SOME ASYMPTOTIC PROPERTIES OF LIKELIHOOD RATIOS ON GENERAL SAMPLE SPACES

R. R. BAHADUR[1]
UNIVERSITY OF CHICAGO
and
M. RAGHAVACHARI[2]
CARNEGIE-MELLON UNIVERSITY

1. Introduction

It is shown in [2] (see also [3], [4], [14]) under certain conditions that in the case of independent and identically distributed observations likelihood ratios are asymptotically optimal test statistics in the sense of exact slopes. The present paper points out that many of the arguments and conclusions of the papers just cited extend to general sampling frameworks, and also develops certain refinements of these conclusions. The present generalizations and refinements seem worthwhile for the following reasons. They enable us to construct asymptotically optimal tests in problems such as testing independence in Markov chains and in exchangeable sequences. Secondly, they provide a useful method of finding the exact slope of a statistic which is equivalent, on some sample space, to the likelihood ratio on that sample space. It suffices in this case to evaluate the limit, in the nonnull case, of the normalized log likelihood ratio; it is not necessary to obtain estimates of the relevant large deviation probabilities in the null case; indeed, the latter estimates are implicit in the initial evaluation. Finally, the present elaborations throw some light on what sort of conditioning is advantageous in making conditional tests. It seems that a conditioning statistic is helpful if it produces an exact conditional null distribution for the contemplated test statistic *and* if in the testing problem on hand the conditioning statistic is useless by itself.

The following Sections 2 to 4 describe the general theory. Some examples illustrative of the theory are given in Sections 5 to 7.

[1] Research supported in part by National Science Foundation Research Grant NSF FP 16071 and the Office of Naval Research Grant N00014-67-A-028-009.
[2] Research supported by National Science Foundation Research Grant NSF GP 22595.

2. Notation and preliminary lemmas

Let S be a space of points s, and let $\mathscr{A}$ be a σ-field of sets of S. Let P and Q be probability measures on $\mathscr{A}$, and consider testing the null hypothesis that P obtains against the alternative that Q obtains.

In typical cases (see Sections 5 to 7 below) all $\mathscr{A}$ measurable procedures are not available to the statistician. Suppose then that we are given a σ-field $\mathscr{B} \subset \mathscr{A}$. and that we are restricted to $\mathscr{B}$ measurable procedures. Suppose for simplicity that P and Q are mutually absolutely continuous on $\mathscr{B}$, and let $r(s)$ be a $\mathscr{B}$ measurable function on S such that $0 < r(s) < \infty$ and

$$dQ = r(s)\, dP \text{ on } \mathscr{B}. \tag{2.1}$$

This r is, of course, the likelihood ratio statistic for testing P against Q when the sample space is $(S, \mathscr{B})$.

Throughout the paper, if $\pi(s)$ is a statement which is either true or false for any given s in S and M is a measure on $\mathscr{A}$, "$\pi(s)$ a.e. $[M]$" will mean that there exists an $\mathscr{A}$ measurable set N of M measure zero such that $\{s: \pi(s) \text{ is false}\} \subset N$.

Let there be given a σ-field $\mathscr{C} \subset \mathscr{B}$, and let $\rho(s)$ be a $\mathscr{C}$ measurable function, $0 < \rho < \infty$, such that

$$dQ = \rho(s)\, dP \text{ on } \mathscr{C}, \tag{2.2}$$

so that ρ is the likelihood ratio statistic when the sample space is $(S, \mathscr{C})$. If $E_P(r(s)|\mathscr{C})$ is a version of the conditional expectation function of r given $\mathscr{C}$ when P obtains, then $\rho(s) = E_P(r(s)|\mathscr{C})$ a.e. P.

In applications, the field $\mathscr{C}$ just introduced is the σ-field induced by a (not necessarily real valued) statistic, say $y = U(s)$, and $\mathscr{C}$ plays two distinct roles. We may be studying the loss of information, if any, when the available sample space is that of the statistic, that is, $(S, \mathscr{C})$, rather than that of s, that is, $(S, \mathscr{B})$. Or we may wish to make conditional tests given y based on s, that is, $\mathscr{C}$ may be the conditioning field. In the remainder of this section $\mathscr{C}$ plays this second role of a conditioning field.

Let $T(s)$ be a real valued $\mathscr{B}$ measurable function, to be thought of as a test statistic, large values of T being significant. The conditional level attained by T given $\mathscr{C}$, say L, is defined as follows. Let $F(t, s)$ be a function defined for $-\infty < t < \infty$ and s in S such that $F(\cdot, s)$ is a left continuous probability distribution function for each s, and $F(t, \cdot)$ is $\mathscr{C}$ measurable for each t, and $F(t, s)$ is a version of the conditional probability function $P(T(s) < t|\mathscr{C})$ for each t. Then

$$L(s) = 1 - F(T(s), s). \tag{2.3}$$

For a given s, L as just defined is the conditional probability given $\mathscr{C}$ of T being as large or larger than the observed value $T(s)$ if the hypothesis P is true. We shall usually refer to L as the level attained by $T|\mathscr{C}$. It is readily seen that if L and L^0 are two versions of the level attained by $T|\mathscr{C}$ then $L(s) = L^0(s)$ a.e. P.

LEMMA 1. *The level L is a $\mathscr{B}$ measurable function of s, $0 \leqq L \leqq 1$.*

LEMMA 2. *For $a > 0$, $P(L(s) < a|\mathscr{C}) \leqq a$ a.e. P.*

The proofs of Lemmas 1 and 2 are virtually the same as the proofs of Propositions 3 and 4, respectively, of [4], and so are omitted.

We are interested mainly in the behavior of L in the nonnull case, that is, under Q. The following lemma gives a crude bound for the distribution of $L(r/\rho)$; this bound is of interest only because it is valid for all $\mathscr{B}$ measurable statistics T.

Let

$$\lambda(s) = \min\{1, \rho(s)/r(s)\} \tag{2.4}$$

and

$$D^2 = E_P\{[r(s) - \rho(s)]^2/\rho(s)\}, \tag{2.5}$$

and assume for the moment that $D^2 < \infty$. Neither this condition nor Lemma 3 are used in subsequent sections.

LEMMA 3. *The inequality $Q(L(s) < a\lambda(s)) \leqq 3(a[1 + D^2])^{1/3}$ holds for $a > 0$.*

PROOF. Let ε and δ be positive constants. Then by (2.4)

$$\begin{aligned} Q(L < a\lambda) &\leqq Q(L < a\rho/r) \\ &\leqq Q(r < \varepsilon\rho) + Q(r > \delta\rho) + Q(L < a\rho/r, \varepsilon\rho \leqq r \leqq \delta\rho). \end{aligned} \tag{2.6}$$

Now, according to (2.1) and (2.2),

$$\begin{aligned} Q(r < \varepsilon\rho) &= \int \{r < \varepsilon\rho\}\, r\, dP \\ &\leqq \varepsilon \int \{r < \varepsilon\rho\}\, \rho\, dP \leqq \varepsilon. \end{aligned} \tag{2.7}$$

Next,

$$\begin{aligned} Q(r > \delta\rho) &\leqq \delta^{-1}E_Q(r/\rho) = \delta^{-1}E_P(r^2/\rho) \\ &= \delta^{-1}[1 + D^2] \end{aligned} \tag{2.8}$$

by Markov's inequality and (2.5). Finally, using the definitions (2.1), (2.2) and Lemma 2,

$$\begin{aligned} Q(L < a\rho/r, \varepsilon\rho \leqq r \leqq \delta\rho) &\leqq Q(L < a/\varepsilon, r \leqq \delta\rho) \\ &= \int \{L < a/\varepsilon, r \leqq \delta\rho\}\, r\, dP \leqq \delta \int \{L < a/\varepsilon\}\, \rho\, dP \\ &= \delta \int_S \rho P(L < a/\varepsilon|\mathscr{C})\, dP \leqq (\delta a/\varepsilon) \int_S \rho\, dP = \delta a/\varepsilon. \end{aligned} \tag{2.9}$$

It follows from (2.6) to (2.9) that $Q(L < a\lambda) \leqq \varepsilon + (\delta a)\varepsilon^{-1} + [1 + D^2]\delta^{-1}$ for all ε and $\delta > 0$. Choosing $\varepsilon = (\delta a)^{1/2}$ and then minimizing the bound with respect to δ completes the proof.

Now regard $r(s)$ as a test statistic and let $\hat{L}(s)$ be the level attained by $r|\mathscr{C}$.

LEMMA 4. *The inequality* $\hat{L}(s) \leqq \lambda(s)$ *holds a.e.* Q.

PROOF. Let $\hat{F}(t, s)$ be a version of the left continuous distribution function of r given $\mathscr{C}$ when P obtains. Consider a fixed $t > 0$, and let C be a $\mathscr{C}$ measurable set. Then from (2.1) and (2.2),

$$\int_C [1 - \hat{F}(t, s)]\, dP = \int_C P(r \geqq t|\mathscr{C})\, dP = \int \{C \cap \{r \geqq t\}\}\, dP \tag{2.10}$$

$$\leqq \int_C (r/t)\, dP = t^{-1}Q(C) = \int_C (\rho/t)\, dP.$$

Since $1 - \hat{F}$ and ρ/t are $\mathscr{C}$ measurable functions and since C is arbitrary it follows from (2.10) that $1 - \hat{F}(t, s) \leqq \rho(s)/t$ a.e. P. Since t is arbitrary, it now follows by a familiar argument that

$$1 - \hat{F}(t, s) \leqq \rho(s)/t \quad \text{for all} \quad t > 0 \text{ a.e. } P. \tag{2.11}$$

It follows from (2.11) that $\hat{L}(s) \equiv 1 - \hat{F}(r(s), s) \leqq \rho(s)/r(s)$ a.e. P. Hence the lemma, since $\hat{L} \leqq 1$ in any case, and $Q \ll P$ on $\mathscr{B}$.

Lemmas 3 and 4 suggest that if Q obtains, the conditional level attained by an optimal statistic is likely to be of the order $\rho(s)/r(s)$. If the sample space $(S, \mathscr{C})$ is itself highly informative for discriminating between P and Q, ρ is likely to be much larger than 1 when Q obtains; then $\rho(s)/r(s)$ is much larger than $1/r(s)$. By temporarily putting $\mathscr{C}$ = the trivial field in Lemmas 3 and 4, it is seen that the unconditional level attained by an optimal statistic on $(S, \mathscr{B})$ is likely to be of the order $1/r(s)$. These heuristic considerations suggest that conditioning is harmful, unless the sample space $(S, \mathscr{C})$ is useless for discriminating between P and Q. This suggestion is discussed more precisely in asymptotic terms in the following sections.

It may be worthwhile to look at some of the details of the special case when $\mathscr{C}$ is induced by a statistic and there exist regular conditional probabilities on $\mathscr{B}$ given $\mathscr{C}$. Suppose then that $y = U(s)$ is a measurable transformation of $(S, \mathscr{B})$ into a space $(Y, \mathscr{D})$. Assume that P admits a regular conditional probability measure on $\mathscr{B}$ given $U(s) = y$, that is, there exists a function $P_y(B)$ on $\mathscr{B} \times Y$ such that $P_y(\cdot)$ is a probability measure on $\mathscr{B}$ for each y and such that, for each B in $\mathscr{B}$, $P_y(B)$ is a version of $P(B|U(s) = y)$. Let $\mathscr{C} = U^{-1}(\mathscr{D})$. Then, for any statistic T, the level attained by $T|\mathscr{C}$ may be defined as $\{P_y(T(s) \geqq t)\}_{t=T(s), y=U(s)}$. In particular, the level attained by $r|\mathscr{C}$ is

$$\hat{L}(s) = \{P_y(r(s) \geqq t)\}_{t=r(s), y=U(s)}. \tag{2.12}$$

Now let

$$r_y(s) = r(s)\Big/\int_S r(s)\, dP_y, \qquad 0 \leqq r_y(s) \leqq \infty, \tag{2.13}$$

and for each y in Y, let the measure $Q_y(\cdot)$ be defined by $Q_y(B) = \int_B r_y(s)\, dP_y$ for B in $\mathscr{B}$. It is readily seen that Q_y is a regular conditional probability on $\mathscr{B}$ given $U(s) = y$ when Q obtains. Consequently, $r_y(s)$ is the conditional likelihood ratio given $U(s) = y$. It is also readily seen that, as an alternative to formula (2.12), we have

$$\hat{L}(s) = \{\{P_y(r_y(s) \geqq t)\}_{t = r_y(s)}\}_{y = U(s)}. \tag{2.14}$$

In subsequent sections, phrases such as "$\mathscr{C}$ is induced by U" or "$\mathscr{C}$ is induced by the mapping: $s \to U(s)$" mean that $\mathscr{C} = U^{-1}(\mathscr{D})$ where $\mathscr{D}$ is determined by the context as follows. If Y is a finite set, $\mathscr{D}$ is the field of all sets of Y; if Y is (or may be taken to be) k dimensional Euclidean space, $\mathscr{D}$ is the class of Borel sets of Y.

3. Limit theorems in the simplest case

In this section and the following ones we consider a space S of points s, a σ-field $\mathscr{A}$ of sets of S, and two sequences $\{\mathscr{B}_n : n = 1, 2, \cdots\}$ and $\{\mathscr{C}_n : n = 1, 2, \cdots\}$ of σ-fields such that

$$\mathscr{C}_n \subset \mathscr{B}_n \subset \mathscr{A}, \qquad n = 1, 2, \cdots. \tag{3.1}$$

Let P and Q be probability measures on $\mathscr{A}$. We assume that P and Q are mutually absolutely continuous on $\mathscr{B}_n$, and let

$$dQ = r_n(s)\, dP \text{ on } \mathscr{B}_n, \qquad dQ = \rho_n(s)\, dP \text{ on } \mathscr{C}_n, \tag{3.2}$$

where r_n is $\mathscr{B}_n$ measurable, ρ_n is $\mathscr{C}_n$ measurable, and $0 < r_n, \rho_n < \infty$; $n = 1, 2, \cdots$.

Let

$$K_n^{(1)}(s) = n^{-1} \log r_n(s), \qquad K_n^{(2)}(s) = n^{-1} \log \rho_n(s), \tag{3.3}$$

and let

$$\Delta_n(s) = K_n^{(1)}(s) - K_n^{(2)}(s). \tag{3.4}$$

Let

$$\underline{\Delta}(s) = \liminf_{n \to \infty} \Delta_n(s), \qquad \overline{\Delta}(s) = \limsup_{n \to \infty} \Delta_n(s). \tag{3.5}$$

THEOREM 1. *The function* $\underline{\Delta}$ *satisfies* $0 \leqq \underline{\Delta}(s) \leqq \infty$ *a.e.* Q.

PROOF. Choose $\varepsilon > 0$. Since the event $\Delta_n < -\varepsilon$ is, by (3.3) and (3.4), identical with the event $r_n < \rho_n \exp\{-n\varepsilon\}$, we have by (3.2)

$$\begin{aligned} Q(\Delta_n < -\varepsilon) &= \int_{\Delta_n < -\varepsilon} r_n\, dP \\ &\leqq \exp\{-n\varepsilon\} \int_{\Delta_n < -\varepsilon} \rho_n\, dP \\ &\leqq \exp\{-n\varepsilon\} \int_S \rho_n\, dP = \exp\{-n\varepsilon\}. \end{aligned} \tag{3.6}$$

Hence $\Sigma_n Q(\Delta_n < -\varepsilon) < \infty$. Hence $\underline{\Delta}(s) \geqq -\varepsilon$, a.e. Q. Since ε is arbitrary, $\underline{\Delta}(s) \geqq 0$ a.e. Q. *Q.E.D.*

Let

$$\underline{K}^{(i)}(s) = \liminf_{n\to\infty} K_n^{(i)}(s), \qquad \bar{K}^{(i)}(s) = \limsup_{n\to\infty} K_n^{(i)}(s), \qquad i = 1, 2. \tag{3.7}$$

COROLLARY 1. *The inequalities* $0 \leqq \underline{K}^{(2)}(s) \leqq \underline{K}^{(1)}(s) \leqq \infty$ *and* $0 \leqq \bar{K}^{(2)}(s) \leqq \bar{K}^{(1)}(s) \leqq \infty$ *hold a.e. Q.*

PROOF. In view of Theorem 1, we need only show that $0 \leqq \underline{K}^{(2)}(s)$ a.e. Q. To this end, for each n let $\mathscr{C}_n =$ the trivial field in Theorem 1. Then $\rho_n(s) \equiv 1$, $K_n^{(2)}(s) \equiv 0$; hence $\underline{K}^{(1)}(s) \geqq 0$ a.e. Q, by (3.3) to (3.7) and Theorem 1. Since $\{\mathscr{B}_n\}$ is an arbitrary sequence, this last conclusion continues to hold when $\{\mathscr{B}_n\}$ is replaced by the initially given $\{\mathscr{C}_n\}$, so $\underline{K}^{(2)}(s) \geqq 0$ a.e. Q. *Q.E.D.*

It is pointed out at the outset of Section 7 below that $\underline{K}^{(i)}$, $\bar{K}^{(i)}$, $i = 1, 2$, are, roughly speaking, generalizations of the Kullback-Leibler information numbers. and that Corollary 1 is a generalization of a well-known inequality concerning these numbers.

For each n let T_n be a real valued $\mathscr{B}_n$ measurable function, and let L_n be the level attained by $T_n|\mathscr{C}_n$. The following theorem is essentially an extension and refinement of Theorem 1 of [2] and of the main theorems of [4], [14].

THEOREM 2. *The inequalities* $\liminf_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\bar{\Delta}(s)$ and $\limsup_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\underline{\Delta}(s)$ *hold a.e. Q.*

PROOF. Choose and fix $\varepsilon > 0$. Then, for any constant $b \geqq 0$ and any n,

$$\begin{aligned} Q(L_n < \exp(-n[\Delta_n + 3\varepsilon]), b - \varepsilon < \Delta_n < b + \varepsilon) \\ \leqq Q(L_n < \exp(-n[b + 2\varepsilon]), \Delta_n < b + \varepsilon) \end{aligned} \tag{3.8}$$

$$= \int \{L_n < \exp(-n[b + 2\varepsilon]), \Delta_n < b + \varepsilon\}\, r_n\, dP \tag{3.9}$$

$$\begin{aligned} &= \int \{L_n < \exp(-n[b + 2\varepsilon]), \Delta_n < b + \varepsilon\} \exp(n\Delta_n)\rho_n\, dP \\ &\leqq \exp(n[b + \varepsilon]) \int \{L_n < \exp(-n[b + 2\varepsilon]), \Delta_n < b + \varepsilon\}\rho_n dP \\ &\leqq \exp(n[b + \varepsilon]) \int \{L_n < \exp(-n[b + 2\varepsilon])\} \end{aligned} \tag{3.10}$$

$$\leqq \exp(-n\varepsilon). \tag{3.11}$$

Here (3.9) follows from (3.2), (3.10) from (3.3) and (3.4), and (3.11) from Lemma 2, as in (2.9).

Now let $b_1, b_2, \cdots$ be an enumeration of the rational points of $[0, \infty)$, and let

$$A_n(i) = \{s: L_n < \exp\{-n[\Delta_n + 3\varepsilon]\}, b_i - \varepsilon < \Delta_n < b_i + \varepsilon\}, \tag{3.12}$$

(3.13) $$B(i) = \limsup_{n\to\infty} A_n(i)$$

and

(3.14) $$C = \bigcup_{i=1}^{\infty} B(i).$$

It follows from (3.8) to (3.11) with $b = b_i$ that $\Sigma_n Q(A_n(i)) < \infty$; hence $Q(B(i)) = 0$; hence $Q(C) = 0$. Let $D_\varepsilon = S - C - \{s: \underline{\Delta}(s) < 0\}$. Then $Q(D_\varepsilon) = 1$. We shall show that, for every s in D_ε,

(3.15) $$\liminf_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\bar{\Delta}(s) - 3\varepsilon$$

and

(3.16) $$\limsup_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\underline{\Delta}(s) - 3\varepsilon.$$

This will establish the theorem, since ε is arbitrary. Choose and fix an s in D_ε.

To establish (3.15) for the given s, we may suppose $\bar{\Delta} < \infty$ for otherwise (3.15) holds trivially. Let ν denote the left side of (3.15) and let $j_1 < j_2 < \cdots$ be a sequence of positive integers such that $j_m^{-1} \log L_{j_m} \to \nu$ as $m \to \infty$. Since $0 \leqq \underline{\Delta} \leqq \bar{\Delta} < \infty$, there exists a subsequence $k_1 < k_2 < \cdots$ of $\{j_m\}$ such that $\Delta_{k_m} \to$ a limit Δ, say, where $0 \leqq \Delta < \infty$. There exists a rational $b_i \geqq 0$ such that $b_i - \varepsilon < \Delta < b_i + \varepsilon$; hence $b_i - \varepsilon < \Delta_{k_m} < b_i + \varepsilon$ for all sufficiently large m. It now follows from (3.12) to (3.14) and the choice of s that $L_{k_m} \geqq \exp\{-k_m(\Delta_{k_m} + 3\varepsilon)\}$ for all sufficiently large m. Since $\{k_m\}$ is a subsequence of $\{j_m\}$ it now follows from the choice of $\{j_m\}$ and $\{k_m\}$ that

(3.17) $$\nu \geqq \lim_{m\to\infty} (-\Delta_{k_m} - 3\varepsilon) = -\Delta - 3\varepsilon \geqq -\bar{\Delta} - 3\varepsilon.$$

so (3.15) holds.

To establish (3.16) we may suppose that $0 \leqq \underline{\Delta} < \infty$. There exists a sequence $j_1 < j_2 < \cdots$ such that $\Delta_{j_m} \to \underline{\Delta}$ as $m \to \infty$. Hence there exists a rational $b_i \geqq 0$ such that $b_i - \varepsilon < \Delta_{j_m} < b_i + \varepsilon$ for all sufficiently large m. Hence, by (3.12) to (3.14) and the choice of s, $L_{j_m} > \exp\{-j_m[\Delta_{j_m} + 3\varepsilon]\}$ for all sufficiently large m. Since the left side of (3.16) is not less than $\limsup_{m\to\infty} \{j_m^{-1} \log L_m\}$, it now follows from the present choice of $\{j_m\}$ that (3.16) holds. *Q.E.D.*

Now let $\hat{L}_n(s)$ be the level attained by $r_n|\mathscr{C}_n$. The following theorem is a partial generalization and refinement of Theorem 2 of [2].

THEOREM 3. *The equalities* $\liminf_{n\to\infty} \{n^{-1} \log \hat{L}_n(s)\} = -\bar{\Delta}(s)$ *and* $\limsup_{n\to\infty} \{n^{-1} \log \hat{L}_n(s)\} = -\underline{\Delta}(s)$ *hold a.e. Q.*

PROOF. It follows from (3.3) and (3.4) by Lemma 4 that

(3.18) $$n^{-1} \log \hat{L}_n(s) \leqq -\Delta_n(s) \quad \text{for all } n \text{ a.e. } Q.$$

It follows from (3.5) and (3.18) that $\liminf_{n\to\infty} \{n^{-1} \log \hat{L}_n\} \leqq -\bar{\Delta}$ and $\limsup_{n\to\infty} \{n^{-1} \log \hat{L}_n\} \leqq -\underline{\Delta}$ a.e. Q. Theorem 3 now follows from Theorem 2 applied to $\{r_n | \mathscr{C}_n\}$. *Q.E.D.*

We shall say that the sequence $\{T_n | \mathscr{C}_n\}$ has exact slope $c(s)$ when Q obtains (see [3]) if $n^{-1} \log L_n(s) \to -\frac{1}{2}c(s)$ as $n \to \infty$ a.e. Q. Let $\Delta(s)$ be an $\mathscr{A}$ measurable function, $0 \leqq \Delta \leqq \infty$. As an immediate consequence of Theorems 2 and 3 we have,

COROLLARY 2. *The sequence* $\{r_n|\mathscr{C}_n\}$ *has exact slope* $2\Delta(s)$ *when* Q *obtains if and only if* $\lim_{n\to\infty} \Delta_n(s) = \Delta(s)$ *a.e.* Q. *In the latter case, if* $c(s)$ *is the exact slope of any sequence* $\{T_n|\mathscr{C}_n\}$ *then* $c(s) \leqq 2\Delta(s)$ *a.e.* Q.

Corollary 2 is perhaps the main conclusion of this paper. The elaborations given in Theorems 2 and 3 are, however, useful on occasion.

Suppose that $\lim_{n\to\infty} \Delta_n(s) = \Delta(s)$ *and* $\Delta(s) < \infty$ a.e. Q. Then, by Corollary 2,

$$\hat{L}_n(s) = \exp\{-n\Delta_n(s) + o(n)\} \quad \text{as} \quad n \to \infty \text{ a.e. } Q. \tag{3.19}$$

The estimate $\exp\{-n\Delta_n\}$ is $\mathscr{B}_n$ measurable, that is, based on the same data as $\hat{L}_n$, but is often much easier to compute. The formulation (3.19) makes sense in the general case, but (3.19) is not valid unless the conditions stated are satisfied. To consider an example, suppose that $s = (x_1, x_2, \cdots)$ where the x_n are independent real valued random variables; under P each x_n is $N(0, 1)$; under Q, x_n is $N(\mu_n, 1)$ where $\mu_n = \exp\{n^2\}$. For each n let $\mathscr{B}_n$ be the field induced by the mapping: $s \to x_n$, and let $\mathscr{C}_n$ be the trivial field. Then $n^{-1} \log \hat{L}_n(s) + \Delta_n(s) \to -\infty$ a.e. Q. The verification is omitted.

Suppose for the moment that $K_n^{(i)}(s) \to K^{(i)}(s)$ as $n \to \infty$ a.e. Q where $0 \leqq K^{(i)}(s) < \infty$, $i = 1, 2$. It then follows from Corollary 2 that the unconditional exact slope of $\{r_n\}$ is $2K^{(1)}(s)$, that the exact slope of $\{r_n | \mathscr{C}_n\}$ is $2[K^{(1)}(s) - K^{(2)}(s)]$, and that these are the maximum available unconditional and conditional slopes, respectively. Since $K^{(2)}(s) \geqq 0$, it is plain that there is certainly no advantage in conditioning and that there is no disadvantage if and only if $K^{(2)}(s) = 0$ a.e. Q. Now, $2K^{(2)}(s)$ is the maximum available unconditional exact slope when $\{\mathscr{B}_n\}$ is replaced by $\{\mathscr{C}_n\}$, that is, $2K^{(2)}$ is the exact slope of $\{\rho_n\}$. It is thus seen that conditioning is asymptotically harmless if and only if the conditioning σ-field or statistic is asymptotically useless for testing P against Q. However, as is pointed out in the following section, if we are testing a *composite* null hypothesis there may exist an asymptotically harmless conditioning which has the following feature: the conditional distribution of a contemplated test statistic does not depend on which null P obtains. This last feature is very convenient, for practical as well as theoretical purposes.

Suppose that, for given $\{\mathscr{B}_n\}$ and $\{\mathscr{C}_n\}$, the assumptions of the preceding paragraph are satisfied and that $0 < K^{(1)}(s)$ a.e. Q. Since the exact slope of the optimal $\mathscr{B}_n$ measurable sequence $\{r_n\}$ is $2K^{(1)}$, and that of the optimal $\mathscr{C}_n$ measurable sequence $\{\rho_n\}$ is $2K^{(2)}$, and since the ratio of slopes is a measure of asymptotic efficiency (see [3]), it is seen that the asymptotic efficiency of $(S, \mathscr{C}_n)$ relative to

$(S, \mathscr{B}_n)$ in testing P against Q is $K^{(2)}(s)/K^{(1)}(s)$. As may be seen from examples in Sections 5 to 7 the $K^{(i)}(s)$ are usually, but not always, independent of s.

In concluding this section we state a partial analogue of Corollary 2 in terms of the size of the optimal test of P which has a given power against Q. (See [3], Section 5, for a description of the main relation between exact slopes and power.) Let β be given, $0 < \beta < 1$, and for each n let $\hat{\alpha}_n = \hat{\alpha}_n(\beta)$ be the size of the (possibly randomized) test of P based on r_n which has power $1 - \beta$ against Q.

THEOREM 4. *Suppose that $K_n^{(1)}(s) \to K^{(1)}$ as $n \to \infty$ a.e. Q, where $K^{(1)}$ is a constant, $0 < K^{(1)} < \infty$. Then, for each β, $n^{-1} \log \hat{\alpha}_n(\beta) \to -K^{(1)}$ as $n \to \infty$.*

This theorem is a generalization of a lemma of Stein. The proof is essentially the same as the proof of Stein's lemma on pp. 316–317 of [3] and so is omitted. It should be noted that, under the hypothesis of the theorem, the exact slope of $\{r_n\}$ against Q is $2K^{(1)}$, by Corollary 2. It should also be noted that (the dual of) Theorem 4 can be used to obtain the Hodges-Lehmann index [11] of likelihood ratio tests on an arbitrary sequence of sample spaces. The theorem is valid provided that $K_n^{(1)}(s) \to K^{(1)}$ in Q probability.

4. The general case

In this section we consider the framework $S = \{s\}$, $\mathscr{A}$, $\{\mathscr{B}_n\}$, and $\{\mathscr{C}_n\}$ of the preceding section, and suppose that we are given two disjoint sets $\mathscr{P}_0$ and $\mathscr{P}_1$ of probability measures on $\mathscr{A}$. The null hypothesis is that some P in $\mathscr{P}_0$ obtains; the alternative is that some Q in $\mathscr{P}_1$ obtains.

We assume that, for given $P \in \mathscr{P}_0$, and $Q \in P_1$, P and Q are mutually absolutely continuous on $\mathscr{B}_n$, and we write $r_n(s)$, $\rho_n(s)$, and $K_n^{(i)}(s)$ of the preceding section as $r_n(s: Q, P)$, $\rho_n(s: Q, P)$, and $K_n^{(i)}(s: Q, P)$ to indicate their dependence on P and Q, $i = 1, 2$; $n = 1, 2, \cdots$.

As a matter of economy, and without much loss of generality, we assume throughout this section that the following condition is satisfied: Given $P \in \mathscr{P}_0$ and $Q \in \mathscr{P}_1$, there exists a *constant* $K^{(1)}(Q, P)$, $0 \leqq K^{(1)} \leqq \infty$, such that

$$\lim_{n\to\infty} K_n^{(1)}(s: Q, P) = K^{(1)}(Q, P) \text{ a.e. } Q. \tag{4.1}$$

Let

$$J^{(1)}(Q) = \inf\{K^{(1)}(Q, P): P \in \mathscr{P}_0\}, \qquad 0 \leqq J^{(1)} \leqq \infty. \tag{4.2}$$

Now let T_n be a real valued $\mathscr{B}_n$ measurable statistic, and let $L_n(s;P)$ be the level attained by $T_n|\mathscr{C}_n$ in testing a given P. Let

$$L_n^*(s) = \sup\{L_n(s;P): P \in \mathscr{P}_0\}. \tag{4.3}$$

Then L_n^* is, by definition, the level attained by $T_n|\mathscr{C}_n$ in testing that some P in $\mathscr{P}_0$ obtains. As noted on p. 29 of [4], it is not necessary to assume that L_n^* is $\mathscr{A}$ measurable or even that there exists an $\mathscr{A}$ measurable version thereof.

COROLLARY 3. *For each Q in $\mathscr{P}_1$,*

$$\liminf_{n\to\infty} \{n^{-1} \log L_n^*(s)\} \geqq -J^{(1)}(Q) \text{ a.e. } Q. \tag{4.4}$$

This corollary is a straightforward consequence of Theorem 2, Corollary 1, (4.1), (4.2), and (4.3). In view of Corollary 3, we shall say that, $\{T_n|\mathscr{C}_n\}$ is an asymptotically optimal sequence (for testing $\mathscr{P}_0$ against $\mathscr{P}_1$) if, for each Q in $\mathscr{P}_1$, it has exact slope $2J^{(1)}(Q)$ against Q; more precisely, if there exists a version of $\{L_n^*\}$ such that $n^{-1} \log L_n^*(s) \to -J^{(1)}(Q)$ as $n \to \infty$ a.e. Q, for each Q.

Let us say that $T_n|\mathscr{C}_n$ has an exact null distribution if there exists $F_n(t, s)$ such that F_n is a left continuous distribution function for each s and such that, for each P in $\mathscr{P}_0$ and for each t, $F_n(t, s)$ is a version of $P(T_n(s) < t|\mathscr{C}_n)$. In this case $1 - F_n(T_n(s), s)$ is a version of $L_n^*(s)$, and the maximization in (4.3) is avoided; this maximization is often inconvenient or impractical. The following Corollary 4 shows in part that a conditioning which produces an exact null distribution for T_n might also have the theoretical advantage of reducing the testing problem to the case considered in Section 3 and thereby producing an optimal testing sequence $\{T_n|\mathscr{C}_n\}$.

COROLLARY 4. *Suppose that $T_n|\mathscr{C}_n$ has an exact null distribution, $n = 1, 2, \cdots$. Suppose also that for each $Q \in \mathscr{P}_1$ there exists a P_Q in $\mathscr{P}_0$ such that* (a) $J^{(1)}(Q) = K^{(1)}(Q, P_Q)$, (b) $\lim_{n\to\infty} K_n^{(2)}(s: Q, P_Q) = 0$, a.e. Q, *and* (c) *for each n, $r_n(s: Q, P_Q)$ is a strictly increasing function of $T_n(s)$. Then $\{T_n|\mathscr{C}_n\}$ is an asymptotically optimal sequence.*

PROOF. For each n, let F_n be a function such that the conditions stated in the paragraph preceding Corollary 4 are satisfied, and let $L_n^*(s) = 1 - F_n(T_n(s), s)$. Choose and fix $Q \in \mathscr{P}_1$. By assumption there exists $P_Q \in \mathscr{P}_0$ such that the stated conditions (a) to (c) are satisfied. It follows from the present definition of L_n^* that, for each n, $L_n^*(s)$ is a version of $L_n(s; P_Q)$. It follows from condition (c) that, with $r_n = r_n(s: Q, P_Q)$, and with $\hat{L}_n(s; P_Q)$ and $L_n(s; P_Q)$ the levels attained by $r_n|\mathscr{C}_n$ and $T_n|\mathscr{C}_n$, respectively, in testing P_Q, we have $L_n(s; P_Q) = \hat{L}_n(s; P_Q)$, a.e. P_Q. Hence $L_n^*(s) = \hat{L}_n(s: P_Q)$ a.e. Q for all n. It now follows from Corollary 2 with $P = P_Q$, (4.1), and condition (b) that $n^{-1} \log L_n^*(s) \to -K^{(1)}(Q, P_Q)$ a.e. Q. It follows hence by condition (a) that $\{T_n|\mathscr{C}_n\}$ has exact slope $2J^{(1)}(Q)$ against Q. *Q.E.D.*

Although Corollary 4 is phrased in terms of asymptotic optimality, it can sometimes be used to find the exact slope of a given sequence $\{T_n|\mathscr{C}_n\}$ against a given Q by *defining* $\{\mathscr{B}_n\}$ suitably. To consider a special case, suppose we are given a measurable space $(S, \mathscr{A})$, a null set $\mathscr{P}_0$ of probability measures on $\mathscr{A}$, and a single nonnull probability measure Q on $\mathscr{A}$. For each n let T_n be an $\mathscr{A}$ measurable function on S into the real line such that $P(T_n < t) = F_n(t)$ for all P in $\mathscr{P}_0$ and all real t. For each n let $\mathscr{B}_n$ be the σ-field induced by T_n. Suppose that, with the present definition of $\mathscr{B}_n$, (4.1) holds for each P in $\mathscr{P}_0$, with some $K^{(1)}$. Let $J^{(1)}$ be defined by (4.2). Suppose there exists P_Q in $\mathscr{P}_0$ such that conditions (a) and (c) of Corollary 4 are satisfied. Since condition (b) is automatically satisfied when $\mathscr{C}_n$ is the trivial field for each n, we conclude that $n^{-1} \log [1 - F_n(T_n(s))] \to$

$-J^{(1)}(Q)$ as $n \to \infty$ a.e. Q, that is, with the present definition of $J^{(1)}$, $\{T_n\}$ has exact slope $2J^{(1)}$ against Q.

The mechanism described in Corollary 4 is rather special and is likely to be available only in rather special cases. Corollary 5 below describes a related but different method of constructing optimal sequences. For simplicity, we consider only unconditional test procedures.

CONDITION 1. *For each Q in $\mathscr{P}_1$,*

$$\liminf_{n\to\infty} T_n(s) \geqq J^{(1)}(Q) \text{ a.e. } Q. \tag{4.5}$$

CONDITION 2. *Given ε and τ, $0 < \varepsilon < 1$ and $0 < \tau < 1$, for each n there exists a positive constant $k_n(\varepsilon, \tau)$ such that*

$$P(T_n(s) \geqq t) \leqq \exp\{-n\tau t\}\,(1 + \varepsilon)^n\, k_n(\varepsilon, \tau) \tag{4.6}$$

for all $t > 0$ and all P in $\mathscr{P}_0$, and such that

$$n^{-1} \log k_n(\varepsilon, \tau) \to 0 \quad \text{as} \quad n \to \infty. \tag{4.7}$$

COROLLARY 5. *Suppose that $\{T_n\}$ satisfies Conditions 1 and 2. Then* (i) *$\{T_n\}$ is an asymptotically optimal sequence,* (ii) *for each Q in $\mathscr{P}_1$,*

$$\lim_{n\to\infty} T_n(s) = J^{(1)}(Q) \text{ a.e. } Q. \tag{4.8}$$

and (iii) *with $G_n(t) = \inf\{P(T_n(s) < t) : P \in \mathscr{P}_0\}$,*

$$\lim_{n\to\infty} n^{-1} \log [1 - G_n(t)] = -t \tag{4.9}$$

for each $t \in \operatorname{int}\{J^{(1)}(Q) : Q \in \mathscr{P}_1\}$.

PROOF. Choose ε and τ, $0 < \varepsilon < 1$ and $0 < \tau < 1$. By replacing k_n with $\max\{k_n, 1\}$ we may suppose that (4.6) holds for all real t, all P in $\mathscr{P}_0$, and all n, and that (4.7) is still satisfied. Hence, by the definition of G_n,

$$L_n^*(s) = 1 - G_n(T_n(s)) \leqq (1 + \varepsilon)^n k_n(\varepsilon, \tau) \exp\{-n\tau T_n(s)\} \tag{4.10}$$

for all n and s. It follows from (4.5), (4.7), and (4.10) that, for any Q in $\mathscr{P}_1$,

$$\limsup_{n\to\infty} n^{-1} \log L_n^*(s) \leqq \tau J^{(1)}(Q) + \log(1 + \varepsilon) \text{ a.e. } Q, \tag{4.11}$$

and that

$$\liminf_{n\to\infty} n^{-1} \log L_n^*(s) \leqq \log(1 + \varepsilon) - \tau \limsup_{n\to\infty} T_n(s). \tag{4.12}$$

Since ε, τ and Q are arbitrary, it follows from (4.11), (4.12) and Corollary 3 that parts (i) and (ii) of Corollary 5 are valid.

Part (iii) follows from parts (i) and (ii) as a special case of the following proposition. Suppose that, for each Q in $\mathscr{P}_1$, $\{T_n\}$ has exact slope $c(Q)$ against Q, and that $T_n(s) \to b(Q)$ as $n \to \infty$ a.e. Q, where b and c are (possibly infinite) constants. Let G_n be defined as in part (iii) of Corollary 5. Then, for any finite t such that

$$f(t) = \sup \{\tfrac{1}{2}c(Q): Q \in \mathscr{P}_1, b(Q) < t\}$$
$$g(t) = \inf \{\tfrac{1}{2}c(Q): Q \in \mathscr{P}_1, b(Q) > t\} \tag{4.13}$$

are well defined, we have

$$-g(t) \leqq \liminf_{n \to \infty} \{n^{-1} \log [1 - G_n(t)]\} \leqq \limsup_{n \to \infty} \{n^{-1} \log [1 - G_n(t)]\} \leqq -f(t). \tag{4.14}$$

This proposition is a straightforward consequence of the monotonicity of G_n for each n; we omit the details.

It seems that Conditions 1 and 2 are satisfied in a great variety of examples by the likelihood ratio statistic

$$\hat{T}_n(s) = \inf_{P \in \mathscr{P}_0} \sup_{Q \in \mathscr{P}_1} \{K_n^{(1)}(s: Q, P)\} \tag{4.15}$$

of Neyman and Pearson. The underlying reason is the following. *Conditions* 1 *and* 2 *are always satisfied by* $\{\hat{T}_n\}$ *if* $\mathscr{P}_0 \cup \mathscr{P}_1$ *is a finite set.* In many cases, although $\mathscr{P}_0 \cup \mathscr{P}_1$ is infinite, compactification devices which reduce the problem to the finite case are applicable, so that Conditions 1 and 2 do hold. The proof in [2] of Theorem 2 consists in verifying that, in the context of [2], Conditions 1 and 2 are satisfied by $\{\hat{T}_n\}$ provided a suitable compactification of $\mathscr{P}_0 \cup \mathscr{P}_1$ exists and certain other conditions are satisfied. In fact, Corollary 5 is a statement in general terms of the essential elements of the proof just cited. We think this generalized statement, however trivial or even tautological it may seem, is useful; see Sections 5 and 6.

We conclude this section with a sort of converse to Corollary 4.

COROLLARY 6. *Suppose that there exists an asymptotically optimal sequence* $\{T_n|\mathscr{C}_n\}$. *Then* $Q \in \mathscr{P}_1, P \in \mathscr{P}_0, K^{(1)}(Q, P) = J^{(1)}(Q) < \infty$ *imply that* $K_n^{(2)}(s: Q, P) \to 0$ *as* $n \to \infty$, *a.e.* Q.

PROOF. Let $\{T_n|\mathscr{C}_n\}$ be optimal, and for each n let $L_n^*(s)$ be a version of the level attained by $T_n|\mathscr{C}_n$ such that $n^{-1} \log L_n^*(s) \to -J^{(1)}(Q)$ a.e. Q, for each $Q \in \mathscr{P}_1$. Choose and fix Q and P such that the conditions stated are satisfied. For each n let $L_n(s)$ be the level attained by $T_n|\mathscr{C}_n$ in testing the simple hypothesis P. Then $L_n^*(s) \geqq L_n(s)$ a.e. P; hence $L_n^*(s) \geqq L_n(s)$ a.e. Q, for all n. Consequently,

$$\begin{aligned} -K^{(1)}(Q, P) &= -J^{(1)}(Q) \\ &\geqq \limsup_{n \to \infty} \{n^{-1} \log L_n(s)\} \\ &\geqq -\liminf_{n \to \infty} \{K_n^{(1)}(s: Q, P) - K_n^{(2)}(s: Q, P)\} \\ &= -K^{(1)}(s: Q, P) + \bar{K}^{(2)}(s: Q, P) \text{ a.e. } Q \end{aligned} \tag{4.16}$$

by Theorem 2 and (4.1). Hence $\bar{K}^{(2)}(s: Q, P) \leqq 0$ a.e. Q, so $\bar{K}^{(2)}(s: Q, P) = 0$ a.e. Q. *Q.E.D.*

5. Examples concerning exchangeable sequences

Let X be a finite set, say $X = \{a_1, \cdots, a_k\}$, $k \geqq 2$, and let $\mathscr{M}$ be the field of all subsets of X. Let $S = X^{(\infty)}$ be the set of all sequences $s = (x_1, x_2, \cdots)$ with each $x_n \in X$, and let $\mathscr{A} = \mathscr{M}^{(\infty)}$. For each n, let $\mathscr{B}_n$ be the field induced by the mapping: $s \to (x_1, \cdots, x_n)$.

Let V be the set of all points $v = (v_1, \cdots, v_k)$ with $v_i \geqq 0$ and $\Sigma_1^k v_i = 1$. For any v in V, let $M(\cdot|v)$ be the probability measure on $\mathscr{A}$ such that $M(s: x_1 = a_{i_1}, \cdots, x_n = a_{i_n}|v) = v_{i_1} \cdots v_{i_n}$ for all n and $a_{i_1}, \cdots, a_{i_n}$ in X. Let θ be a probability measure on V and let P_θ be defined by $P_\theta(A) = \int_V M(A|v)\, d\theta$ for $A \in \mathscr{A}$.

EXAMPLE 5.1. Let $\pi = (\pi_1, \cdots, \pi_k)$ be a given point in V with $\pi_i > 0$ for $i = 1, \cdots, k$, let θ_0 be the probability measure degenerate at π, and let Θ be the set of all measures θ such that $P_\theta\{s: x_1 = a_i\} = \pi_i$ for $i = 1, \cdots, k$. Let $\mathscr{P}_0$ consist of the one measure P_{θ_0}, and let $\mathscr{P}_1 = \{P_\theta: \theta \in \Theta\} - \{P_{\theta_0}\}$. In other words, we wish to test independence against exchangeability, the common marginal distribution of the x_n being known.

In the following, for $v = (v_1, \cdots, v_k)$ and $u = (u_1, \cdots, u_k)$ in V let

$$\phi(v, u) = \sum_{i=1}^{k} v_i \log (v_i/u_i), \tag{5.1}$$

with $0/0 = 1$ (say) and $0 \log 0 = 0$. Then $0 \leqq \phi \leqq \infty$, and $\phi = 0$ if and only if $v = u$.

For each $s = (x_1, x_2, \cdots)$ let $f_i^{(n)}(s)$ denote the number of $x_j = a_i$ for $1 \leqq j \leqq n$, $i = 1, \cdots, k$; $n = 1, 2, \cdots$, and let $\xi^{(n)}(s) = \big(f_1^{(n)}(s)/n, \cdots, f_k^{(n)}(s)/n\big)$. Let $z(s) = \lim_{n\to\infty} \xi^{(n)}(s)$ if the limit exists and let $z(s) = (1/k, \cdots, 1/k)$ otherwise. Then z is an $\mathscr{A}$ measurable function on S into V. It is known that, for each θ,

$$\lim_{n\to\infty} \xi^{(n)}(s) = z(s) \text{ a.e. } P_\theta, \tag{5.2}$$

and that $P_\theta\big(z(s) \in B\big) = \theta(B)$ for all Borel sets $B \subset V$. Now let

$$T_n(s) = \phi\big(\xi^{(n)}(s), \pi\big). \tag{5.3}$$

We shall show that $\{T_n\}$ is an optimal sequence and that its exact slope is $2\phi\big(z(s), \pi\big)$. It is interesting to note that here the optimal slope is a random variable, and that this slope depends on which alternative P_θ obtains only to the extent that the probability distribution of z is then θ.

It follows from the easy part of Sanov's theorem (see [12]) that, for $t \geqq 0$,

$$P_{\theta_0}(T_n \geqq t) = P_{\theta_0}\big(\phi\big(\xi^{(n)}(s), \pi\big), \geqq t\big) \leqq n^k \exp\{-n\alpha(t)\}, \tag{5.4}$$

where $\alpha(t) = \inf\{\phi(v, \pi): v \in V,\ \phi(v, \pi) \geqq t\} \geqq t$; thus (see Condition 2 of Corollary 5), $1 - G_n(t) \leqq n^k \exp\{-nt\}$. Now choose and fix a nonnull θ. A straightforward calculation shows that

$$r_n(s: P_\theta, P_{\theta_0}) = \exp\{nT_n(s)\} \int_V \exp\{-n\phi\big(\xi^{(n)}(s), v\big)\}\, d\theta. \tag{5.5}$$

Since $\phi \geqq 0$, it follows from (3.3) and (5.5) that

$$K_n^{(1)}(s: P_\theta, P_{\theta_0}) \leqq T_n(s) \tag{5.6}$$

for all s and n. Since $\phi(v, \pi)$ is continuous in v, it follows from (5.2) and (5.3) that

$$\lim_{n\to\infty} T_n(s) = \phi(z(s), \pi) \text{ a.e. } P_\theta. \tag{5.7}$$

It follows from the estimate of the null distribution of T_n obtained at the outset of this paragraph that $n^{-1} \log L_n(s) \leqq -T_n(s) + kn^{-1} \log n$ for all n and s. It follows hence from (5.6) and (5.7) that

$$\liminf_{n\to\infty} \{n^{-1} \log L_n(s)\} \leqq -\phi(z(s), \pi) \leqq -\bar{K}^{(1)}(s: P_\theta, P_{\theta_0}) \text{ a.e. } P_\theta \tag{5.8}$$

and

$$\limsup_{n\to\infty} \{n^{-1} \log L_n(s)\} \leqq -\phi(z(s), \pi) \text{ a.e. } P_\theta. \tag{5.9}$$

It follows from Theorem 2 with $P = P_{\theta_0}$, $Q = P_\theta$, and $\mathscr{C}_n$ the trivial field that

$$\liminf_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\bar{K}^{(1)}(s: P_\theta, P_{\theta_0}) \text{ a.e. } P_\theta \tag{5.10}$$

and

$$\limsup_{n\to\infty} \{n^{-1} \log L_n(s)\} \geqq -\underline{K}^{(1)}(s: P_\theta, P_{\theta_0}) \text{ a.e. } P_\theta. \tag{5.11}$$

It follows from (5.8) to (5.11) first that

$$\lim_{n\to\infty} K_n^{(1)}(s: P_\theta, P_{\theta_0}) = \phi(z(s), \pi) \text{ a.e. } P_\theta. \tag{5.12}$$

and next that

$$\lim_{n\to\infty} \{n^{-1} \log L_n(s)\} = -\phi(z(s), \pi) \text{ a.e. } P_\theta. \tag{5.13}$$

Since θ is arbitrary, it is plain from (5.13) that $\{T_n\}$ has slope 2ϕ and from (5.12) and Corollary 2 with $\mathscr{C}_n$ trivial that $\{T_n\}$ is optimal.

The preceding argument, which is an elaboration of the argument of Corollary 5, could be greatly simplified if we could deduce (5.12) directly from (5.5) and (5.7), but this direct deduction seems difficult for arbitrary θ.

It may be noted that the T_n discussed above is not quite the statistic $\hat{T}_n$ defined by (4.15). It can be shown that in the present case $\{\hat{T}_n\}$ is also an optimal sequence, but we are unable to compute $\hat{T}_n$ explicitly.

EXAMPLE 5.2. In the same framework as that of the preceding example, let Θ_0 be the set of all θ_0 which are degenerate at some point in the interior of V, and let Θ be the set of all measures on V such that $P_\theta\{s: x_1 = a_i\} > 0$ for $i = 1, \cdots, k$. Let $\mathscr{P}_0 = \{P_{\theta_0}: \theta_0 \in \Theta_0\}$ and $\mathscr{P}_1 = \{P_\theta: \theta \in \Theta\} - \mathscr{P}_0$. In other words, we wish to test independence versus exchangeability, the common marginal distribution of the x_n being unknown.

For each n let T_n be any $\mathscr{B}_n$ measurable statistic, let $\mathscr{C}_n$ be any field $\subset \mathscr{B}_n$, and let L_n^* be the level attained by $T_n|\mathscr{C}_n$, $n = 1, 2, \cdots$. We shall show that, for each nonnull P_θ,

$$\lim_{n\to\infty} n^{-1} \log L_n^*(s) = 0 \text{ a.e. } P_\theta. \tag{5.14}$$

Thus there exists no test sequence for which the level attained goes to zero exponentially fast. It would be interesting to know whether there exists a sequence for which $\lim_{n\to\infty} L_n^*(s) = 0$ a.e. P_θ, for all nonnull P_θ, and if so, to determine the fastest possible rate of convergence.

To establish (5.14), let $\{v^{(1)}, v^{(2)}, \cdots\}$ be a countable everywhere dense subset of V with each $v^{(j)}$ in the interior of V. Let θ_j be the probability measure degenerate at $v^{(j)}$. Then $\{\theta_1, \theta_2, \cdots\} \subset \Theta_0$. For each n and j, let $L_n(s; P_{\theta_j})$ be the level attained by $T_n|\mathscr{C}_n$ in testing P_{θ_j}. Then, by (4.3), $L_n^*(s) \geqq L_n(s; P_{\theta_j})$ for all s, n, and j. Now choose and fix a nonnull P_θ. Since the proof of (5.12) involves no conditions on θ (except perhaps the mutual absolute continuity of P_θ and P_{θ_0} on $\mathscr{B}_n$ for $n = 1, 2, \cdots$), (5.12) holds with $\theta_0 = \theta_j$ and $\pi = v^{(j)}$. It now follows from Theorem 2 with $P = P_{\theta_j}$, $Q = P_\theta$, and trivial $\mathscr{C}_n$ that

$$\liminf_{n\to\infty} \{n^{-1} \log L_n(s; P_{\theta_j})\} \geqq -\phi(z(s), v^{(j)}) \text{ a.e. } P_\theta, \tag{5.15}$$

for each j. Hence $\liminf_{n\to\infty} \{n^{-1} \log L_n^*(s)\} \geqq -\phi(z(s), v^{(j)})$ for all j a.e. P_θ. Since $0 = \phi(v, v) = \inf \{\phi(v, v^{(j)}) : j = 1, 2, \cdots\}$ for each $v \in V$, we conclude that $\liminf_{n\to\infty} n^{-1} \log L_n^*(s) \geqq 0$ a.e. P_θ. Since $L_n^* \leqq 1$, it follows that (5.14) holds.

6. Examples concerning Markov chains

Let $X = \{a_1, \cdots, a_k\}$ be a finite set, and let $S = X^{(\infty)}$ and $\mathscr{A}$ be defined as in the preceding section. It is assumed now that $s = (x_1, x_2, \cdots)$ is a Markov chain. Let $\mathscr{B}_n$ be the field induced by the mapping: $s \to (x_1, \cdots, x_{n+1})$, $n = 1, 2, \cdots$. In order to effect certain simplifications we shall assume that with probability one the sequence s starts off with $x_1 =$ a given point of X, say a_1. In effect, then, we shall be considering conditional tests given x_1, but this conditioning will not require explicit attention.

EXAMPLE 6.1. Let $\theta = \{\theta_{i,j}\}$ denote a $k \times k$ matrix with $\theta_{i,j} > 0$ and $\sum_{j=1}^k \theta_{i,j} = 1$ for $i = 1, \cdots, k$. Let P_θ be the measure on $\mathscr{A}$ determined by $P_\theta\{s : x_1 = a_1\} = 1$ and

$$P_\theta\{s : x_2 = a_{i_2}, \cdots, x_{n+1} = a_{i_{n+1}}\} = \theta_{1,i_2}\theta_{i_2,i_3} \cdots \theta_{i_n,i_{n+1}} \tag{6.1}$$

for all n and all $a_{i_2}, \cdots, a_{i_{n+1}}$ in X. Let Θ be a given set of transition probability matrices θ; let Θ^0 be a given subset of Θ; let $\mathscr{P}_0 = \{P_\theta : \theta \in \Theta^0\}$, and let $\mathscr{P}_1 = \{P_\theta : \theta \in \Theta^1\}$ where $\Theta^1 = \Theta - \Theta^0$.

For each $s = (x_1, x_2, \cdots)$ and n, let $g_{i,j}^{(n)}(s)$ denote the number of indices m with $1 \leqq m \leqq n$ such that $x_m = a_i$ and $x_{m+1} = a_j$, for $i, j = 1, \cdots, k$. The matrix $g^{(n)}(s) = \{g_{i,j}^{(n)}(s)\}$ is called the transition count matrix. Let $f_i^{(n)}(s) = \Sigma_{j=1}^k g_{i,j}^{(n)}(s)$ be the total frequency of a_i in $\{x_1, \cdots, x_n\}$. Then $\Sigma_{i=1}^k f_i^{(n)}(s) = n$. Define $\gamma_{i,j}^{(n)}(s) = g_{i,j}^{(n)}(s)/f_i^{(n)}(s)$ if $f_i^{(n)}(s) > 0$ and $\gamma_{i,j}^{(n)}(s) = 1/k$ (say) otherwise, for $i, j = 1, \cdots, k$. Let $\gamma^{(n)}(s) = \{\gamma_{i,j}^{(n)}(s)\}$. The matrix $\gamma^{(n)}$ is an estimate of θ; in fact, it is a maximum likelihood estimate based on $(x_1, \cdots, x_{n+1})$ if θ is entirely unknown. Let $\xi^{(n)}(s) = (f_1^{(n)}(s), \cdots, f_k^{(n)}(s))/n$. It is known that, for each θ,

$$\lim_{n\to\infty} \gamma^{(n)}(s) = \theta \text{ a.e. } P_\theta, \tag{6.2}$$

and that

$$\lim_{n\to\infty} \xi^{(n)}(s) = \pi(\theta) \text{ a.e. } P_\theta, \tag{6.3}$$

where $\pi(\theta) = (\pi_1(\theta), \cdots, \pi_k(\theta))$ is the stationary distribution over X corresponding to θ, that is, $\pi(\theta)$ is the unique solution of $\Sigma_{i=1}^k \pi_i\theta_{i,j} = \pi_j, j = 1, \cdots, k$, with $\pi_i > 0$, $\Sigma_{i=1}^k \pi_i = 1$.

In the following let $\gamma_i^{(n)}(s) = (\gamma_{i,1}^{(n)}(s), \cdots, \gamma_{i,k}^{(n)}(s))$ and $\theta_i = (\theta_{i,1}, \cdots, \theta_{i,k})$ denote the ith rows of $\gamma^{(n)}(s)$ and θ, respectively, and let

$$U_n(s;\theta) = \sum_{i=1}^k \xi_i^{(n)}(s)\phi(\gamma_i^{(n)}(s), \theta_i), \tag{6.4}$$

where ϕ is given by (5.1), and $\xi_i^{(n)} = f_i^{(n)}/n$ is the ith component of $\xi^{(n)}$. Then U_n is a sort of squared distance between $\gamma^{(n)}$ and θ. It is readily seen that, for any θ and θ^0,

$$K_n^{(1)}(s: P_\theta, P_{\theta^0}) = U_n(s;\theta^0) - U_n(s;\theta). \tag{6.5}$$

It follows from (6.2), (6.3), (6.4), and (6.5) that (4.1) holds, with

$$K^{(1)}(P_\theta, P_{\theta^0}) = \sum_{i=1}^k \pi_i(\theta)\phi(\theta_i, \theta_i^0). \tag{6.6}$$

Hence

$$J^{(1)}(P_\theta) = \inf\left\{\sum_{i=1}^k \pi_i(\theta)\phi(\theta_i, \theta_i^0) : \theta^0 \in \Theta^0\right\}. \tag{6.7}$$

Now let

$$T_n(s) = \inf\{U_n(s: \theta^0) : \theta^0 \in \Theta^0\}, \tag{6.8}$$

with U_n defined by (6.4). We shall show, by means of Corollary 5, that $\{T_n\}$ is asymptotically optimal.

To show that Condition 1 of Corollary 5 is satisfied, choose and fix P_θ. Let s be a point such that $\gamma^{(n)}(s) \to \theta$ and $\xi^{(n)}(s) \to (\pi_1(\theta), \cdots, \pi_k(\theta))$. In view of (6.2) and (6.3) it will suffice to show that $\alpha(s) \equiv \lim\inf_{n\to\infty} T_n(s) \geqq J^{(1)}(P_\theta)$ for this s. We may suppose that $\alpha < \infty$. For each n, let $\delta^{(n)}$ be a point in Θ^0 such that $T_n \geqq U_n(s;\delta^{(n)}) - n^{-1}$; such a $\delta^{(n)}$ exists, by (6.8). There exists a sequence $m_1 < m_2 < \cdots$, of positive integers m_r and a probability matrix δ such that $T_{m_r} \to \alpha$ and $\delta^{(m_r)} \to \delta$ as $r \to \infty$. Let n be restricted to $\{m_r\}$. Then

(6.9) $$\alpha = \lim_{n\to\infty} T_n \geqq \liminf_{n\to\infty} \{U_n(s;\delta^{(n)})\} \geqq \sum_{i=1}^{k} \pi_i(\theta) \liminf_{n\to\infty} \phi(\gamma_i^{(n)}, \delta_i^{(n)})$$

by (6.4). Since $\pi_i(\theta) > 0$ and $\theta_{i,j} > 0$ for all i and j, since $\alpha < \infty$, and since $\gamma^{(n)} \to \theta$ and $\delta^{(n)} \to \delta$, it follows first that $\delta_{i,j} > 0$ for all i and j and next that $\alpha \geqq \sum_{i=1}^{k} \pi_i(\theta)\phi(\theta_i, \delta_i)$. This last lower bound cannot be less than $J^{(1)}(P_\theta)$ defined by (6.6), (6.7) since δ is in the closure of Θ^0. Thus $\alpha \geqq J^{(1)}(P_\theta)$.

In order to show that Condition 2 of Corollary 5 also holds, let M_n be the set of all $k \times k$ matrices $m = \{m_{i,j}\}$ with $m_{i,j} = 0, 1, 2, \cdots$ and $\sum_{i=1}^{k} \sum_{i=1}^{k} m_{i,j} = n$. It is readily seen that for any $m \in M_n$ and any θ, $P_\theta(g^{(n)}(s) = m) = \nu(m) \prod_{i,j=1}^{k} [\theta_{i,j}]^{m_{i,j}}$, where $\nu(m)$ is the number of distinct sequences $(x_1, \cdots, x_{n+1})$, possibly zero, with $x_1 = a_1$ and transition count $g^{(n)} = m$. Now, $U_n(s;\theta)$ defined by (6.4) depends on s only through $g^{(n)}(s)$, say $U_n(s;\theta) \equiv U(g^{(n)}(s);\theta)$, and $\gamma^{(n)}(s)$ is also a function of $g^{(n)}(s)$, say $\gamma^{(n)}(s) \equiv G(g^{(n)}(s))$. An easy calculation shows that

(6.10) $$\begin{aligned} P_\theta(g^{(n)}(s) = m) &= \exp\{-nU(m;\theta)\} P_{G(m)}(g^{(n)}(s) = m) \\ &\leqq \exp\{-nU(m;\theta)\}. \end{aligned}$$

Now choose $t \geqq 0$ and $\theta^0 \in \Theta^0$, and let $A = A(n;t;\theta^0)$ be the set $\{m: m \in M_n, U(m,\theta^0) \geqq t\}$. Then

(6.11) $$\begin{aligned} P_{\theta^0}(U_n(s;\theta^0) \geqq t) &= P_{\theta^0}(U(g^{(n)}(s);\theta^0) \geqq t) \\ &= \sum_A P_{\theta_0}(g^{(n)} = m) \leqq \sum_A \exp\{-nU(m;\theta^0)\} \\ &\leqq \exp\{-nt\} \sum_A 1 \leqq \exp\{-nt\} \sum_{M_n} 1 \\ &\leqq n^{k^2} \exp\{-nt\}. \end{aligned}$$

It now follows from (6.8) that $P_{\theta^0}(T_n \geqq t) \leqq n^{k^2} \exp\{-nt\}$. Since θ^0 is arbitrary, $1 - G_n(t) \leqq n^{k^2} \exp\{-nt\}$ for all n and t, and Condition 2 is satisfied.

It is shown in [8] that if Θ^0 consists of a single point θ^0 then T_n is asymptotically optimal even in the sense of [12].

The sequence $\{T_n\}$ considered above does not depend on what the given set Θ is. We now show that, with $T_n(s;\Theta^0)$ defined by the right side of (6.8), and with

(6.12) $$\hat{T}_n(s) = T_n(s;\Theta^0) - T_n(s;\Theta),$$

$\{\hat{T}_n\}$ is also asymptotically optimal. For any θ, $0 \leqq T_n(s;\Theta) \leqq U_n(s;\theta)$, and $U_n(s;\theta) \to 0$ a.e. P_θ. Hence, by (6.8) and (6.12), $T_n(s) - \hat{T}_n(s) \to 0$ a.e. P_θ. Secondly, $\hat{T}_n(s) \leqq T_n(s)$ for all n and s, by (6.4) and (6.12). Since $\{T_n\}$ satisfies Conditions 1 and 2 of Corollary 5, we see that $\{\hat{T}_n\}$ also satisfies these conditions. In general, that is, for arbitrary Θ^0 and $\Theta^1 = \Theta - \Theta^0$, T_n and $\hat{T}_n$ are quite different. Presumably $\hat{T}_n$ is not only asymptotically optimal but actually better than T_n (see [2], pp. 16–17) for testing Θ^0 against Θ^1 in cases where $T_n \not\equiv \hat{T}_n$.

EXAMPLE 6.2. Suppose now that Θ is the set of all θ with positive elements, and that Θ^0 is the set of all θ in Θ with identical rows. In other words, we wish to test independence against stationary Markovian dependence, the actual distribution in either case being unspecified.

In the present case, $T_n(s;\Theta) \equiv 0$, so the statistics $T_n(s)$ and $\hat{T}_n(s)$ defined by (6.8) and (6.12) are identical. In order to express $\{\hat{T}_n\}$ and its slope in explicit form we require the following easily verified proposition. Let $v = (v_1, \cdots, v_k)$ be a fixed point in V (that is, $v_i \geqq 0$, $\Sigma_{i=1}^k v_i = 1$); let $\theta_1, \cdots, \theta_k$ be fixed points in V, say $\theta_i = (\theta_{i,1}, \cdots, \theta_{i,k})$; and let u be a variable point in V. Then, with the convention that $0 \cdot \infty = 0$, $\Sigma_{i=1}^k v_i \phi(\theta_i, u)$ is minimized by $u = (\Sigma_{i=1}^k v_i\theta_{i,1}, \cdots, \Sigma_{i=1}^k v_i \theta_{i,k})$. Now for each n let $h_i^{(n)}(s)$ = the number of $x_j = a_i$ for $2 \leqq j \leqq n+1$, and let $\eta^{(n)}(s) = n^{-1}(h_1^{(n)}(s), \cdots, h_k^{(n)}(s))$, that is, $\eta^{(n)}$ is the vector of relative frequency counts in $\{x_2, \cdots, x_{n+1}\}$. It follows from the stated proposition that

$$\hat{T}_n(s) = \sum_{i=1}^k \xi_i^{(n)}(s)\phi(\gamma_i^{(n)}(s), \eta^{(n)}(s)). \tag{6.13}$$

It also follows that

$$J^{(1)}(P_\theta) = \sum_{i=1}^k \pi_i(\theta)\phi(\theta_i, \pi(\theta)). \tag{6.14}$$

According to Example 6.1, $\{\hat{T}_n\}$ is asymptotically optimal and has exact slope $2J^{(1)}$.

Now for $n \geqq 2$ let $W^{(n)}(s) = (W_1^{(n)}(s), \cdots, W_k^{(n)}(s))$ be the vector of frequency counts in $\{x_2, \cdots, x_n\}$ and let $\mathscr{C}_n$ be the field induced by the mapping: $s \to (x_1;\ W^{(n)}(s);\ x_{n+1})$. It is known (see [7]) that the conditional distribution of $(x_1, \cdots, x_{n+1})$ given $\mathscr{C}_n$ is the same for every θ in Θ^0. In particular, $\hat{T}_n | \mathscr{C}_n$ has an exact null distribution function, say $\hat{F}_n(t, s)$. We proceed to show, essentially by an extension of Corollary 5 to conditional tests, that $\{\hat{T}_n|\mathscr{C}_n\}$ is also an optimal sequence. It is plain from (6.13) and (6.14) that $\hat{T}_n(s) \to J^{(1)}(P_\theta)$ a.e. P_θ, for every θ. We shall show that $1 - \hat{F}_n(t, s) \leqq \exp\{-nt\} f_n(s)$ for all n, t, and s, where $n^{-1} \log f_n(s) \to 0$ a.e. P_θ, for each θ. It will then follow from Corollary 3, as desired, that with $L_n^*(s) = 1 - \hat{F}_n(\hat{T}_n(s), s)$, $n^{-1} \log L_n^*(s) \to -J^{(1)}(P_\theta)$ a.e. P_θ for every nonnull θ.

Nonnull sets in $\mathscr{C}_n$ are unions of sets of the form

$$\{s: x_1 = a_1,\ W^{(n)}(s) = b,\ x_{n+1} = a_j\} = C_n(b; a_j), \tag{6.15}$$

say, where a_j is a point of X and b is $(b_1, \cdots, b_k)$ with each b_i a nonnegative integer and $\Sigma_{i=1}^k b_i = n - 1$. For $t \geqq 0$ let $B_n(t) = \{s: \hat{T}_n(s) \geqq t\}$. Let θ^0 be a point in Θ^0. Then $1 - \hat{F}_n(t, s)$ equals $P_{\theta^0}(B_n(t) | C_n(b, a_j))$ evaluated at $b = W^{(n)}(s)$ and $a_j = x_{n+1}$. Now, $P_{\theta^0}(B_n | C_n) \leqq P_{\theta^0}(B_n)/P_{\theta^0}(C_n)$. We have seen in Example 6.1 that $P_{\theta^0}(B_n) \leqq n^{k^2} \exp\{-nt\}$. If θ^0 has $(\pi_1, \cdots, \pi_k)$ as each row, then

$$P_{\theta^0}(C_n(b; a_j)) = \pi_j \left[\prod_{i=1}^k \left(\frac{\pi_i^{b_i}}{b_i!}\right)\right] (n-1)! = \pi_j \psi_n(b; \theta^0), \tag{6.16}$$

say. Let $\delta(\theta^0) = \min\{\pi_1, \cdots, \pi_k\}$. Then $\delta > 0$, and $P_{\theta^0}(C_n(b; a_j)) \geqq \delta(\theta^0)\psi_n(b; \theta^0)$ for all j. Since θ^0 is arbitrary, we see that $1 - \hat{F}_n(t, s) \leqq \exp\{-nt\} f_n(s)$, where

$$f_n(s) = n^{k^2} \inf\{[\delta(\theta^0)\psi_n(W^{(n)}(s); \theta^0)]^{-1}: \theta^0 \in \Theta^0\}. \tag{6.17}$$

Suppose now that P_θ obtains. Since $W^{(n)}(s) = f^{(n)}(s) - (1, 0, \cdots, 0)$, it follows from (6.3) that $n^{-1} W^{(n)}(s) \to \pi(\theta)$ a.e. P_θ. Let θ^* be the point in Θ^0 which has $\pi(\theta)$ for each row. It follows from Stirling's formula that $n^{-1} \log \psi_n(W^{(n)}(s); \theta^*) \to 0$ a.e. P_θ, as $n \to \infty$. Since $f_n(s) \geqq 1$, and since θ^* is a point in Θ^0, it follows from (6.17) that $n^{-1} \log f_n(s) \to 0$ a.e. P_θ.

EXAMPLE 6.3. Let Θ^0 be the same set as in the preceding example, let θ be a given transition probability matrix, $\theta \notin \Theta^0$, and suppose that we wish to test Θ^0 against θ. Then Examples 6.1 and 6.2 already provide three different asymptotically optimal test sequences. Another optimal sequence for the present case is $\{T_n^* \mid \mathscr{C}_n\}$, where $\mathscr{C}_n$ is the field considered above, and

$$(6.18) \qquad T_n^*(s) = \sum_{i=1}^{k} \xi_i^{(n)}(s) \left[\phi\left(\gamma_i^{(n)}(s), \pi(\theta)\right) - \phi\left(\gamma_i^{(n)}(s), \theta_i\right)\right].$$

That $\{T_n^* \mid \mathscr{C}_n\}$ has slope $2J^{(1)}(P_\theta)$ against θ may be seen from Corollary 4, as follows. Since T_n^* is $\mathscr{B}_n$ measurable, $T_n^* \mid \mathscr{C}_n$ has an exact null distribution. Let θ^* be the matrix with $\pi(\theta)$ as each row. Then $K^{(1)}(P_\theta, P_{\theta^*}) = J^{(1)}(P_\theta)$, and it follows from (6.4), (6.5), and (6.18) that $T_n^*(s) \equiv n^{-1} \log r_n(s: P_\theta, P_{0^*})$. It remains therefore to verify that condition (b) of Corollary 4 is satisfied with $Q = P_\theta$ and $P_Q = P_{\theta^*}$. This verification can be made by a direct calculation, but is immediately available from Corollary 6 since $\{\hat{T}_n \mid \mathscr{C}_n\}$ is an optimal sequence.

7. Examples concerning independent and identically distributed observations

In this section X is a Borel set of a Euclidean space of points x, $\mathscr{M}$ is the field of Borel sets of X, $S = X^{(\infty)}$ is the space of points $s = (x_1, x_2, \cdots)$, and $\mathscr{A} = \mathscr{M}^{(\infty)}$. For each n, $\mathscr{B}_n$ is the σ-field induced by the mapping: $s \to (x_1, \cdots, x_n)$. The set Θ is an index set of points θ, and Θ_0 is a subset of Θ. For each θ in Θ, p_θ is a probability measure on $\mathscr{M}$, and $P_\theta = P_\theta^{(\infty)}$. The null set of measures is $\mathscr{P}_0 = \{P_\theta: \theta \in \Theta_0\}$; the nonnull set is $\mathscr{P}_1 = \{P_\theta: \theta \in \Theta\} - \mathscr{P}_0$. It is assumed that, for any θ and θ_0 in Θ, p_θ and p_{θ_0} are mutually absolutely continuous on $\mathscr{M}$. Consequently, P_θ and P_{θ_0} are mutually absolutely continuous on $\mathscr{B}_n$ for each n, and (4.1) is satisfied with

$$(7.1) \qquad K^{(1)}(P_\theta, P_{\theta_0}) = \int_X \log \left(dp_\theta / dp_{\theta_0}\right) dp_\theta.$$

In accordance with the notation of [2], [3] we shall write the integral in (7.1) as $K(\theta, \theta_0)$, and write $J^{(1)}(P_\theta)$ defined by (4.2) as $J(\theta)$.

Let $y(x)$ be a Borel measurable transformation of X into a Euclidean space Y and for each n let $\mathscr{C}_n$ be the σ-field induced by: $s \to \left(y(x_1), \cdots, y(x_n)\right)$. It is then readily seen that $K_n^{(2)}(s: P_\theta, P_{\theta_0}) \to K^*(\theta, \theta_0)$ a.e. P_θ, where

$$(7.2) \qquad K^* = \int_Y \log \left(dp_\theta y^{-1} / dp_{\theta_0} y^{-1}\right) dp_\theta y^{-1}.$$

With the present choice of $\mathscr{C}_n$ Corollary 1 reduces to the statement that $0 \leqq K^*(\theta, \theta_0) \leqq K(\theta, \theta_0) \leqq \infty$. This choice of $\mathscr{C}_n$ is not used elsewhere in this section.

In the following examples we give the exact slopes and related large deviation probability estimates for various likelihood ratio statistics. Most of these results have been obtained previously by other methods; our main object in reconsidering these examples here is to point out that the first part of Corollary 2 offers a simple method for all such examples. This method does not require explicit estimation of large deviation probabilities in the null case; indeed the estimates referred to can be obtained, if needed, after the exact slope is found.

EXAMPLE 7.1. Suppose X is the real line and p_0 is a nondegenerate probability measure on X. Suppose that the moment generating function $\phi(\theta) = \int_X \exp\{\theta x\}\, dp_0$ is finite for $\theta \in \Theta = [0, \delta)$ where $0 < \delta \leqq \infty$. For each θ in Θ let p_θ be defined by $dp_\theta = [\phi(\theta)]^{-1} \exp\{\theta x\}\, dp_0$ and let Θ_0 consist of the single point $\theta_0 = 0$. For $0 < \theta < \delta$ let $b(\theta) = E_\theta(x)$. Then $-\infty < b(\theta) < \infty$, and $K(\theta, \theta_0) = \theta b(\theta) - \log \phi(\theta) = J(\theta)$. For each n, let $T_n(s) = (x_1 + \cdots + x_n)/n$. Consider a particular $\theta > 0$. Then $T_n(s)$ is equivalent to $r_n(s: P_\theta, P_{\theta_0})$. Since (4.1) holds with $K^{(1)} = J$, it follows from Corollary 2 that $\{T_n\}$ has exact slope $2J(\theta)$ when θ obtains.

We observe next that $T_n(s) \to b(\theta)$ a.e. P_θ. Since b and J are continuous and strictly increasing over $(0, \delta)$ it follows that for given $t \in \{b(\theta): 0 < \theta < \delta\}$ there exists a unique θ_t such that $b(\theta_t) = t$, and f and g of (4.13) are both equal to $J(\theta_t) = \theta_t t - \log \phi(\theta_t)$. It follows hence from the conclusion of the preceding paragraph that, for each $t \in \{b(\theta): 0 < \theta < \delta\}$,

$$n^{-1} \log P_0(x_1 + \cdots + x_n \geqq nt) \to -J(\theta_t) \tag{7.3}$$

as $n \to \infty$. It is thus seen that Chernoff's theorem [9] is deducible from Corollary 2 and the law of large numbers. A different proof, based on the central limit theorem, is given in [5].

EXAMPLE 7.2. Let $x = (y, z)$ where y and z are zero-one variables. Each possible value of x has positive probability but the distribution of x is otherwise entirely unknown. The null hypothesis is that y and z are independent. For each n let $\hat{T}_n$ be the likelihood ratio statistic (4.15). It follows from Theorem 2 of [2] that $\{\hat{T}_n\}$ is optimal. Write $x_n = (y_n, z_n)$, and let $\mathscr{C}_n$ be the field induced by $U_n(s) = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n z_i\right)$. Then $\{\hat{T}_n | \mathscr{C}_n\}$ is also optimal. The level attained by $\hat{T}_n | \mathscr{C}_n$ equals the level attained by $T_n | \mathscr{C}_n$, where $T_n = \left|\sum_{i=1}^n y_i z_i - M_n(s)\right|$ and M_n is a complicated function of U_n and $\Sigma_1^n y_i z_i$;

$$M_n \doteq n^{-1}\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n z_i\right). \tag{7.4}$$

We omit the verification.

EXAMPLE 7.3. Suppose X is the k dimensional Euclidean space of points $x = (y_1, \cdots, y_k)$, and that Θ is the space of all points $\theta = (\mu_1, \cdots, \mu_k; \sigma)$ with $-\infty < \mu_i < \infty$ for each i and $0 < \sigma < \infty$. Suppose that when θ obtains

$y_1, \cdots, y_k$ are independent normally distributed variables with $E_\theta(y_i) = \mu_i$ and $\mathrm{Var}_\theta(y_i) = \sigma^2$ for $i = 1, \cdots, k$. The null hypothesis is that $\mu_1 = \cdots = \mu_k = 0$. It is readily seen that in this case

$$J(\theta) = \tfrac{1}{2}k \log [1 + k^{-1}\delta^2(\theta)], \qquad \delta^2(\theta) = \sum_{i=1}^{k} \left(\frac{\mu_i}{\sigma}\right)^2. \tag{7.5}$$

Now for each $n \geqq 2$, let $T_n(s)$ be n^{-1} times the appropriate F statistic based on $(x_1, \cdots, x_n)$. It has been shown by Abrahamson [1] that $\{T_n\}$ has exact slope $2J(\theta)$ against every θ and so is asymptotically optimal. The method used in [1] is to note that

$$T_n(s) \to k^{-1}\delta^2(\theta) \text{ a.e. } P_\theta, \tag{7.6}$$

and then to show that, with $G_n(t)$ the null distribution function of T_n,

$$n^{-1} \log [1 - G_n(t)] \to -\tfrac{1}{2}k \log [1 + t], \tag{7.7}$$

for each $t > 0$. Since the limit in (7.7) is continuous in t, it follows (see [3], pp. 309–310) from (7.6) that $\{T_n\}$ has slope $2J$ defined by (7.5). A second possible method of establishing Abrahamson's result is by means of Theorem 2 of [2] since T_n is equivalent to $\hat{T}_n$, but the verifications required seem formidable. We now show that Corollary 2 can be used to obtain first the slope of $\{T_n\}$ and then (7.7).

Choose and fix a nonnull $\theta = (\mu_1, \cdots, \mu_k; \sigma)$. Let $\theta^0 = (0, \cdots, 0; \sigma)$. Then $J(\theta) = K(\theta, \theta^0)$. Let $f_n(t)$ and $g_n(t)$ be the probability densities of $T_n(s)$ under θ and θ^0 respectively, and let $h_n(t) = f_n(t)/g_n(t)$ for $t > 0$. Let $\mathscr{C}_n$ be the σ-field induced by $T_n(s)$. Then

$$\rho_n(s: P_\theta, P_{\theta^0}) = h_n(T_n(s)). \tag{7.8}$$

It follows from known results (see [13], p. 312) that

$$h_n(t) = \sum_{j=0}^{\infty} \gamma_j(n)\pi_j(\tfrac{1}{2}n\delta^2)\left[\frac{nt}{n(1 + t) - 1}\right]^j \tag{7.9}$$

where

$$\gamma_j(n) = \frac{\Gamma(\frac{1}{2}nk + j)\Gamma(\frac{1}{2}k)}{\Gamma(\frac{1}{2}nk)\Gamma(\frac{1}{2}k + j)}, \tag{7.10}$$

and $\pi_j(\lambda)$ denotes the Poisson probability $\exp\{-\lambda\}\, \lambda^j/j!$. Since h_n is a strictly increasing function of t, we see from (7.8) that T_n and ρ_n are equivalent statistics. Consequently, $\{T_n\}$ has exact slope $2J(\theta)$ against θ if $\{\rho_n\}$ has exact slope $2J(\theta)$. We shall show that this last is the case by showing that, with $K_n^{(2)} = n^{-1} \log \rho_n$,

$$K_n^{(2)}(s) \to J(\theta) \text{ a.e. } P_\theta, \tag{7.11}$$

for then Corollary 2 (with $\mathscr{B}_n$ and $\mathscr{C}_n$ of the corollary replaced by the present $\mathscr{C}_n$ and the trivial field, respectively) applies.

For each n let j_n be the positive integer such that $\frac{1}{2}n\delta^2 < j_n \leqq \frac{1}{2}n\delta^2 + 1$. Since each term in the series in (7.9) is positive,

$$h_n(t) \geqq \gamma_{j_n}(n)\pi_{j_n}(\tfrac{1}{2}n\delta^2)[nt/(n(1+t)-1)]^{j_n}. \tag{7.12}$$

It now follows from (7.10) by an application of Stirling's formula that, for each t,

$$\liminf_{n\to\infty} n^{-1}\log h_n(t) \geqq \tfrac{1}{2}k\log(1+k^{-1}\delta^2) + \tfrac{1}{2}\delta^2\log\left[\frac{kt(1+k^{-1}\delta^2)}{\delta^2(1+t)}\right]. \tag{7.13}$$

Since the left side of (7.13) is nondecreasing in t and the right side is continuous in t, it follows from (7.6), (7.8) and (7.13) that $\underline{K}^{(2)}(s) \geqq \frac{1}{2}k\log(1+k^{-1}\delta^2)$ a.e. P_θ. But $\frac{1}{2}k\log(1+k^{-1}\delta^2) = J(\theta) = K(\theta,\theta^0) = K^{(1)}(P_\theta, P_{\theta^0})$. It therefore follows from Corollary 1, as desired, that (7.11) holds. Since $\{T_n\}$ is shown to have exact slope $2J(\theta)$ against each θ, it follows from (7.5) and (7.6) that (7.7) holds for each $t > 0$. Incidentally, it follows from (7.5), (7.6), (7.8), and (7.11) that $n^{-1}\log h_n(t) \to \frac{1}{2}k\log(1+t)$ for each $t > 0$ and $\delta^2 > 0$.

EXAMPLE 7.4. Let (a, b) be an interval on the real line, $-\infty \leqq a < b \leqq \infty$, and let $\mathscr{F}$ denote the set of all probability distribution functions F on the real line such that F assigns probability 1 to (a, b) and $F'(t)$ exists and is continuous and positive over (a, b). Let $\theta = (F, G)$ be a point of $\mathscr{F} \times \mathscr{F}$. Let X be the set of all points $x = (y, z)$ with real y and z and let p_θ be the measure on X corresponding to y and z being independent random variables with distribution functions F and G, respectively. Let $\Theta_0 = \{(H, H): H \in \mathscr{F}\}$ be the null set. The nonnull set is a single point $\theta = (F, G)$ with $F \not\equiv G$.

It is readily seen that, with $\theta_0 = (H, H)$ and $K(\theta, \theta_0)$ defined by the right side of (7.1),

$$K(\theta,\theta_0) = \int_{-\infty}^{+\infty}\log(dF/dH)\,dF + \int_{-\infty}^{+\infty}\log(dG/dH)\,dG. \tag{7.14}$$

It follows easily from (7.14) that

$$J(\theta) = K(\theta,\theta^0), \tag{7.15}$$

where $\theta^0 = (\bar{H}, \bar{H})$ and

$$\bar{H}(t) = \tfrac{1}{2}[F(t) + G(t)]. \tag{7.16}$$

For each n let $x_n = (y_n, z_n)$, let N denote $2n$, and let $u_{1,N}(s) \leqq \cdots \leqq u_{N,N}(s)$ be the ordered values in the set $\{y_1, z_1; \cdots; y_n, z_n\}$. Let $v_{i,N}(s) = 1$ if $u_{i,N}(s) = z_j$ for some $j = 1, \cdots, n$ and $v_{i,N}(s) = 0$ otherwise, and let $V_n(s) = (v_{1,N}(s), \cdots, v_{N,N}(s))$.

It has been shown by Hájek [10] that the rank statistic vector V_n is asymptotically fully informative in the following sense: there exist weights $\{a_{i,N}: i = 1, \cdots, N: N = 2, 4, \cdots\}$ such that, with $T_n(s) = \sum_{i=1}^N a_{i,N}v_{i,N}(s)$, $\{T_n\}$ has exact slope $2J(\theta)$ against the given θ; the weights depend of course on θ. As noted in

[10], this remarkable result implies that the likelihood ratio statistic for testing Θ_0 against θ based on V_n also has slope $2J(\theta)$.

Let $\mathscr{C}_n$ be the field induced by V_n. Then the likelihood ratio statistic based on V_n is $\rho_n(s) = \rho_n(s: P_\theta, P_{\theta^0})$. We have

$$\rho_n(s) = \binom{N}{n}\{P_\theta(V_n(s) = b_n)\}_{b_n = V_n(s)}, \tag{7.17}$$

where b_n denotes an N long vector of zeros and ones. It follows from Corollary 2 (with $\mathscr{B}_n$ and $\mathscr{C}_n$ of the corollary replaced with the present $\mathscr{C}_n$ and the trivial field, respectively) that the second part of Hájek's result is equivalent to

$$\lim_{n\to\infty} K_n^{(2)}(s) = J(\theta) \text{ a.e. } P_\theta. \tag{7.18}$$

where $K_n^{(2)} = n^{-1} \log \rho_n$. That (7.18) does hold can be seen from [6]. It is shown in [6] that $K_n^{(2)}(s) \to I^*$ a.e. P_θ where I^* is a constant, and it follows from the formulae for I^* given in [6], [15] that in fact $I^*(\theta) = \inf\{K(\theta, \theta_0): \theta_0 \in \Theta_0\} = \mathrm{J}(\theta)$; we omit the details.

Now let $r_n(s) = r_n(s: P_\theta, P_{\theta^0})$ be the density of P_θ with respect to P_{θ^0} on $\mathscr{B}_n$. Let $\mathscr{D}_n$ be the σ-field induced by $U_n(s) = (u_{1,N}(s), \cdots, u_{N,N}(s))$, where the $u_{i,N}$ are the order statistics as above. We shall show that $\{r_n|\mathscr{D}_n\}$ is also an optimal sequence. For each null θ_0, the conditional distribution of $(y_1, z_1; \cdots; y_n, z_n)$ given $U_n = (a_1, \cdots, a_N)$, with $a_1 < \cdots < a_N$, is uniform over the permutations of $(a_1, \cdots, a_N)$; hence $r_n | \mathscr{D}_n$ has an exact null distribution. It follows from (7.15) that conditions (a) and (c) of Corollary 4 (with $\mathscr{C}_n$ of the corollary replaced by $\mathscr{D}_n$) are satisfied with $Q = P_\theta$ and $P_Q = P_{\theta^0}$. It therefore remains to show that condition (b) is satisfied by $\{\mathscr{D}_n\}$. Let $\xi_n(s)$ be a $\mathscr{D}_n$ measurable function such that $dP_\theta = \xi_n(s)\, dP_{\theta^0}$ on $\mathscr{D}_n$, $0 < \xi_n < \infty$, and let $K_n^{(3)}(s) = n^{-1} \log \xi_n(s)$. We have to show that

$$K_n^{(3)}(s) \to 0 \text{ a.e. } P_\theta. \tag{7.19}$$

It seems difficult to establish (7.19) directly, but one can argue as follows. In the null case, U_n and V_n are independent random vectors. Since ρ_n is a function of V_n, it follows that the level attained by ρ_n is a version of the level attained by $\rho_n|\mathscr{D}_n$. Consequently $\{\rho_n|\mathscr{D}_n\}$ is an asymptotically optimal sequence. Since $J < \log 4$, it follows from Corollary 6 with $P = P_{\theta^0}$ and $Q = P_\theta$ (with $\mathscr{C}_n$ of the corollary replaced by $\mathscr{D}_n$) that (7.19) holds.

Let $f = F'$, $g = G'$, and $h = \frac{1}{2}(f + g)$. Then

$$r_n(s) = \prod_{i=1}^n f(y_i)g(z_i) \Big/ \prod_{i=1}^n h(y_i)h(z_i). \tag{7.20}$$

It follows from this formula, as is well known, that if f is a normal density and $g(t) = f(t - c)$ with $c > 0$, then the level attained by $r_n|\mathscr{D}_n$ equals the level attained by the Fisher-Pitman test $\sum_1^n z_i - \sum_1^n y_i|\mathscr{D}_n$. Since the latter test does not depend on $\theta = (F, G)$, we conclude that the Fisher-Pitman test is asymptotically optimal against all one sided normal translation alternatives.

REFERENCES

[1] I. G. Abrahamson, "On the stochastic comparison of tests of hypotheses," Ph.D. dissertation, University of Chicago, 1965.

[2] R. R. Bahadur, "An optimal property of the likelihood ratio statistics," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 1, pp. 13–26.

[3] ———, "Rates of convergence of estimates and test statistics," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 303–324.

[4] R. R. Bahadur and P. J. Bickel, "On conditional test levels in large samples," *Essays in Prob. Statist.*, University of North Carolina Monograph Series No. 3 (1970), pp. 25–34.

[5] R. R. Bahadur and R. Ranga Rao, "On deviations of the sample mean," *Ann. Math. Statist.*, Vol. 31 (1960), pp. 1015–1027.

[6] R. H. Berk and I. R. Savage, "The information in rank-order and the stopping time of some associated SPRT's," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 1661–1674.

[7] P. Billingsley, "Statistical methods in Markov chains," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 12–40.

[8] L. B. Boza, "Asymptotically optimal tests for finite Markov chains," Ph.D. thesis, University of California, Berkeley, 1970.

[9] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 493–507.

[10] J. Hájek, "Asymptotic sufficiency of the vector of ranks in the Bahadur sense," *Ann. Math. Statist.*, Vol. 42 (1971).

[11] J. L. Hodges, Jr. and E. L. Lehmann, "The efficiency of some nonparametric competitors of the t-test," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 324–335.

[12] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 369–408.

[13] E. L. Lehmann, *Testing Statistical Hypotheses*, New York, Wiley, 1959.

[14] M. Raghavachari, "On a theorem of Bahadur on the rate of convergence of test statistics," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 1695–1699.

[15] I. R. Savage, "Nonparametric statistics: a personal view," *Sankhyā, Ser. A*, Vol. 31 (1969), pp. 107–144.