_p) = \lim_{n' \to \infty} (Eh_p(\mathbf{X}_{n'})^* 1_{K_m^\delta}(\mathbf{X}_{n'})_* + Eh_p(\mathbf{X}_{n'})^* 1_{(K_m^\delta)^c}(\mathbf{X}_{n'})^*)$$

$$\leq 2\varepsilon + \|h_1\|_\infty \frac{1}{m} ,$$

proving C and (3).

(4). By the Daniell-Stone theorem (Bauer (1983), theorem 7.1.4, page 197), there exists a probability measure L on $\mathrm{U}(C_b(M))$ with

$$T(h) = \int h \, dL , \qquad h \in C_b(M) .$$

Of course $\mathrm{U}(C_b(M))$ = Borel $\sigma-$field of M ; see e.g. Bauer (1983), theorem 7.2.4, page 206.

(5). $L(K_m) \geq 1 - \frac{1}{m}$.

Proof of (5). Indeed there exist $\{h_p\} \subset C_b(M)$ with

$$1_{K_m} \leq h_p \leq 1 \qquad \text{and} \qquad h_p \downarrow 1_{K_m} \qquad \text{as } p \to \infty ;$$

for example, take $h_p(x) = (1 - p\, d(x, K_m))^+$. Thus

$$L(K_m) = \lim_{p \to \infty} \int h_p \, dL = \lim_{p \to \infty} \lim_{n' \to \infty} E^* h_p(\mathbf{X}_{n'}) .$$

Let $0 < r < 1$. Since $h_p = 1$ on K_m, there exists a $\delta \equiv \delta(p, r) > 0$ such that

$$\sup_{y \in K_m^\delta} |h_p(y)| > r .$$

Thus

$$\lim_{n' \to \infty} E^* h_p(\mathbf{X}_{n'}) \geq \limsup_{n' \to \infty} r\, E^* 1_{K_m^\delta}(\mathbf{X}_{n'})$$

$$\geq \liminf_{n \to \infty} r\, E_* 1_{K_m^\delta}(\mathbf{X}_{n'}) \geq r\,(1 - \frac{1}{m}) , \qquad \text{by (i)} ,$$

and the claim follows. □

The proofs of corollaries 1.1 - 1.3 involve use of the following lemma to verify the second hypothesis of the generalized Prohorov theorem 1.1.

Lemma 5.1. Suppose that $\mathbf{H} \subset C_b(M)$ satisfies:

(i) $\mathbf{H}$ is an algebra; i.e. it is a vector space and $g, h \in \mathbf{H}$ implies $g\, h \in \mathbf{H}$;

(ii) $1 \in \mathbf{H}$;

(iii) $\mathbf{H}$ separates points of M;

and one of the following conditions:

(iv) for every $h \in \mathbf{H}$, $h \wedge 1 \in \mathbf{H}$;

(iv′) for every compact set $K_1 \subset M$ and $\delta > 0$ there exists a compact set $K \supset K_1$ and a function $\chi_\delta \in \mathbf{H}$ with $1_K \le \chi_\delta \le 1$ and $\chi_\delta = 0$ on $(K^\delta)^c$.

Then $\mathbf{H}$ approximates the unit ball of $C_b(M)$ on compacts; i.e. (ii) of theorem 1.1 is satisfied.

Proof. By the Stone - Weierstrass theorem (Jameson (1974), page 266), the restrictions $h|_K$ of the elements $h \in \mathbf{H}$ to a compact K are uniformly dense in $C(K)$. Let $h \in C_b(M)$ with $\|h\|_\infty \le 1$. There exists $\bar{h}_0 \in \mathbf{H}$ with $\bar{h}_0|_K$ uniformly close to $h|_K$; i.e.

$$\sup_{x \in K} |h(x) - \bar{h}_0(x)| \le \eta .$$

Now under (iv) take $h_0 = (\bar{h}_0 \wedge 1) \vee (-1)$. For the case of (iv′), first note that by the argument as in the proof of theorem 1.1, there exists $\delta > 0$ with

$$\sup_{x \in K^\delta} |h(x) - \bar{h}_0(x)| \le 2\eta .$$

Now set $h_0(x) \equiv \bar{h}_0(x)\chi_\delta(x)$; note that $h_0 \in \mathbf{H}$ since $\bar{h}_0$ and $\chi_\delta \in \mathbf{H}$ and $\mathbf{H}$ is an algebra. Then $h_0|_K = \bar{h}_0|_K$ and

$$\begin{aligned} \|h_0\|_\infty &\le \sup_{x \in K^\delta} |h_0(x)| + \sup_{x \in (K^\delta)^c} |h_0(x)| \\ &\le 2\eta + \sup_{x \in K^\delta} |h(x)| + 0 \\ &\le 2\eta + \|h\|_\infty . \quad \square \end{aligned}$$

Proof of corollary 1.1. By lemmas 1.4 and 1.5, hypothesis (i) of theorem 1.1 is satisfied. Thus we only need to check (ii) and (iii) of theorem 1.1.

Take $\mathbf{H}$ to be the linear space spanned by all functions of the form

$$(x, y) \rightarrow f(x)g(y), \quad f \in C_b(M_1), \qquad g \in C_b(M_2).$$

(iii): For $f \ge 0$ and $g \ge 0$

$$\text{(a)} \qquad f(\boldsymbol{X}_n)_* \, g(\boldsymbol{Y}_n)_* \le (f(\boldsymbol{X}_n) g(\boldsymbol{Y}_n))_* \le f(\boldsymbol{X}_n) g(\boldsymbol{Y}_n) \le (f(\boldsymbol{X}_n) g(\boldsymbol{Y}_n))^* \le f^*(\boldsymbol{X}_n) g^*(\boldsymbol{Y}_n) .$$

Thus

$$E^* f(\mathbf{X}_n) g(\mathbf{Y}_n) - E_* f(\mathbf{X}_n) g(\mathbf{Y}_n)$$
$$\le E f(\mathbf{X}_n)^* g(\mathbf{Y}_n)^* - E f(\mathbf{X}_n)_* g(\mathbf{Y}_n)_*$$
$$\le E|f(\mathbf{X}_n)^* - f(\mathbf{X}_n)_*||g(\mathbf{Y}_n)^*| + E|f(\mathbf{X}_n)_*||g(\mathbf{Y}_n)^* - g(\mathbf{Y}_n)_*|$$
$$\le \|g\|_\infty E(f(\mathbf{X}_n)^* - f(\mathbf{X}_n)_*) + \|f\|_\infty E(g(\mathbf{Y}_n)^* - g(\mathbf{Y}_n)_*)$$

(b) $\to 0$ by (1.4) and (1.15) .

For $f \ge 0$, $g \ge 0$ and arbitrary $a, b \in R$,

$$E^*(a+f)(\mathbf{X}_n)(b+g)(\mathbf{Y}_n) - E_*(a+f)(\mathbf{X}_n)(b+g)(\mathbf{Y}_n)$$
$$\le a + E^*(b f(\mathbf{X}_n)) + E^*(a\, g(\mathbf{Y}_n)) + E^*(f(\mathbf{X}_n) g(\mathbf{Y}_n))$$
$$- (a + E_*(b f(\mathbf{X}_n)) + E_*(a\, g(\mathbf{Y}_n)) + E_*(f(\mathbf{X}_n) g(\mathbf{Y}_n))$$

using lemma 1.2.i and 1.2.iv

(c) $\to 0$ using lemma 1.2.vii and the ≥ 0 case .

Finally, for linear combinations,

$$E^* \sum_i f_i(\mathbf{X}_n) g_i(\mathbf{Y}_n) - E_* \sum_i f_i(\mathbf{X}_n) g_i(\mathbf{Y}_n)$$
$$\le \sum_i (E^* f_i(\mathbf{X}_n) g_i(\mathbf{Y}_n) - E_* f_i(\mathbf{X}_n) g_i(\mathbf{Y}_n)) \qquad \text{by lemma 1.2(i) and (iv)}$$
$$\to 0 \qquad \text{using (c) .}$$

(ii): We apply lemma 5.1. It is easily seen that **H** is an algebra separating points of $M = M_1 \times M_2$, and $1 \in \mathbf{H}$. Thus (i) - (iii) of lemma 5.1 hold. To verify (iv)′, let $K \subset M$ be compact. Then $K \subset K_1 \times K_2$ where K_1 and K_2 are compacts in M_1 and M_2, and the function

$$\chi_\delta(x, y) = (1 - \delta^{-1} d_1(x, K_1))^+ (1 - \delta^{-1} d_2(y, K_2))^+ .$$

satisfies hypothesis (iv). □

Proof of corollary 1.2. We again verify the hypotheses of theorem 1.1. Hypothesis (i) of theorem 1.1 is obviously satisfied in view of (i). If we take

(a) $\mathbf{H} \equiv \{ h : M \to R : h(x) = g(x(f_1), \cdots, x(f_m))$

for some $f_1, \cdots, f_m \in \mathbf{F}$ and $g \in C_b(R^m)$ for some $m \ge 1 \}$,

then (ii) implies that (iii) of theorem 1.1 holds.

Moreover **H** satisfies (i) - (iv) of lemma 5.1, implying the validity of (ii) of theorem 1.1. Hence the conclusion follows from theorem 1.1. □

Proof of corollary 1.3. Use theorem 1.1 applied with $\mathbf{H}$ equal to the set of measurable $f \in C_b(M)$. The validity of (ii) of theorem 1.1 follows from lemma 5.1 (i) - (iv). □

Proof of proposition 1.2. Let $h : M \to R$ be bounded and continuous, and suppose $\mathbb{X}_n \Rightarrow_{Dudley} \mathbb{X}_0$. Then

$$\text{(a)} \qquad E^* h(\boldsymbol{X}_n) \equiv \inf\{ \int g \, dP_n : g : \mathbf{X}_n \to R \, , \, g \ge h \circ \boldsymbol{X}_n \, , \, g \text{ measurable} \}$$
$$\le \inf\{ \int g \, dP_n : g = f \circ \boldsymbol{X}_n \, , \, f \ge h \, , \, f : M \to R \, , \, f \ \mathbf{M}_b - \text{measurable} \}$$
$$= \inf\{ \int f \, dP_n \circ \boldsymbol{X}_n^{-1} : f : M \to R \, , \, f \ge h \, , \, f \ \mathbf{M}_b - \text{measurable} \}$$
$$= \int^* h \, dP_n \circ \boldsymbol{X}_n^{-1} .$$

Hence

$$\text{(b)} \qquad \limsup_{n \to \infty} E^* h(\boldsymbol{X}_n) \le \lim_{n \to \infty} \int^* h \, dP_n \circ \boldsymbol{X}_n^{-1} = \int h \, dP_0 \circ \boldsymbol{X}_0^{-1} = Eh(\boldsymbol{X}_0) ,$$

and this implies $\boldsymbol{X}_n \Rightarrow \boldsymbol{X}_0$ by (9).

On the other hand, if $\mathbb{X}_n \Rightarrow \mathbb{X}_0$, then

$$E^* h(\boldsymbol{X}_n) \to Eh(\boldsymbol{X}_0) \qquad \text{for all } h \in C_b(M) ,$$

so, in particular, for $h \in C_b(M, \mathbf{M}_b) \subset C_b(M)$, $\mathbf{A}_n - \mathbf{M}_b$ measurability of $\mathbb{X}_n$ implies that

$$Eh(\boldsymbol{X}_n) = E^* h(\boldsymbol{X}_n) \to Eh(\boldsymbol{X}_0) ;$$

i.e. $\mathbb{X}_n \Rightarrow_{Dudley} \mathbb{X}_0$. □

Proof of proposition 1.4. A: Let $\varepsilon > 0$; for $k = 1, 2, \cdots$ set

$$B_k \equiv \{ x \in M : \text{for some } y \in M \text{ with } d(x, y) < 1/k \, , \ d'(g(y), g(x)) > \varepsilon \} .$$

Then $C_k \equiv B_k \cap M_0$ is relatively open in M_0, and hence $C_k \in \mathbf{M}$. Moreover, the sets $C_k \downarrow \varnothing$ as $k \to \infty$ by continuity of g on M_0. Choose k so large that $P(\mathbb{X}_0 \in C_k) < \varepsilon$. Then

$$[d'(g(\boldsymbol{X}_n), g(\boldsymbol{X}_0)) > \varepsilon]$$
$$= [d'(g(\boldsymbol{X}_n), g(\boldsymbol{X}_0)) > \varepsilon] \cap ([\boldsymbol{X}_0 \in C_k^c] \cap [d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* < 1/k])^c$$
$$\cup [d'(g(\boldsymbol{X}_n), g(\boldsymbol{X}_0)) > \varepsilon] \cap [\boldsymbol{X}_0 \in C_k^c] \cap [d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* < 1/k]$$
$$\text{(a)} \qquad \subset [\boldsymbol{X}_0 \in C_k] \cup [d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* \ge 1/k] \cup [\boldsymbol{X}_0 \in M_0^c]$$

by the definition of C_k, and hence

$$P^*(d'(g(\boldsymbol{X}_n), g(\boldsymbol{X}_0)) > \varepsilon) \le P(\boldsymbol{X}_0 \in C_k) + P(d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* \ge 1/k)$$
$$\text{(b)} \qquad \le \varepsilon + \varepsilon = 2\varepsilon$$

for $n \ge$ some N_ε.

B: Let C_k, $k = 1, 2, \cdots$ be the sets in the proof of A. Then, by (a),

$$\text{(c)} \qquad [\sup_{m \geq n} d'(g(\boldsymbol{X}_m), g(\boldsymbol{X}_0)) > \varepsilon] = \bigcup_{m=n}^{\infty} [d'(g(\boldsymbol{X}_m), g(\boldsymbol{X}_0)) > \varepsilon]$$

$$\subset [\boldsymbol{X}_0 \in C_k] \cup \bigcup_{m=n}^{\infty} [d(\boldsymbol{X}_m, \boldsymbol{X}_0)^* \geq 1/k] \cup [\boldsymbol{X}_0 \in M_0^c]$$

$$\subset [\boldsymbol{X}_0 \in C_k] \cup [\sup_{m \geq n} d(\boldsymbol{X}_m, \boldsymbol{X}_0)^* \geq 1/k] \cup [\boldsymbol{X}_0 \in M_0^c],$$

and hence

$$\text{(d)} \qquad P^*(\sup_{m \geq n} d'(g(\boldsymbol{X}_m), g(\boldsymbol{X}_0)) > \varepsilon) \leq P(\boldsymbol{X}_0 \in C_k) + P(\sup_{m \geq n} d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* \geq 1/k)$$

$$\leq \varepsilon + \varepsilon = 2\varepsilon$$

for $n \geq$ some N_ε. □

Proof of proposition 1.5. First we show that $g|_{M_0}$ is continuous. Note that (ii) is equivalent to:

$$\text{(a)} \qquad \begin{array}{l} \text{for every } x \in M_0 \text{ both } g_n(x) \to g(x) \\ \text{and } \lim_{\delta \to 0} \limsup_{n \to \infty} \omega_{g_n}(S(x, \delta)) = 0 \end{array}$$

where $S(x, \delta) \subset M$ is the open sphere of radius δ centered at $x \in M_0$ and $\omega_g(S)$ denotes the oscillation of g on the set S. Hence for $x, y \in M_0$,

$$d'(g(y), g(x)) \leq d'(g(y), g_n(y)) + d'(g_n(y), g_n(x)) + d'(g_n(x), g(x))$$

$$\leq d'(g(y), g_n(y)) + \omega_{g_n}(S(x, d(x,y))) + d'(g_n(x), g(x))$$

$$\to 0$$

as $n \to \infty$ and then $d(x,y) \downarrow 0$.

Proof of A. Let $G \subset M'$ be open. By lemma 1.3, it suffices to show that

$$\text{(b)} \qquad \liminf_{n \to \infty} P_{n*}(g_n(\boldsymbol{X}_n) \in G) \geq P_0(g(\boldsymbol{X}_0) \in G).$$

Now (ii) implies that for every $x \in M_0$ and every $\varepsilon > 0$ there exist k, $\delta > 0$ such that $i \geq k$ and $d(x, y) < \delta$ imply $d'(gx, g_i y) < \varepsilon$. Set

$$T_k \equiv \bigcap_{i \geq k} \{ x \in M : g_i(x) \in G \} = \bigcap_{i \geq k} g_i^{-1} G.$$

Then, with $T_k^0 \equiv \text{interior}(T_k)$,

$$\text{(c)} \qquad T_k^0 \subset T_k \subset g_n^{-1} G \quad \text{for } n \geq k.$$

Note that $x \in M_0 \cap g^{-1}G$ implies that for δ sufficiently small and k sufficiently large, whenever $d(x, y) < \delta$ and $i \geq k$ it follows that $g_i(y) \in G$; hence the open sphere of radius δ about x, $S(x, \delta) \subset T_k$, so that $x \in T_k^0$, an open set. Thus

(d) $$g^{-1}G = (g^{-1}G \cap M_0) \cup (g^{-1}G \cap M_0^c) \subset \bigcup_k T_k^0 \cup M_0^c .$$

Since $P_0(\mathbf{X}_0 \in M_0^c) = 0$ by (i),

(e) $$\begin{aligned} P_0(g(\mathbf{X}_0) \in G) &\le P_0(\mathbf{X}_0 \in \bigcup_k T_k^0) \quad \text{by (d)} \\ &\le P_0(\mathbf{X}_0 \in T_k^0) + \varepsilon \quad \text{for } k \ge \text{some } K_\varepsilon \\ &\quad \text{since } T_k \uparrow \text{ implies } T_k^0 \uparrow \\ &\le \liminf_{n \to \infty} P_{n*}(\mathbf{X}_n \in T_k^0) + \varepsilon \\ &\quad \text{by } \mathbf{X}_n \Rightarrow \mathbf{X}_0 \text{ and lemma 1.3(ii)} \\ &\le \liminf_{n \to \infty} P_{n*}(g_n(\mathbf{X}_n) \in G) + \varepsilon \quad \text{by } (c). \end{aligned}$$

Since ε is arbitrary, (b) holds.

Proof of B. As noted in the proof of A, the following assertion is equivalent to the hypothesis (ii):

(f) for every $x \in M_0$ and every $\varepsilon > 0$ there exist $k \equiv k(x,\varepsilon)$, $\delta = \delta(x,\varepsilon) > 0$ such that $n \ge k$ and $d(x,y) < \delta$ imply $d'(g_n(y), g(x)) \le \varepsilon$.

This is equivalent to:

(g) for every $x \in M_0$ and every $\varepsilon > 0$ there exists a $k \equiv k(x,\varepsilon)$ such that $n \ge k$ and $d(x,y) < \frac{1}{k}$ imply $d'(g_n(y), g(x)) \le \varepsilon$.

We can therefore define

(h) $$k(x,\varepsilon) \equiv \min\left\{ k : \begin{array}{l} \text{for all } y \text{ with } d(x,y) < 1/k, \text{ for all } n \ge k, \\ d'(g_n(y), g(x)) \le \varepsilon \end{array} \right\}.$$

Claim. $k(\cdot,\varepsilon)$ is a measurable function. This will follow from

$$k(x,\varepsilon) \le \liminf_{m \to \infty} k(x_m,\varepsilon)$$

for any sequence $\{x_m\} \subset M_0$ with $x_m \to x \in M_0$.

Proof. Since k is integer - valued, the liminf is achieved for some subsequence $\{x_{m'}\}$, and in fact $\liminf_m k(x_m,\varepsilon) = \lim_{m'} k(x_{m'},\varepsilon) \equiv k'$ for all m' sufficiently large. Suppose that the right hand side is finite for some subsequence (if not, then the inequality is trivially true). If $d(x, y) < 1/k'$, then there exists an m_0 such that $d(x_m, y) < 1/k'$ for all $m \geq m_0$. Hence

(i) $$d'(g_n(y), g(x_m)) \leq \varepsilon \quad \text{for all } n \geq k' .$$

Since $g|_{M_0}$ is continuous as a consequence of (ii), we can let $m \to \infty$ in (d) to obtain

(j) $$d'(g_n(y), g(x)) \leq \varepsilon \quad \text{for all } n \geq k' .$$

Hence

(k) $$k(x, \varepsilon) \leq k' = \liminf_{m \to \infty} k(x_m, \varepsilon) .$$

It follows from (k) that for any fixed integer K the set $\{x : k(x, \varepsilon) \leq K\}$ is closed, and hence $k(\cdot, \varepsilon)$ is measurable, proving the claim.

Now we are ready to prove B. Let $\varepsilon > 0$. By the claim, $k(\mathbb{X}_0, \varepsilon)$ is a (proper) random variable, and there exists a $k_0 = k_0(\varepsilon)$ such that

(l) $$P(k(\boldsymbol{X}_0, \varepsilon) > k_0) < \frac{\varepsilon}{2} .$$

Since $d(\mathbb{X}_n, \mathbb{X}_0)^* \to_p 0$, for all $n \geq n_0(\varepsilon)$,

(m) $$P(d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* \geq \frac{1}{k_0}) < \frac{\varepsilon}{2} .$$

Then, with

(n) $$\begin{aligned} B_n &\equiv \{ d'(g_n(\boldsymbol{X}_n), g(\boldsymbol{X}_0))^* > \varepsilon \} \\ C_n &\equiv \{ d(\boldsymbol{X}_n, \boldsymbol{X}_0)^* \geq \frac{1}{k_0} \} \\ D &\equiv \{ k(\boldsymbol{X}_0, \varepsilon) > k_0 \} , \end{aligned}$$

for $n \geq \max\{n_0, k_0\}$,

$$\begin{aligned} P(B_n) &\leq P(B_n \cap (C_n^c \cap D^c)) + P(B_n \cap (C_n^c \cap D^c)^c) \\ &\leq P(\varnothing) + P(C_n) + P(D) \qquad \text{by (g) and (h)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad \text{by (l) - (n)} , \end{aligned}$$

and B holds.

Proof of C. Let $\varepsilon > 0$ and write

(o) $$\tilde{B}_n \equiv \{ \sup_{m \geq n} d'(g_m(\boldsymbol{X}_m), g(\boldsymbol{X}_0))^* > \varepsilon \} = \bigcup_{m=n}^{\infty} \{ d'(g_m(\boldsymbol{X}_m), g(\boldsymbol{X}_0))^* > \varepsilon \}$$

and

(p) $\qquad \tilde{C}_n \equiv \{ \sup_{m \geq n} d(\mathbf{X}_m, \mathbf{X}_0)^* \geq \frac{1}{k_0} \} \supset \bigcup_{m=n}^{\infty} \{ d(\mathbf{X}_m, \mathbf{X}_0)^* \geq \frac{1}{k_0} \}$.

By hypothesis, there is an $n_0 = n_0(\varepsilon)$ such that

(q) $\qquad P(\tilde{C}_n) < \frac{\varepsilon}{2} \quad$ for all $n \geq n_0$.

Hence

$$P(\tilde{B}_n) \leq P(\tilde{B}_n \cap (\tilde{C}_n^c \cap D^c)) + P(\tilde{B}_n \cap (\tilde{C}_n^c \cap D^c)^c)$$
$$\leq P(\emptyset) + P(\tilde{C}_n) + P(D) \qquad \text{by (g) and (h)}$$
$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad \text{by (l), (p), and (q)},$$

for all $n \geq \max\{n_0, k_0\}$, and C holds. □

Proof of lemma 1.8. Let $F \subset M$ be closed and $\varepsilon > 0$. Then

$$P_n^*(\mathbf{Y}_n \in F) \leq P_n^*(\mathbf{Y}_n \in F \text{ and } d(\mathbf{X}_n, \mathbf{Y}_n)^* < \varepsilon) + P_n(d(\mathbf{X}_n, \mathbf{Y}_n)^* \geq \varepsilon)$$
$$\leq P_n^*(\mathbf{X}_n \in F^\varepsilon) + o(1).$$

Thus $\limsup P_n^*(\mathbf{Y}_n \in F) \leq P(\mathbf{X}_0 \in F^\varepsilon)$ for every $\varepsilon > 0$. Finally let $\varepsilon \downarrow 0$; the conclusion then follows from (iii) implies (i) of the portmanteau lemma 1.3. □

Proof of lemma 1.7. Apply lemma 1.8 with $\mathbb{X}_n = \mathbb{X}_0$ and $\mathbb{Y}_n = \mathbb{X}_n$, $n = 1, 2, \cdots$. □

Proof of lemma 1.9. Let $\tilde{\mathbb{X}}_n$ and $\tilde{\mathbb{X}}_0$ be as in theorem 1.2. It suffices to show that

$$\sup_{h \in \mathbf{H}} |E^* h(\tilde{\mathbf{X}}_n) - Eh(\tilde{\mathbf{X}}_0)| \to 0 .$$

First, from $\tilde{\mathbf{X}}_n \to \tilde{\mathbf{X}}_0$ almost uniformly and equi-continuity of $\mathbf{H}$, it follows that

$$\sup_{h \in \mathbf{H}} |h(\tilde{\mathbf{X}}_n) - h(\tilde{\mathbf{X}}_0)| \to 0 \qquad \text{almost uniformly}.$$

This follows from a minor extension of the argument for B of proposition 1.4. It suffices to replace the B_k there by

$$B_k = \{x : \text{there exists } y \text{ with } d(x, y) < 1/k \text{ and } \sup_{h \in \mathbf{H}} |h(y) - h(x)| > \varepsilon \}.$$

Next,

$$\sup_{h \in \mathbf{H}} |E^* h(\tilde{\mathbf{X}}_n) - Eh(\tilde{\mathbf{X}}_0)| \leq E(\sup_{h \in \mathbf{H}} |h(\tilde{\mathbf{X}}_n) - h(\tilde{\mathbf{X}}_0)|)^* \to 0$$

since the integrand is uniformly bounded by assumption and converges to zero a.s. □

Proof of lemma 1.10. It suffices to show that $\liminf_{n \to \infty} E_* h(\mathbb{X}_n) \geq Eh(\mathbb{X}_0)$ for every $h \in C_b(M)$ with $h \geq 0$. For such h there exists a nondecreasing sequence f_m in $\mathbf{H}_0 \equiv \mathbf{H}_1$ with $f_m \leq h$ everywhere on M and $f_m \uparrow h$ on M_0. To see this,

note first that, since $\mathbf{H}_0$ is countable, this statement follows from

$$h(y) = \sup\{f(y) : f \in \mathbf{H}_S ,\ f \le h\}, \quad \text{every} \quad y \in M_0 .$$

Next fix $y \in M_0$. Assume without loss of generality that $h(y) > 0$. Fix $\varepsilon \in (0,1)$ with $h(y) - \varepsilon \in \mathbf{Q}^+$. There exists $\delta > 0$ such that $h(x) > h(y) - \varepsilon$ for all $x \in B(y, 2\delta)$. Now take $f \in \mathbf{H}_S$ as above with $d(s,y) \le \delta\varepsilon$, $p = 1/\delta$, and $q = h(y) - \varepsilon$. Then f vanishes outside $B(y, 2\delta)$ and has maximum value $h(y) - \varepsilon$. Thus $f \le h$. Moreover,

$$f(y) = (h(y) - \varepsilon)(1 - d(y,s)/\delta) \ge (h(y) - \varepsilon)(1 - \varepsilon) .$$

Now since $f_m \in \mathbf{H}_0$ or $-f_m \in \mathbf{H}_1$, it follows from (22) that $E^*(-f_m(\mathbf{X}_n) \to E(-f_m(\mathbf{X}_0)$ by (22), and hence

$$\liminf_{n \to \infty} E_* h(\mathbf{X}_n) \ge \liminf_{n \to \infty} E_* f_m(\mathbf{X}_n) = E f_m(\mathbf{X}_0) ,$$

for every m. Finally let $m \to \infty$ and apply the monotone convergence theorem. □

Proof of corollary 1.4. Every $h \in \mathbf{H}_0$ satisfies

$$|h(x) - h(y)| \le M\, d(x, y) ,$$

for some $M < \infty$, and hence is uniformly continuous. The assertion then follows from lemma 1.10. □

Proof of corollary 1.5. BL_1 is a bounded equicontinuous subset of $C_b(M)$, so $\mathbb{X}_n \Rightarrow \mathbb{X}_0$ implies (24) by lemma 1.9. Conversely, (24) immediately implies $E^* h(\mathbf{X}_n) \to Eh(\mathbf{X}_0)$ for all $h \in BL_1$, which implies that (22) holds since $h \in \mathbf{H}_0$ implies that $h/M \in BL_1$ for some finite constant M. Thus the conclusion follows from lemma 1.10. □

Proof of corollary 1.6. This follows immediately from lemma 1.10. □

Proof of theorem 1.3. This is analogous to the proof of theorem 1.1. The only step that needs a substantial change is step (1).

(1). There exists $\{\alpha'\}$ such that $\lim_{\alpha'} E^* h(\mathbb{X}_{\alpha'})$ exists for every $h \in C_b(M)$.

Proof of (1). Consider the net $\{E^* h(\mathbb{X}_\alpha)\}_{h \in C_b(M)}$ as a net in the product space

$$\prod_{h \in C_b(M)} [-\|h\|_\infty, \|h\|_\infty] .$$

This space is compact in the product topology, by Tychonov's theorem. Hence there exists a converging subnet. □

Acknowledgment: We owe thanks to R. M. Dudley for the proofs of parts B and C of proposition 1.5 during discussions at Oberwolfach in March 1989. We also owe thanks to Peter Gaenssler for a number of corrections of the first version, a suggestion concerning the formulation of lemma 1.10 and corollary 1.6, and for his encouragement.

BIBLIOGRAPHY

Andersen, N. T. and Dobrić, V., (1987). "The central limit theorem for stochastic processes," *Ann. Probability*, vol. 15, pp. 164 - 177.

Anderson, T. W., (1955). "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proc. Amer. Math. Soc.*, vol. 6, pp. 170 - 176.

Bauer, H., (1972). *Probability Theory and Elements of Measure Theory*, Holt, Rinehart, and Winston, Inc., New York.

Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A., (1983). "Information and asymptotic efficiency in parametric - nonparametric models," *Ann. Statist.*, vol. 11, pp. 432 - 452.

Beran, R., (1977). "Estimating a distribution function," *Ann. Statist.*, vol. 5, pp. 400 - 404.

Bickel, P. J., Ritov, Y., and Wellner, J. A., (1988). "Efficient estimation of linear functionals of a probability measure P with known marginal distributions," *Technical Report*, vol. 135, Department of Statistics, University of Washington.

Billingsley, P. , (1968). *Convergence of Probability Measures*, Wiley, New York.

Billingsley, P., (1971). "Weak Convergence of Measures: Applications in Probability," *Regional Conference Series in Applied Mathematics*, vol. 5, SIAM, Philadelphia.

Dudley, R. M., (1966). "Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces," *Illinois J. Math.*, vol. 10, pp. 109 - 126.

Dudley, R. M., (1976). "Probabilities and metrics: convergence of laws on metric spaces with a view to statistical testing," *Lecture Notes Series*, vol. 45, Matematish Institute, Aarhus Universitet.

Dudley, R. M., (1978). "Central limit theorems for empirical measures," *Ann. Probability*, vol. 6, pp. 899-929.

Dudley, R. M., (1984). "A course on empirical processes. Ecole d'eté de probabilités de St.-Flour, 1982.," *Lecture Notes in Math.*, vol. 1097, pp. 2 - 142.

Dudley, R. M., (1985). "An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions," *Lecture Notes in Math.*, vol. 1153, pp. 141 - 178, Springer - Verlag, New York.

Dudley, R. M. and Philipp, W., (1983). "Invariance principles for sums of Banach space valued random elements and empirical processes," *Z. Wahrsch. verw. Geb.*, vol. 62, pp. 509 - 552.

Gaenssler, P., (1983). "Empirical Processes," *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, vol. 3, Institute of Mathematical Statistics, Hayward.

Gaenssler, P. and Schneemeier, W., (1987). "On functional limit theorems for a class of stochastic processes indexed by pseudo-metric parameter spaces (with applications to empirical processes)," *Preprint, Mathematics Institute, University of Munich.*

Giné, E. and Zinn, J. , (1984). "Some limit theorems for empirical processes," *Ann. Probability*, vol. 12, pp. 929 - 989.

Giné, E. and Zinn, J., (1986). "Lectures on the central limit theorem for empirical processes," *Lect. Notes in Math.*, vol. 1221, pp. 50 - 113, Springer, Berlin.

Hájek, J., (1970). "A characterization of limiting distributions of regular estimates," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 14, pp. 323 - 330.

Hájek, J., (1972). "Local asymptotic minimax and admissibility in estimation," *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 175 - 194, University of California Press, Berkeley.

Hoffmann-Jørgensen, J., (1984). "Stochastic Processes on Polish Spaces," *Unpublished manuscript.*

Jain, N., (1977). "Central limit theorem and related questions in Banach space," *Proc. 1976 A.M.S. Probability Symp. Urbana*, vol. 31, pp. 55 - 65, American Mathematical Society, Providence.

Jameson, G. J. O., (1974). *Topology and Normed Spaces*, Chapman and Hall, London.

Kelley, J. , (1955). *General Topology*, Springer Verlag, New York.

Koshevnik, Yu. A. and Levit, B. Ya., (1976). "On a non-parametric analogue of the information matrix," *Theor. Prob. Applic.*, vol. 21, pp. 738 - 753.

Kuelbs, J., (1976). "A strong convergence theorem for Banach space valued random variables," *Ann. Probability*, vol. 4, pp. 684 - 689.

Le Cam, L., (1953). "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates," *University of California Publ. Statist.*, vol. 1, pp. 277 - 330.

Le Cam, L., (1972). "Limits of experiments," *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 245 - 261, University of California Press, Berkeley.

Levit, B. Ya., (1978). "Infinite - dimensional informational inequalities," *Theory Prob. Appl.* , vol. 23, pp. 371 - 377.

Millar, P. W., (1979.). "Asymptotic minimax theorems for the sample distribution function," *Z. Wahrsch. verw. Geb.* , vol. 48, pp. 233 - 252.

Millar, P. W. , (1983). "The minimax principle in asymptotic statistical theory. Ecole d'Eté de Probabilités de St. Flour XI - 1981. ," *Lecture Notes in Math.*, vol. 876, pp. 76 - 262.

Millar, P. W., (1985). "Nonparametric applications of an infinite dimensional convolution theorem," *Z. Wahrsch. Th. verw. Gebiete*, vol. 68, pp. 545-556.

Nummelin, E. , (1984). *General irreducible Markov chains and non-negative operators*, Cambridge University Press, Cambridge.

Parthasarathy, K. R., (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.

Penev, S., (1988). "On the local asymptotic optimality of the empirical distribution function for the stationary distribution of a strongly stable Markov sequence," *Preprint*, Applied Mathematics Center, Sofia, Bulgaria.

Pfanzagl, J. (with W. Wefelmeyer), (1985). "Asymptotic Expansions for General Statistical Models ," *Lecture Notes in Statistics* , vol. 31, Springer Verlag, New York.

Pollard, D., (1984). *Convergence of Stochastic Processes*, Springer - Verlag, New York.

Pollard, D., (1988). "Empirical Processes: Theory and Applications," *Lectures at the University of Iowa.*

Roussas, G. G., (1972). *Contiguity of Probability Measures*, Cambridge University Press, Cambridge.

Sheehy, A. and Wellner, J. A., (1988). "Uniformity in P of some limit theorems for empirical measures and processes," *Technical Report*, vol. 134, p. Department of Statistics, University of Washington, Seattle.

Skorohod, A. V., (1956). "Limit theorems for stochastic processes," *Theor. Probability Appl.*, vol. 1, pp. 261 - 290.

Topsøe, Flemming, (1967). "Preservation of weak convergence under mappings," *Ann. Math. Statist.*, vol. 38, pp. 1661 - 1665.

Vaart, A. W. van der, (1988a). "Statistical estimation in large parameter spaces," *CWI Tract*, vol. 44, Centrum voor Wiskunde en Informatica, Amsterdam.

Vaart, A. W. van der, (1988b). "Efficiency and Hadamard Differentiability," *Technical Report*, vol. 143, p. Department of Statistics, University of Washington, Seattle.

Wellner, J. A., (1989). ''Discussion of: "Non- and semiparametric maximum likelihood estimators and the von Mises method (Part I)" by Richard Gill,'' *Scand. J. Statist.*, vol. 16, pp. 97 - 128.

Wichura, M. J., (1968). ''On the weak convergence of non-Borel probabilities on a metric space.,'' *Ph.D. dissertation* , Columbia University.

Wichura, M. J., (1970). ''On the construction of almost uniformly convergent random variables with given weakly convergent image laws,'' *Ann. Math. Statist.* , vol. 41, pp. 284 - 291.

APPENDIX: Statements and Proofs of the Net Convergence Results

Here we give complete statements and proofs for the net convergence generalizations of lemmas 1.3 - 1.8, proposition 1.1, and corollary 1.1 which were briefly stated in section 1.7.

As in section 1.1 suppose that (M, d) is a metric space (nonseparable in general), $\{(\mathbf{X}_\alpha, \mathbf{A}_\alpha, P_\alpha)\}_{\alpha \in A}$ is a net of probability spaces, and

(1) $\qquad \mathbb{X}_\alpha : \mathbf{X}_\alpha \to M, \quad \text{for} \quad \alpha \in A$

are arbitrary maps. As before, $C_b(M)$ is the collection of bounded, continuous functions h from M to R.

Definition 1.2.net. We say that $\mathbb{X}_\alpha$ *converges weakly to a random element* $\mathbb{X}_0$ in $(M, \mathbf{M})$, and write $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$, if for every $h \in \mathbf{C}_b(M)$,

(2) $\qquad \lim_\alpha E^* h(\mathbb{X}_\alpha) = E\, h(\mathbb{X}_0)$.

Definition 1.3.net. The net $\{\mathbb{X}_\alpha\}_{\alpha \in A}$ is *tight* if and only if for every $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon \subset M$ such that for all $\delta > 0$

(3) $\qquad \liminf_\alpha P_{\alpha,*}(\mathbb{X}_\alpha \in K^\delta) \geq 1 - \varepsilon;$

where $K^\delta \equiv \{x \in M : d(x, K) < \delta\}$, or, equivalently

(4) $\qquad \limsup_\alpha P_\alpha^*(\mathbb{X}_\alpha \notin K^\delta) < \varepsilon$.

Lemma 1.3.net. (Portmanteau theorem). The following statements are equivalent:

(i) $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$.

(ii) $\liminf_\alpha P_{\alpha,*}(\mathbb{X}_\alpha \in G) \geq P_0(\mathbb{X}_0 \in G)$ for every open set $G \subset M$.

(iii) $\limsup_\alpha P_\alpha^*(\mathbb{X}_\alpha \in F) \leq P_0(\mathbb{X}_0 \in F)$ for every closed set $F \subset M$.

(iv) $\liminf_\alpha E_* h(\mathbb{X}_\alpha) \geq Eh(\mathbb{X}_0)$ for every bounded, lower semi-continuous function h.

(v) $\limsup_\alpha E^* h(\mathbb{X}_\alpha) \leq Eh(\mathbb{X}_0)$ for every bounded, upper semi-continuous function h.

(vi) $\lim_\alpha P_\alpha^*(\mathbb{X}_\alpha \in A) = P(\mathbb{X}_0 \in A)$ for every Borel set A with $P_0(\mathbb{X}_0 \in \partial A) = 0$.

Lemma 1.4.net. (Weak convergence to a tight limit implies tightness). Suppose that $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$ where $\mathbb{X}_0$ is tight. Then $\{\mathbb{X}_\alpha\}$ is tight.

For $\alpha \in A$, let $\mathbb{X}_\alpha : \mathbf{X}_\alpha \to M_1$ and $\mathbb{Y}_\alpha : \mathbf{X}_\alpha \to M_2$ be maps into metric spaces M_1 and M_2. Equip $M_1 \times M_2$ with the metric

$$d((x_1,x_2),(y_1,y_2)) = d_1(x_1,y_1) \vee d_2(x_2,y_2).$$

Lemma 1.5.net. (Marginal tightness implies joint tightness). If both $\{\mathbb{X}_\alpha\}$ and $\{\mathbb{Y}_\alpha\}$ are tight, then $\{(\mathbb{X}_\alpha, \mathbb{Y}_\alpha)\}$ is tight too.

Proposition 1.1.net. (Continuous mapping theorem; CM). Suppose that:

(i) $g : M \to M'$ is continuous on a Borel set $M_0 \subset M$.

(ii) $\mathbb{X}_0$ is Borel measurable and $P_0(\mathbb{X}_0 \in M_0) = 1$.

Then $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$ implies that $g(\mathbb{X}_\alpha) \Rightarrow g(\mathbb{X}_0)$.

Theorem 1.1.net. (= **Theorem 1.3;** Extension of Prohorov's theorem for nets). Suppose that:

(i) $\{\mathbb{X}_\alpha\}_{\alpha \in A}$ is tight.

There exists a subset $\mathbf{H} \subset C_b(M)$ such that:

(ii) $\mathbf{H}$ aproximates the unit ball of $C_b(M)$ on compacts.

(iii) $\lim_\alpha \{ E^* h(\mathbb{X}_\alpha) - E_* h(\mathbb{X}_\alpha)\} = 0$ for all $h \in \mathbf{H}$.

Then there exists a subnet $\{\mathbb{X}_{\alpha'}\}_{\alpha' \in A'} \subset \{\mathbb{X}_\alpha\}_{\alpha \in A}$ such that $\mathbb{X}_{\alpha'} \Rightarrow$ some tight $\mathbb{X}_0$.

Corollary 1.1.net. (= **Corollary 1.7.**) For $\alpha \in A$, suppose that $\mathbb{X}_\alpha : \mathrm{X}_\alpha \to M_1$ and $\mathbb{Y}_\alpha : \mathrm{X}_\alpha \to M_2$ are maps with

$$\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0 \quad \text{and} \quad \mathbb{Y}_\alpha \Rightarrow \mathbb{Y}_0,$$

where $\mathbb{X}_0$ and $\mathbb{Y}_0$ are tight Borel measurable maps into M_1 and M_2 respectively. Then there exists a subnet $\{(\mathbb{X}_{\alpha'}, \mathbb{Y}_{\alpha'})\}_{\alpha' \in A'} \subset \{(\mathbb{X}_\alpha, \mathbb{Y}_\alpha)\}_{\alpha \in A}$ such that

$$(5) \qquad (\mathbb{X}_{\alpha'}, \mathbb{Y}_{\alpha'}) \Rightarrow (\mathbb{X}_0, \mathbb{Y}_0)$$

for some tight joint law $\mathbf{L}(\mathbb{X}_0, \mathbb{Y}_0)$ on the product space $(M_1 \times M_2, \mathbf{M}_1 \times \mathbf{M}_2)$.

To prepare for the net version of Le Cam's third lemma, for $\alpha \in A$, let P_α and Q_α be probability measures on measurable spaces $(\mathbf{X}_\alpha, \mathbf{A}_\alpha)$. Let Λ_α be the log-likelihood ratio of Q_α with respect to P_α; i.e. given densities q_α and p_α with respect to a $\sigma-$finite dominating measure μ_α (e.g. $\mu_\alpha = P_\alpha + Q_\alpha$), let

$$\Lambda_\alpha = \log\left(\frac{q_\alpha}{p_\alpha}\right),$$

where $\log \frac{a}{b} = -\infty$ if $a = 0 < b$, $+\infty$ if $b = 0 < a$, and 0 if $a = b$.

Let $\mathbb{X}_\alpha : \mathrm{X}_\alpha \to M$ as before. Equip $M \times \bar{R}$ with the metric

$$d((x,r),(y,s)) = d(x,y) \vee \arctan |r - s| .$$

Lemma 1.6.net. (An extension of Le Cam's third lemma). Let P_α and Q_α be contiguous, and suppose that

(6) $\qquad (\mathbb{X}_\alpha, \Lambda_\alpha) \Rightarrow (\mathbb{X}, \Lambda) \qquad$ under P_α

where $(\mathbb{X}, \Lambda) : \mathrm{X}_0 \to M \times R$ is Borel measurable. Then

(7) $\qquad \mathbb{X}_\alpha \Rightarrow \mathbb{Z} \qquad$ under Q_α

where $\mathbb{Z} : \mathrm{X}_0 \to M$ is Borel measurable and

(8) $\qquad P(\mathbb{Z} \in B) = E\, 1_B(\mathbb{X}) e^{\Lambda}.$

Furthermore, if $\mathbb{X}$ is separable (or tight), then $\mathbb{Z}$ may be taken to be separable (or tight) too.

Lemma 1.7.net. Let every $\mathbb{X}_\alpha$ be defined on the same probability space $(\mathrm{X}, \mathrm{A}, P)$ and assume $d(\mathbb{X}_\alpha, \mathbb{X}_0)^* \to_p 0$ where $\mathbb{X}_0$ is Borel measurable. Then $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$.

This lemma is actually a special case of the following lemma:

Lemma 1.8.net. For $\alpha \in A$, let $\mathbb{X}_\alpha : \mathrm{X}_\alpha \to M$ and $\mathbb{Y}_\alpha : \mathrm{X}_\alpha \to M$ be arbitrary maps. Suppose that $\mathbb{X}_\alpha \Rightarrow \mathbb{X}_0$ where $\mathbb{X}_0$ is Borel measurable and $d(\mathbb{X}_\alpha, \mathbb{Y}_\alpha)^* \to_p 0$. Then $\mathbb{Y}_\alpha \Rightarrow \mathbb{X}_0$.

Proofs

Proof of lemma 1.3.net. The equivalence of (ii) and (iii) is trivial; similarly, the equivalence of (iv) and (v) is also trivial. That (iv) and (v) together imply (i) is easy since a continuous function is both upper and lower semicontinuous.

Now we show that (i) implies (ii). First, there exists a sequence $\{h_m\} \subset C_b(M)$ with $0 \le h_m \le 1_G$ and $h_m \uparrow 1_G$. Now for every $m = 1, 2, \cdots$

$$\liminf_\alpha P_{\alpha *}(\mathbb{X}_\alpha \in G) \ge \liminf_\alpha E_* h_m(\mathbb{X}_\alpha) \ge E_* h_m(\mathbb{X}_0) .$$

By monotone convergence $Eh_m(\mathbb{X}_0) \uparrow P(\mathbb{X}_0 \in G)$ as $m \to \infty$. To show that (iii) implies (v), assume without loss of generality that $0 \le f \le 1$. Fix r. For $p = 0, \cdots, r$, set $F_p = \{x : f(x) \ge p/r\}$ (a closed set by upper semicontinuity of f) and $f_r(x) = r^{-1} \sum_{p=0}^{r} 1_{F_p}(x)$. Then $f_r \ge f$ and $\|f_r - f\|_\infty \le 1/r$. Now

$$\limsup_\alpha E^* f(\mathbb{X}_\alpha) \le \limsup_\alpha E^* f_r(\mathbb{X}_\alpha) \le \frac{1}{r} \sum_{p=0}^{r} \limsup_\alpha P^*_\alpha(\mathbb{X}_\alpha \in F_r)$$

$$\le \frac{1}{r} \sum_{p=0}^{r} P_0(\mathbb{X}_0 \in F_p) \qquad \text{by (iii)}$$

$$= E f_r(\boldsymbol{X}_0).$$

Let $r \to \infty$ to get the conclusion.

Now we show that (vi) implies (iii). We have $\partial F^{\varepsilon} = \{x : d(x, F) = \varepsilon\}$. Thus we have $\partial F^{\varepsilon_1} \cap \partial F^{\varepsilon_2} \neq \emptyset$ if $\varepsilon_1 \neq \varepsilon_2$, so that $P_0(\boldsymbol{X}_0 \in \partial F^{\varepsilon})$ can be nonzero for at most countably many $\varepsilon > 0$. Choose $\varepsilon_m \downarrow 0$ such that $P_0(\boldsymbol{X}_0 \in \partial F^{\varepsilon_m}) = 0$, $m = 1, 2, \cdots$. Then

$$\limsup_{\alpha} P_{\alpha}^{*}(\boldsymbol{X}_{\alpha} \in F) \leq \limsup_{\alpha} P_{\alpha}^{*}(\boldsymbol{X}_{\alpha} \in F^{\varepsilon_m}) = P(\boldsymbol{X}_0 \in F^{\varepsilon_m}).$$

Letting $m \to \infty$ yields (iii).

To complete the proof, we show that (ii) and (iii) together imply (vi). For $A \subset M$ with $P_0(\boldsymbol{X}_0 \in \partial A) = 0$,

$$\begin{aligned} P_0(\boldsymbol{X}_0 \in \operatorname{Int} A) &\leq \liminf_{\alpha} P_{\alpha *}(\boldsymbol{X}_{\alpha} \in \operatorname{Int} A) \\ &\leq \liminf_{\alpha} P_{\alpha}^{*}(\boldsymbol{X}_{\alpha} \in A) \\ &\leq \limsup_{\alpha} P_{\alpha}^{*}(\boldsymbol{X}_{\alpha} \in A) \\ &\leq \limsup_{\alpha} P_{\alpha}^{*}(\boldsymbol{X}_{\alpha} \in \bar{A}) \leq P_0(\boldsymbol{X}_0 \in \bar{A}). \end{aligned}$$

But the left and right sides are equal since $P_0(\boldsymbol{X}_0 \in \partial A) = 0$, and hence (vi) follows.
□

Proof of lemma 1.4.net. For every $\varepsilon > 0$ there exists a compact set $K = K_{\varepsilon}$ with $P_0(\boldsymbol{X}_0 \in K) \geq 1 - \varepsilon$. By lemma 1.3.net,

$$\liminf_{\alpha} P_{\alpha *}(\boldsymbol{X}_{\alpha} \in K^{\delta}) \geq P_0(\boldsymbol{X}_0 \in K^{\delta}) \geq P_0(\boldsymbol{X}_0 \in K) \geq 1 - \varepsilon. \quad \square$$

Proof of lemma 1.5.net. Let $\varepsilon > 0$, and let K_1 and K_2 be compact sets in M_1 and M_2 respectively with $\liminf_{\alpha} P_{\alpha *}(\boldsymbol{X}_{\alpha} \in K_1^{\delta}) \geq 1 - \varepsilon$ and $\liminf_{\alpha} P_{\alpha *}(\boldsymbol{Y}_{\alpha} \in K_2^{\delta}) \geq 1 - \varepsilon$ for every $\delta > 0$. Then $(K_1 \times K_2)^{\delta} = K_1^{\delta} \times K_2^{\delta}$ and

$$\begin{aligned} &\limsup_{\alpha} P_{\alpha}^{*}((\boldsymbol{X}_{\alpha}, \boldsymbol{Y}_{\alpha}) \in (K_1^{\delta} \times K_2^{\delta})^c) \\ &\quad \leq \limsup_{\alpha} P_{\alpha}^{*}((\boldsymbol{X}_{\alpha}, \boldsymbol{Y}_{\alpha}) \in (K_1^{\delta})^c \times M_2 \cup M_1 \times (K_2^{\delta})^c) \\ &\quad \leq \limsup_{\alpha} \{ P_{\alpha}^{*}((\boldsymbol{X}_{\alpha}, \boldsymbol{Y}_{\alpha}) \in (K_1^{\delta})^c \times M_2) + P_{\alpha}^{*}((\boldsymbol{X}_{\alpha}, \boldsymbol{Y}_{\alpha}) \in M_1 \times (K_2^{\delta})^c \} \\ &\quad \leq 2\varepsilon. \quad \square \end{aligned}$$

Proof of Proposition 1.1.net. This is almost as in Billingsley (1968): Let $F \subset M'$ be closed and let $C(g) \subset M$ denote the continuity set of g. Then, since $\overline{g^{-1}(F)} \subset C(g)^c \cup g^{-1}(F)$ and $P_0(\boldsymbol{X}_0 \in C^c(g)) = 0$,

$$\limsup_\alpha P_\alpha^*(g(\boldsymbol{X}_\alpha) \in F) = \limsup_\alpha P_\alpha^*(\boldsymbol{X}_\alpha \in g^{-1}(F))$$

$$\leq \limsup_\alpha P_\alpha^*(\boldsymbol{X}_\alpha \in \overline{g^{-1}(F)})$$

$$\leq P_0^*(\boldsymbol{X}_0 \in \overline{g^{-1}(F)})$$

by (i) implies (iii) of lemma 1.3.net,

$$\leq P_0^*(\boldsymbol{X}_0 \in C(g)^c \cup g^{-1}(F))$$

$$= P_0(\boldsymbol{X}_0 \in g^{-1}(F))$$

$$= P_0(g(\boldsymbol{X}_0) \in F)$$

which yields the conclusion C by (iii) implies (i) of lemma 1.3.net. □

Proof of lemma 1.6.net. By contiguity, $E\, e^\Lambda = 1$. Furthermore, there exists a compact $K_\varepsilon \subset R$ with

$$Q_\alpha(\Lambda_\alpha \in (K_\varepsilon)^c) \leq \varepsilon, \qquad n = 1, 2, \cdots .$$

Here $(K_\varepsilon)^c = M - K_\varepsilon$, the complement in M of K_ε. Since $1 = E\exp(\Lambda) < \infty$, K_ε can be chosen such that

$$E\, 1_{K_\varepsilon^c}(\Lambda)e^\Lambda \leq \varepsilon .$$

Let g be continuous, with compact support such that $0 \leq 1_{K_\varepsilon} \leq g \leq 1$. Hence $|g - 1| \leq 1_{K_\varepsilon^c}$. Let $h \in C_b(M)$. Since $h(\boldsymbol{X}_\alpha)^* \leq \|h\|_\infty$,

$$|E_{Q_\alpha}^* h(\boldsymbol{X}_\alpha) - Eh(\boldsymbol{X})e^\Lambda|$$

$$\leq |E_{Q_\alpha} h(\boldsymbol{X}_\alpha)^*(1 - g(\Lambda_\alpha))| + |E_{Q_\alpha} h(\boldsymbol{X}_\alpha)^* g(\Lambda_\alpha) - Eh(\boldsymbol{X})g(\Lambda)e^\Lambda|$$

$$+ |Eh(\boldsymbol{X})(g(\Lambda) - 1)e^\Lambda|$$

$$\leq \|h\|_\infty E_{Q_\alpha}|1 - g(\Lambda_\alpha)| + |E_{Q_\alpha} h(\boldsymbol{X}_\alpha)^* g(\Lambda_\alpha) - Eh(\boldsymbol{X})g(\Lambda)e^\Lambda|$$

$$+ \|h\|_\infty E|g(\Lambda) - 1|e^\Lambda$$

(a) $$\leq 2\|h\|_\infty \varepsilon + |E_{Q_\alpha} h(\boldsymbol{X}_\alpha)^* g(\Lambda_\alpha) - Eh(\boldsymbol{X})g(\Lambda)e^\Lambda| .$$

But

$$E_{Q_\alpha} h(\boldsymbol{X}_\alpha)^* g(\Lambda_\alpha) = E_{P_\alpha} h(\boldsymbol{X}_\alpha)^* g(\Lambda_\alpha)e^{\Lambda_\alpha},$$

$$= E_{P_\alpha}(h(\mathbf{X}_\alpha)\, g(\Lambda_\alpha)\, e^{\Lambda_\alpha})^* \qquad \text{by lemma 1.2.vii}$$

(b) $$\to E h(\mathbf{X})\, g(\Lambda)\, e^{\Lambda},$$

since $(x,\lambda) \to h(x)\, g(\lambda) e^{\lambda} \in C_b(M \times R)$. Combining (a) and (b) completes the proof of (8).

The second assertion of the lemma is obvious. □

Proof of theorem 1.1.net. The only step that needs a substantial change is step (1).

For $m = 1, 2, \cdots,$ let K_m be a compact such that $\liminf_\alpha P_{\alpha,*}(\mathbf{X}_\alpha \in K_m^\delta) \geq 1 - 1/m$ for every $\delta > 0$.

(1). There exists a subnet $\{\alpha'\}$ such that $\lim_{\alpha'} E^* h(\mathbf{X}_{\alpha'})$ exists for every $h \in C_b(M)$.

Proof of (1). Consider the net $\{E^* h(\mathbf{X}_\alpha)\}_{h \in C_b(M)}$ as a net in the product space

$$\prod_{h \in C_b(M)} [-\|h\|_\infty, \|h\|_\infty].$$

This space is compact in the product topology, by Tychonov's theorem. Hence there exists a converging subnet. □

(2). For every $h \in C_b(M)$

$$\lim_\alpha \{E^* h(\mathbf{X}_\alpha) - E_* h(\mathbf{X}_\alpha)\} = 0.$$

Proof of (2). Fix $h \in C_b(M)$ with $\|h\|_\infty \leq 1$, m, and $\varepsilon > 0$. By (ii) and an argument as above, there exists a $h_0 \in \mathbf{H}$ with $\|h_0\|_\infty \leq 2$ and $\delta > 0$ such that

$$\sup_{y \in K_m^\delta} |h(y) - h_0(y)| \leq 2\varepsilon.$$

Then, by lemma 1.2.x and 1.2.vi,

$$|E_* h(\mathbf{X}_\alpha) - E_* h_0(\mathbf{X}_\alpha)|$$

$$\leq E\, |h(\mathbf{X}_\alpha)_* - h_0(\mathbf{X}_\alpha)_*|\, 1_{K_m^\delta}(\mathbf{X}_\alpha)_*$$

$$+ E\, |h(\mathbf{X}_\alpha)_*\, 1_{(K_m^\delta)^c}(\mathbf{X}_\alpha)^*|$$

$$+ E\, |h_0(\mathbf{X}_\alpha)_*\, 1_{(K_m^\delta)^c}(\mathbf{X}_\alpha)^*|$$

$$\leq E\, |h(\mathbf{X}_\alpha) - h_0(\mathbf{X}_\alpha)|^*\, 1_{K_m^\delta}(\mathbf{X}_\alpha)_* + 3 P_\alpha^*(\mathbf{X}_\alpha \in (K_m^\delta)^c).$$

(a) $$\leq 2\varepsilon + \frac{3}{m} \qquad \text{for } \alpha \geq \text{some } \alpha_m.$$

Similarly, by lemma 1.2.x and 1.2.iii,

$$|E^* h(\mathbf{X}_\alpha) - E^* h_0(\mathbf{X}_\alpha)|$$

(b) $$\leq E\, |h(\boldsymbol{X}_\alpha) - h_0(\boldsymbol{X}_\alpha)|^* \, 1_{K_m^\delta}(\boldsymbol{X}_\alpha)_* + 3P^*(\boldsymbol{X}_\alpha \in (K_m^\delta)^c)\,.$$

For $\alpha \geq \alpha_m$ the latter expression is less than $2\varepsilon + (3/m)$, just as in (a). Since

(c) $$E^* h(\boldsymbol{X}_\alpha) - E_* h(\boldsymbol{X}_\alpha)$$
$$= E^* h_0(\boldsymbol{X}_\alpha) - E_* h_0(\boldsymbol{X}_\alpha) + E^* h(\boldsymbol{X}_\alpha) - E^* h_0(\boldsymbol{X}_\alpha)$$
$$+ E_* h_0(\boldsymbol{X}_\alpha) - E_* h(\boldsymbol{X}_\alpha)\,,$$

(2) follows from (a), (b), and (iii). ☐

The remaining parts (3) - (5) of the proof are exactly the same as in the proof of theorem 1.1. ☐

Proof of corollary 1.1.net. By lemmas 1.4.net and 1.5.net, hypothesis (i) of theorem 1.1.net is satisfied. Thus we only need to check (ii) and (iii) of theorem 1.1.net.

Take **H** to be the linear space spanned by all functions of the form

$$(x, y) \to f(x) g(y)\,, \quad f \in C_b(M_1)\,, \quad g \in C_b(M_2)\,.$$

(iii): For $f \geq 0$ and $g \geq 0$

(a) $$f(\boldsymbol{X}_\alpha)_* \, g(\boldsymbol{Y}_\alpha)_* \leq (f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha))_* \leq f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha)$$
$$\leq (f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha))^* \leq f^*(\boldsymbol{X}_\alpha) g^*(\boldsymbol{Y}_\alpha)\,.$$

Thus

$$E^* f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha) - E_* f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha)$$
$$\leq E f(\boldsymbol{X}_\alpha)^* g(\boldsymbol{Y}_\alpha)^* - E f(\boldsymbol{X}_\alpha)_* g(\boldsymbol{Y}_\alpha)_*$$
$$\leq E|f(\boldsymbol{X}_\alpha)^* - f(\boldsymbol{X}_\alpha)_*||g(\boldsymbol{Y}_\alpha)^*| + E|f(\boldsymbol{X}_\alpha)_*||g(\boldsymbol{Y}_\alpha)^* - g(\boldsymbol{Y}_\alpha)_*|$$
$$\leq \|g\|_\infty E(f(\boldsymbol{X}_\alpha)^* - f(\boldsymbol{X}_\alpha)_*) + \|f\|_\infty E(g(\boldsymbol{Y}_\alpha)^* - g(\boldsymbol{Y}_\alpha)_*)$$

(b) $\to 0$ by (1.4) and (1.15).

For $f \geq 0$, $g \geq 0$ and arbitrary $a, b \in R$,

$$E^*(a+f)(\boldsymbol{X}_\alpha)(b+g)(\boldsymbol{Y}_\alpha) - E_*(a+f)(\boldsymbol{X}_\alpha)(b+g)(\boldsymbol{Y}_\alpha)$$
$$\leq a + E^*(b f(\boldsymbol{X}_\alpha)) + E^*(a\, g(\boldsymbol{Y}_\alpha)) + E^*(f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha))$$
$$- (a + E_*(b f(\boldsymbol{X}_\alpha)) + E_*(a\, g(\boldsymbol{Y}_\alpha)) + E_*(f(\boldsymbol{X}_\alpha) g(\boldsymbol{Y}_\alpha))$$

using lemma 1.2.i and 1.2.iv

(c) $\to 0$ using lemma 1.2.vii and the ≥ 0 case.

Finally, for linear combinations,

$$E^* \sum_i f_i(\boldsymbol{X}_\alpha) g_i(\boldsymbol{Y}_\alpha) - E_* \sum_i f_i(\boldsymbol{X}_\alpha) g_i(\boldsymbol{Y}_\alpha)$$
$$\leq \sum_i (E^* f_i(\boldsymbol{X}_\alpha) g_i(\boldsymbol{Y}_\alpha) - E_* f_i(\boldsymbol{X}_\alpha) g_i(\boldsymbol{Y}_\alpha)) \quad \text{by lemma 1.2(i) and (iv)}$$

$\to 0$ using (c) .

(ii): We apply lemma 5.1. It is easily seen that $\mathbf{H}$ is an algebra separating points of $M = M_1 \times M_2$, and $1 \in \mathbf{H}$. Thus (i) - (iii) of lemma 5.1 hold. To verify (iv)′, let $K \subset M$ be compact. Then $K \subset K_1 \times K_2$ where K_1 and K_2 are compacts in M_1 and M_2, and the function

$$\chi_\delta(x, y) = (1 - \delta^{-1} d_1(x, K_1))^+ (1 - \delta^{-1} d_2(y, K_2))^+ .$$

satisfies hypothesis (iv). □

Proof of lemma 1.7.net. Apply lemma 1.8.net with $\mathbb{X}_\alpha = \mathbb{X}_0$ and $\mathbb{Y}_\alpha = \mathbb{X}_\alpha$, $\alpha \in A$. □

Proof of lemma 1.8.net. Let $F \subset M$ be closed and $\varepsilon > 0$. Then

$$P_\alpha^*(\boldsymbol{Y}_\alpha \in F) \le P_\alpha^*(\boldsymbol{Y}_\alpha \in F \text{ and } d(\boldsymbol{X}_\alpha, \boldsymbol{Y}_\alpha)^* < \varepsilon) + P_\alpha(d(\boldsymbol{X}_\alpha, \boldsymbol{Y}_\alpha)^* \ge \varepsilon)$$
$$\le P_\alpha^*(\boldsymbol{X}_\alpha \in F^\varepsilon) + P_\alpha(d(\boldsymbol{X}_\alpha, \boldsymbol{Y}_\alpha)^* \ge \varepsilon) .$$

Thus $\limsup_\alpha P_\alpha^*(\boldsymbol{Y}_\alpha \in F) \le P(\boldsymbol{X}_0 \in F^\varepsilon)$ for every $\varepsilon > 0$. Finally let $\varepsilon \downarrow 0$; the conclusion then follows from (iii) implies (i) of the portmanteau lemma 1.3.net. □