ASYMPTOTIC SUFFICIENCY AND SOME RELATED NOTIONS

BY

TADAYUKI MATSUDA

TECHNICAL REPORT NO. 362

December 1990

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Asymptotic Sufficiency and Some Related Notions

By

Tadayuki Matsuda

Technical Report No. 362
December 1990

Prepared Under the Auspices
Of
National Science Foundation Grant DMS89–05874

Department of Statistics
Stanford University
Stanford, California
ASYMPTOTIC SUFFICIENCY AND SOME RELATED NOTIONS

by

Tadayuki Matsuda
Wakayama University
Visiting Stanford University

Abstract

The concept of sufficiency plays an important role in statistical inference. Roughly speaking, a statistic is said to be sufficient if the statistic does not lose the information contained in the sample. In this paper, we consider some notions which are related to the concept of sufficiency as the sample size tends to infinity and investigate the relations between these notions.

Key Words: Asymptotic Sufficiency ; Asymptotic Bayes risk sufficiency ; Fisher information ; Likelihood ratio ; Locally uniform contiguity ; Local asymptotic normality.
1 Introduction

Let \((\mathcal{X}, \mathcal{A}_n), n \in N = \{1, 2, \ldots\}\), be a sequence of measurable spaces. For each \(n \in N\) let \(\mathcal{P}_n = \{P_{\theta, n}; \theta \in \Theta\}\) be a family of probability measures defined on \(\mathcal{A}_n\), where \(\Theta\) is an open subset of the \(k\)-dimensional Euclidean space \(\mathbb{R}^k\).

Throughout this paper, we consider a sequence \(\{\mathcal{B}_n\}\) of sub \(\sigma\)-fields satisfying \(\mathcal{B}_n \subseteq \mathcal{A}_n\) for all \(n \in N\). \(\mathcal{A}_n\) and \(\mathcal{B}_n\) are interpreted as the \(\sigma\)-field generated by the sample and the one generated by a statistic, respectively, where \(n\) means the sample size.

We present the concept of asymptotic sufficiency including higher order asymptotic sufficiency. The concept of asymptotic sufficiency of order \(o(1)\) is due to LeCam [2].

**Definition 1.1** Let \(\{a_n\}\) be a sequence of positive numbers. \(\{\mathcal{B}_n\}\) is said to be **asymptotically sufficient for** \(\{\mathcal{A}_n\}\) **up to** \(o(a_n)\) **if** for each \(n\) there is a family of probability measures \(\mathcal{Q}_n = \{Q_{\theta, n}; \theta \in \Theta\}\) on \(\mathcal{A}_n\) such that (a) for every \(n\), \(\mathcal{B}_n\) is sufficient for \(\mathcal{Q}_n\) and that (b) \(\sup_{\theta \in K} \|P_{\theta, n} - Q_{\theta, n}\| = o(a_n)\) for every compact subset \(K\) of \(\Theta\), where \(\|\cdot\|\) means the total variation norm over \(\mathcal{A}_n\).

Let us consider the problem of testing a sequence of null hypotheses \(H_{n0} : P_{\theta, n}\) against a sequence of alternative \(H_{n1} : P_{\tau, n}, \tau \neq \theta\). For any real number \(c \geq 0\) let

\[
r_n(c, \tau, \theta; \mathcal{A}_n) = \inf \left\{ c \frac{E_{P_{\tau, n}}[\phi_n]}{1 + c} + \frac{1}{1 + c}(1 - E_{P_{\tau, n}}[\phi_n]) \right\}.
\]

where the infimum is taken over all \(\mathcal{A}_n\)-measurable statistical test functions and \(E_{P_{\theta, n}}[\phi_n]\) stands for the expectation of \(\phi_n\) under \(P_{\theta, n}\). It is noticed that \(r_n(c, \tau, \theta; \mathcal{A}_n)\) means the Bayes risk of statistical problem of testing \(H_{n0}\) against \(H_{n1}\) relative to a prior probability distribution \(\{c/(1 + c), 1/(1 + c)\}\) on \(\{\theta, \tau\}\) provided that the loss function is simple. The difference of the Bayes risks between \(\mathcal{B}_n\)-measurable test functions and \(\mathcal{A}_n\)-measurable test functions

\[
R_n(c, \tau, \theta) = r_n(c, \tau, \theta; \mathcal{B}_n) - r_n(c, \tau, \theta; \mathcal{A}_n)
\]

may then be taken as a measure of the information loss. It is obvious from the definition of Bayes risk that \(R_n(c, \tau, \theta)\) is nonnegative.

**Definition 1.2** Let \(\{a_n\}\) be a sequence of positive numbers. \(\{\mathcal{B}_n\}\) is said to be **asymptotically Bayes risk sufficient for** \(\{\mathcal{A}_n\}\) **up to** \(o(a_n)\) **with respect**
to the statistical problem of testing $H_{n0}$ against $H_{n1}$ if for each $n$ and $\theta \in \Theta$ there is a neighborhood $U_n(\theta) \subset \Theta$ of $\theta$ such that for every compact subset $K$ of $\Theta$

$$\sup_{c>0} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} R_n(c, \tau, \theta) = o(a_n).$$

Moreover, there are some notions considered as a measure of the information loss in an asymptotic sense. In this paper, we deal with the likelihood ratio function, the score of the sample and the Fisher information as such a measure.

In Section 2 we introduce how a sequence of asymptotically sufficient $\sigma$-fields is useful for the problems of statistical inference. Section 3 discusses the relation between Bayes risks and likelihood ratio functions. Section 4 investigates the case that the Fisher information is taken as a measure of the information loss. In the final Section 5 we find the relation between the concept of asymptotic sufficiency and the concept of asymptotic Bayes risk sufficiency.

2 Asymptotic Sufficiency

We shall state a result due to LeCam [2] concerning asymptotic sufficiency in the framework of decision theory. For each $n$ let the space $\mathcal{D}_n$ of available decisions be a Borel subset of a Euclidean space. It is assumed that, for each $n$, the loss function $L_n$ defined on $\mathcal{D}_n \times \Theta$ is measurable in $u \in \mathcal{D}_n$ for every $\theta \in \Theta$ and that there exists a positive number $M$ such that $|L_n(u, \theta)| \leq M$ for every triplet $(n, u, \theta)$. For each $n$ let $\mathcal{B}_n$ be a sub $\sigma$-field of $\mathcal{A}_n$. $\delta_n$ is called a $\mathcal{B}_n$-measurable decision function if, for every Borel $B$ of $\mathcal{D}_n$, $\delta_n(B(\cdot))$ is a $\mathcal{B}_n$-measurable function of $x$ and, for every $x \in \mathcal{X}$, $\delta_n(\cdot|x)$ is a probability measure defined on the Borel subsets of $\mathcal{D}_n$. To a decision function $\delta_n$, correspond the risk function $r_n(\delta_n, \theta)$ by

$$r_n(\delta_n, \theta) = \int_\mathcal{X} \int_{\mathcal{D}_n} L_n(u, \theta) \delta_n(du|x) dP_{\theta,n}(x).$$

LeCam showed that if $\{\mathcal{B}_n\}$ is asymptotically sufficient for $\{\mathcal{A}_n\}$ up to $o(a_n)$ then for every sequence $\{\delta_n\}$ of $\mathcal{A}_n$-measurable decision functions there is a sequence $\{\delta_n^*\}$ of $\mathcal{B}_n$-measurable decision functions such that for every compact subset $K$ of $\Theta$

$$\sup_{\delta \in K} |r_n(\delta_n, \theta) - r_n(\delta_n^*, \theta)| = o(a_n).$$
In other words, we can make use of a sequence of \( B_n \)-measurable decision functions with information loss of order \( o(a_n) \).

This problem in the estimation theory with unbounded loss function was studied by Suzuki [11], Suzuki-Kusama [15] and Matsuda [6].

We restrict our attention to the class of estimators with the following property.

**Definition 2.1** Let \( \{T_n\} \) be a sequence of \( A_n \)-measurable functions from \( X \) into \( R^k \). \( \{T_n\} \) is said to be locally uniformly consistent estimator of order \( o(a_n) \) if for every \( \varepsilon > 0 \) and every compact subset \( K \) of \( \Theta \)

\[
\sup_{\theta \in K} P_{\theta,n}\{|T_n - \theta| > \varepsilon\} = o(a_n).
\]

We now state a result which means Rao-Blackwell theorem in an asymptotic sense (Matsuda [6]).

**Theorem 2.1** Let \( L(t) \) be a convex function from \( R^k \) into \([0, \infty)\). If \( \{B_n\} \) is asymptotically sufficient for \( \{A_n\} \) up to \( o(a_n) \) and \( \{T_n\} \) is locally uniformly consistent estimator of order \( o(a_n) \), then there exists a sequence \( \{T^*_n\} \) of \( B_n \)-measurable functions such that

(a) \( \{T^*_n\} \) is locally uniformly consistent estimator of order \( o(a_n) \),

(b) for every compact subset \( K \) of \( \Theta \)

\[
\limsup_{n \to \infty} \sup_{\theta \in K} a_n^{-1} \{E_{P_{\theta,n}}[L(T^*_n - \theta)] - E_{P_{\theta,n}}[L(T_n - \theta)]\} \leq 0.
\]

In particular, the theorem implies the following result due to Suzuki [11] and Suzuki-Kusama [15].

**Corollary 2.1** For \( r \geq 1 \) let \( L_r(t) = |t|^r \) be the loss function defined on \( R^k \). Suppose that \( \{B_n\} \) is asymptotically sufficient for \( \{A_n\} \) up to \( o(n^{-r/2}) \) and that \( \{T_n\} \) is locally uniformly consistent estimator of order \( o(n^{-r/2}) \).

Furthermore suppose that for each \( \theta \in \Theta \) there exists a probability measure \( \lambda_\theta \) on \( R^k \) which is weakly continuous relative to \( \theta \) such that for any compact subset \( K \) of \( \Theta \) the distribution of \( \sqrt{n}(T_n - \theta) \) converges weakly to \( \lambda_\theta \) uniformly relative to \( \theta \) in \( K \) and that \( \lambda_\theta(\{0\}) \neq 1 \) for every \( \theta \in \Theta \). Then there exists a sequence \( \{T^*_n\} \) of \( B_n \)-measurable functions such that

(a) \( \{T^*_n\} \) is locally uniformly consistent estimator of order \( o(n^{-r/2}) \),

(b) for every compact subset \( K \) of \( \Theta \)

\[
\limsup_{n \to \infty} \sup_{\theta \in K} \frac{E_{P_{\theta,n}}[L_r(T^*_n - \theta)]}{E_{P_{\theta,n}}[L_r(T_n - \theta)]} \leq 1.
\]
Pfanzagl [8] proved that a sequence of estimators with properties analogous to those of maximum likelihood estimator is asymptotically sufficient up to $O(n^{-1/2})$. Higher order asymptotically sufficient statistics have been constructed by Michel [7], Suzuki [11],[12] and Matsuda [4], [5].

3 Asymptotic Bayes risk sufficiency

Hereafter we assume that for each $n \in N$, $\mathcal{P}_n = \{P_{\theta_n}; \theta \in \Theta \}$ is dominated by a $\sigma$-finite measure $\mu_n$ on $(\mathcal{X}, \mathcal{A}_n)$. Let $p_n(x, \theta) = dP_{\theta_n}/d\mu_n$ and $S_n(\theta) = \{x \in \mathcal{X}; p_n(x, \theta) > 0\}$. Define the likelihood ratio function by

$$h_n(x; \tau, \theta) = \begin{cases} p_n(x, \tau)/p_n(x, \theta), & \text{if } x \in S_n(\theta), \\ \infty, & \text{if } x \in S_n(\theta)^c \cap S_n(\tau), \\ 1, & \text{if } x \in S_n(\theta)^c \cap \overline{S_n(\tau)}^c. \end{cases} \quad (3)$$

Let $\{B_n\}$ be a sequence of sub $\sigma$-fields satisfying $B_n \subset A_n$ for all $n \in N$. For each $\theta \in \Theta$ and $n \in N$ define $\bar{p}_n(x, \theta) \equiv E\mu_n[p_n(x, \theta)|B_n]$ which is the conditional expectation of $p_n(x, \theta)$ given $B_n$ with respect to $\mu_n$ and put $\bar{S}_n(\theta) = \{x \in \mathcal{X}; \bar{p}_n(x, \theta) > 0\}$. Similarly, we can define $B_n$-measurable likelihood ratio function $\bar{h}_n(x; \tau, \theta)$. This function is considered as the likelihood ratio restricted to sub $\sigma$-field $B_n$.

Consider the problem of testing a null hypothesis $H_{n0} : P_{\theta_n}$ against an alternative $H_{n1} : P_{\tau_n}$ with a prior distribution $\{c/(1 + c), 1/(1 + c)\}$ on $\{\theta, \tau\}$. Define the sequences $\{\phi_n^*\}$ and $\{\bar{\phi}_n^*\}$ of test functions by

$$\phi_n^*(x) = \begin{cases} 1, & \text{if } p_n(x, \tau) > cp_n(x, \theta), \\ 0, & \text{if } p_n(x, \tau) < cp_n(x, \theta), \end{cases} \quad (4)$$

and

$$\bar{\phi}_n^*(x) = \begin{cases} 1, & \text{if } \bar{p}_n(x, \tau) > c\bar{p}_n(x, \theta), \\ 0, & \text{if } \bar{p}_n(x, \tau) < c\bar{p}_n(x, \theta). \end{cases} \quad (5)$$

It is clear that $\phi_n^*(x)$ is $A_n$-measurable and $\bar{\phi}_n^*(x)$ is $B_n$-measurable. By the definition (1) of Bayes risk it is easy to see that

$$(1 + c)r_n(c, \tau, \theta; A_n) = 1 - \int_{\mathcal{X}} \phi_n^*(x)(p_n(x, \tau) - cp_n(x, \theta))d\mu_n,$$

$$(1 + c)r_n(c, \tau, \theta; B_n) = 1 - \int_{\mathcal{X}} \bar{\phi}_n^*(x)(\bar{p}_n(x, \tau) - c\bar{p}_n(x, \theta))d\mu_n.$$
and hence
\[(1 + c)R_n(c, \tau, \theta) = \int_X (\phi_n(x) - \tilde{\phi}_n(x))(p_n(x, \tau) - cp_n(x, \theta))d\mu_n. \quad (6)\]

We now state a result due to Suzuki [13] which gives relations between Bayes risks and likelihood ratio functions.

**Lemma 3.1** For every $c > 0$ and $\eta > 0$

\begin{align*}
(a) & \quad P_{\theta,n}\{\tilde{h}_n < c < c + \eta < h_n\} \leq \frac{1 + c}{\eta}R_n(c, \tau, \theta), \\
(b) & \quad P_{\theta,n}\{h_n < c - \eta < c < \tilde{h}_n\} \leq \frac{1 + c}{\eta}R_n(c, \tau, \theta),
\end{align*}

where $h_n = h_n(x; \tau, \theta)$ and $\tilde{h}_n = \tilde{h}_n(x; \tau, \theta)$.

**Proof.** We divide the integral (6) into the following parts:

\[(1 + c)R_n(c, \tau, \theta) = \int_{p_n(x, \tau) > cp_n(x, \theta)} (1 - \tilde{\phi}_n(x))(p_n(x, \tau) - cp_n(x, \theta))d\mu_n \]
\[+ \int_{p_n(x, \tau) < cp_n(x, \theta)} \tilde{\phi}_n(x)(cp_n(x, \theta) - p_n(x, \tau))d\mu_n \]
\[= \tilde{R}_{n1} + \tilde{R}_{n2}\text{ (say)}. \quad (7)\]

Since $\tilde{R}_{n1}$ and $\tilde{R}_{n2}$ are positive, we have

\[(1 + c)R_n(c, \tau, \theta) \geq \int_{p_n(x, \tau) > cp_n(x, \theta) > 0} (1 - \tilde{\phi}_n(x))(p_n(x, \tau) - cp_n(x, \theta))d\mu_n \]
\[\geq \int_{h_n < c < c + \eta < h_n} (1 - \tilde{\phi}_n(x))(h_n - c)dP_{\theta,n} \]
\[\geq \eta P_{\theta,n}\{h_n < c < c + \eta < h_n, \tilde{p}_n(x, \tau) < cp_n(x, \theta)\} \]
\[= \eta P_{\theta,n}\{h_n < c < c + \eta < h_n\},\]

which is the first part (a) stated in the lemma. In a similar way we can obtain the second part (b) of the lemma.

**Lemma 3.2** For every $\varepsilon > 0$ and $M \geq \max(\varepsilon, 1)$

\[P_{\theta,n}\{|h_n - \tilde{h}_n| > \varepsilon, h_n \leq M, \tilde{h}_n \leq M\} \leq 16M^2\varepsilon^{-2}\sup_{c > 0} R_n(c, \tau, \theta).\]
Proof. Observe that
\[
P_{\theta,n}\{|h_n - \bar{h}_n| > \varepsilon, h_n \leq M, \bar{h}_n \leq M\}
\leq \sum_{i=1}^{L} P_{\theta,n}\{h_n < \frac{i\varepsilon}{2}, h_n > \frac{i+1}{2}\varepsilon\} + \sum_{i=1}^{L} P_{\theta,n}\{\bar{h}_n < \frac{i\varepsilon}{2}, h_n > \frac{i+1}{2}\varepsilon\},
\]
where \(L = \lfloor 2(M - \varepsilon)/\varepsilon \rfloor + 1 \) ([ \cdot ] denotes the Gauss symbol). It follows from Lemma 3.1 that
\[
P_{\theta,n}\{|h_n - \bar{h}_n| > \varepsilon, h_n \leq M, \bar{h}_n \leq M\}
\leq \sum_{i=1}^{L} \frac{2}{\varepsilon} \left\{(1 + \frac{i + 1}{2}\varepsilon)R_n(\frac{i + 1}{2}\varepsilon, \tau, \theta) + (1 + \frac{i}{2}\varepsilon)R_n(\frac{i}{2}\varepsilon, \tau, \theta)\right\}
\leq \left\{\frac{4L}{\varepsilon} + 2L + L^2\right\} \sup_{c>0} R_n(c, \tau, \theta).
\]
This with the fact that \(L \leq 2M/\varepsilon\) leads to the desired result.

We obtain the following proposition from the lemma above.

**Proposition 3.1** Suppose that for every \(\theta \in \Theta\) and every \(n \in N\) there exists a neighborhood \(U_n(\theta)\) of \(\theta\) such that for every compact subset \(K\) of \(\Theta\)
\[
\lim_{n \to \infty} \sup_{c>0} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} R_n(c, \tau, \theta) = 0.
\]
Then for every \(\varepsilon > 0\) and every compact subset \(K\) of \(\Theta\) we have
\[
\lim_{n \to \infty} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} P_{\theta,n}\{|h_n(x; \tau, \theta) - \bar{h}_n(x; \tau, \theta)| > \varepsilon\} = 0.
\]

Proof. According to Lemma 3.2 it is sufficient to notice that for \(M > 0\)
\[
P_{\theta,n}\{h_n > M\} \leq E_{\theta,n}[h_n(x; \tau, \theta)]/M \leq 1/M
\]
and
\[
P_{\theta,n}\{\bar{h}_n > M\} \leq E_{\theta,n}[\bar{h}_n(x; \tau, \theta)]/M \leq 1/M.
\]

The condition stated in Proposition 3.1 is expected to hold with a neighborhood \(U_n(\theta; b) = \{\tau \in \Theta; \sqrt{n}|\tau - \theta| \leq b\}, b > 0\), of \(\theta\) in regular cases. We shall give an easy example in order to mention it.
Example 3.1 Let \( \{X_n\} \) be a sequence of independent and identically distributed random variables according to a normal distribution \( N(\theta, 1) \) with \(-\infty < \theta < \infty\). For \( n \in N \), let \( A_n = \sigma(X_1, \ldots, X_n) \) and \( B_n = \sigma(X_1, \ldots, X_m) \) with \( m = m(n) \leq n \). Then we have

\[
p_n(x, \theta) = \left( \frac{1}{2\pi} \right)^{n/2} \exp \left\{ -\frac{\sum_{i=1}^{n} (x_i - \theta)^2}{2} \right\}, \quad x = (x_1, x_2, \ldots),
\]

\[
\tilde{p}_n(x, \theta) = \left( \frac{1}{2\pi} \right)^{m/2} \exp \left\{ -\frac{\sum_{i=1}^{m} (x_i - \theta)^2}{2} \right\}, \quad x = (x_1, x_2, \ldots).
\]

It follows from easy computation that if \( 1 - m/n = o(n^{-\gamma}) \) for some \( \gamma > 0 \), then for every finite \( b > 0 \)

\[
\sup_{c > 0} \sup_{\theta \in \hat{R}} \sup_{\tau \in U_n(\theta, b)} R_n(c, \tau, \theta) = o(n^{-\gamma}).
\]

Moreover, it holds that for every \( b > 0 \) and \( s > 0 \)

\[
\sup_{\theta \in \hat{R}} \sup_{\tau \in U_n(\theta, b)} \mathbb{E}_{P_{\theta, n}}[h_n(x; \tau, \theta)^s] < \infty.
\]

Since

\[
\bar{h}_n(x; \tau, \theta) = \mathbb{E}_{P_{\theta, n}}[h_n(x; \tau, \theta)|B_n], \quad (8)
\]

the above inequality still holds by replacing \( h_n(x; \tau, \theta) \) by \( \bar{h}_n(x; \tau, \theta) \). Hence, a similar argument to Proposition 3.1 shows that for every \( \varepsilon > 0 \) and \( b > 0 \)

\[
\sup_{\theta \in \hat{R}} \sup_{\tau \in U_n(\theta, b)} P_{\theta, n}\{|h_n(x; \tau, \theta) - \bar{h}_n(x; \tau, \theta)| > \varepsilon\} = o(n^{-\beta})
\]

with every \( \beta \) satisfying \( 0 < \beta < \gamma \). This example shows that we can estimate the speed of convergence for Proposition 3.1 if the supports \( S_n(\theta) \) of \( P_{\theta, n} \) are independent of \( \theta \) and the likelihood ratio functions have higher order moments. Here the former condition is needed to hold the relation (8).

Next, we shall show that if the difference between the likelihood ratio functions \( h_n \) and \( \bar{h}_n \) is tending to zero in probability as \( n \to \infty \), then the difference of the corresponding Bayes risks is tending to zero. We start with the following definition which is a uniform version of the notion of contiguity introduced by LeCam [3].

8
Definition 3.1 The family \( \{P_{\theta,n} : \theta \in \Theta\} \) of probability measures is said to be \textit{locally uniformly contiguous} if for every \( \theta \in \Theta \) and every \( n \in N \) there exists a neighborhood \( U_n(\theta) \) of \( \theta \) such that
(a) \( \theta \in U_n(\tau) \) implies \( \tau \in U_n(\theta) \),
(b) for every compact subset \( K \) of \( \Theta \) there exists compact \( \tilde{K} \) such that for sufficiently large \( n \)
\[
\bigcup_{\theta \in K} U_n(\theta) \subset \tilde{K} \subset \Theta,
\]
(c) for compact subset \( K \) of \( \Theta \) and \( A_n \in A_n, n \in N \),
\[
\lim_{n \to \infty} \sup_{\theta \in K} P_{\theta,n} \{A_n\} = 0 \implies \lim_{n \to \infty} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} P_{\tau,n} \{A_n\} = 0.
\]
It is well known in the independent and identically distributed case with the same support that
\[
\lim_{n \to \infty} P_{\theta,n} \{A_n\} = 0 \implies \lim_{n \to \infty} \sup_{\tau \in U_n(\theta;b)} P_{\tau,n} \{A_n\} = 0
\]
for any \( b > 0 \), where \( U_n(\theta;b) \) is the same as in Example 3.1. The above definition gives a uniform contiguity on compact sets.
Moreover, in this case, if
\[
\sup_{\theta \in K} \sup_{\tau \in U_n(\theta;b)} E_{P_{\theta,n}}[h_n(x;\tau,\theta)^2] < \infty,
\]
then it follows from Schwarz's inequality that
\[
(P_{\tau,n} \{A_n\})^2 \leq P_{\theta,n} \{A_n\} E_{P_{\theta,n}}[h_n(x;\tau,\theta)^2].
\]
This inequality guarantees the condition (c) in Definition 3.1.
We show the reverse relation of Proposition 3.1.

Proposition 3.2 Let \( \{P_{\theta,n} : \theta \in \Theta\} \) be locally uniformly contiguous with a sequence \( \{U_n(\theta)\} \) of neighborhoods of \( \theta \). Assume that for every \( \varepsilon > 0 \) and every compact subset \( K \) of \( \Theta \)
\[
\lim_{n \to \infty} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} P_{\theta,n} \{|h_n(x;\tau,\theta) - \bar{h}_n(x;\tau,\theta)| > \varepsilon\} = 0.
\]
Then for every compact subset \( K \) of \( \Theta \)
\[
\lim_{n \to \infty} \sup_{c > 0} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} R_n(c,\tau,\theta) = 0.
\]
Proof. For every $\epsilon > 0$ let $A_n = \{x; |h_n(x; \tau, \theta) - \bar{h}_n(x; \tau, \theta)| \leq \epsilon \}$. By virtue of (7) we have

$$
\bar{R}_{n1} \leq \int_{p_n(x, \tau) > c} (1 - \bar{p}_n^*(x))(p_n(x, \tau) - cp_n(x, \theta))d\mu_n
+ \int_{p_n(x, \theta) = 0} p_n(x, \tau)d\mu_n
\leq \int_{p_n(x, \tau) > c} (1 - \bar{p}_n^*(x))(p_n(x, \tau) - cp_n(x, \theta))d\mu_n
+ P_{\tau, n}\{A_n^c\} + P_{\tau, n}\{S_n(\theta)^c\}
\leq \int_{h_n > c - \epsilon} (1 - \bar{p}_n^*(x))(h_n + \epsilon - c)p_n(x, \theta)d\mu_n
+ P_{\tau, n}\{A_n^c\} + P_{\tau, n}\{S_n(\theta)^c\}.
$$

The first term on the right side of the last inequality is less than

$$
\int_{c - \epsilon < h_n \leq c} \epsilon \bar{p}_n(x, \theta)d\mu_n \leq \epsilon,
$$

which leads to

$$
\bar{R}_{n1} \leq \epsilon + P_{\tau, n}\{A_n^c\} + P_{\tau, n}\{S_n(\theta)^c\}.
$$

A similar argument shows that

$$
\bar{R}_{n2} \leq \epsilon + cP_{\theta, n}\{A_n^c\} + cP_{\theta, n}\{S_n(\tau)^c\}.
$$

Consequently, for any $c > 0$

$$
R_n(c, \tau, \theta) \leq 2\epsilon + P_{\tau, n}\{A_n^c\} + P_{\tau, n}\{S_n(\theta)^c\} + P_{\theta, n}\{A_n^c\} + P_{\theta, n}\{S_n(\tau)^c\}.
$$

Let $K$ be a compact subset of $\Theta$. By assumption $P_{\theta, n}\{A_n^c\}$ tends to 0 uniformly in $\theta \in K$ and $\tau \in U_n(\theta)$ as $n \to \infty$ and so $P_{\tau, n}\{A_n^c\}$ tends to 0 uniformly in $\theta \in K$ and $\tau \in U_n(\theta)$ as $n \to \infty$ because of locally uniform contiguity. Similarly, $P_{\theta, n}\{S_n(\theta)^c\} = 0$ for $\theta \in K$ implies $P_{\tau, n}\{S_n(\theta)^c\}$ tends to 0 uniformly in $\theta \in K$ and $\tau \in U_n(\theta)$ as $n \to \infty$. To estimate $P_{\theta, n}\{S_n(\tau)^c\}$ choose compact set $\tilde{K}$ such that $\cup_{\theta \in K} U_n(\theta) \subset \tilde{K} \subset \Theta$ for sufficiently large $n$. $P_{\tau, n}\{S_n(\tau)^c\} = 0$ for $\tau \in \tilde{K}$ implies $P_{\theta, n}\{S_n(\tau)^c\}$ tends to 0 uniformly in $\tau \in \tilde{K}$ and $\theta \in U_n(\tau)$ as $n \to \infty$. Thus we have

$$
\lim_{n \to \infty} \sup_{c > 0} \sup_{\theta \in K} \sup_{\tau \in U_n(\theta)} R_n(c, \tau, \theta) \leq 2\epsilon.
$$

Since $\epsilon$ is arbitrary, letting $\epsilon \to 0$ gives the desired result.
It is remarked that if $\mathcal{P}_n$ has the same support, then the conditions (a) and (b) in Definition 3.1 are omitted in proving Proposition 3.2.

We illustrate the proposition with an example of exponential family including the normal distribution in Example 3.1.

**Example 3.2** Let $\{X_n\}$ be a sequence of independent and identically distributed random variables according to $k$-parameter exponential family whose density can be expressed as

$$p(x, \theta) = \exp\{\theta'x - \psi(\theta)\}$$

with respect to some common measure $\mu$. Let $\Theta$ be an open set included in the natural parameter space of the exponential family. For $n \in N$, let $\mathcal{A}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{B}_n = \sigma(X_1, \ldots, X_m)$ with $m = m(n) \leq n$. Since for every $b > 0$ and every compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} \sup_{\tau \in \mathcal{U}_n(\theta, b)} E_{P_{\theta, n}}[h_n(x; \tau, \theta)^2] < \infty,$$

exponential family is locally uniformly contiguous with a sequence $\{U_n(\theta; b)\}$ of neighborhoods of $\theta$. Moreover, we can easily show that if $m/n = 1 - o(n^{-\gamma})$ for some $\gamma \geq 0$, then for every $\varepsilon > 0$, $b > 0$ and every compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} \sup_{\tau \in \mathcal{U}_n(\theta, b)} P_{\theta, n}\{|h_n(x; \tau, \theta) - \tilde{h}_n(x; \tau, \theta)| > \varepsilon\} = o(n^{-\gamma}).$$

Hence, from Proposition 3.2 we obtain

$$\lim_{n \to \infty} \sup_{c > 0} \sup_{\theta \in K} \sup_{\tau \in \mathcal{U}_n(\theta, b)} R_n(c, \tau, \theta) = 0.$$

### 4 Fisher information

In this section, we assume that the support $S_n(\theta)$ of probability measure $P_{\theta, n}$ is independent on $\theta \in \Theta$. Without loss of generality, it is assumed that $p_n(x, \theta) > 0$ for all $x \in \mathcal{X}$, $\theta \in \Theta$ and $n \in N$. Since $\theta$ is vector-valued, we put $\theta = (\theta_1, \ldots, \theta_k)$. The Fisher information contained in the sample of size $n$ is defined by

$$i_n(\theta; \mathcal{A}_n) = \left(E_{P_{\theta, n}}[\theta_i \log p_n(x, \theta)] \frac{\partial}{\partial \theta_j} \log p_n(x, \theta) ; 1 \leq i, j \leq k\right). \quad (9)$$
Here it is assumed that all the elements of the Fisher information matrix are finite and the matrix is positive definite.

Let \( \{ \mathcal{B}_n \} \) be a sequence of sub-\( \sigma \)-fields such that \( \mathcal{B}_n \subset \mathcal{A}_n \). Then, the Fisher information matrix restricted in sub-\( \sigma \)-field \( \mathcal{B}_n \) is given by

\[
i_n(\theta; \mathcal{B}_n) = \left( E_{P_{\theta,n}} [\frac{\partial}{\partial \theta_i} \log \bar{p}_n(x, \theta) \frac{\partial}{\partial \theta_j} \log \bar{p}_n(x, \theta)] ; 1 \leq i, j \leq k \right).
\] (10)

In order for the information to be well defined, we must assume that the conditional expectation \( \bar{p}_n(x, \theta) \) of \( p_n(x, \theta) \) given \( \mathcal{B}_n \) is positive for all \( x \in \mathcal{X} \) and \( \theta \in \Theta \), and differentiable with respect to \( \theta \in \Theta \) for all \( x \in \mathcal{X} \). Moreover, if differentiation with respect to \( \theta \) of \( p_n(x, \theta) \) under the integral sign is permitted in \( E_{\mu_n}[p_n(x, \theta)|\mathcal{B}_n] \), then we have

\[
E_{P_{\theta,n}} [\frac{\partial}{\partial \theta_i} \log p_n(x, \theta)|\mathcal{B}_n] = \frac{\partial}{\partial \theta_i} \log \bar{p}_n(x, \theta) \quad \text{a.s., } i = 1, \ldots, k,
\] (11)

which imply

\[
E_{P_{\theta,n}} [\frac{\partial}{\partial \theta_i} \log p_n(x, \theta) \frac{\partial}{\partial \theta_j} \log \bar{p}_n(x, \theta)] = E_{P_{\theta,n}} [\frac{\partial}{\partial \theta_i} \log \bar{p}_n(x, \theta) \frac{\partial}{\partial \theta_j} \log \bar{p}_n(x, \theta)].
\]

Let

\[
\left( \frac{\partial}{\partial \theta_1} \log p_n(x, \theta), \ldots, \frac{\partial}{\partial \theta_k} \log p_n(x, \theta), \frac{\partial}{\partial \theta_1} \log \bar{p}_n(x, \theta), \ldots, \frac{\partial}{\partial \theta_k} \log \bar{p}_n(x, \theta) \right)
\]

be 2\( k \)-dimensional random vector. Then the covariance matrix of the random vector is given by

\[
\begin{pmatrix}
i_n(\theta; \mathcal{A}_n) & i_n(\theta; \mathcal{B}_n) \\
i_n(\theta; \mathcal{B}_n) & i_n(\theta; \mathcal{B}_n)
\end{pmatrix}.
\]

Since the covariance matrix is nonnegative definite, the matrix defined by

\[
\begin{pmatrix}I & -I \\
O & I\end{pmatrix} \begin{pmatrix}i_n(\theta; \mathcal{A}_n) & i_n(\theta; \mathcal{B}_n) \\
i_n(\theta; \mathcal{B}_n) & i_n(\theta; \mathcal{B}_n)\end{pmatrix} \begin{pmatrix}I & O \\
-I & I\end{pmatrix}
\]

\[
= \begin{pmatrix}i_n(\theta; \mathcal{A}_n) - i_n(\theta; \mathcal{B}_n) & O \\
O & i_n(\theta; \mathcal{B}_n)\end{pmatrix}
\]

is also nonnegative definite, where \( I \) and \( O \) denote the identity matrix and the zero matrix, respectively. Thus submatrix \( i_n(\theta; \mathcal{A}_n) - i_n(\theta; \mathcal{B}_n) \) is also
nonnegative definite. It is well known that a necessary and sufficient condition for \( i_n(\theta; A_n) = i_n(\theta; B_n) \) is that \( B_n \) is sufficient for \( P_n \). It suffices to note that \( B_n \) is sufficient for \( P_n \) iff
\[
\frac{\partial}{\partial \theta_i} \log p_n(x, \theta) = \frac{\partial}{\partial \theta_i} \log \bar{p}_n(x, \theta) \text{ a.s., } i = 1, \ldots, k.
\]

Let
\[
\Delta_n(\theta) = \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta_1} \log p_n(x, \theta), \ldots, \frac{\partial}{\partial \theta_k} \log p_n(x, \theta) \right)'
\]
and
\[
\bar{\Delta}_n(\theta) = \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta_1} \log \bar{p}_n(x, \theta), \ldots, \frac{\partial}{\partial \theta_k} \log \bar{p}_n(x, \theta) \right)'
\]
be \( k \times 1 \) column vectors. \( \sqrt{n} \Delta_n(\theta) \) is called the score of the sample. It is noticed that \( \bar{\Delta}_n(\theta) = E_{P_\theta,n}[\Delta_n(\theta)|B_n] \) because of (11). Since
\[
E_{P_\theta,n}[\Delta_n(\theta) - \bar{\Delta}_n(\theta)]^2 = \frac{1}{n} tr(i_n(\theta; A_n) - i_n(\theta; B_n)),
\]
it follows from Schwarz's inequality that
\[
\frac{1}{\sqrt{k}} E_{P_\theta,n}[\Delta_n(\theta) - \bar{\Delta}_n(\theta)]^2 \leq \frac{1}{n} ||i_n(\theta; A_n) - i_n(\theta; B_n)||
\leq E_{P_\theta,n}[\Delta_n(\theta) - \bar{\Delta}_n(\theta)]^2,
\]
where the norm \( || \cdot || \) of matrix means the usual Euclidean norm on \( R^{k^2} \). This inequality yields that \( B_n \) is sufficient for \( P_n \) iff \( \Delta_n(\theta) = \bar{\Delta}_n(\theta) \) a.s.

The following definition is due to Rao [10] in which the concept is called asymptotic efficiency.

**Definition 4.1** \( B_n \) is said to be *asymptotically sufficient for \( \{A_n\} \) in the sense of Fisher information* if for every compact subset \( K \) of \( \Theta \)
\[
\lim_{n \to \infty} \sup_{\theta \in K} \frac{1}{n} ||i_n(\theta; A_n) - i_n(\theta; B_n)|| = 0.
\]

We shall give a condition equivalent to asymptotic sufficiency in the sense of Fisher information.

**Lemma 4.1** Suppose the following conditions hold:
(1) The distribution of $\Delta_n(\theta)$ under $P_{\theta,n}$ converges weakly to a distribution $\lambda_\theta$ uniformly on compact sets of $\theta$, where $\lambda_\theta$ is weakly continuous on $\Theta$.

(2) For any $M > 0$, there exists $M_0 > M$ such that for all $\theta \in \Theta$
$$\lambda_\theta\{|y| = M_0\} = 0.$$ 

(3) A family of measures defined on the Borel $\sigma$-field of $\mathbb{R}^k$
$$\nu_\theta(A) = \int_A |y|^2 \lambda_\theta(dy)$$ 

is uniformly tight on compact sets of $\theta$.

(4) For every compact subset $K$ of $\Theta$
$$\lim_{n \to \infty} \sup_{\theta \in K} \left| \int_{|x| < M} |\Delta_n(\theta)|^2 dP_{\theta,n} - \int_{\mathbb{R}^k} |y|^2 \lambda_\theta(dy) \right| = 0.$$ 

Then for every compact subset $K$ of $\Theta$ there exists positive integer $n_0$ such that a family \{|$\Delta_n(\theta)|^2; n \geq n_0, \theta \in K$\} of random variables is uniformly integrable.

Proof. Let $\varepsilon > 0$ be given and $K$ be a compact subset of $\Theta$. According to the assumption (3), for sufficiently large $M > 0$
$$\sup_{\theta \in K} \int_{|y| \geq M} |y|^2 \lambda_\theta(dy) < \frac{\varepsilon}{3}.$$ 

Applying Theorem 8 in Ibragimov and Has'minskii [1] from assumptions (1) and (2), there exist $M_0 > M$ and sufficiently large $n_0$ such that for $n \geq n_0$
$$\sup_{\theta \in K} \left| \int_{|\Delta_n(\theta)| < M_0} |\Delta_n(\theta)|^2 dP_{\theta,n} - \int_{|y| < M_0} |y|^2 \lambda_\theta(dy) \right| < \frac{\varepsilon}{3}.$$ 

These inequalities together with assumption (4) imply that
$$\sup_{n \geq n_0} \sup_{\theta \in K} \int_{|\Delta_n(\theta)| \geq M_0} |\Delta_n(\theta)|^2 dP_{\theta,n} < \varepsilon,$$

which is the desired result.

Let $X_1, \ldots, X_n$ be independent and identically distributed sample. In this case, it is well known that the distribution of $\Delta_n(\theta)$ under $P_{\theta,n}$ converges weakly to the normal distribution $N(0, i_1(\theta; A_1))$ uniformly on compact sets of $\theta$ under certain regularity conditions. Recall that the matrix $i_1(\theta; A_1)$ is the Fisher information contained in the sample of size 1. If $i_1(\theta; A_1)$ is positive definite and continuous on $\Theta$, then it is easy to see that all the conditions stated in Lemma 4.1 are satisfied.
Lemma 4.2 If the assumptions of Lemma 4.1 are satisfied and $\Delta_n(\theta) - \bar{\Delta}_n(\theta)$ tends to zero in $P_{\theta,n}$ probability uniformly on compact sets of $\theta$ as $n \to \infty$, then for every compact subset $K$ of $\Theta$ there exists a positive integer $n_0$ such that a family of random variables $\{\Delta_n(\theta) - \bar{\Delta}_n(\theta); n \geq n_0, \theta \in K\}$ is uniformly integrable.

Proof. According to Lemma 4.1, it suffices to prove the uniform integrability for a family $\{\Delta_n(\theta)|^2; n \geq n_0, \theta \in K\}$ with sufficiently large $n_0$. By Jensen's inequality for conditional expectations, for $0 < \eta < M$

$$\int_{|\Delta_n(\theta)| \geq M} |\Delta_n(\theta)|^2 dP_{\theta,n}$$

$$\leq \int_{|\Delta_n(\theta)| \geq M} |\Delta_n(\theta)|^2 dP_{\theta,n} + \int_{|\Delta_n(\theta)| - |\Delta_n(\theta)| > \eta} |\Delta_n(\theta)|^2 dP_{\theta,n}$$

$$\leq \int_{|\Delta_n(\theta)| \geq M - \eta} |\Delta_n(\theta)|^2 dP_{\theta,n} + \int_{|\Delta_n(\theta)| > M} |\Delta_n(\theta)|^2 dP_{\theta,n} + \int_{\Delta_n(\theta) - |\Delta_n(\theta)| < \eta} |\Delta_n(\theta)|^2 dP_{\theta,n}$$

$$\leq 2\int_{|\Delta_n(\theta)| \geq M - \eta} |\Delta_n(\theta)|^2 dP_{\theta,n} + M^2 P_{\theta,n} \{ |\Delta_n(\theta) - \bar{\Delta}_n(\theta)| > \eta \}.$$ 

This completes the proof of the lemma because of the assumptions and Lemma 4.1.

Lemma 4.2 together with inequality (14) yields the following proposition.

Proposition 4.1 Under the assumptions of Lemma 4.1, the following statements are equivalent:
(a) $\lim_{n \to \infty} \frac{1}{n} \|i_n(\theta; A_n) - i_n(\theta; B_n)\| = 0$ uniformly on compact sets of $\theta$.
(b) $\Delta_n(\theta) - \bar{\Delta}_n(\theta) \to 0$ in $P_{\theta,n}$ probability uniformly on compact sets of $\theta$ as $n \to \infty$.

We introduce a uniform version of the concept of local asymptotic normality due to LeCam [3].

Definition 4.2 The family $\{P_{\theta,n}; \theta \in \Theta\}$ of probability measures is said to satisfy local asymptotic normality if the likelihood ratio has the following representation:

$$h_n(x; \theta + t/\sqrt{n}, \theta) = \exp\{t'\Delta_n(\theta) - \frac{1}{2} t'T(\theta)t + Z_n(\theta, t)\}$$
for $\theta \in \Theta$ and $\theta + t/\sqrt{n} \in \Theta$, where $\Delta_n(\theta)$ and $Z_n(\theta, t)$ satisfy that (1) the distribution of $\Delta_n(\theta)$ under $P_{\theta,n}$ converges weakly to a normal distribution $N(0, \Gamma(\theta))$ uniformly on compact sets of $\theta$ as $n \to \infty$, where the covariance matrix $\Gamma(\theta)$ is continuous on $\Theta$, (2) $Z_n(\theta, t)$ converges to 0 in $P_{\theta,n}$ probability uniformly on compact sets of $\theta$ and bounded sets of $t$ as $n \to \infty$.

The following proposition gives a relation between the asymptotic sufficiency in the sense of Fisher information and the asymptotic Bayes risk sufficiency.

**Proposition 4.2** Suppose that the family $\{ P_{\theta,n}; \theta \in \Theta \}$ satisfies local asymptotic normality. Moreover, assume that the following conditions are fulfilled:

1. For every $b > 0$ and every compact subset $K$ of $\Theta$

$$
\sup_{\theta \in K} \sup_{|t| \leq b} \sup_{n \in \mathbb{N}} \int_X \exp\{t'\Delta_n(\theta)\} dP_{\theta,n} < \infty.
$$

2. There exists $s > 2$ such that for every compact subset $K$ of $\Theta$

$$
\sup_{\theta \in K} \sup_{|t| \leq b} \sup_{n \in \mathbb{N}} \int_X \exp\{sZ_n(\theta, t)\} dP_{\theta,n} < \infty.
$$

If $\Delta_n(\theta) - \bar{\Delta}_n(\theta)$ converges to zero in $P_{\theta,n}$ probability uniformly on compact sets of $\theta$ as $n \to \infty$, then for any $\varepsilon > 0$ and $b > 0$

$$
\lim_{n \to \infty} \sup_{\theta \in K} \sup_{|t| \leq b} \left\{ |h_n(x; \tau, \theta) - \bar{h}_n(x; \tau, \theta)| > \varepsilon \right\} = 0,
$$

where $U_n(\theta; b) = \{ \tau \in \Theta; \sqrt{n} |\theta - \tau| \leq b \}$.

Proof. Let $\varepsilon > 0$, $b > 0$ and $K$ be a compact subset of $\Theta$. Noting that

$$
\bar{h}_n(x; \tau, \theta) = E_{P_{\theta,n}}[h_n(x; \tau, \theta)|B_n],
$$

it follows from the local asymptotic normality condition that

$$
\bar{h}_n(x; \theta + t/\sqrt{n}, \theta) = \exp\{t'\Delta_n(\theta) - \frac{1}{2} t'\Gamma(\theta) t\} E_{P_{\theta,n}}[\exp\{t'\bar{\Delta}_n + \bar{Z}_n\}|B_n]
$$

for $\theta \in K$ and $|t| \leq b$, where $\Delta_n = \Delta_n(\theta) - \bar{\Delta}_n(\theta)$ and $\bar{Z}_n = Z_n(\theta, t)$ for notational convenience. Since $\{\Delta_n(\theta); n \geq n_0, \theta \in K\}$ is stochastically bounded for sufficiently large $n_0$ because of Theorem 8 in [1], so is $\{\Delta_n; n \geq n_0, \theta \in K\}$. Hence, it is enough to show that $E_{P_{\theta,n}}[\exp\{t'\Delta_n + \bar{Z}_n\}|B_n] \to 1$ in $P_{\theta,n}$ probability uniformly in $\theta \in K$ and $|t| \leq b$ as $n \to \infty$. Choose $\delta > 0$
such that \(|\exp(x) - 1| < \varepsilon\) for \(|x| < \delta\). We have for \(0 < \delta < M\) and \(r\) with \(2 < r < s\)

\[
\int_X [\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 dP_{\theta,n} \\
= \int_{|t'\tilde{\Delta}_n + \tilde{Z}_n| < \delta} [\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 dP_{\theta,n} \\
+ \int_{\delta \leq |t'\tilde{\Delta}_n + \tilde{Z}_n| < M} [\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 dP_{\theta,n} \\
+ \int_{t'\tilde{\Delta}_n + \tilde{Z}_n \leq -M} [\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 dP_{\theta,n} \\
+ \int_{t'\tilde{\Delta}_n + \tilde{Z}_n \geq M} [\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 dP_{\theta,n} \\
\leq \varepsilon^2 + \varepsilon^2 P_{\theta,n} \{ |t'\tilde{\Delta}_n + \tilde{Z}_n| \geq \delta \} + P_{\theta,n} \{ |t'\tilde{\Delta}_n + \tilde{Z}_n| \geq M \} \\
+ \varepsilon^{-M(r-2)} \int_X \exp(r(t'\tilde{\Delta}_n + \tilde{Z}_n)) dP_{\theta,n}.
\]

By virtue of conditions (1) and (2), the integral in the last term of the inequality above is bounded in \(\theta \in K, |t| \leq b\) and \(n \in N\). This implies

\[
\lim_{n \to \infty} \sup_{\theta \in K} \sup_{|t| \leq b} E_{P_{\theta,n}}[\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 = 0,
\]

whence

\[
\sup_{\theta \in K} \sup_{|t| \leq b} P_{\theta,n} \{ |E_{P_{\theta,n}}[\exp(t'\tilde{\Delta}_n + \tilde{Z}_n)|B_n| - 1| > \varepsilon \} \\
\leq \frac{1}{\varepsilon^2} \sup_{\theta \in K} \sup_{|t| \leq b} E_{P_{\theta,n}}[\exp(t'\tilde{\Delta}_n + \tilde{Z}_n)|B_n| - 1]^2 \\
\leq \frac{1}{\varepsilon^2} \sup_{\theta \in K} \sup_{|t| \leq b} E_{P_{\theta,n}}[\exp(t'\tilde{\Delta}_n + \tilde{Z}_n) - 1]^2 \\
\to 0 \quad \text{as} \ n \to \infty.
\]

It is easy to see that the exponential family stated in Example 3.2 satisfies the conditions (1) and (2) in Proposition 4.2.

5 Relation between asymptotic sufficiency and Bayes risk sufficiency

The definitions of asymptotic sufficiency and Bayes risk sufficiency are given by Definition 1.1 and Definition 1.2, respectively. It is easy to show in the
similar way stated in Section 2 that if \( \{ B_n \} \) is asymptotically sufficient for \( \{ A_n \} \) up to \( o(a_n) \) then \( \{ B_n \} \) is asymptotically Bayes risk sufficient for \( \{ A_n \} \) up to \( o(a_n) \). In this section, we shall give a sufficient condition which guarantees the reverse relation. Here it is not assumed that \( P_n \) has the same support.

We begin with the following lemma.

**Lemma 5.1** Let \( 0 < \varepsilon \leq 1 \leq M < \infty \). Then for any positive integer \( m \)

\[
\int_{|h_n - \tilde{h}_n| > \varepsilon, h_n \leq M, \tilde{h}_n \leq M} |h_n - \tilde{h}_n| dP_{\theta, n} \\
\leq 16 M M^{(2m+1)/2m-1} \varepsilon^{-2m/(2m-1)} \sup_{c > 0} R_n(c, \tau, \theta),
\]

where \( h_n = h_n(x; \tau, \theta) \) and \( \tilde{h}_n = \tilde{h}_n(x; \tau, \theta) \).

Proof. Divide the left side of the inequality in the lemma into the following parts \( I_1 \) and \( I_2 \):

\[
I_1 = \sum_{i=1}^{m-1} \int_{A_i, h_n \leq M, \tilde{h}_n \leq M} |h_n - \tilde{h}_n| dP_{\theta, n}, \\
I_2 = \int_{A_m, h_n \leq M, \tilde{h}_n \leq M} |h_n - \tilde{h}_n| dP_{\theta, n},
\]

where

\[
A_i = \{ x; M^{2m^{-1}-1} \varepsilon^{2m-2m^{-1}} < |h_n - \tilde{h}_n| \leq M^{2m^{-1}-1} \varepsilon^{2m-2m^{-1}}, i \leq m - 1, \\
A_m = \{ x; |h_n - \tilde{h}_n| > M^{2m^{-1}-1} \varepsilon^{2m-2m^{-1}} \}.
\]

According to Lemma 3.2 we have

\[
I_1 \leq 16(m - 1) M^{(2m+1)/2m-1} \varepsilon^{-2m/(2m-1)} \sup_{c > 0} R_n(c, \tau, \theta), \\
I_2 \leq 16 M^{(2m+1)/2m-1} \varepsilon^{-2m/(2m-1)} \sup_{c > 0} R_n(c, \tau, \theta),
\]

which leads to the desired result.

The log-likelihood ratio function is defined by

\[
\Lambda_n(x; \tau, \theta) = \log h_n(x; \tau, \theta). \quad (15)
\]

We state a result about higher order asymptotic sufficiency in a local sense.
Proposition 5.1 Suppose that there exist sequences \( \{\theta_n\} \) and \( \{U_n\} \) with \( \theta_n \in U_n \subset \Theta \) satisfying that

(a) \[ \sup_{\tau \in U_n} P_{\tau,n} \{ A_n(x; \tau, \theta_n) \geq \delta \log n \} = o(n^{-\gamma}), \]

(b) \[ \sup_{\tau \in U_n} P_{\tau,n} \{ S_n(\theta_n)^c \} = o(n^{-\gamma}), \]

(c) \[ \sup_{c > 0} \sup_{\tau \in U_n} R_n(c, \tau, \theta_n) = o(n^{-\alpha}), \]

where \( \alpha, \delta \) and \( \gamma \) are positive constants. If \( \alpha > 2\delta \), then there exists a sequence of families of probability measures \( Q_n = \{ Q_{\theta,n}; \theta \in \Theta \}, n \in N, \) such that

(1) for each \( n \in N \) and \( \tau \in \Theta \), \( dQ_{\tau,n}/d\mu_n = q_n(x; \tau, \theta_n)p_n(x, \theta_n) \), where \( q_n(x, \tau, \theta_n) \) is \( \mathcal{B}_n \)-measurable function,

(2) for every positive number \( \beta \) satisfying \( \beta < \alpha/2 - \delta \) and \( \beta \leq \gamma \)

\[ \sup_{\tau \in U_n} \| P_{\tau,n} - Q_{\tau,n} \| = o(n^{-\beta}). \]

Proof. Let \( 0 < \beta < \alpha/2 - \delta \) and \( \beta \leq \gamma \). Define \( \mathcal{B}_n \)-measurable function by

\[ \tilde{q}_n(x; \tau, \theta_n) = I_{W_n,\tau}(x)h_n(x; \tau, \theta_n), \]

where \( W_n,\tau = \{ x; h_n(x; \tau, \theta_n) \leq 2n^2 \} \) and \( I_W(\cdot) \) means the indicator function of a set \( W \). For \( \tau \in \Theta \) and \( n \in N, \) we denote by \( \tilde{Q}_{\tau,n} \) the measure given by

\[ \tilde{Q}_{\tau,n}/d\mu_n = \tilde{q}_n(x; \tau, \theta_n)p_n(x, \theta_n). \]

Then we have

\[ \| P_{\tau,n} - \tilde{Q}_{\tau,n} \| = \int_{\mathcal{X}} | p_n(x, \tau) - \tilde{q}_n(x; \tau, \theta_n)p_n(x, \theta_n) | d\mu_n \]

\[ = \int_{W_{n,\tau}} | h_n - \tilde{h}_n | dP_{\theta_n,n} + \int_{W_{n,\tau}^c} h_n dP_{\theta_n,n} + P_{\tau,n} \{ S_n(\theta_n)^c \} \]

\[ = I_1 + I_2 + P_{\tau,n} \{ S_n(\theta_n)^c \} \) (say), \] (16)

where \( h_n = h_n(x; \tau, \theta_n) \) and \( \tilde{h}_n = \tilde{h}_n(x; \tau, \theta_n) \). Choose positive integer \( m \) such that \( \xi = (2^m - 1)\alpha/(2^{m+1}) - \delta > \beta \). In order to estimate the first term \( I_1 \) on the right side of (16), observe that

\[ I_1 \leq \int_{|h_n - \tilde{h}_n| \leq \xi} | h_n - \tilde{h}_n | dP_{\theta_n,n} + \int_{h_n \leq 2n^\delta, h_n > 2n^\delta} | h_n - \tilde{h}_n | dP_{\theta_n,n} \]

\[ + \int_{|h_n - \tilde{h}_n| > \xi, h_n \leq 2n^\delta, h_n \leq 2n^\delta} | h_n - \tilde{h}_n | dP_{\theta_n,n} \]

\[ = I_{11} + I_{12} + I_{13} \) (say). \]
It is obvious that
\[ \sup_{\tau \in \mathcal{U}_n} I_{11} = o(n^{-\beta}). \]

By the assumption (a)
\[ \sup_{\tau \in \mathcal{U}_n} I_{12} \leq \sup_{\tau \in \mathcal{U}_n} \int_{h_n > 2n^\delta} h_n dP_{\theta_n,n} \leq \sup_{\tau \in \mathcal{U}_n} P_{\tau,n}(\Lambda_n(x; \tau, \theta) \geq \delta \log n) = o(n^{-\gamma}). \]

Taking \( \varepsilon = n^{-\delta}, M = 2n^\delta \) in Lemma 5.1 yields
\[ \sup_{\tau \in \mathcal{U}_n} I_{13} = o(n^{-(\alpha/2 - \delta)}) \]
and so, combining these estimates,
\[ \sup_{\tau \in \mathcal{U}_n} I_1 = o(n^{-\beta}). \]

Next, in order to estimate \( I_2 \) we put
\[ I_2 = \int_{h_n < n^\delta, \bar{h}_n > 2n^\delta} h_n dP_{\theta_n,n} + \int_{h_n > n^\delta, \bar{h}_n > 2n^\delta} h_n dP_{\theta_n,n} = I_{21} + I_{22} \text{ (say)}. \]

It follows Lemma 3.1 that
\[ \sup_{\tau \in \mathcal{U}_n} I_{21} \leq n^\delta \sup_{\tau \in \mathcal{U}_n} P_{\theta_n,n}(h_n < n^\delta < 2n^\delta < \bar{h}_n) = o(n^{-(\alpha - \delta)}). \]

Also, a similar way to estimate \( I_{12} \) shows
\[ \sup_{\tau \in \mathcal{U}_n} I_{22} = o(n^{-\gamma}). \]

Thus we have
\[ \sup_{\tau \in \mathcal{U}_n} I_2 = o(n^{-\beta}). \]

It follows from these estimates and the assumption (b) that
\[ \sup_{\tau \in \mathcal{U}_n} \|P_{\tau,n} - \tilde{Q}_{\tau,n}\| = o(n^{-\beta}). \quad (17) \]
If $\bar{Q}_{\tau,n}$ is not a probability measure, then define the probability measure $Q_{\tau,n}$ having density $q_n(x;\tau,\theta_n)p_n(x,\theta_n)$ with respect to $\mu_n$ such that

$$q_n(x;\tau,\theta_n) = \begin{cases} \tilde{q}_n(x;\tau,\theta_n)/a_n(\tau,\theta_n), & \text{if } a_n(\tau,\theta_n) > 0, \\ 1, & \text{if } a_n(\tau,\theta_n) = 0, \end{cases}$$  \hspace{1cm} (18)

where $a_n(\tau,\theta_n) = \int_X \tilde{q}_n(x;\tau,\theta_n)d\mu_{\theta_n}$.

In other words, $Q_{\tau,n}$ means the normalizing of $\bar{Q}_{\tau,n}$. It is easy to see that the sequence $\{Q_{\tau,n}\}$ satisfies the property (2) in the proposition because of (17). This completes the proof.

It is noted that for every $n \in N$, $B_n$ is sufficient $\sigma$-field for the family $Q_n$ by the Neyman factorization theorem.

We shall impose the following Conditions (A), (B) and (C). Let $\gamma > 0$ be a positive number.

(A) There exists a sequence $\{\hat{\theta}_n\}$ of $B_n$-measurable estimates having the following property: There exist $c > 0$ and positive definite symmetric matrices $J(\theta)$ such that for every compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} P_{\theta,n}\{\sqrt{n}|J(\theta)^{1/2}(\hat{\theta}_n - \theta)| \geq c\sqrt{\log n}\} = o(n^{-\gamma}),$$

where $J(\theta) = (J(\theta)^{1/2}J(\theta)^{1/2})$. Moreover, it is assumed that for every compact subset $K$ of $\Theta$

$$0 < \inf_{\theta \in K} \min_{|u|=1} u'J(\theta)u \leq \sup_{\theta \in K} \max_{|u|=1} u'J(\theta)u < \infty.$$  \hspace{1cm} (19)

(B) For $\theta \in \Theta$ and $a > 0$ let $V_n(\theta; a) = \{\tau \in \Theta; \sqrt{n}|J(\tau)^{1/2}(\tau - \theta)| \leq a\sqrt{\log n}\}$. For any $a > 0$ there exists $\delta(a, \gamma) > 0$ such that for any $\delta > \delta(a, \gamma)$ and any compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} \sup_{\tau \in V_n(\theta; a)} P_{\tau,n}\{\Lambda_n(x;\tau,\theta) \geq \delta \log n\} = o(n^{-\gamma}).$$

Here $\delta(a, \gamma)$ is strictly increasing, continuous in $a$ and $\lim_{a \to \infty} \delta(a, \gamma) = \infty$.

(C) For every $a > 0$ and every compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} \sup_{\tau \in V_n(\theta; a)} P_{\tau,n}\{S_n(\theta)^a\} = o(n^{-\gamma}).$$

**Theorem 5.1** Suppose that Conditions (A)–(C) hold and that for any $a > 0$ and any compact subset $K$ of $\Theta$

$$\sup_{\theta \in K} \sup_{\tau \in V_n(\theta; a)} \sup_{c > 0} R_n(c, \tau, \theta) = o(n^{-a}).$$
If \( \alpha > 2 \delta(c, \gamma) \), then \( \{ B_n \} \) is asymptotically sufficient for \( \{ A_n \} \) up to \( o(n^{-\beta}) \) for every positive number \( \beta \) satisfying \( \beta < \alpha/2 - \delta(c, \gamma) \) and \( \beta \leq \gamma \).

Proof. Let \( 0 < \beta < \alpha/2 - \delta(c, \gamma) \) and choose \( \delta > \delta(c, \gamma) \) such that \( \beta < \alpha/2 - \delta \).

From Condition (B), there exists \( a > c \) such that \( \delta = \delta(a, \gamma) \). For each \( n \in N \) we divide the Euclidean space \( R^k \) into countably many disjoint cubes \( \{ V_n(\theta_{ni}); i \in N \} \) where

\[
V_n(\theta_{ni}) = \{ \theta; (\theta_{ni})_j - \frac{a - c}{\sqrt{nkJ}} \sqrt{\log n} < \theta_j \leq (\theta_{ni})_j + \frac{a - c}{\sqrt{nkJ}} \sqrt{\log n}; j \leq k \}
\]

with \( J = \sup_{\tau \in K} \max_{|h| = 1} u^T \Gamma(\tau) u, 0 < J < \infty \). Here \( (\theta_{ni})_j \) denotes the j-th coordinate of \( \theta_{ni} \). Let \( \theta(t) \) be a measurable map from \( R^k \) to \( \Theta \) such that

\[
\theta(t) = \begin{cases} 
  t, & \text{if } t \in \Theta, \\
  \theta_0, & \text{if } t \in \Theta^c \ (\theta_0 \in \Theta \text{ is fixed}).
\end{cases}
\]

Let \( \hat{\theta}_n(x) = \theta_{ni} \) if \( \hat{\theta}_n(x) \in V_n(\theta_{ni}) \). Then define

\[
\theta_n(x) = \theta(\hat{\theta}_n(x)).
\]

Obviously \( \theta_n(x) \) is a \( B_n \)-measurable map from \( \mathcal{X} \) into \( \Theta \). Let \( D_n(\tau) = \{ x; \sqrt{n} |J(\tau)|^{1/2}(\hat{\theta}_n - \tau) | < c \sqrt{\log n} \} \) and define a measure \( Q_{\tau,n}^* \) having a density \( q_n^*(x; \tau, \theta_n(x)) \) with respect to \( \mu_n \) such that

\[
q_n^*(x; \tau, \theta_n(x)) = I_{D_n(\tau)}(x) q_n(x; \tau, \theta_n(x)) p_n(x, \theta_n(x)),
\]

where \( q_n(x; \tau, \theta_n(x)) \) is given by (18). Then we have

\[
\sup_{\tau \in K} \| P_{\tau,n} - Q_{\tau,n}^* \| \leq \sup_{\tau \in K} \int_{D_n(\tau)} |p_n(x, \tau) - q_n(x; \tau, \theta_n(x)) p_n(x, \theta_n(x))| d\mu_n
\]

\[
+ \sup_{\tau \in K} P_{\tau,n} \{ D_n(\tau)^c \}.
\]

Since Condition (A) implies the second term on the right side of the inequality above is \( o(n^{-\beta}) \), it remains to estimate the first term. Let \( I_1 \) be the first term and \( I_n(\tau) = \{ i \in N; \tau \in V_n(\theta_{ni}; a) \} \). Choose \( \eta > 0 \) such that \( K_\eta = \{ \theta \in R^k; \inf_{\tau \in K} |\theta - \tau| \leq \eta \} \subset \Theta \). If \( \tau \in K \) and \( x \in D_n(\tau) \), then there exists \( i \in N \) such that \( i \in I_n(\tau) \) and \( \hat{\theta}_n(x) \in V_n(\theta_{ni}) \). By virtue of Condition (A), \( \{ i \in N; i \in I_n(\tau) \} \) is independent of \( \tau \in K \) and finite. If \( d_K \) denotes this number, then for sufficiently large \( n \)

\[
I_1 \leq \sup_{\tau \in K} \sum_{i \in I_n(\tau)} \int_{\hat{\theta}_n(x) \in V_n(\theta_{ni})} |p_n(x, \tau) - q_n(x; \tau, \theta_n(x)) p_n(x, \theta_n(x))| d\mu_n
\]
\[ d_K \sup_{\tau \in K} \sup_{i \in I_n(\tau)} \int_{\hat{\delta}_n(x) \in V_n(\delta_{ni})} \left| p_n(x, \tau) - q_n(x; \tau, \theta_n(x)) p_n(x, \theta_n(x)) \right| d\mu_n \]

\[ \leq d_K \sup_{\theta \in K} \sup_{\tau \in V_n(\theta; \alpha)} \int_X \left| p_n(x, \tau) - q_n(x; \tau, \theta) p_n(x, \theta) \right| d\mu_n. \]

By Proposition 5.1 we have \( I_1 = o(n^{-\beta}) \), which implies \( \sup_{\tau \in K} \| P_{\tau,n} - Q_{\tau,n}^* \| = o(n^{-\beta}) \). Finally, if \( Q_{\tau,n} \) is defined by the normalizing of \( Q_{\tau,n}^* \), it follows from the Neyman factorization theorem that \( B_n \) is sufficient for \( Q_n \) for each \( n \). This completes the proof.

We shall consider about Condition (B) in the independent and identically distributed case with the same support. Let \( f_n(x, \theta) = \log p_n(x, \theta) \). Denote the \( m \)-th derivative relative to \( \theta \) of \( f_n(x, \theta) \) by

\[ f_n^{(m)}(x, \theta) = \left( \frac{\partial^m}{\partial \theta_{i_1} \cdots \partial \theta_{i_m}} f_n(x, \theta) \right) ; \quad i_1, \ldots, i_m \in \{1, \ldots, k\} \).

The Euclidean norm \( \| \cdot \| \) of \( f_n^{(m)}(x, \theta) \) is defined by

\[ \| f_n^{(m)}(x, \theta) \|^2 = \sum_{i_1, \ldots, i_m=1}^k \left( \frac{\partial^m}{\partial \theta_{i_1} \cdots \partial \theta_{i_m}} f_n(x, \theta) \right)^2. \]

For any \( \tau = (\tau_1, \ldots, \tau_k) \in \Theta \) define

\[ f_n^{(m)}(x, \theta)^\tau_m = \sum_{i_1, \ldots, i_m=1}^k \frac{\partial^m}{\partial \theta_{i_1} \cdots \partial \theta_{i_m}} f_n(x, \theta) \prod_{p=1}^m \pi_{i_p}. \]

It is assumed that the log-likelihood ratio function has the following Taylor expansion:

\[ \Lambda_n(x; \tau, \theta) = - \sum_{m=1}^2 \frac{1}{m!} f_n^{(m)}(x, \tau)(\theta - \tau)^m - \frac{1}{6} f_n^{(3)}(x, \theta^*)(\theta - \tau)^3, \quad (20) \]

where \( \max(|\theta^* - \theta|, |\theta^* - \tau|) \leq |\theta - \tau| \). Moreover, suppose that the following assumptions (a)–(c) are satisfied:

(a) There exists \( c(\gamma, k) > 0 \) such that for every compact subset \( K \) of \( \Theta \)

\[ \sup_{\tau \in K} P_{\tau,n} \{ |(i_1(\tau; A_1)^{1/2} y) f_n^{(1)}(x, \tau)| \geq c(\gamma, k) \sqrt{n \log n} \} = o(n^{-\gamma}), \]

where \( 0 < \inf_{\tau \in K} \min_{|u|=1} u_i(\tau; A_1) u \).
(b) For every compact subset $K$ of $\Theta$ there exists $c_K > 0$ such that
\[
\sup_{r \in K} P_{\tau,n}\{||f_n^{(2)}(x, \tau) - i_1(\tau; A_1)|| \geq c_K \sqrt{n \log n}\} = o(n^{-\gamma}).
\]

(c) Let $V_n^*(\theta; a) = \{\tau \in \Theta; \sqrt{n}|i_1(\tau; A_1)^{1/2}(\tau - \theta)| \leq a \sqrt{\log n}\}$. For every $a > 0$ and every compact subset $K$ of $\Theta$ there exists $d_{a, K} > 0$ such that
\[
\sup_{\theta \in K} \sup_{r \in V_n^*(\theta; a)} P_{\tau,n}\{||f_n^{(3)}(x, \theta)|| \geq nd_{a, K}\} = o(n^{-\gamma}).
\]

Then it follows from (20) that the relation (19) holds with $J(\tau) = i_1(\tau; A_1)$ and $\delta(a, \gamma) = a\alpha(\gamma, k) + \frac{a^2}{2}$. An estimate for $c(\gamma, 1)$ is given by Lemma 1 in Pfanzagl [9], where $c(\gamma, 1) = \sqrt{2\gamma}$.

Suzuki [13] showed the equivalence of order $o(1)$ between asymptotic sufficiency and Bayes risk sufficiency. In [14] he also investigated the problem in higher order case under a local situation. Theorem 5.1 gives a solution for his open problem in a non-local situation.

Acknowledgements
This work has been done during a visit to the Department of Statistics at Stanford University. I wish to express my hearty thanks to Professor Bradley Efron for giving me this opportunity. I am also grateful to Professor Takeru Suzuki for helpful comments.

References


