Ann. Inst. Statist. Math.
Vol. 45, No. 4, 703–719 (1993)

ON THE CONSISTENCY OF CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATORS

J. PFANZAGL

Mathematical Institute, University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany

(Received March 10, 1992; revised November 2, 1992)

Abstract. Let $\{P_{\vartheta,\eta} : \vartheta \in \Theta,\ \eta \in \mathrm{H}\}$ be a family of probability measures admitting a sufficient statistic for the nuisance parameter η. The paper presents conditions for consistency of (asymptotic) conditional maximum likelihood estimators for ϑ. An application to the Rasch-model (a stochastic model for psychological tests) yields a condition on the sequence of nuisance parameters which is sufficient for strong consistency of conditional maximum likelihood estimators, and necessary for the existence of any weakly consistent estimator-sequence.

Key words and phrases: Estimation, consistency, nuisance parameters, logistic distribution.

1. Introduction

Let $(X, \mathcal{A})$ be a measurable space and $\mu \mid \mathcal{A}$ a σ-finite measure. For $\vartheta \in \Theta$ (a Hausdorff space) and $\eta \in \mathrm{H}$ let $P_{\vartheta,\eta} \mid \mathcal{A}$ be a p (= probability)-measure, equivalent to μ, with μ-density

$$p(x, \vartheta, \eta) = q(x, \vartheta) p_0(S(x), \vartheta, \eta), \qquad x \in X, \tag{1.1}$$

where S is a measurable function from X into some measurable space $(Y, \mathcal{B})$.

$q(\cdot, \vartheta)$ and $p_0(\cdot, \vartheta, \eta)$ are assumed to be measurable and positive.

Our problem is to estimate ϑ from observations $x_1, \ldots, x_n$ which are independently distributed according to $P_{\vartheta,\eta_1}, \ldots, P_{\vartheta,\eta_n}$, with the nuisance parameters $\eta_1, \ldots, \eta_n$ unknown.

If the representation (1.1) is suitably standardized (see (2.1)), $q(\cdot, \vartheta)$ may be interpreted as conditional density of x, given S. It can, therefore, be used to define a CML (= conditional maximum likelihood)-estimator for ϑ by $\sum_1^n \log q(x_\nu, \vartheta) = \max$. The purpose of this paper is to give conditions for the consistency of CML-estimators under mild conditions on the sequence of nuisance parameters η_ν, $\nu \in \mathbb{N}$. To account for the possibility that ϑ might be identifiable from the conditional distribution of x, given $S(x) = y$, only for certain values y, say for $y \in Y_0 \subset Y$, the definition of the CML-estimator is based only on observations x_ν with $S(x_\nu) \in Y_0$, i.e. on the condition $\sum_1^n 1_{Y_0}(S(x_\nu)) \log q(x_\nu, \vartheta) = \max$.

Theorem 2.1 in Section 2 gives conditions on the conditional distribution of x, given S, which guarantee that such a sequence of restricted CML-estimators is consistent under a mild condition on the sequence of nuisance parameters, namely $\sum_1^\infty P_{\vartheta_0,\eta_\nu}(S^{-1}Y_0) = \infty$. According to the Lemma of Borel-Cantelli this condition is equivalent to the requirement that the sequence of observations x_ν, $\nu \in \mathbb{N}$, contains with probability 1 infinitely many elements which are suitable for the computation of the restricted CML-estimator, i.e. elements x_ν fulfilling $S(x_\nu) \in Y_0$. If the nuisance parameters are independent realizations from a prior distribution over Y, say M, this condition amounts to the requirement that $\int P_{\vartheta,\eta}(S^{-1}Y_0)M(d\eta) > 0$.

Section 2 will be concluded by a discussion of these results in their relation to earlier results by Andersen.

After conditions for strong consistency (= convergence with probability 1) of restricted CML-estimators have been obtained in Section 2, Section 3 gives a condition on the sequence of nuisance parameters which is necessary for the existence of any weakly consistent (= stochastically convergent) estimator-sequence.

Section 4 is on the problem of obtaining consistent estimators for item difficulties in the so-called Rasch-model for psychological tests. The results of Sections 2 and 3 yield a condition on the sequence of the unknown ability parameters which (i) guarantees strong consistency of any sequence of restricted CML-estimators and which is (ii) necessary for the existence of even weakly consistent estimator-sequences.

Some more technical auxiliary results are collected in Section 5.

2. A consistency theorem

To define the CML-estimator we need the following canonical version of the representation (1.1):

$$p(x, \vartheta, \eta) = q(x, \vartheta) p_0(S(x), \vartheta, \eta) \qquad \text{for} \quad \vartheta \in \Theta, \ \ \eta \in \mathrm{H}, \tag{2.1}$$

where

(a) $p(\cdot, \vartheta, \eta)$ is a density of $P_{\vartheta,\eta} \mid \mathcal{A}$ with respect to some σ-finite measure $\mu \mid \mathcal{A}$,

(b) $p_0(\cdot, \vartheta, \eta)$ is a density of $P_{\vartheta,\eta} \circ S \mid \mathcal{B}$ with respect to some σ-finite measure $\nu \mid \mathcal{B}$,

(c) the measures $\mu \mid \mathcal{A}$ and $\nu \mid \mathcal{B}$ are connected by a measure kernel $M \mid Y \times \mathcal{A}$, i.e.

$$\mu(A \cap S^{-1}B) = \int 1_B(y) M(y, A) \nu(dy) \qquad \text{for} \quad A \in \mathcal{A}, \ \ B \in \mathcal{B}.$$

If X is a complete separable metric space and $\mathcal{A}$ its Borel algebra, then the existence of a representation (1.1) implies the existence of a canonical representation fulfilling (2.1). This can be seen as follows: W.l.g. we may assume that μ is a p-measure. Then there exists (see, e.g., Ash (1972), p. 266) a Markov-kernel $M \mid Y \times \mathcal{A}$ such that

$$\mu(A \cap S^{-1}B) = \int 1_B(y) M(y, A) \nu(dy) \qquad \text{with} \quad \nu = \mu \circ S.$$

From (1.1) we obtain (see Lemma 5.3)

$$\begin{aligned} P_{\vartheta,\eta} \circ S(B) &= \int p(x,\vartheta,\eta) 1_B(S(x))\mu(dx) \\ &= \int \left(\int q(x,\vartheta) M(y,dx) \right) p_0(y,\vartheta,\eta) 1_B(y) \nu(dy) \qquad \text{for} \quad B \in \mathcal{B}. \end{aligned}$$

Hence $\hat{p}_0(y,\vartheta,\eta) := \int q(x,\vartheta)M(y,dx)p_0(y,\vartheta,\eta)$ is a ν-density of $P_{\vartheta,\eta} \circ S$. Rewriting (1.1) as

$$p(x,\vartheta,\eta) = \frac{q(x,\vartheta)}{\int q(\xi,\vartheta)M(y,d\xi)} \cdot \hat{p}_0(S(x),\vartheta,\eta)$$

yields the canonical representation.

Starting from the canonical representation (2.1), we define a measure $Q_{\vartheta,y} \mid \mathcal{A}$ by

$$Q_{\vartheta,y}(A) := \int 1_A(x) q(x,\vartheta) M(y,dx) \qquad \text{for} \quad A \in \mathcal{A}. \tag{2.2}$$

Using (2.1) and the conditions (a), (c) and (b) we obtain by Lemma 5.3

$$\begin{aligned} P_{\vartheta,\eta}(A \cap S^{-1}B) &= \int \left(\int 1_A(x) q(x,\vartheta) M(y,dx) \right) 1_B(y) p_0(y,\vartheta,\eta)\nu(dy) \\ &= \int Q_{\vartheta,y}(A) 1_B(y) P_{\vartheta,\eta} \circ S(dy) \qquad \text{for} \quad A \in \mathcal{A}, \;\; B \in \mathcal{B}. \end{aligned} \tag{2.3}$$

Applied for $A = X$ this yields $Q_{\vartheta,y}(X) = 1$ for $P_{\vartheta,\eta} \circ S$-a.a. $y \in Y$, i.e. $Q_{\vartheta,y}$ is a p-measure for $P_{\vartheta,\eta} \circ S$-a.a. $y \in Y$. Moreover, $y \to Q_{\vartheta,y}(A)$ is $\mathcal{B}$-measurable for every $A \in \mathcal{A}$. After modifying the definition of $Q_{\vartheta,y}$ on the exceptional null set to achieve $Q_{\vartheta,y}(X) = 1$ for *all* $y \in Y$, $(y, A) \to Q_{\vartheta,y}(A)$ is a Markov-kernel on $Y \times \mathcal{A}$. Because of (2.3), this Markov-kernel is a (regular) conditional probability of x, given S, under $P_{\vartheta,\eta}$.

By (2.2), $q(\cdot,\vartheta)$ is a density of $Q_{\vartheta,y} \mid \mathcal{A}$ with respect to $M(y,\cdot) \mid \mathcal{A}$. Since $q(\cdot,\vartheta)$ is positive for every $\vartheta \in \Theta$, $q(\cdot,\vartheta)/q(\cdot,\vartheta_0)$ is a density of $Q_{\vartheta,y} \mid \mathcal{A}$ with respect to $Q_{\vartheta_0,y} \mid \mathcal{A}$. This implies

$$Q_{\vartheta_0,y}\left(\frac{q(\cdot,\vartheta)}{q(\cdot,\vartheta_0)} \right) = 1 \qquad \text{for all } \vartheta, \vartheta_0 \in \Theta \text{ and all } y \in Y, \tag{2.4}$$

hence

$$Q_{\vartheta_0,y}\left(\log \frac{q(\cdot,\vartheta)}{q(\cdot,\vartheta_0)} \right) \leq 0 \qquad \text{for} \quad \vartheta, \vartheta_0 \in \Theta \text{ and } y \in Y, \tag{2.5}$$

with strict inequality unless $Q_{\vartheta_0,y}\{q(\cdot,\vartheta) \neq q(\cdot,\vartheta_0)\} = 0$.

(Here and in the following we write $P(f)$ for $\int f dP$, if convenient.)

Even if ϑ is identifiable from $P_{\vartheta,\eta}$ in the family $\{P_{\vartheta',\eta'} : \vartheta' \in \Theta, \; \eta' \in \mathrm{H}\}$ (i.e. if $P_{\vartheta',\eta'} = P_{\vartheta,\eta}$ implies $\vartheta' = \vartheta$), it may occur that—for some $y \in S(X)$—the parameter ϑ is not identifiable any more from $Q_{\vartheta,y}$ in the family $\{Q_{\vartheta',y} : \vartheta' \in \Theta\}$ (i.e. there are $\vartheta' \neq \vartheta$ with $Q_{\vartheta',y} = Q_{\vartheta,y}$). (This occurs, in particular, in the Rasch-model considered in Section 4.)

DEFINITION 1. $y \in Y$ is *non-contracting* if every $\vartheta \in \Theta$ is identifiable from $Q_{\vartheta,y}$ in the family $\{Q_{\vartheta',y} : \vartheta' \in \Theta\}$ (i.e. if $\vartheta' \neq \vartheta$ implies $Q_{\vartheta',y} \neq Q_{\vartheta,y}$).

Throughout the following, Y_0 denotes a nonempty set of non-contracting elements in Y.

To obtain an estimator which is simple enough to lend itself to a mathematical treatment, we introduce the concept of a *restricted* CML-estimator which uses only observations x_ν for which $S(x_\nu) \in Y_0$.

DEFINITION 2. An estimator $\vartheta^{(n)}$ is restricted CML if

$$\sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu)) \log q(x_\nu, \vartheta^{(n)}(\boldsymbol{x})) = \sup_{\vartheta\in\Theta} \sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu)) \log q(x_\nu, \vartheta) \qquad \text{for all} \quad \boldsymbol{x} \in X^n. \tag{2.6}$$

For consistency of such estimator-sequences the following (weaker) technical condition suffices.

DEFINITION 3. An estimator-sequence $\vartheta^{(n)}$, $n \in \mathbb{N}$, is restricted ACML (= asymptotic CML) if

$$\lim_{n\to\infty} \frac{\inf_{\vartheta\in\Theta} \sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu)) \log[q(x_\nu, \vartheta^{(n)}(\boldsymbol{x}))/q(x_\nu, \vartheta)]}{\sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu))} = 0 \tag{2.7}$$

for all $\boldsymbol{x} \in X^{\mathbb{N}}$.

(We leave the left hand side of (2.7) undefined if $\sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu)) = 0$ for all $n \in \mathbb{N}$, since this is an event of $\times_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu}$-probability zero under condition (2.8).)

In this definition, $\vartheta^{(n)}$ is considered as a map from $X^{\mathbb{N}}$ into Θ, depending on $\boldsymbol{x} \in X^{\mathbb{N}}$ through $x_1, \ldots, x_n$ only. It is natural to think of $\vartheta^{(n)}(x_1, \ldots, x_n)$ as an estimator depending only on those elements x_ν, $\nu = 1, \ldots, n$, for which $S(x_\nu) \in Y_0$ (i.e. the elements x_ν which really enter condition (2.7)). This is, however, not required.

Remark 1. A function $\vartheta^{(n)}$ fulfilling (2.7) obviously exists. If Θ is a Hausdorff space with countable base, $\vartheta^{(n)}$ can always be defined such that it becomes measurable, provided $\vartheta \to q(x, \vartheta)$ is lower semicontinuous for every $x \in X$.

This can be seen as follows. Let $f : X \times \Theta \to \mathbb{R}$ be measurable in x and lower semicontinuous in ϑ, and let $\Theta_0 = \{\vartheta_\nu : \nu \in \mathbb{N}\}$ be a dense subset of Θ. Let

$$U := \left\{ (x, \vartheta) \in X \times \Theta : f(x, \vartheta) > \sup_{\tau\in\Theta} f(x, \tau) - \varepsilon \right\}.$$

Since the section $U_x := \{\vartheta \in \Theta : (x, \vartheta) \in U\}$ is nonempty and open for every $x \in X$, it contains elements of Θ_0. Let $\varphi(x) = \vartheta_n$ if $\vartheta_n \in U_x$ and $\vartheta_\nu \in U_x^c$ (the complement of U_x) for $\nu = 1, \dots, n-1$. We have

$$f(x, \varphi(x)) > \sup_{\tau \in \Theta} f(x, \tau) - \varepsilon,$$

and

$$\{x \in X : \varphi(x) = \vartheta_n\} = U^c_{\vartheta_1} \cap \cdots \cap U^c_{\vartheta_{n-1}} \cap U_{\vartheta_n} \qquad \text{for every} \quad n \in \mathbb{N}.$$

Since Θ is a Hausdorff space with countable base, and $\vartheta \to f(x, \vartheta)$ is lower semicontinuous for every $x \in X$, $x \to \sup_{\tau \in \Theta} f(x, \tau)$ is measurable (see, e.g., Pfanzagl and Wefelmeyer (1985), p. 451, Lemma 13.1.1). Hence $U_\vartheta \in \mathcal{A}$ for every $\vartheta \in \Theta$, and measurability of φ follows.

From these considerations, applied with $f(x, \vartheta)$ replaced by $\sum_1^n 1_{Y_\nu}(S(x_\nu)) \cdot \log q(x_\nu, \vartheta)$ and $\varepsilon = (1/n) \sum_1^n 1_{Y_0}(S(x_\nu))$, the existence of a sequence of measurable estimators $\vartheta^{(n)}$ fulfilling (2.7) follows easily.

The Consistency Theorem uses the following

Condition C. A measurable function $h : X \to \mathbb{R}$ fulfills Condition C for the family of p-measures $\{Q_y \mid \mathcal{A} : y \in Y\}$ if the following holds true.

(i) $\sup_{y \in Y} Q_y(h) < 0$,

(ii) $\sup_{y \in Y} Q_y(h_+^{1+\delta}) < \infty$ for some $\delta > 0$,

(iii) h_- is uniformly integrable, i.e. $\lim_{c \to \infty} \sup_{y \in Y} Q_y(h_- 1_{(c,\infty)} \circ h_-) = 0$.

(As usual, h_+ and h_- are the positive and the negative part of h, respectively.)

The uniformity in y, required in Condition C, is not unrealistic if the integrals $Q_y(f)$ are continuous functions of y, and if Y is compact. Conditions similar to Condition C have been used by Hoadley ((1971), p. 1981) for consistency of ML-estimators in the case of known varying nuisance parameters.

Throughout the following we write $\sup f(B) := \sup\{f(\vartheta) : \vartheta \in B\}$.

Theorem 2.1. *Let Θ be a Hausdorff space with countable base, $\vartheta \to q(x, \vartheta)$ for every $x \in X$ a lower semicontinuous function, and $Y_0 \in \mathcal{B}$ a given subset of Y containing non-contracting elements only.*

Assume that the following conditions are fulfilled for some $\vartheta_0 \in \Theta$.

(i) *For every neighborhood U of ϑ_0 there exists a cover of U^c by a finite collection of sets B_j, $j \in \{1, \dots, J\}$, such that, for every $j \in \{1, \dots, J\}$, the function* $\log[\sup q(\cdot, B_j)/q(\cdot, \vartheta_0)]$ *fulfills Condition C for $\{Q_{\vartheta_0, y} : y \in Y_0\}$.*

(ii) *The sequence $\eta_\nu \in \mathrm{H}$, $\nu \in \mathbb{N}$, fulfills*

$$\sum_{\nu=1}^{\infty} P_{\vartheta_0, \eta_\nu}(S^{-1} Y_0) = \infty. \tag{2.8}$$

Under these assumptions, any estimator-sequence fulfilling (2.7) *is strongly consistent at ϑ_0, i.e.*

$$\lim_{n \to \infty} \vartheta^{(n)}(\boldsymbol{x}) = \vartheta_0 \quad \textit{for} \quad \mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_0, \eta_\nu}\textit{-a.a. } \boldsymbol{x} \in X^{\mathbb{N}}. \tag{2.9}$$

Remark 2. Condition (i) follows if it holds true for some *compact* neighborhood of ϑ_0, and if for every $\vartheta \neq \vartheta_0$ there is an open neighborhood V_ϑ such that

$$\sup_{y \in Y_0} Q_{\vartheta_0,y}(\log[\sup q(\cdot, V_\vartheta)/q(\cdot, \vartheta_0)]) < \infty.$$

PROOF. Let U be an arbitrary neighborhood of ϑ_0, and let $\boldsymbol{x} \in X^{\mathbb{N}}$ be a fixed element. If $\vartheta^{(n)}(\boldsymbol{x}) \in U^c$ infinitely often, there exists $\ell \in \{1, \ldots, J\}$ such that $\mathbb{N}_\ell := \{n \in \mathbb{N} : \vartheta^{(n)}(\boldsymbol{x}) \in B_\ell\}$ is infinite. This implies

$$\log q(x_\nu, \vartheta^{(n)}(\boldsymbol{x})) \leq \log \sup q(x_\nu, B_\ell) \quad \text{for} \quad \nu \in \mathbb{N}, \ n \in \mathbb{N}_\ell,$$

hence also

$$\begin{aligned} &\frac{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu)) \log[q(x_\nu, \vartheta^{(n)}(\boldsymbol{x}))/q(x_\nu, \vartheta_0)]}{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu))} \\ &\qquad \leq \frac{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu)) \log[\sup q(x_\nu, B_\ell)/q(x_\nu, \vartheta_0)]}{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu))} \end{aligned}$$

for $n \in \mathbb{N}_\ell$. Since $\mathbb{N}_\ell$ is infinite, this implies

$$\begin{aligned} &\varlimsup_{n \to \infty} \frac{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu)) \log[q(x_\nu, \vartheta^{(n)}(\boldsymbol{x}))/q(x_\nu, \vartheta_0)]}{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu))} \\ &\quad \leq \varlimsup_{n \to \infty} \frac{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu)) \log[\sup q(x_\nu, B_\ell)/q(x_\nu, \vartheta_0)]}{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu))} \\ &\quad \leq \max_{j \in \{1, \ldots, J\}} \varlimsup_{n \to \infty} \frac{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu)) \log[\sup q(x_\nu, B_j)/q(x_\nu, \vartheta_0)]}{\sum_{\nu=1}^n 1_{Y_0}(S(x_\nu))}. \end{aligned}$$

By Lemma 5.2, applied with $h(x) = \log[\sup q(x, B_j)/q(x, \vartheta_0)]$ there exists a $\times_{\nu=1}^\infty P_{\vartheta_0, \eta_\nu}$-null set $N \subset X^{\mathbb{N}}$ such that the last expression is negative for $\boldsymbol{x} \in N^c$. If the estimator-sequence $\vartheta^{(n)}$, $n \in \mathbb{N}$, fulfills (2.7), we obtain $\boldsymbol{x} \in N$ if $\vartheta^{(n)}(\boldsymbol{x}) \in U^c$ for infinitely many $n \in \mathbb{N}$. The set N depends on U. Since Θ has a countable base, the assertion follows. □

In applications it is usually difficult to verify conditions on a sequence of unknown nuisance parameters. The intention of Theorem 2.1 is, therefore, to keep these conditions as weak as possible. In fact, (2.8) is the only condition left for the sequence of nuisance parameters. If it happens that no $y \in Y$ is contracting, one is free to choose $Y_0 = Y$, in which case $P_{\vartheta_0, \eta_\nu}(Y_0) = 1$ for $\nu \in \mathbb{N}$. Then condition (2.8) holds true for *any* sequence η_ν, $\nu \in \mathbb{N}$. There are, however, applications of practical relevance where contracting values of S do occur. One such case will be discussed in Section 4. Even if every $y \in Y$ is non-contracting, restricting the computation of the estimator to observations x_ν with $S(x_\nu)$ in some subset $Y_0 \subset Y$ opens the possibility of excluding x_ν with $S(x_\nu)$ in a "nasty" part of Y. By restriction to an appropriate subset Y_0 for which condition C is fulfilled, one may obtain a consistent estimator-sequence in cases where the sequence of unrestricted CML-estimators fails to be consistent.

The following corollary refers to the situation where η_ν, $\nu \in \mathbb{N}$, is not just an arbitrary sequence in $\mathrm{H}^{\mathbb{N}}$, but a sequence of i.i.d.-realizations from a p-measure M on $(\mathrm{H}, \mathcal{C})$. In the following, $P_{\vartheta_0,M} \mid \mathcal{A}$ is the p-measure defined by $P_{\vartheta_0,M}(A) = \int P_{\vartheta_0,\eta}(A)M(d\eta)$, $A \in \mathcal{A}$.

COROLLARY 2.1. *Let now* $(\mathrm{H}, \mathcal{C})$ *be a measurable space, and* $(x,\eta) \to p(x,\vartheta,\eta)$ *measurable for every* $\vartheta \in \Theta$. *Assume Condition* (i) *from Theorem* 2.1, *and*

$$(2.8') \qquad P_{\vartheta_0,M}(S^{-1}Y_0) > 0.$$

Then

$$(2.9') \qquad \lim_{n\to\infty} \vartheta^{(n)}(\boldsymbol{x}) = \vartheta_0 \quad \textit{for} \quad P^{\mathbb{N}}_{\vartheta_0,M}\textit{-a.a. } \boldsymbol{x} \in X^{\mathbb{N}}.$$

PROOF. For any M-integrable function $h : \mathrm{H} \to \mathbb{R}$ we have

$$\lim_{n\to\infty} n^{-1} \sum_{\nu=1}^{n} h(\eta_\nu) = \int h(\eta)M(d\eta)$$

for $M^{\mathbb{N}}$-a.a. $(\eta_\nu)_{\nu\in\mathbb{N}} \in \mathrm{H}^{\mathbb{N}}$. Hence $\int h(\eta)M(d\eta) > 0$ implies $\lim_{n\to\infty} \sum_{\nu=1}^{n} h(\eta_\nu) = \infty$ for $M^{\mathbb{N}}$-a.a. $(\eta_\nu)_{\nu\in\mathbb{N}} \in \mathrm{H}^{\mathbb{N}}$. Applied for $h(\eta) = P_{\vartheta_0,\eta}(S^{-1}Y_0)$ this entails that Condition (2.8′) implies Condition (2.8) for $M^{\mathbb{N}}$-a.a. $(\eta_\nu)_{\nu\in\mathbb{N}} \in \mathrm{H}^{\mathbb{N}}$. □

A number of consistency theorems for CML-estimators has been given by Andersen ((1970), p. 286, Theorem 1, (1971), p. 44, Theorem 1, (1973*a*), p. 45, Theorem 2.1, (1980), p. 80, Theorem 3.7) under varying conditions and with proofs of varying force. All these theorems are restricted to the case that the likelihood equation has a unique solution. Andersen's conditions concerning the nuisance parameter amount to something like continuity of $\eta \to P_{\vartheta_0,\eta}$ $((\log[\sup q(\cdot,B)/q(\cdot,\vartheta_0)])^2) < \infty$, and compactness of H.

3. A condition for non-existence of consistent estimator-sequences

The Consistency Theorem 2.1 is restricted to sequences of nuisance parameters η_ν, $\nu \in \mathbb{N}$, for which Condition (2.8) holds true. If this condition fails, this does not necessarily exclude the existence of some consistent estimator-sequences—even such ones fulfilling (2.7).

In Theorem 3.1 we give under (3.3) a condition which excludes the existence of any (even weakly) consistent estimator-sequences. The usefulness of this condition is obviously confined to families $\{P_{\vartheta,\eta} : \vartheta \in \Theta,\ \eta \in \mathrm{H}\}$ with finite or countable support. Even in this case it looks somewhat strange. Yet it suffices to answer the question about the existence of consistent estimator-sequences for the particular case considered in Section 4.

Let $P \ll Q$ denote absolute continuity of P with respect to Q.

THEOREM 3.1. (i) *If*

$$\mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu} \ll \mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu}, \tag{3.1}$$

then there is no estimator-sequence $\vartheta^{(n)}$, $n \in \mathbb{N}$, *which is weakly consistent at* ϑ_0 *and* ϑ_1, *i.e. which fulfills*

$$\lim_{n\to\infty} \vartheta^{(n)} = \vartheta_i \textit{ stochastically under } \mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_i,\eta_\nu} \quad \textit{for} \quad i = 0, 1. \tag{3.2}$$

(ii) *Assume there exists a sequence* $\hat{x}_\nu \in X$, $\nu \in \mathbb{N}$, *such that*

$$\sum_{\nu=1}^{\infty} (1 - P_{\vartheta_i,\eta_\nu}\{\hat{x}_\nu\}) < \infty \quad \textit{for} \quad i = 0, 1. \tag{3.3}$$

Under this condition, $P_{\vartheta_1,\eta_\nu} \ll P_{\vartheta_0,\eta_\nu}$ *for* $\nu \in \mathbb{N}$ *implies* (3.1).

PROOF. (i) Relation (3.2) for $i = 0$ implies $\lim_{n\to\infty} \vartheta^{(n)} = \vartheta_0$ stochastically under $\times_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu}$. To see this, consider an arbitrary subsequence $\mathbb{N}' \subset \mathbb{N}$. Since $\vartheta^{(n)}$, $n \in \mathbb{N}'$, converges to ϑ_0 stochastically under $\times_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu}$, there exists a subsequence $\mathbb{N}'' \subset \mathbb{N}'$ such that $\vartheta^{(n)}$, $n \in \mathbb{N}''$, converges to ϑ_0 $\times_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu}$-a.e. Because of (3.1), this implies convergence of $\vartheta^{(n)}$, $n \in \mathbb{N}''$, to ϑ_0 $\times_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu}$-a.e. Hence every subsequence of $\vartheta^{(n)}$, $n \in \mathbb{N}$, contains a subsequence converging to ϑ_0 $\times_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu}$-a.e. This implies stochastic convergence of $\vartheta^{(n)}$, $n \in \mathbb{N}$, to ϑ_0 under $\times_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu}$.

(ii) Let

$$\mathcal{X}_0 := \liminf_{\nu\to\infty} \{\boldsymbol{x} \in X^{\mathbb{N}} : x_\nu = \hat{x}_\nu\}$$

and

$$\mathcal{X}_\lambda := \{\boldsymbol{x} \in X^{\mathbb{N}} : x_\nu = \hat{x}_\nu \text{ for } \nu > \lambda\}.$$

We have $\mathcal{X}_\lambda \subset \mathcal{X}_{\lambda+1}$, and $\mathcal{X}_0 = \bigcup_{\lambda=1}^{\infty} \mathcal{X}_\lambda$. Relation (3.3) implies $(\times_{\nu=1}^{\infty} P_{\vartheta_i,\eta_\nu})(\mathcal{X}_0) = 1$, hence

$$\lim_{\lambda\to\infty} \left(\mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_i,\eta_\nu} \right) (\mathcal{X}_\lambda) = 1. \tag{3.4}$$

For any set $A \in \mathcal{A}^{\mathbb{N}}$, $A \cap \mathcal{X}_\lambda = A_\lambda \times \{(\hat{x}_\nu)_{\nu>\lambda}\}$. The set $A_\lambda \subset X^\lambda$, being a section of the measurable set $A \cap \mathcal{X}_\lambda$, is in $\mathcal{A}^\lambda$. We have

$$\left(\mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_i,\eta_\nu} \right) (A \cap \mathcal{X}_\lambda) = \left(\mathop{\times}_{\nu=1}^{\lambda} P_{\vartheta_i,\eta_\nu} \right) (A_\lambda) \cdot \left(\mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_i,\eta_\nu} \right) (\mathcal{X}_\lambda). \tag{3.5}$$

Because of (3.4), $(\times_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu})(\mathcal{X}_\lambda) > 0$ for $\lambda \geq \lambda_0$, say. Hence $(\times_{\nu=1}^{\infty} P_{\vartheta_0,\eta_\nu})(A) = 0$ implies $(\times_{\nu=1}^{\lambda} P_{\vartheta_0,\eta_\nu})(A_\lambda) = 0$ for $\lambda \geq \lambda_0$. From $P_{\vartheta_1,\eta_\nu} \ll P_{\vartheta_0,\eta_\nu}$ for $\nu \in \mathbb{N}$, we have $(\times_{\nu=1}^{\lambda} P_{\vartheta_1,\eta_\nu})(A_\lambda) = 0$ for $\lambda \geq \lambda_0$, hence, by (3.5),

$$\left(\mathop{\times}_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu} \right) (A \cap \mathcal{X}_\lambda) = 0 \quad \text{for} \quad \lambda \geq \lambda_0.$$

Because of (3.4), this implies $(\times_{\nu=1}^{\infty} P_{\vartheta_1,\eta_\nu})(A) = 0$. This proves (3.1). $\square$

4. An application to the Rasch-model

In this section we apply the results of Sections 2 and 3 to a special case, the so-called Rasch-model for psychological tests. (For the psychological background, see, e.g., Lord and Novick ((1968), Chapter 17), contributed by A. Birnbaum.)

From the formal point of view we have for $\delta = (\delta_1, \dots, \delta_m) \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$ a discrete p-measure defined for $x = (x_0, x_1, \dots, x_m) \in \{0,1\}^{1+m}$ by

$$P_{\delta,\alpha}(x) := \frac{\exp[\sum_{i=0}^m (\alpha - \delta_i)x_i]}{\prod_{i=0}^m (1 + \exp[\alpha - \delta_i])}. \tag{4.1}$$

In this definition of $P_{\delta,\alpha}$ read $\delta_0 = 0$. (For typographical reasons we write $P_{\delta,\alpha}(x)$ instead of $P_{\delta,\alpha}(\{x\})$.)

The family $\{P_{\delta,\alpha} : \alpha \in \mathbb{R}\}$ admits the sufficient statistic

$$S(x_0, x_1, \dots, x_m) := \sum_{i=0}^m x_i$$

which attains the values $0, 1, \dots, 1+m$.

For $k \in \{0, 1, \dots, 1+m\}$ let $\Delta_k(\delta) := \sum \exp[-\sum_{i=1}^m \delta_i x_i]$, where the summation extends over all $x \in \{0,1\}^{1+m}$ with $S(x) = k$.

We have

$$P_{\delta,\alpha} \circ S\{k\} = \Delta_k(\delta) \exp[\alpha k] \Big/ \prod_{i=0}^m (1 + \exp[\alpha - \delta_i]). \tag{4.2}$$

Let $D_k := S^{-1}\{k\}$. For $k \in \{0, 1, \dots, 1+m\}$ and $\delta \in \mathbb{R}^m$ we define a p-measure $Q_k(\cdot, \delta)$ over $\{0,1\}^{1+m}$ by

$$Q_k(x, \delta) := \frac{1}{\Delta_k(\delta)} \exp\left[-\sum_{i=1}^m \delta_i x_i\right] \qquad \text{if} \quad x \in D_k \tag{4.3}$$

and

$$Q_k(x, \delta) := 0 \qquad \text{elsewhere.}$$

With this notation, representation (2.1) can be written as

$$P_{\delta,\alpha}(x) = Q_{S(x)}(x, \delta) P_{\delta,\alpha} \circ S\{S(x)\}.$$

In this representation, $Q_{S(x)}(x, \delta)$ corresponds to $q(x, \delta)$, the density of the conditional distribution of x, given $S(x) = k$, with respect to the measure kernel $M(k, x) = 1_{\{k\}}(S(x))$. Moreover, μ and ν are the counting measures over $\{0,1\}^{1+m}$ and $\{0, 1, \dots, 1+m\}$, respectively.

Among the values $0, 1, \dots, 1+m$ attained by the sufficient statistic S, the values $1, \dots, m$ are non-contracting. For the values $k = 0$ and $k = 1+m$ the p-measures $Q_k(\cdot, \delta)$ are identical for all $\delta \in \mathbb{R}^m$, since D_0 and D_{1+m} consist of a single $(1+m)$-tuple only, $(0, 0, \dots, 0)$ and $(1, 1, \dots, 1)$, respectively. Hence

$\log[Q_k(x_\nu, \delta^{(n)}(\boldsymbol{x}))/Q_k(x_\nu, \delta)] = 0$ for $k = 0$ and $k = 1+m$, so that CML-estimators in the usual sense are in this case identical with restricted CML-estimators fulfilling (2.6) with $Y_0 = \{1, \dots, m\}$.

An estimator $\delta^{(n)}$ is restricted CML if (see (2.6))

$$\sum_{\nu=1}^{n} 1_{\{1,\dots,m\}}(S(x_\nu)) \log Q_{S(x_\nu)}(x_\nu, \delta^{(n)}(\boldsymbol{x})) = \sup_{\delta \in \mathbb{R}^m} \sum_{\nu=1}^{n} 1_{\{1,\dots,m\}}(S(x_\nu)) \log Q_{S(x_\nu)}(x_\nu, \delta) \quad \text{for all } \boldsymbol{x} \in (\{0,1\}^{1+m})^{\mathbb{N}}. \tag{4.4}$$

Sufficient for consistency is the following weaker restricted ACML-condition, corresponding to (2.7),

$$\varliminf_{n\to\infty} \frac{\inf_{\delta \in \mathbb{R}^m} \sum_{\nu=1}^{n} 1_{\{1,\dots,m\}}(S(x_\nu)) \log[Q_{S(x_\nu)}(x_\nu, \delta^{(n)}(\boldsymbol{x}))/Q_{S(x_\nu)}(x_\nu, \delta)]}{\sum_{\nu=1}^{n} 1_{\{1,\dots,m\}}(S(x_\nu))} = 0 \quad \text{for all } \boldsymbol{x} \in (\{0,1\}^{1+m})^{\mathbb{N}}. \tag{4.5}$$

THEOREM 4.1. (i) *For the family of p-measures given by* (4.1), *all estimator-sequences fulfilling* (4.5) *are strongly consistent at each* $\delta \in \mathbb{R}^m$ *if*

$$\sum_{\nu=1}^{\infty} \exp[-|\alpha_\nu|] = \infty. \tag{4.6}$$

(ii) *Condition* (4.6) *is necessary for the existence of weakly consistent estimator-sequences.*

Condition (4.6) is certainly fulfilled if $\sup_{\nu\in\mathbb{N}} |\alpha_\nu| < \infty$, but it establishes consistency of restricted ACML-estimators also for certain sequences α_ν, $\nu \in \mathbb{N}$, with $\overline{\lim}_{\nu\to\infty} |\alpha_\nu| = \infty$. It fails to guide our intuition if we think of α_ν, $\nu \in \mathbb{N}$, as a sequence with $|\alpha_\nu|$, $\nu \in \mathbb{N}$, increasing: The distinction between $\alpha_\nu = \log \nu$ (which implies strong consistency of restricted ACML-estimator-sequences) and $\alpha_\nu = 2\log\nu$ (which excludes the existence of any weakly consistent estimator-sequence) is not very clear from the intuitive point of view. A reassuring aspect of Condition (4.6) is that it guarantees consistency as long as there are enough small values $|\alpha_\nu|$, not withstanding the occurrence of many large values $|\alpha_\nu|$.

For the particular model (4.1) Andersen ((1973*a*), p. 159, Theorem 5.1) asserts consistency under the condition that—in our notations—the nuisance parameters α_ν are restricted to a compact set, in Andersen ((1973*b*), p. 127, Theorem 2 and (1980), p. 247, Theorem 6.2) under conditions equivalent to

$$\sum_{\nu=1}^{\infty} 1_{\{1,\dots,m\}}(S(x_\nu)) = \infty \tag{4.7}$$

(together with the redundant condition $\delta_i \neq \delta_j \neq 0$ for $i \neq j \neq 0$).

Andersen's assertion is presumably to be understood in the sense that the sequence of CML-estimators converges to δ for $\times_{\nu=1}^{\infty} P_{\delta,\alpha_\nu}$-a.a. $\boldsymbol{x} \in (\{0,1\}^{1+m})^{\mathbb{N}}$ for which (4.7) holds true. Since this condition refers to an *infinite* sequence of *observations*, it appears to be meaningless. We may, however, interpret Andersen's result as a consistency theorem if conditions on the sequence α_ν, $\nu \in \mathbb{N}$, make sure that (4.7) holds for $\times_{\nu=1}^{\infty} P_{\delta,\alpha_\nu}$-a.a. $\boldsymbol{x} \in (\{0,1\}^{1+m})^{\mathbb{N}}$. This is the case iff condition (4.6) holds true.

The consistency results of Andersen are, again, based on the assumption that the CML-estimator is unique. In Andersen ((1980), p. 248, Theorem 6.3) Andersen states that the CML-estimator is unique if $0 < \sum_{\nu=1}^{n} x_{\nu i} < n$ for $i = 0, 1, \ldots, m$. As remarked by Fischer ((1981), p. 60) this is not true. (Andersen's use of his Theorem 3.2 is not justified.) If (3.4) holds true, the (necessary and sufficient) condition for existence and uniqueness of the CML-estimator discovered by Fischer ((1981), pp. 66–69, Theorems 3 and 4) is fulfilled for eventually all $n \in \mathbb{N}$ for $\times_{\nu=1}^{\infty} P_{\delta,\alpha_\nu}$-a.a. $\boldsymbol{x} \in (\{0,1\}^{1+m})^{\mathbb{N}}$. Hence scholars who are willing to accept Andersen's consistency-proof for solutions of the likelihood equation might arrive at the conclusion that the consistency result stated in Theorem 4.1(i) is not entirely new. The first mathematically satisfying proof was obtained in the diploma-thesis of Hillgruber ((1990), p. 35, Satz 2.3.36). Our proof of Theorem 4.1 uses ideas from this thesis.

PROOF OF THEOREM 4.1. (i) Let $\mathbb{N}^+ := \{j \in \mathbb{N} : \alpha_j > 0\}$ and $\mathbb{N}^- := \{j \in \mathbb{N} : \alpha_j \leq 0\}$. We need the following relations.

$$(4.8') \qquad \begin{aligned} &\sum_{j\in\mathbb{N}^+} P_{\delta,\alpha_j}(S^{-1}\{m\}) < \infty \quad \text{implies,} \\ &\sum_{j\in\mathbb{N}^+} \exp[-|\alpha_j|] < \infty \quad \text{implies,} \\ &\sum_{j\in\mathbb{N}^+} P_{\delta,\alpha_j}(S^{-1}\{k\}) < \infty \quad \text{for all } k \in \{0,\ldots,m\}. \end{aligned}$$

$$(4.8'') \qquad \begin{aligned} &\sum_{j\in\mathbb{N}^-} P_{\delta,\alpha_j}(S^{-1}\{1\}) < \infty \quad \text{implies,} \\ &\sum_{j\in\mathbb{N}^-} \exp[-|\alpha_j|] < \infty \quad \text{implies,} \\ &\sum_{j\in\mathbb{N}^-} P_{\delta,\alpha_j}(S^{-1}\{k\}) < \infty \quad \text{for all } k \in \{1,\ldots,m+1\}. \end{aligned}$$

We shall prove (4.8′). The proof for (4.8″) runs similarly.

Since $\alpha \to \exp[(m+1)\alpha]/\prod_{i=0}^{m}(1+\exp[\alpha-\delta_i])$ is continuous and positive on $\mathbb{R}$, and

$$0 < \lim_{\alpha\to\infty} \exp[(m+1)\alpha] \Big/ \prod_{i=0}^{m}(1+\exp[\alpha-\delta_i]) < \infty,$$

there exist $0 < c_1(\delta) < c_2(\delta) < \infty$ such that

$$c_1(\delta) \le \exp[(m+1)\alpha] \Big/ \prod_{i=0}^{m}(1+\exp[\alpha-\delta_i]) \le c_2(\delta) \qquad \text{for} \quad \alpha \in [0,\infty),$$

i.e.

$$c_1(\delta)\exp[-|\alpha|] \le \exp[m\alpha] \Big/ \prod_{i=0}^{m}(1+\exp[\alpha-\delta_i]) \le c_2(\delta)\exp[-|\alpha|] \qquad \text{for} \quad \alpha \in [0,\infty).$$

Hence (4.8′) follows from (4.2).

(ii) That consistent estimator-sequences do not exist if (4.6) fails, follows from Theorem 3.1. To see this, let $\hat{x}_j \in \{0,1\}^{1+m}$ be defined by

$$\hat{x}_j = \begin{cases} (0,\dots,0) & \text{if } \alpha_j \le 0 \\ (1,\dots,1) & \text{if } \alpha_j > 0. \end{cases}$$

Since

$$1 - P_{\delta,\alpha_j}(\hat{x}_j) = \begin{cases} P_{\delta,\alpha_j}(S^{-1}\{1,\dots,m+1\}) & \text{if } \alpha_j \le 0 \\ P_{\delta,\alpha_j}(S^{-1}\{0,\dots,m\}) & \text{if } \alpha_j > 0, \end{cases}$$

$\sum_{j=1}^{\infty}\exp[-|\alpha_j|] < \infty$ implies (3.3), i.e. $\sum_{j=1}^{\infty}(1-P_{\delta,\alpha_j}(\hat{x}_j)) < \infty$, by means of (4.8).

The sufficiency-part (i) of Theorem 4.1 follows from Theorem 2.1, applied with $Y_0 = \{1,\dots,m\}$.

(iii) To begin with we remark that relations (4.8) imply the equivalence between (2.8) and (4.6). This establishes condition (ii) of Theorem 2.1.

Now we establish condition (i) of Theorem 2.1. Let $k \in \{1,\dots,m\}$ be fixed. At first we shall show that for every neighborhood U of $\delta_0 \in \mathbb{R}^m$ there is a cover of U^c by a finite collection of sets B_j, $j = 1,\dots,J$, fulfilling

$$\sum_{x \in D_k} Q_k(x,\delta_0)\log[\sup Q_k(x,B_j)/Q_k(x,\delta_0)] < 0. \tag{4.9}$$

(iv) Following the suggestion of Remark 2 we first show that this holds true for the compact set

$$U_0(r) := \{(\delta_1,\dots,\delta_m) \in \mathbb{R}^m : |\delta_i| \le r \text{ for } i = 1,\dots,m\}, \tag{4.10}$$

provided r is sufficiently large.

For $j \in \{1,\dots,m\}$ let

$$B_j^+(r) := \{(\delta_1,\dots,\delta_m) \in \mathbb{R}^m : \delta_j > r\}$$

and

$$B_j^-(r) := \{(\delta_1,\dots,\delta_m) \in \mathbb{R}^m : \delta_j < -r\}.$$

We have

$$U_0(r)^c \subset \bigcup_{j=1}^{m} (B_j^+(r) \cup B_j^-(r)).$$

Now we shall show that (4.9) holds for $B_j = B_j^+(r)$ and $B_j = B_j^-(r)$, provided r is sufficiently large.

The following relation holds for arbitrary $x, y \in D_k$ and arbitrary $\delta \in \mathbb{R}^m$

$$(4.11) \qquad Q_k(x,\delta) = \frac{1}{\Delta_k(\delta)} \exp\left[-\sum_{j=1}^{m} \delta_j x_j\right] \leq \exp\left[-\sum_{j=1}^{m} \delta_j (x_j - y_j)\right].$$

For any $k \in \{1,\dots,m\}$ there is ${}_jx^+ \in D_k$ with ${}_jx_0^+ = 0$, and ${}_jx_j^+ = 1$. Let $y \in \mathbb{R}^{1+m}$ be defined by $y_0 = 1$, $y_j = 0$, and $y_i = {}_jx_i^+$ elsewhere. Since $y \in D_k$ we obtain from (4.11) that

$$Q_k({}_jx^+, \delta) \leq \exp[-\delta_j],$$

hence

$$\log \sup Q_k({}_jx^+, B_j^+(r)) \leq -r.$$

(The proof for $B_j^-(r)$ uses ${}_jx^- \in D_k$ with ${}_jx_0^- = 1$ and ${}_jx_j^- = 0$.)

Since $\log\sup Q_k(x, B_j^+(r)) \leq 0$ for every $x \in D_k$, we have for every $r > 0$

$$\begin{aligned}\sum_{x \in D_k} & Q_k(x,\delta_0) \log \sup Q_k(x, B_j^+(r)) \\ & \leq Q_k({}_jx^+, \delta_0) \log\sup Q_k({}_jx^+, B_j^+(r)) \leq -rQ_k({}_jx^+, \delta_0).\end{aligned}$$

Since $Q_k(x,\delta_0) > 0$ for every $x \in D_k$, we have

$$-rQ_k({}_jx^+, \delta_0) < \sum_{x \in D_k} Q_k(x,\delta_0) \log Q_k(x,\delta_0)$$

if r is sufficiently large. Hence (4.9) follows with $B_j^+(r)$ in place of B_j.

(v) $\sup Q_k(x, \mathbb{R}^m) \leq 1$ for every $x \in \{0,1\}^{1+m}$, implies

$$\begin{aligned}\sum_{x \in D_k} & Q_k(x,\delta_0) \log[\sup Q_k(x, \mathbb{R}^m)/Q_k(x,\delta_0)] \\ & \leq -\sum_{x \in D_k} Q_k(x,\delta_0) \log Q_k(x,\delta_0) < \infty\end{aligned}$$

(since $Q_k(x,\delta_0) > 0$ for $x \in D_k$).

Since $k \in \{1,\dots,m\}$ is non-contracting, relation (2.5) implies

$$\sum_{x \in D_k} Q_k(x,\delta_0) \log[Q_k(x,\delta)/Q_k(x,\delta_0)] < 0$$

for every $\delta \neq \delta_0$. Since $\delta \to Q_k(x,\delta)$ is continuous, there exists (see Pfanzagl (1990), p. 90, Lemma L.7) an open neighborhood V_δ of δ such that

$$\sum_{x \in D_k} Q_k(x,\delta_0) \log[\sup Q_k(x,V_\delta)/Q_k(x,\delta_0)] < 0.$$

Let U_0 denote the compact neighborhood of δ_0 determined in (iv). If U is an arbitrary neighborhood of δ_0, $\{V_\delta : \delta \in U_0 \cap U^c\}$ is an open cover of the compact set $U_0 \cap U^c$ and contains, therefore, a finite subcover. Together with the sets $B_j^+(r)$, $B_j^-(r)$, $j = 1, \ldots, m$ determined in (iv) this provides a finite cover of U^c fulfilling condition (4.9).

(vi) Since $Q_k(x,\delta_0) = 0$ for $x \notin D_k$, relation (4.9) may be rewritten as

$$\sum_{x \in \{0,1\}^{1+m}} Q_k(x,\delta_0) \log[\sup q(x,B_j)/q(x,\delta_0)] < 0, \tag{4.12}$$

with $q(x,\delta) := Q_{S(x)}(x,\delta)$.

To stress that the cover of U^c by B_j, $j = 1, \ldots, J$, still depends on k we now write $B_{k,j}$, $j = 1, \ldots, J_k$. Condition C requires a finite cover of U^c which fulfills (i), (ii) and (iii) simultaneously for all $k \in \{1, \ldots, m\}$. Such one is given by

$$\left\{ \bigcap_{k=1}^{m} B_{k,j_k} : j_k \in \{1, \ldots, J_k\} \text{ for } k = 1, \ldots, m \right\}.$$

Condition C(i) now follows from (4.12). Conditions C(ii) and C(iii) follow from the fact that $0 < \sup q(x,B_j) \le 1$ and $0 < q(x,\delta_0) < 1$ for every $x \in \{0,1\}^{1+m}$, so that

$$x \to \log[\sup q(x,B_j)/q(x,\delta_0)]$$

is bounded on $\{0,1\}^{1+m}$. □

5. Auxiliary results

In this section we collect technical lemmas which are used in Section 2.

LEMMA 5.1. *If h fulfills Condition C for $\{Q_y : y \in Y\}$, then the following holds true for every sequence $y_\nu \in Y$, $\nu \in \mathbb{N}$:*

$$\overline{\lim_{n\to\infty}}\, n^{-1} \sum_{\nu=1}^{n} h(x_\nu) < 0 \quad \textit{for} \quad \times_{\nu=1}^{\infty} Q_{y_\nu}\text{-}a.a.\ \boldsymbol{x} \in X^{\mathbb{N}}.$$

PROOF. By Condition C(i), $Q_y(h) \le -2s$ for some $s > 0$. By Condition C(iii) there is $c > 0$ such that $Q_y(h_- 1_{(c,\infty)} \circ h_-) < s$ for $y \in Y$. With $h_c := h_- 1_{(0,c]} \circ h_-$ we obtain $Q_y(h_+) - Q_y(h_c) < -s$ for $y \in Y$. Therefore,

$$\begin{aligned} n^{-1} \sum_{\nu=1}^{n} h(x_\nu) &\le n^{-1} \sum_{\nu=1}^{n} (h_+(x_\nu) - h_c(x_\nu)) \\ &< n^{-1} \sum_{\nu=1}^{n} (h_+(x_\nu) - Q_{y_\nu}(h_+)) - n^{-1} \sum_{\nu=1}^{n} (h_c(x_\nu) - Q_{y_\nu}(h_c)) - s. \end{aligned}$$

From a suitable version of the strong law of large numbers we obtain for $\times_{\nu=1}^{\infty} Q_{y_\nu}$-a.a. $\boldsymbol{x} \in X^{\mathbb{N}}$

$$\lim_{n\to\infty} n^{-1} \sum_{\nu=1}^{n} (h_+(x_\nu) - Q_{y_\nu}(h_+)) = 0$$

and

$$\lim_{n\to\infty} n^{-1} \sum_{\nu=1}^{n} (h_c(x_\nu) - Q_{y_\nu}(h_c)) = 0.$$

(For the first equation use, e.g., Loève ((1977), Section 17.4, p. 253) with $b_n = n$ and $r_n = 1 + \delta$.) Hence the assertion follows. □

LEMMA 5.2. *For $\nu \in \mathbb{N}$ let $P_\nu \mid \mathcal{A}$ be a p-measure. Assume there exists a measurable function $S : X \to Y$ and a conditional distribution of x, given S, i.e. a Markov-kernel $Q_y \mid \mathcal{A}$, $y \in Y$, such that*

$$P_\nu(A \cap S^{-1}B) = \int Q_y(A) 1_B(y) P_\nu \circ S(dy) \quad \textit{for} \quad A \in \mathcal{A},\ B \in \mathcal{B}. \tag{5.1}$$

Assume that the function $h : X \to \mathbb{R}$ fulfills Condition C for $\{Q_y : y \in Y_0\}$ for some set $Y_0 \in \mathcal{B}$.

If $\sum_{\nu=1}^{\infty} P_\nu(S^{-1}Y_0) = \infty$, then

$$\overline{\lim_{n\to\infty}} \frac{\sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu)) h(x_\nu)}{\sum_{\nu=1}^{n} 1_{Y_0}(S(x_\nu))} < 0 \quad \textit{for} \quad \mathop{\times}_{\nu=1}^{\infty} P_\nu\textit{-a.a. } \boldsymbol{x} \in X^{\mathbb{N}}. \tag{5.2}$$

PROOF. Let $\mathcal{Y}_0$ denote the set of all sequences $y_\nu \in Y$, $\nu \in \mathbb{N}$, such that $\{\nu \in \mathbb{N} : y_\nu \in Y_0\}$ is infinite. Let $\boldsymbol{y} \in \mathcal{Y}_0$ be fixed. From Lemma 5.1, applied with Y_0 in place of Y, and with $\mathbb{N}_0 := \{\nu \in \mathbb{N} : y_\nu \in Y_0\}$ in place of $\mathbb{N}$, we obtain that there exists a $\times_{\nu\in\mathbb{N}_0} Q_{y_\nu}$-null set $N_{\boldsymbol{y}} \subset X^{\mathbb{N}_0}$ such that

$$\overline{\lim_{n\to\infty}} \frac{\sum_{\nu=1}^{n} 1_{Y_0}(y_\nu) h(x_\nu)}{\sum_{\nu=1}^{n} 1_{Y_0}(y_\nu)} < 0 \quad \text{for} \quad (x_\nu)_{\nu\in\mathbb{N}_0} \in N_{\boldsymbol{y}}^c. \tag{5.3}$$

Let $\hat{N}_{\boldsymbol{y}}$ denote the cylinder in $X^{\mathbb{N}}$ with base $N_{\boldsymbol{y}}$. We have $(\times_{\nu=1}^{\infty} Q_{y_\nu})(\hat{N}_{\boldsymbol{y}}) = 0$, and relation (5.3) (with the left hand side viewed as a function of $\boldsymbol{x} \in X^{\mathbb{N}}$) holds for all $\boldsymbol{x} \notin \hat{N}_{\boldsymbol{y}}$.

With

$$C(\boldsymbol{y}) := \left\{ \boldsymbol{x} \in X^{\mathbb{N}} : \overline{\lim_{n\to\infty}} \frac{\sum_{\nu=1}^{n} 1_{Y_0}(y_\nu) h(x_\nu)}{\sum_{\nu=1}^{n} 1_{Y_0}(y_\nu)} < 0 \right\},$$

relation (5.3) may be rewritten as

$$\left(\mathop{\times}_{\nu=1}^{\infty} Q_{y_\nu} \right)(C(\boldsymbol{y})) = 1 \quad \text{for every } \boldsymbol{y} \in \mathcal{Y}_0. \tag{5.4}$$

We remark that the left hand side of (5.4), considered as a function of $\boldsymbol{y}$, is $\mathcal{B}^{\mathbb{N}}$-measurable by Lemma 5.3 (applied with $1_{C(\boldsymbol{y})}(\boldsymbol{x})$ in place of $f(x,y)$) and that $\mathcal{Y}_0 \in \mathcal{B}^{\mathbb{N}}$ (since $\boldsymbol{y} \to 1_{Y_0}(y_\nu)$ is $\mathcal{B}^{\mathbb{N}}$-measurable and $\mathcal{Y}_0 = \{\boldsymbol{y} \in Y^{\mathbb{N}} : \sum_{\nu=1}^{\infty} 1_{Y_0}(y_\nu) = \infty\}$).

If $\sum_{\nu=1}^{\infty} P_\nu(S^{-1}Y_0) = \infty$, we obtain from the Lemma of Borel-Cantelli that

$$\left(\mathop{\times}_{\nu=1}^{\infty} P_\nu\right)\{\boldsymbol{x} \in X^{\mathbb{N}} : (S(x_\nu))_{\nu\in\mathbb{N}} \in \mathcal{Y}_0\} = 1.$$

Together with (5.4) this implies that

$$\left(\mathop{\times}_{\nu=1}^{\infty} Q_{S(x_\nu)}\right)(C((S(x_\nu))_{\nu\in\mathbb{N}})) = 1 \quad \text{for} \quad \mathop{\times}_{\nu=1}^{\infty} P_\nu\text{-a.a. } \boldsymbol{x} \in X^{\mathbb{N}}. \tag{5.5}$$

Using Lemma 5.3 we obtain that

$$\begin{aligned}\int 1_{C((S(x_\nu))_{\nu\in\mathbb{N}})}(\boldsymbol{x}) & \left(\mathop{\times}_{\nu=1}^{\infty} P_\nu\right)(d\boldsymbol{x}) \\ &= \int \left(\mathop{\times}_{\nu=1}^{\infty} Q_{S(x_\nu)}\right)(C((S(x_\nu))_{\nu\in\mathbb{N}}))\left(\mathop{\times}_{\nu=1}^{\infty} P_\nu\right)(d\boldsymbol{x}) = 1,\end{aligned}$$

i.e. $\boldsymbol{x} \in C((S(x_\nu))_{\nu\in\mathbb{N}})$ for $\times_{\nu=1}^{\infty} P_\nu$-a.a. $\boldsymbol{x} \in X^{\mathbb{N}}$. This is the same as (5.2). □

LEMMA 5.3. *Let μ and ν be σ-finite measures on measurable spaces $(X,\mathcal{A})$ and $(Y,\mathcal{B})$, respectively. Let $S : X \to Y$ be a measurable function and $M \mid Y \times \mathcal{A}$ a measure kernel such that*

$$\mu(A \cap S^{-1}B) = \int 1_B(y)M(y,A)\nu(dy) \qquad \textit{for} \quad A \in \mathcal{A}, \;\; B \in \mathcal{B}.$$

Let $f : X \times Y \to \mathbb{R}$ be an $\mathcal{A}\times\mathcal{B}$-measurable function such that $x \to f\big(x, S(x)\big)$ is μ-integrable. Then $\int f(x,y)M(y,dx)$ exists for ν-a.a. $y \in Y$,

$$y \to \int f(x,y)M(y,dx) \ \textit{is } \mathcal{B}\textit{-measurable} \tag{5.6}$$

and

$$\int f(x,S(x))\mu(dx) = \int \left(\int f(x,y)M(y,dx)\right)\nu(dy). \tag{5.7}$$

PROOF. Since f is approximable by $\mathcal{A}\times\mathcal{B}$-measurable elementary functions, it suffices to prove the assertion for $f = 1_C$, with $C \in \mathcal{A}\times\mathcal{B}$. Let $\mathcal{S}$ denote the class of all sets $C \in \mathcal{A}\times\mathcal{B}$ such that (5.6) and (5.7) hold true for $f = 1_C$. The class $\mathcal{S}$ contains $A \times B$ for $A \in \mathcal{A}$ and $B \in \mathcal{B}$, because $\int 1_{A\times B}(x,y)M(y,dx) = 1_B(y)M(y,A)$. Since $\mathcal{S}$ is a Dynkin-system, this implies $\mathcal{S} = \mathcal{A}\times\mathcal{B}$ (see Ash (1972), p. 169, Theorem 4.1.2). □

Acknowledgments

The author wishes to thank Mr. R. Hamböker. He checked the proofs and contributed to an improved version by various critical remarks. The referees suggested various improvements, including the more general framework given under (2.1).

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators, *J. Roy. Statist. Soc. Ser. B*, **32**, 283–301.

Andersen, E. B. (1971). A strictly conditional approach in estimation theory, *Skand. Aktuarietidskr.*, **54**, 39–49.

Andersen, E. B. (1973*a*). *Conditional Inference and Models for Measuring*, Publ. Nr. 5, Mentalhygiejnisk Forskningsinstitut, Mentalhygiejnisk Forlag.

Andersen, E. B. (1973*b*). A goodness of fit test for the Rasch model, *Psychometrika*, **28**, 123–140.

Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*, North-Holland, Amsterdam.

Ash, R. B. (1972). *Real Analysis and Probability*, Academic Press, New York.

Fischer, G. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model, *Psychometrika*, **46**, 59–77.

Hillgruber, G. (1990). Schätzung von Parametern in psychologischen Testmodellen, Diploma-thesis, University of Cologne.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case, *Ann. Math. Statist.*, **42**, 1977–1991.

Loève, M. (1977). *Probability Theory I*, 4th ed., Springer, New York.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, Massachusetts.

Pfanzagl, J. (1990). Estimation in semiparametric models; some recent developments, *Lecture Notes in Statist.*, **63**, Springer, New York.

Pfanzagl, J. and Wefelmeyer, W. (1985). Asymptotic expansions for general statistical models, *Lecture Notes in Statist.*, **31**, Springer, Berlin.