Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 1, pp. 57-83
Bangladesh

ASYMPTOTICALLY OPTIMAL TESTS UNDER A GENERAL DEPENDENCE SET-UP

George G. Roussas
Department of Statistics, University of California, Davis
Email: ggroussas@ucdavis.edu

Debasis Bhattacharya
Department of Statistics, University of California, Davis
Email: debasis_us@yahoo.com

summary

Let the random variables $X_0, X_1, \ldots, X_n$ be $(n+1)$ observations from a general discrete parameter stochastic process $\{X_n\}$, $n \geq 0$, whose probability laws are of a known functional form, but dependent on a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^k$, $k \geq 1$. Asymptotically optimal tests for testing a null hypothesis $H_0: \ \theta = \theta_0$ against a composite alternative for Locally Asymptotically Normal (LAN) and Locally Asymptotically Mixture of Normal (LAMN) models are derived, using the results on asymptotic expansion of the log-likelihood ratio statistic (in the probability sense), its asymptotic distribution, asymptotic distribution of certain random quantities which are closely related to the log-likelihood ratios, and an exponential approximation result on the log-likelihood ratio statistic. The concepts of contiguity, differentially equivalent probability measures and differentially sufficient statistics play a key role in deriving the results. The testing hypothesis problem is restricted to the case that $k = 1$, although all other underlying results hold for $k \geq 1$. The general case $(k \geq 1)$ will be discussed elsewhere.

Keywords and phrases: asymptotic distribution, asymptotic expansion, exponential approximation, local asymptotic normality, local asymptotic mixed normality, differentially sufficient statistics, asymptotically optimal tests.

AMS Classification: Primary 62H15, 62G20, 62F05; Secondary 62E20, 62M99.

1 Introduction

Let $X_0, X_1, \ldots, X_n$ be $(n+1)$ observations form a general discrete parameter stochastic process $\{X_n\}$, $n \geq 0$. The random variables (r.v.'s) X_n are defined on the probability space $(\mathcal{X}, \mathcal{A}, P_\theta)$, $\theta \in \Theta$, where Θ is an open subset of $\mathbb{R}^k$, $k \geq 1$, and takes values in $(S, \mathcal{S})$, where S is a Borel subset of Euclidean space and $\mathcal{S}$ is the σ-field of Borel subsets of S. Let $\mathcal{A}_n$

be the σ-field induced by the r.v.'s $X_0, X_1, \ldots, X_n$, $\mathcal{A}_n = \sigma(X_0, X_1, \ldots, X_n)$, and let $P_{n,\theta}$ be the restriction of P_θ to $\mathcal{A}_n$. The process considered needs not be even stationary, in the strict or in the wide sense of the term. The purpose of the paper is to restrict ourselves presently to the case $k = 1$ and derive optimal tests (in an asymptotic sense) for testing the hypothesis $H_0: \ \theta = \theta_0$ against a one-sided alternative. Wald (1941, 1943) considered the problem of deriving optimal tests when the underlying process consists of i.i.d. observations, and the tests formulated by him are based on the maximum likelihood estimate (MLE) of the parameter involved. Moreover, the regularity conditions on the population density used to derive the tests are quite stringent. Later, Johnson and Roussas (1969, 1970) extend the tests formulated by Wald to stationary Markov processes, under substantially weaker conditions on the population density, and the proposed tests need not be based on the maximum likelihood estimate of the parameter. Here in this paper, we further extend their results to the case of general stochastic processes $\{X_n\}$, $n \geq 0$.

Under suitable conditions on the process, the asymptotic expansion of the log-likelihood ratio for general stochastic processes has been obtained in Roussas and Bhattacharya (2007), for $k \geq 1$. In the same paper, it has also been established that, in the neighborhood of θ and for large n, the likelihood function behaves as if it were (approximately) an exponential family. The exponential approximation of the likelihood ratio statistic derived therein, along with some other auxiliary results, will play a key role in deriving the main results of this paper. Let θ_0 be an arbitrary but fixed point in Θ, and let $\Delta_n(\theta_0)$ be a k-dimensional vector defined in terms of the random variables $X_0, X_1, \ldots, X_n$ (see relation (2.6)). Let $\Delta_n^*(\theta_0)$ be a suitable truncated version of $\Delta_n(\theta_0)$ for which the exponential approximation result (stated in Theorem 7) holds, and which plays the all-important role of the statistic appearing in the exponent of an exponential family. Using $\Delta_n^*(\theta_0)$ and for $h \in \mathbb{R}^k$, we define probability measures $R_{n,h}$, $n \geq 0$, as in (3.4). The first two main results that are used in the sequel are the following: (i) the sequences $\{P_{n,\theta}\}$ and $\{R_{n,h}\}$ of probability measures with $h = n^{\frac{1}{2}}(\theta - \theta_0)$, $\theta \in \Theta$, are differentially equivalent (see Definition 5), and (ii) the sequence $\{\Delta_n^*(\theta_0)\}$ is differentially sufficient (see Definition 6) at θ_0 for the family $\{P_{n,\theta}$, $\theta \in \Theta\}$. These results are established in Propositions 2 and 3, respectively. Next, let $\{h_n\}$ be a bounded sequence of h's in $\mathbb{R}^k$ and set $\theta_n = \theta_0 + h_n n^{-\frac{1}{2}}$. Then, for testing hypotheses about θ, we prove Theorem 8 which allows us to consider tests depending on $\Delta_n(\theta_0)$ alone in search of an optimal test (in the sense of maximizing the asymptotic power, say), under the type of alternatives considered above; and Theorem 9, which asserts that one needs not be concerned with tests not based on $\Delta_n(\theta_0)$. In statistical applications, especially for obtaining the power of a test, the asymptotic distribution of $\Delta_n(\theta_0)$ under P_{n,θ_n} is also needed, which is established in Theorem 6.

Throughout the entire paper, we use the following notation: for a vector $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{y}'$ denotes the transpose of $\mathbf{y}$, and for a square matrix D, $|D|$ denotes the determinant of D, $\|D\|$ denotes the norm of D, defined by the square root of the sum of the squares of its elements. The symbol '$\Longrightarrow$' denotes convergence in distribution, whereas the symbol '$\xrightarrow{P}$' denotes convergence in probability. Unless otherwise stated, expectation of a random variable is to be understood under θ. Also, in order to avoid repetitions, it is stated that all the limits are taken as n or a subsequence of $\{n\}$ tend to infinity.

The main theorem established here depends on results derived in Roussas and Bhattacharya (2007), and will be freely cited here.

The paper is organized as follows: Section 2 introduces the technical notation and the relevant assumptions required to develop the main result. In the same section, some remarks on the stated assumptions are made. In the following section, some preliminary results used to derive the main theorem are stated and their justifications are given. In Section 4, the case $k = 1$ is considered and asymptotically optimal tests for testing the hypothesis $\theta = \theta_0$ against a one-sided alternative are constructed. They are discussed in Theorem 10 and Theorem 11.

2 Notation and Assumptions

Consider the first $(n+1)$ r.v.'s $X_0, X_1, \ldots, X_n$, and let $\mathcal{A}_n = \sigma(X_0, X_1, \ldots, X_n)$ be the σ-field induced by $X_0, X_1, \ldots, X_n$. Let $P_{n,\theta}$ be the restriction of P_θ to $\mathcal{A}_n$. It will be assumed that, for $n \geq 0$, the probability measures $P_{n,\theta}$ and P_{n,θ_n} are mutually absolutely continuous for all θ, $\theta_n \in \Theta$. Then the Radon-Nikodym derivative (the likelihood) of P_{n,θ_n} with respect to $P_{n,\theta}$ is

$$\frac{dP_{n,\theta_n}}{dP_{n,\theta}} = L_n(\mathbf{X}_n;\theta,\theta_n) = L_n(\theta,\theta_n) = q_n(\mathbf{X}_n;\theta,\theta_n) = q_n(\theta,\theta_n), \text{ say,} \tag{2.1}$$

where $\mathbf{X}_n = (X_0, X_1, \ldots, X_n)$.

For $n \geq 1$, set

$$\phi_n(\mathbf{X}_n;\theta,\theta_n) = \phi_n(\theta,\theta_n) = \left[\frac{q_n(\theta,\theta_n)}{q_{n-1}(\theta,\theta_n)}\right]^{\frac{1}{2}} = \left[q_n(X_n \mid \mathbf{X}_{n-1};\theta,\theta_n)\right]^{\frac{1}{2}}, \tag{2.2}$$

so that

$$L_n(\theta,\theta_n) = \left[\prod_{j=1}^{n} q_j(X_j \mid \mathbf{X}_{j-1};\theta,\theta_n)\right] q_0(X_0;\theta,\theta_n) = q_0(\theta,\theta_n)\prod_{j=1}^{n}\phi_j^2(\theta,\theta_n), \tag{2.3}$$

and

$$\Lambda_n(\theta,\theta_n) = \log L_n(\theta,\theta_n) = \log q_0(\theta,\theta_n) + 2\sum_{j=1}^{n}\log\phi_j(\theta,\theta_n). \tag{2.4}$$

Clearly, $\Lambda_n(\theta,\theta_n)$ is well-defined with P_θ-probability 1 for all $\theta, \theta_n \in \Theta$.

It will be assumed in the following that, for each $\theta \in \Theta$, the random functions $\phi_j(\theta;\cdot)$, $j \geq 1$, are differentiable in quadratic mean (q.m.) at θ when the probability measure P_θ is used. Let $\dot{\phi}_j(\theta)$, $j \geq 1$, be the derivatives in q.m. involved. Next, set

$$\Gamma_j(\theta) = 4\mathcal{E}_\theta\left[\dot{\phi}_j(\theta)\dot{\phi}_j'(\theta)\right], \; j \geq 1, \;\; \overline{\Gamma}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}\Gamma_j(\theta), \tag{2.5}$$

and

$$\Delta_n(\theta) = 2n^{-\frac{1}{2}} \sum_{j=1}^{n} \dot{\phi}_j(\theta), n \geq 1. \tag{2.6}$$

The assumptions listed below are taken from Roussas and Bhattacharya (2007).

Assumptions

(A1) For each $n \geq 0$, the (finite-dimensional) probability measures $\{P_{n,\theta};\ \theta \in \Theta\}$ are mutually absolutely continuous.

(A2) (i) For each $\theta \in \Theta$, the random functions $\phi_j(\theta;\cdot)$ are differentiable in q.m. $[P_\theta]$ uniformly in $j \geq 1$. That is, there are k-dimensional random vectors $\dot{\phi}_j(\theta)$, the q.m. derivatives of $\phi_j(\theta,\theta^*)$ with respect to θ^* at θ, such that

$$\frac{1}{\lambda}\left| \, [\phi_j(\theta;\theta+\lambda h) - 1] - \lambda h'\dot{\phi}_j(\theta)\right| \to 0 \tag{2.7}$$

in q.m. $[P_\theta]$, as $\lambda \to 0$, uniformly on bounded sets of $h \in \mathbb{R}^k$ and uniformly in $j \geq 1$.

(ii) For $j \geq 1$, $\dot{\phi}_j(\theta)$ is $\mathcal{A}_j \times \mathcal{C}$-measurable, where $\mathcal{C}$ is the σ-field of Borel subsets of Θ.

(A3) (i) For each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$, $\left[t'\dot{\phi}_j(\theta)\right]^2$, $j \geq 1$, are uniformly integrable with respect to P_θ. That is, uniformly in $j \geq 1$,

$$\int_{\{\ [t'\dot{\phi}_j(\theta)]^2 > \ a\}} \left[t'\dot{\phi}_j(\theta)\right]^2 \quad dP_\theta \quad \to 0, \text{ as } a \to \infty. \tag{2.8}$$

(ii) For each $\theta \in \Theta$ and $n \geq 1$, let the $k \times k$ covariance matrix $\overline{\Gamma}_n(\theta)$ be defined by (2.5). Then $\overline{\Gamma}_n(\theta) \to \overline{\Gamma}(\theta)$ (in any one of the standard norms in $\mathbb{R}^k$) and $\overline{\Gamma}(\theta)$ is positive definite, for each $\theta \in \Theta$.

(iii) For each $\theta \in \Theta$ and for the probability measure P_θ, the WLLN holds for the sequence of r.v.'s

$$\left\{\left[t'\dot{\phi}_j(\theta)\right]^2\right\}, \quad j \geq 1, \text{ for each } t \in \mathbb{R}^k.$$

(iv) For each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$,

$$\frac{1}{n}\sum_{j=1}^{n}\left[E_\theta\left[t'\dot{\phi}_j(\theta)\right]^2 |\mathcal{A}_{j-1}\right] - \left[t'\dot{\phi}_j(\theta)\right]^2 \to 0, \tag{2.9}$$

in P_θ-probability.

(A4) For each $\theta \in \Theta$, let $q_0(\theta;\cdot)$ be defined by (2.1). Then $q_0(\theta;\cdot)$ is $\mathcal{A}_0 \times \mathcal{C}$-measurable and continuous in P_θ-probability.

Some Comments on the Assumptions

In the first place, assumption (A3)(iii) means that, for each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$,

$$\frac{1}{n}\sum_{j=1}^{n}\left\{\left[t'\dot{\phi}_j(\theta)\right]^2 - E_\theta\left[t'\dot{\phi}_j(\theta)\right]^2\right\} \to 0 \tag{2.10}$$

in P_θ-probability. Then, on the basis of assumption (A3)(ii) and relation (2.5), relation (2.10) may be reformulated equivalently as follows:

(A3)(iii′) For each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$,

$$\frac{1}{n}\sum_{j=1}^{n}\left[t'\dot{\phi}_j(\theta)\right]^2 \to \frac{1}{4}t'\overline{\Gamma}(\theta)t \text{ in } P_\theta\text{-probability.} \tag{2.11}$$

Next, on the basis of relations (2.9) and (2.10), it is clear that assumption (A3)(iv) may be reformulated equivalently as follows:

(A3)(iv′) For each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$, the r.v.'s

$$\left\{\mathcal{E}_\theta\left[\left[t'\dot{\phi}_j(\theta)\right]^2 | \mathcal{A}_{j-1}\right]\right\}, \quad j \geq 1,$$

satisfy the WLLN when the probability measure P_θ is used.

Also, from relations (2.9) and (2.11), another equivalent reformulation of assumption (A3)(iv) is the following:

(A3)(iv″) For each $\theta \in \Theta$ and each $t \in \mathbb{R}^k$,

$$\frac{1}{n}\sum_{j=1}^{n}\mathcal{E}_\theta\left\{\left[t'\dot{\phi}_j(\theta)\right]^2 | \mathcal{A}_{j-1}\right\} \to \frac{1}{4}t'\overline{\Gamma}(\theta)t \text{ in } P_\theta\text{-probability.} \tag{2.12}$$

Assumption (A1) is not as strong as it may look. It is made primarily to simplify derivations. All results may be obtained without this assumption; this is so, by Theorem 5.1 in Roussas (1972, also reprinted in a paperback form in 2008), because for any two sequences of probability measures $\{P_n\}$ and $\{P'_n\}$, there exist sequences $\{Q_n\}$ and $\{Q'_n\}$ of probability measures such that Q_n and Q'_n are mutually absolutely continuous for all sufficiently large n, and $\|P_n - Q_n\| + \|P'_n - Q'_n\| \to 0$. It is to be recalled here that

Definition 2.1. For any two probability measures P and Q defined on $(\Omega, \mathcal{A}^*)$, the *sup-norm* $\|P - Q\|$ is defined by $\|P - Q\| = 2\sup\{|P(A) - Q(A)|; A \in \mathcal{A}^*\}$. Also, $\|P - Q\| = \int_\Omega |f - g|\, d\mu$, where $f = dP/d\mu$, $g = dQ/d\mu$ for some σ-finite measure μ dominating P and Q (for example, $\mu = P + Q$). For this reason, $\|P - Q\|$ is also known as the L_1-*norm*.

Assumption (A2), which requires differentiability in q.m. of a certain random function, is a weak assumption and replaces in effect the usual assumptions about existence of pointwise derivatives of several orders.

Some examples, where assumptions (A1)–(A4), as they apply to them, have been checked and found to hold (see Stamatelos (1976)).

3 Preliminary Results

We proceed further by recalling some known results.

To this end, refer to relations (2.1) and (2.4), and set

$$\Lambda_n(\theta) = \Lambda_n(\theta;\theta_n) = \log\left[\frac{dP_{n,\theta_n}}{dP_{n,\theta}}\right], \tag{3.1}$$

where $\theta_n = \theta + h_n n^{-1/2}$, $h_n \in \mathbb{R}^k$ with $h_n \to h \in \Theta$.

Then for all sufficiently large n the following fundamental result holds, regarding the asymptotic expansion of $\Lambda_n(\theta)$ in $P_{n,\theta}$-probability.

Theorem 1. *Let $\Lambda_n(\theta)$ and $\Delta_n(\theta)$ be defined by* (3.1) *and* (2.6)*, respectively. Then, under assumptions* (A1)–(A4) *and for each $\theta \in \Theta$, it holds*

$$\Lambda_n(\theta) - h'\Delta_n(\theta) \to -\frac{1}{2}h'\overline{\Gamma}(\theta)h \textit{ in } P_{n,\theta}\textit{-probability,}$$

where $\overline{\Gamma}(\theta)$ is given in assumption (A3)(ii)*.*

The proof of this theorem is long, and is based on an extensive series of auxiliary results. The interested reader is referred to Roussas (1979), Section 5. Heuristically, Theorem 1 states that in the neighborhood of θ, $\exp\left[-\frac{1}{2}h'\overline{\Gamma}(\theta)h + h'\Delta_n(\theta)\right] dP_{n,\theta}$ approximates $dP_{n,\theta_n}/dP_{n,\theta}$; that is, $L_n(\theta,\theta_n) \simeq \exp\left[h'\Delta_n(\theta) - \frac{1}{2}h'\overline{\Gamma}(\theta)h\right]$. A precise formulation of this heuristic interpretation is given in Theorem 7.

Below, the definition of weak convergence, which is used throughout this paper, is recalled.

Definition 3.1. For $n \geq 1$, let $\mathcal{L}_n$ and $\mathcal{L}$ be two probability measures defined on the m-dimensional Euclidean space $(\mathbb{R}^m, \mathcal{B}^m)$, $(m \geq 1)$. We say that $\{\mathcal{L}_n\}$ converges *weakly* to $\mathcal{L}$ and write $\mathcal{L}_n \Longrightarrow \mathcal{L}$, if $\int_{\mathbb{R}^m} f d\mathcal{L}_n \to \int_{R^m} f d\mathcal{L}$ for all real-valued bounded and continuous functions f defined on $\mathbb{R}^m$.

The assumptions made above also allow us to derive the asymptotic distribution of the k-dimensional random vector $\Delta_n(\theta)$. Specifically, we have the following theorem:

Theorem 2. *Let $\Delta_n(\theta)$ be defined by* (2.6)*, and suppose assumptions* (A1)–(A4) *hold. Then, for each $\theta \in \Theta$,*

$$\mathcal{L}\left[\Delta_n(\theta)|P_{n,\theta}\right] \Longrightarrow N\left(0, \overline{\Gamma}(\theta)\right),$$

where $\overline{\Gamma}(\theta)$ is given in assumption (A3)(ii)*.*

The proof of this theorem can be found in Roussas and Bhattacharya (2007).

Now, combining Theorems 1 and 2, and employing the Slutsky theorem, we obtain the following result:

Theorem 3. *Under assumptions* (A1)–(A4)*, in the notation of Theorems 1 and 2, and for each $\theta \in \Theta$,*

$$\mathcal{L}\left[\Lambda_n(\theta)|P_{n,\theta}\right] \Longrightarrow N\left(-\frac{1}{2}h'\overline{\Gamma}(\theta)h, \quad h'\overline{\Gamma}(\theta)h\right).$$

Results similar to Theorems 1–3 also hold when the "fixed" probability measure $P_{n,\theta}$ is replaced by the "moving" probability measure P_{n,θ_n}. For this purpose, however, the concept of contiguity is needed. This concept was introduced by Le Cam in his seminal paper (1960). See also his book, Le Cam (1986), as well as an easier-to-read source Le Cam and Yang (2000). However, the concept of contiguity was popularized early on by the monograph Roussas (1972, 2008), written in a Markov process framework, where several statistical applications were also discussed.

There are several characterizations of contiguity. The first one to be introduced here allows for an intuitive interpretation, and the second is a working definition.

Definition 3.2. The sequences of probability measures $\{P_n\}$ and $\{P'_n\}$ defined on the measurable spaces $(\mathcal{X}, \mathcal{A}_n)$ are said to be *contiguous* if $P_n(A_n) \to 0$ for A_n in $\mathcal{A}_n$ implies $P'_n(A_n) \to 0$, and vice versa.

Remark 1. It appears as if contiguity is akin to some kind of asymptotic mutual absolute continuity of the pairs P_n and P'_n. It is to be pointed out, however, that $\{P_n\}$ and $\{P'_n\}$ may be contiguous, and yet P_n and P'_n are not mutually absolutely continuous for any n. Also, it is possible that P_n and P'_n be mutually absolutely continuous for all n, and yet $\{P_n\}$ and $\{P'_n\}$ are not contiguous. These points are illustrated by Examples 2.1 and 2.2, pp.9–10, in Roussas (1972, 2008). On the other hand, if $\|P_n - P'_n\| \to 0$ (see Definition 2.1), then $\{P_n\}$ and $\{P'_n\}$ are contiguous, as one would expect. The justification of the result is simple and is given in Lemma 2.1, pp. 8–9, in Roussas (1972, 2008).

Another equivalent working definition of contiguity is the following:

Definition 3.3. In the setting of Definition 3.2, the sequences of probability measures $\{P_n\}$ and $\{P'_n\}$ are said to be *contiguous* if the following is true: for any $\mathcal{A}_n$-measurable r.v.'s T_n, $T_n \to 0$ in P_n-probability if and only if $T_n \to 0$ in P'_n-probability.

The equivalence of the two ways of defining contiguity in Definitions 3.2 and 3.3 is established in Roussas (1972, 2008). There are additional equivalent characterizations of contiguity; the interested reader is referred to pp.8,11,17,31–33, in Roussas (1972, 2008).

This last result allows us to establish the following proposition:

Proposition 3.1. *Under assumptions* (A1)–(A4)*, the sequences of probability measures* $\{P_{n,\theta}\}$ *and* $\{P_{n,\theta_n^*}\}$ *are contiguous, where* $\theta_n^* = \theta + h_n^* n^{-1/2}$ *and* $\{h_n^*\}$ *is bounded.*

The proof of the above proposition can be found in Roussas and Bhattacharya (2007).

We may now proceed with the second installment of theorems.

Theorem 4. *Under assumptions* (A1)–(A4)*,*

$$\Lambda_n(\theta) - h'\Delta_n(\theta) \to -\frac{1}{2}h'\overline{\Gamma}(\theta)h \text{ in } P_{n,\theta_n}\text{-probability.}$$

The proof of the theorem follows from Theorem 1, Proposition 3.1, and Definition 3.2.

Theorem 5. *Under assumptions* (A1)–(A4)*,*

$$\mathcal{L}\left[\Lambda_n(\theta)|P_{n,\theta_n}\right] \Longrightarrow N\left(\frac{1}{2}h'\overline{\Gamma}(\theta)h,\ h'\overline{\Gamma}(\theta)h\right).$$

The proof of the theorem follows from Theorem 3, Proposition 3.1, and Le Cam's third lemma (see Corollary 7.2, p.35, in Roussas 1972, 2008).

Theorem 6. *Under assumptions* (A1)–(A4)*,*

$$\mathcal{L}\left[\Delta_n(\theta)|P_{n,\theta_n}\right] \Longrightarrow N\left(\overline{\Gamma}(\theta)h, \overline{\Gamma}(\theta)\right).$$

The proof follows from Theorem 2, Theorem 1, Proposition 3.1, and Theorem 7.2, p.38, in Roussas (1972, 2008).

It has been stated above (see discussion following Theorem 1) that, for any arbitrary but fixed $\theta \in \Theta$, for all sufficiently large n, and in $P_{n,\theta}$-probability,

$$\mathcal{L}_n(\theta,\theta_n) = \exp\left[\Lambda_n(\theta,\theta_n)\right] \simeq \exp\left[-\frac{1}{2}h'\overline{\Gamma}(\theta)h\right]\exp\left[h'\Delta_n(\theta)\right], \quad h \in \mathbb{R}^k. \tag{3.2}$$

This loose interpretation leads to an exact exponential approximation (in the sup-norm, or in the L_1-norm sense). Looking at the right-hand side of (3.2), it is obvious that $\mathcal{E}_\theta\left[\exp\left[h'\Delta_n(\theta)\right]\right]$ should be a multiple of the norming constant in the approximating exponential family. The problem, however, is that this expectation needs not be finite. This leads to the replacement of $\Delta_n(\theta)$ by a truncated version $\Delta_n^*(\theta)$, which ensures that $\mathcal{E}_\theta \exp\left[h'\Delta_n^*(\theta)\right] < \infty$. The relevant details will be outlined below. In most of what follows, the presence of θ is suppressed to avoid more cumbersome notation as long as it is understood that θ is arbitrary but fixed. Here we recall (2.6),

$$\Delta_n(\theta) = 2n^{-\frac{1}{2}}\sum_{j=1}^{n}\dot{\phi}_j(\theta), \quad n \geq 1,$$

and let Δ_n^* $(= \Delta_n^*(\theta))$ be a truncated version of Δ_n (see (1.5), p.72, in Roussas (1972, 2008)), for which $\mathcal{E}_\theta \exp(h'\Delta_n^*) < \infty$, so that

$$\exp B_n(h) = \mathcal{E}_\theta \exp(h'\Delta_n^*) < \infty, \quad h \in \mathbb{R}^k, \tag{3.3}$$

and such that (see Proposition 2.3(ii), pp.73–74, in Roussas (1972, 2008))

$$P_{n,\theta}(\Delta_n^* \neq \Delta_n) \to 0, \qquad P_{n,\theta_n}(\Delta_n^* \neq \Delta_n) \to 0, \tag{3.4}$$

where $\theta_n = \theta + h_n n^{-1/2}$, with $h_n \in \mathbb{R}^k$ and $\{h_n\}$ bounded. Next, define $R_{n,h}$ by

$$R_{n,h}(A) = \exp[-B_n(h)] \int_A \exp(h'\Delta_n^*)\, dP_{n,\theta}, \quad A \in \mathcal{A}_n, \tag{3.5}$$

so that

$$\begin{aligned} \frac{dR_{n,h}}{dP_{n,\theta}} &= \exp[-B_n(h)] \exp(h'\Delta_n^*), \quad h \in \mathbb{R}^k \\ &= \exp[h'\Delta_n^* - B_n(h)], \quad h \in \mathbb{R}^k \end{aligned} \tag{3.6}$$

is an exponential p.d.f. with parameter $h \in \mathbb{R}^k$. Then we have the following important theorem:

Theorem 7. *Let $\theta_n = \theta + h_n n^{-1/2}$, with $\{h_n\}$ bounded, let $R_{n,h}(A)$ be defined by* (3.5), *and suppose assumptions* (A1)–(A4) *hold. Then*

$$\begin{aligned} \|R_{n,h_n} - P_{n,\theta_n}\| &= (2 \sup |\, R_{n,h_n}(A) - P_{n,\theta_n}(A)\,|\,; \quad A \in \mathcal{A}_n) \\ &= \int_{\mathcal{X}} \left| \frac{dR_{n,h_n}}{dP_{n,\theta}} - \frac{dP_{n,\theta_n}}{dP_{n,\theta}} \right| dP_{n,\theta} \\ &\to 0\,. \end{aligned} \tag{3.7}$$

The proof of the theorem can be found in Roussas and Bhattacharya (2007). The implication of the above theorem is very important in our context. It states that the probability measures P_{n,θ_n} and R_{n,h_n} are differentially (asymptotically) equivalent; in other words, in the neighborhood of θ, either one of the sequences$\{P_{n,\theta_n}\}$ and $\{R_{n,h_n}\}$ is as good as the other. To increase the smooth flow of the paper, the definition of differentially (asymptotically) equivalent probability measures at θ, and the definition of a differentially (asymptotically) sufficient statistic at θ for the family $\{P_{n,\theta}; \theta \in \Theta\}$ are recalled (Roussas 1972, 2008, pp.79–81).

Definition 3.4. For $\theta \in \Theta$, let $\{P_{n,\theta}\}$ and $\{P_{n,\theta}^*\}$ be two sequences of probability measures defined on $\mathcal{A}_n$. Then we say that these two sequences are *differentially* (*asymptotically*) *equivalent* at θ_0 if, for each bounded set C in $\mathbb{R}^k$,

$$\sup \Big[\|P_{n,\theta} - P_{n,\theta}^*\| \quad n^{\frac{1}{2}}(\theta - \theta_0) \in C \Big] \to 0\,. \tag{3.8}$$

The definition states that, in the neighborhood of θ_0, either one of the sequences $\{P_{n,\theta}\}$ and $\{P^*_{n,\theta}\}$ is as good as the other.

By writing $n^{\frac{1}{2}}(\theta-\theta_0)=h$ so that $\theta=\theta_n=\theta_0+hn^{-\frac{1}{2}}$, relation (3.8) becomes

$$\sup\left[\|P_{n,\theta_n}-P^*_{n,\theta_n}\| \quad h\in C, \theta_n=\theta_0+hn^{-\frac{1}{2}}\right]\to 0\ . \tag{3.9}$$

Now, let us introduce the following notation.

For each n and all $\theta\in\Theta$, define the probability measure $R^*_{n,\theta}$ as follows:

$$R^*_{n,\theta}=R_{n,n^{1/2}(\theta-\theta_0)} \tag{3.10}$$

Then the following result holds.

Proposition 3.2. *Let $R^*_{n,\theta}$ be defined by* (3.10). *Then the sequences of probability measures $\{P_{n,\theta}\}$ and $\{R^*_{n,\theta}\}$ are differentially* (*asymptotically*) *equivalent at θ_0.*

Proof. It is to be shown that

$$\sup\left[\|P_{n,\theta}-R^*_{n,\theta}\|\ n^{1/2}(\theta-\theta_0)\in C\right]\to 0,\ \text{for every bounded set } C \text{ in } \mathbb{R}^k.$$

By virtue of (3.10), this is equivalent to proving that

$$\sup\left[\|P_{n,\theta_n}-R^*_{n,\theta_n}\| \quad h\in C,\ \theta_n=\theta_0+hn^{-\frac{1}{2}}\right]\to 0,\ \text{for every bounded set } C \text{ in } \mathbb{R}^k. \tag{3.11}$$

However, (3.11) is true because of (3.7). ∎

Definition 3.5. For $\theta\in\Theta$, let $\{P_{n,\theta}\}$ be a sequence of probability measures defined on $\mathcal{A}_n$ and let $\{T_n\}$ be a sequence of k-dimensional, $\mathcal{A}_n$-measurable random vectors. Denote by $\mathcal{B}_n$ the σ-field induced by T_n. Then the sequence $\{T_n\}$ or $\{\mathcal{B}_n\}$ is said to be *differentially* (*asymptotically*) *sufficient* at θ_0 for the family $\{P_{n,\theta};\theta\in\Theta\}$, if there exists a family of probability measures $\{P^*_{n,\theta};\ \theta\in\Theta\}$, such that, for each n, T_n or $\mathcal{B}_n$ is sufficient for the family $\{P^*_{n,\theta};\ \theta\in\Theta\}$ and $\{P_{n,\theta};\ \theta\in\Theta\},\{P^*_{n,\theta};\ \theta\in\Theta\}$ are differentially *(*asymptotically*)* equivalent at θ_0.

According to the following result, the random vector Δ^*_n, as defined in (3.4), possesses the property of being differentially sufficient at θ_0 for the family $\{P_{n,\theta},\theta\in\Theta\}$.

Proposition 3.3. *Let Δ^*_n $(=\Delta^*_n(\theta_0))$ be a truncated version of Δ_n* (*as described just prior to relation* (3.4)) *satisfying* (3.4). *Then the sequence $\{\Delta^*_n\}$ is differentially* (*asymptotically*) *sufficient at θ_0 for the family $\{P_{n,\theta};\ \theta\in\Theta\}$.*

Proof. For each n and all $\theta\in\Theta$, define the probability measure $R^*_{n,\theta}$ as follows:

$$R^*_{n,\theta}=R_{n,n^{1/2}(\theta-\theta_0)}.$$

From the definition of $R_{n,h}$ in (3.6) and for each n, one has

$$\frac{dR_{n,h}}{dP_{n,\theta_0}} = \exp[-B_n(h)]\exp\left(h'\Delta_n^*\right), \quad h \in \mathbb{R}^k, \quad h = n^{\frac{1}{2}}(\theta - \theta_0). \tag{3.12}$$

Using (3.10) and the definition of $R_{n,h}$ in (3.6), we have

$$\begin{aligned} \frac{dR^*_{n,\theta}}{dP_{n,\theta_0}} &= \frac{dR_{n,n^{1/2}(\theta-\theta_0)}}{dP_{n,\theta_0}} \\ &= \exp\left\{n^{\frac{1}{2}}(\theta-\theta_0)'\Delta_n^* - B_n\left[n^{\frac{1}{2}}(\theta-\theta_0)\right]\right\} \\ &= \exp\left\{-B_n\left[n^{\frac{1}{2}}(\theta-\theta_0)\right]\right\}\exp\left(n^{\frac{1}{2}}\theta'\Delta_n^*\right)\exp\left(-n^{\frac{1}{2}}\theta_0'\Delta_n^*\right). \end{aligned}$$

From the above expression, it is clear that, for each n, Δ_n^* is sufficient for the family $\{R^*_{n,\theta};\ \theta \in \Theta\}$, or equivalently, for the family $\{R_{n,h};\ h \in C\}$ (C a bounded set in $\mathbb{R}^k$) and hence differentially (asymptotic) sufficient at θ_0 for the family $\{P_{n,\theta}; \theta \in \Theta\}$, since $\{P_{n,\theta};\ \theta \in \Theta\}$ and $\{R_{n,h};\ h \in C\}$ are differentially (asymptotically) equivalent at θ_0. ■

Finally, we state the following lemmas which are used to derive our main results:

Lemma 3.1. *Let $\{Y_n\}$ be a sequence of random variables defined on some measurable space $(\Omega, \mathcal{F})$, and for each n, let Q_n be a probability measure on $\mathcal{F}$. It is assumed that*

$$\mathcal{L}(Y_n|Q_n) \Longrightarrow N(\mu, \sigma^2).$$

Then (i) *For any numerical sequence $\{y_n\}$, one has*

$$Q_n(Y_n = y_n) \to 0\ .$$

Let the sequences of numbers $\{c_n\}$ and $\{\gamma_n\}$ with $0 \le \gamma_n \le 1$ for all n, be defined by

$$Q_n(Y_n > c_n) + \gamma_n Q_n(Y_n = c_n) = \alpha \quad (0 < \alpha < 1).$$

Then (ii) *$c_n \to \mu + \sigma\xi_\alpha$, where ξ_α is the upper α^{th} quantile of a $N(0,1)$ variable.*

Lemma 3.2. *Consider the measurable space $(\Omega, \mathcal{F})$ and let Z be the logarithm of the likelihood ratio of Q relative to P, where P and Q are two probability measures on $\mathcal{F}$ such that the density of P is $f = dP/d\mu$ and the density of Q is $g = dQ/d\mu$ for some dominating σ-finite measure μ (e.g., $\mu = P + Q$) and $Z = \log\left(\dfrac{g}{f}\right)$. Then, for every $\varepsilon > 0$,*

$$\|P - Q\| \le 2\left(1 - e^{-\varepsilon}\right) + 2P(|Z| > \varepsilon).$$

For a complete proof of the Lemmas, the reader is referred to Roussas (1972, 2008), Lemmas 2.1 and 2.2, pp.97–99.

4 Asymptotic Properties of Tests Based on $\Delta_n(\theta_0)$

The results discussed in Section 3 can be used to derive optimal tests for the hypothesis $H_0: \ \theta = \theta_0$ (asymptotically most powerful tests or asymptotically most powerful unbiased tests, as the case may be) for a real-valued parameter θ. To proceed further we need to prove two theorems stated below, where $\Delta_n = \Delta_n(\theta_0)$. However, we first formulate the following remark, which will be useful in the sequel.

Remark 2. For each $h \in \mathbb{R}^k$, let $\{x_n(h)\}$ be a sequence of nonnegative numbers, let C be any bounded set in $\mathbb{R}^k$, and set $x_n = \sup[x_n(h); \ h \in C]$. Then $x_n \to 0$ if and only if $x_n(h_n) \to 0$ for any $h_n \in C$.

Indeed, for any $h_n \in C$, we have $x_n(h_n) \le x_n \to 0$, so that $x_n(h_n) \to 0$. Next, let $\varepsilon_n \downarrow 0$. Then, for each n, there exists at least one $h_n \in C$ such that $x_n(h_n) > x_n - \varepsilon_n$. Hence $0 = \lim x_n(h_n) \ge \overline{\lim} x_n - \lim \varepsilon_n = \overline{\lim} x_n = \lim x_n = 0$; that is, $x_n \to 0$.

Theorem 8. *Let $\{Z_n\}$ be a sequence of random variables such that $|Z_n| \le 1$ for all n and set $\overline{Z}_n = \mathcal{E}_{\theta_0}[Z_n|\Delta_n(\theta_0)]$. Then*

$$\sup\left(\left|\mathcal{E}_{\theta_n} Z_n - \mathcal{E}_{\theta_n}\overline{Z}_n\right|\right) \to 0,$$

where $\theta_n = \theta_0 + hn^{-\frac{1}{2}}$ and the sup is taken over all random variables bounded by 1 in absolute value and over all h's in a bounded set C in $\mathbb{R}^k$.

Proof. We have,

$$\mathcal{E}_{\theta_n} Z_n - \mathcal{E}_{\theta_n}\overline{Z}_n = \mathcal{E}\left(Z_n|P_{n,\theta_n}\right) - \mathcal{E}\left(\overline{Z}_n|P_{n,\theta_n}\right) = I_1(n,h) + I_2(n,h) + I_3(n,h),$$

where

$$\begin{aligned} I_1(n,h) &= \mathcal{E}\left(Z_n|P_{n,\theta_n}\right) - \mathcal{E}\left(Z_n|R_{n,h}\right), \\ I_2(n,h) &= \mathcal{E}\left(Z_n|R_{n,h}\right) - \mathcal{E}\left(\overline{Z}_n|R_{n,h}\right), \\ I_3(n,h) &= \mathcal{E}\left(\overline{Z}_n|R_{n,h}\right) - \mathcal{E}\left(\overline{Z}_n|P_{n,\theta_n}\right). \end{aligned}$$

From the assumption that $|Z_n| \le 1$, the definition of $\overline{Z}_n$ and the fact that $R_{n,h}$ and P_{n,θ_n}, as well as P_{n,θ_n} and P_{θ_0} are differentially equivalent, we have that $\overline{Z}_n$ is also bounded by 1 almost surely(a.s.) $[R_{n,h}]$ and $[P_{n,\theta_n}]$. Then with the sup as above, both $\sup|I_1(n,h)|$ and $\sup|I_3(n,h)|$ are bounded by

$$\sup\left[\|P_{n,\theta_n} - R_{n,h}\| \ h \in C\right].$$

Setting $x_n = \sup\left[\|P_{n,\theta_n} - R_{n,h}\| \ h \in C\right]$, we have that $x_n \to 0$, because $x_n(h_n) = \|R_{n,h_n} - P_{n,\theta_n}\| \to 0$ by (3.7); this is so by Remark 2. We next investigate the behavior of $\sup|I_2(n,h)|$. With sup as above, we will show that $\sup|I_2(n,h)| = 0$, which will complete the proof of the theorem. We observe that

$$\begin{aligned}
\mathcal{E}\left(\overline{Z}_n|R_{n,h}\right) &= \mathcal{E}\left[\mathcal{E}_{\theta_0}\left(Z_n|\Delta_n\right)|R_{n,h}\right] \\
&= \int_{\mathcal{X}} \mathcal{E}_{\theta_0}\left(Z_n|\Delta_n\right) dR_{n,h} \\
&= \int_{\mathcal{X}} \mathcal{E}_{\theta_0}\left(Z_n|\Delta_n\right) \exp\left[-B_n(h)\right] \exp\left(h'\Delta_n^*\right) dP_{n,\theta_0} \\
&= \int_{\mathcal{X}} \mathcal{E}_{\theta_0}\left[Z_n \exp\left[-B_n(h)\right] \exp\left(h'\Delta_n^*\right)|\Delta_n\right] dP_{n,\theta_0}, \\
&\qquad\qquad \text{because } \Delta_n^* \text{ is a function of } \Delta_n \\
&= \mathcal{E}_{\theta_0}\left\{\mathcal{E}_{\theta_0}\left[Z_n \exp[-B_n(h)] \exp\left[h'\Delta_n^*|\Delta_n\right]\right]\right\} \\
&= \mathcal{E}_{\theta_0}\left\{Z_n \exp[-B_n(h)] \exp\left(h'\Delta_n^*\right)\right\} \\
&= \int_{\mathcal{X}} Z_n \exp[-B_n(h)] \exp\left(h'\Delta_n^*\right) dP_{n,\theta_0} \\
&= \int_{\mathcal{X}} Z_n dR_{n,h}, \text{ using the definition of } R_{n,h} \\
&= \mathcal{E}\left(Z_n|R_{n,h}\right).
\end{aligned}$$

Therefore $|I_2(n,h)| = 0$, as was to be seen, and hence the proof of the theorem is completed. ■

The implication of the above theorem is that from the viewpoint of asymptotic power, we may confine ourselves to tests which depend on $\Delta_n(\theta_0)$ alone.

Now, the statistic $\Delta_n^* = \Delta_n^*(\theta_0)$ is the all-important statistic appearing in the exponent of the exponential family defined through relations (3.3)–(3.6), and therefore any (optimal) tests would have to be expressed in terms of Δ_n^*. However, Theorem 9 below states that any tests may be based on $\Delta_n (= \Delta_n(\theta_0))$ rather than Δ_n^*.

Theorem 9. *Let $\{Z_n\}$ be a sequence of test functions defined on $\mathbb{R}^k$ and set $\theta_n = \theta_0 + hn^{-\frac{1}{2}}$, $h \in \mathbb{R}^k$. Then for any bounded subset C of $\mathbb{R}^k$, we have*

$$\sup\left[\mid \mathcal{E}_{\theta_n} Z_n(\Delta_n) - \mathcal{E}_{\theta_n} Z_n(\Delta_n^*) \mid \; h \in C\right] \to 0 \,.$$

Proof. With $\theta_n = \theta_0 + h_n n^{-1/2}$ and $h_n \in C$, a bounded set in $\mathbb{R}^k$, we have

$$\begin{aligned}
|\mathcal{E}_{\theta_n} Z_n(\Delta_n) - \mathcal{E}_{\theta_n} Z_n(\Delta_n^*)| &= \left|\int_{\mathcal{X}} Z_n(\Delta_n) dP_{n,\theta_n} - \int_{\mathcal{X}} Z_n(\Delta_n^*) dP_{n,\theta_n}\right| \\
&= \left|\int_{\mathcal{X}} \left[Z_n(\Delta_n) - Z_n(\Delta_n^*)\right] dP_{n,\theta_n}\right| \\
&= \left|\left\{\int_{(\Delta_n = \Delta_n^*)} \left[Z_n(\Delta_n) - Z_n(\Delta_n^*)\right] + \int_{(\Delta_n \neq \Delta_n^*)} \left[Z_n(\Delta_n) - Z_n(\Delta_n^*)\right]\right\} dP_{n,\theta_n}\right|
\end{aligned}$$

$$
\begin{aligned}
&= \left| \int_{(\Delta_n \neq \Delta_n^*)} [Z_n(\Delta_n) - Z_n(\Delta_n^*)] dP_{n,\theta_n} \right| \\
&= \left| \int_{\mathcal{X}} \{[Z_n(\Delta_n) - Z_n(\Delta_n^*)] I(\Delta_n \neq \Delta_n^*)\} \, dP_{n,\theta_n} \right| \\
&\leq 2 \int_{\mathcal{X}} I(\Delta_n \neq \Delta_n^*) dP_{n,\theta_n} \\
&= 2P_{n,\theta_n}(\Delta_n \neq \Delta_n^*) \longrightarrow 0 \text{ by (3.4).}
\end{aligned}
$$

But then

$$
\sup \left[|\mathcal{E}_{\theta_n} Z_n(\Delta_n) - \mathcal{E}_{\theta_n} Z_n(\Delta_n^*)| ; \quad \theta_n = \theta + hn^{-1/2}, \ h \in C \right] \longrightarrow 0
$$

by Remark 2 with x_n being the sup on the left-hand side above, and

$$
x_n(h_n) = |\mathcal{E}_{\theta_n} Z_n(\Delta_n) - \mathcal{E}_{\theta_n} Z_n(\Delta_n^*)|
$$

with $\theta_n = \theta_0 + h_n n^{-1/2}$. The justification of (3.4) may be found in Roussas (1972, 2008), Proposition 2.3(ii), pp.73–74.

Hence

$$
\begin{aligned}
| \, \mathcal{E}_{\theta_n} Z_n(\Delta_n) - \mathcal{E}_{\theta_n} Z_n(\Delta_n^*) \, | \quad &\leq \quad 2P_{n,\theta_n}(\Delta_n \neq \Delta_n^*), \quad h \in C \\
&\to \quad 0,
\end{aligned}
$$

since $P_{n,\theta_n}(\Delta_n^* \neq \Delta_n) \to 0$ as stated in (3.4). Thus the proof of the theorem is complete. ■

While deriving all the results stated above, we have considered moving parameter points θ_n, where $\theta_n = \theta_0 + hn^{-\frac{1}{2}}$. Here as $n \to \infty$, θ_n approaches the fixed parameter point θ_0 at a prescribed rate. If θ_n either does not approach θ_0 at that rate or it stays away from θ_0, equivalently, if $n^{\frac{1}{2}}(\theta_n - \theta_0) \to \pm\infty$, the results stated in the earlier sections need not be true. This suggests the need for further assumptions to accommodate these cases. These assumptions, stated as (A5) and (A5′), are listed below:

Let $\theta_0 \in \Theta$ and define $\omega = \{\theta \in \Theta; \ \theta > \theta_0\}$ and $\omega' = \{\theta \in \Theta; \ \theta < \theta_0\}$. Then

Assumption (A5): Consider a sequence $\{\theta_n\}$, where $\theta_n \in \omega$ for all n. Then

$$
\Delta_n(\theta_0) \to \infty \text{ in } P_{n,\theta_n}\text{-probability whenever } n^{\frac{1}{2}}(\theta_n - \theta_0) \to \infty.
$$

Assumption (A5′): Consider a sequence$\{\theta_n\}$, where $\theta_n \in \omega'$ for all n. Then

$$
\Delta_n(\theta_0) \to -\infty \text{ in } P_{n,\theta_n}\text{-probability whenever } n^{\frac{1}{2}}(\theta_n - \theta_0) \to -\infty.
$$

Here $\Delta_n(\theta_0)$ is obtained, of course, from (2.6) for $\theta = \theta_0$.

Assumptions (A5) and (A5′) are not particularly stringent, and they have been checked and found to be true for a number of interesting cases (Roussas 1972, 2008).

With $\Theta \in \mathbb{R}$, consider the problem of testing the simple hypothesis $H_0: \ \theta = \theta_0$, $\theta_0 \in \Theta$, against the composite alternative $A: \ \theta \in \omega$. We shall restrict to sequences of certain tests based on the r.v. $\Delta_n(\theta_0)$ whose level of significance is α and we shall establish that under assumptions (A1–A4) and (A5), the tests are optimal in the sense of being asymptotically uniformly most powerful (AUMP). Under assumptions (A1–A4) alone, the tests are asymptotically locally most powerful (ALMP).

Here the definition of an asymptotically uniformly most powerful (AUMP) test is recalled.

Definition 4.1. For testing $H_0: \ \theta = \theta_0, \theta_0 \in \Theta \subseteq \mathbb{R}$ against $A: \ \theta \in \omega = \{\theta \in \Theta; \ \theta > \theta_0\}$, the sequence of level α tests $\{\varphi_n\}$ is said to be *asymptotically uniformly most powerful* (AUMP) if, for any other sequence of level α tests $\{\omega_n\}$, one has

$$\limsup[\sup(\mathcal{E}_\theta \omega_n - \mathcal{E}_\theta \varphi_n)] \leq 0 \, . \tag{4.1}$$

A similar expression is required to hold with $\theta > \theta_0$ replaced by $\theta < \theta_0$ when the alternative is $A: \theta \in \omega' = \{\theta \in \Theta; \ \theta < \theta_0\}$.

Let $\Delta_n(\theta_0)$ be given by (2.6) with θ replaced by θ_0, and due to the earlier stated results (Theorems 8 and 9), we define the sequence $\{\varphi_n\}$ as follows:

$$\varphi_n = \varphi_n(\Delta_n(\theta_0)) = \begin{cases} 1 & \text{if } \Delta_n(\theta_0) > c_n \\ \gamma_n & \text{if } \Delta_n(\theta_0) = c_n \\ 0 & \text{if } \Delta_n(\theta_0) < c_n, \end{cases} \tag{4.2}$$

where the sequences $\{c_n\}$ and $\{\gamma_n\}$ are determined by the requirement

$$\mathcal{E}_{\theta_0} \varphi_n = \alpha \text{ for all } n. \tag{4.3}$$

Then we have the following theorem:

Theorem 10. *Under assumptions* (A1–A4) *and* (A5), *the sequence of level-α tests* $\{\varphi_n\}$ *defined by* (4.2) *and* (4.3) *is AUMP for testing $H_0: \ \theta = \theta_0$ against $A: \ \theta \in \omega$.*

Proof. . The proof follows by contradiction. Let the sequence of tests $\{\varphi_n\}$ be not AUMP. Then it is possible to find some sequence of level-α tests $\{\omega_n\}$ for which (4.1) is violated, and let the left-hand side of (4.1) have the value $\delta > 0$, that is,

$$\limsup[\sup(\mathcal{E}_\theta \omega_n - \mathcal{E}_\theta \varphi_n)] = \delta. \tag{4.4}$$

Then there exists a subsequence $\{m\} \subseteq \{n\}$ and a sequence $\{\theta_m\}$ with $\theta_m \in \omega$ for which

$$\mathcal{E}_{\theta_m} \omega_m - \mathcal{E}_{\theta_m} \varphi_m \to \delta. \tag{4.5}$$

Next, it is shown that (4.5) cannot happen, which implies that (4.4) also cannot hold, and that would complete the proof of the theorem. We consider the following two cases: (i) the sequence $\{\theta_m\}$ is such that $\{m^{\frac{1}{2}}(\theta_m - \theta_0)\}$ is unbounded and (ii) the sequence $\{\theta_m\}$ is such that $\{m^{\frac{1}{2}}(\theta_m - \theta_0)\}$ remains bounded.

Consider first that $\{m^{\frac{1}{2}}(\theta_m - \theta_0)\}$ is unbounded. Then there exists a subsequence $\{r\} \subseteq \{m\}$ such that $r^{\frac{1}{2}}(\theta_r - \theta_0) \to \infty$. Then Assumption (A5) implies that

$$\Delta_r(\theta_0) \to \infty \text{ in } P_{r,\theta_r}\text{-probability}. \tag{4.6}$$

Now, Theorem 2 gives

$$\mathcal{L}[\Delta_n(\theta_0)|P_{n,\theta_0}] \Longrightarrow N(0, \overline{\Gamma}(\theta_0)),$$

and apply Lemma 3.1 stated above with

$$Y_n = \Delta_n(\theta_0), \quad Q_n = P_{n,\theta_0}, \quad \mu = 0 \text{ and } \sigma^2(\theta_0) = \overline{\Gamma}(\theta_0),$$

to get that $\{c_n\}$ in (4.2) stays bounded. This together with (4.6) implies that

$$P_{\theta_r}[\Delta_r(\theta_0) > c_r] \to 1 \text{ and } P_{\theta_r}[\Delta_r(\theta_0) = c_r] \to 0,$$

so that $\mathcal{E}_{\theta_r}\varphi_r \to 1$. This result together with (4.5) (with m replaced by r) gives that (4.4) cannot occur.

Now, consider the case where $\{m^{\frac{1}{2}}(\theta_m - \theta_0)\}$ is bounded. Then there exists a subsequence $\{s\} \subseteq \{m\}$ such that $s^{\frac{1}{2}}(\theta_s - \theta_0) \to h \geq 0$. Let us first consider the case that $h > 0$ and set $h_s = s^{\frac{1}{2}}(\theta_s - \theta_0)$, so that $\theta_s = \theta_0 + hs^{-\frac{1}{2}}$ with $h_s \to h$. Then by Theorem 6 of Section 3, we have

$$\mathcal{L}[h\Delta_s(\theta_0)|P_{s,\theta s}] \Longrightarrow N(h^2\sigma^2(\theta_0), h^2\sigma^2(\theta_0)).$$

Then Lemma 3.1 applies again with $\{n\}$ replaced by $\{s\}$ and with

$$Y_s = h\Delta_s(\theta_0), \quad Q_s = P_{s,\theta_s}, \quad \mu = \sigma^2 = h^2\sigma^2(\theta_0).$$

It follows then that

$$P_{\theta_s}[h\Delta_s(\theta_0) = hc_s] \to 0\ . \tag{4.7}$$

Also, by the same Lemma with $Y_s = h\Delta_s(\theta_0)$, $Q_s = P_{s,\theta_0}$ and $\mu = 0$, we get $hc_s \to h\sigma(\theta_0)\xi_\alpha$, where ξ_α is the upper α^{th} quantile of Φ, the distribution function of a standard normal, so that

$$\begin{aligned}
P_{\theta_s}[h\Delta_s(\theta_0) > hc_s] &= 1 - P_{\theta_s}[h\Delta_s(\theta_0) \leq hc_s] \\
&= 1 - P_{\theta_s}\left[\frac{h\Delta_s(\theta_0) - h^2\sigma^2(\theta_0)}{h\sigma(\theta_0)} \leq \frac{hc_s - h^2\sigma^2(\theta_0)}{h\sigma(\theta_0)}\right] \\
&\to 1 - \Phi\left(\frac{h\sigma(\theta_0)\xi_\alpha - h^2\sigma^2(\theta_0)}{h\sigma(\theta_0)}\right) \\
&= 1 - \Phi(\xi_\alpha - h\sigma(\theta_0)),\ \text{since } h\sigma(\theta_0) = \left(\overline{\Gamma}(\theta_0)\right)^{\frac{1}{2}} > 0\ .
\end{aligned}$$

Thus, we have

$$P_{\theta_s}[h\Delta_s(\theta_0) > hc_s] \to 1 - \Phi(\xi_\alpha - h\sigma(\theta_0)). \tag{4.8}$$

Relations (4.7) and (4.8) yield

$$\mathcal{E}_{\theta_s}\varphi_s \to 1 - \Phi(\xi_\alpha - h\sigma(\theta_0)). \tag{4.9}$$

Next, define the sequence of critical functions $\{\psi_s\}$ as follows:

$$\psi_s(X_0, X_1, \ldots, X_s) = \begin{cases} 1 & \text{if } \Lambda(\theta_0, \theta_s) + \frac{1}{2}h^2\sigma^2(\theta_0) > d_s \\ \delta_s & \text{if } \Lambda(\theta_0, \theta_s) + \frac{1}{2}h^2\sigma^2(\theta_0) = d_s \\ 0 & \text{otherwise,} \end{cases} \tag{4.10}$$

where $\Lambda(\theta_0, \theta_s) = \log[dP_{s,\theta_s}/dP_{s,\theta_0}]$ and the sequences $\{\delta_s\}$ and $\{d_s\}$ are determined by the size requirement of the test, that is,

$$\mathcal{E}_{\theta_s}\psi_s = \alpha, \text{ for all } s. \tag{4.11}$$

By Theorem 5, we have

$$\mathcal{L}\left[\Lambda(\theta_0, \theta_s) + \frac{1}{2}h^2\sigma^2(\theta_0)|P_{s,\,\theta_s}\right] \Longrightarrow N(h^2\sigma^2(\theta_0), h^2\sigma^2(\theta_0)),$$

so that by a way similar to the one by which (4.9) has been obtained, we have

$$\mathcal{E}_{\theta_s}\psi_s \to 1 - \Phi(\xi_\alpha - h\sigma(\theta_0)). \tag{4.12}$$

Then, by virtue of (4.9) and (4.12), we have

$$\mathcal{E}_{\theta_s}\varphi_s - \mathcal{E}_{\theta_s}\psi_s \to 0\ . \tag{4.13}$$

Replacing m by s in (4.5) (which can be done since $\{s\} \subseteq \{m\}$), we have

$$\mathcal{E}_{\theta_s}\omega_s - \mathcal{E}_{\theta_s}\varphi_s \to \delta. \tag{4.14}$$

Hence, using (4.13) and (4.14) we have

$$\begin{aligned} &\mathcal{E}_{\theta_s}\omega_s - \mathcal{E}_{\theta_s}\psi_s \to \delta(>0), \text{ so that} \\ &\mathcal{E}_{\theta_s}\psi_s \leq \mathcal{E}_{\theta_s}\omega_s, \text{ for sufficiently large } s. \end{aligned} \tag{4.15}$$

Now, consider the problem of testing $H_0: \theta = \theta_0$ against the simple alternative $A_s: \theta = \theta_s$ $(\theta_s \in \omega)$ at level α. By the Neyman-Pearson fundamental lemma, the test defined by (4.10) and (4.11) is the most powerful one. Accordingly, relation (4.15) cannot be true. It is to be noted that we arrived at (4.15) by assuming (4.5) and the existence of $\{s\} \subseteq \{m\}$ such that $s^{\frac{1}{2}}(\theta_s - \theta_0) \to h \geq 0$. Therefore (4.5) and hence (4.4) cannot be true.

Finally, it remains to show that the relation given in (4.4) also cannot be true for the case that $\{m^{\frac{1}{2}}(\theta_m - \theta_0)\}$ is bounded and the only convergent subsequence converges to zero. To prove the above, let there be a subsequence $\{s\} \subseteq \{m\}$ such that $s^{\frac{1}{2}}(\theta_s - \theta_0) = h_s \to h = 0$. Then Theorem 1 implies that

$$\Lambda(\theta_0,\theta_s) \to 0 \text{ in } P_{s,\theta_s}\text{-probability},$$

so that, for every $\varepsilon > 0$,

$$P_{\theta_0}(|\Lambda(\theta_0,\theta_s)| > \varepsilon] \to 0\ . \tag{4.16}$$

Now, Lemma 3.2 applies with

$$P = P_{\theta_0}, \quad Q = P_{\theta_s} \text{ and } Z = \Lambda(\theta_0, \theta_s)$$

and gives

$$\|P_{\theta_0} - P_{\theta_s}\| \leq 2(1 - e^{-\varepsilon}) + 2P_{\theta_0}\left[|\Lambda(\theta_0, \theta_s)| > \varepsilon\right]. \tag{4.17}$$

Relation (4.17) together with (4.16) implies that

$$\|P_{\theta_0} - P_{\theta_s}\| \to 0\ . \tag{4.18}$$

Next, we observe that

$$\begin{aligned}
& |\alpha - \mathcal{E}_{\theta_s}\varphi_s| \\
= \quad & |\{P_{\theta_0}[\Delta_s(\theta_0) > c_s] + \gamma_s P_{\theta_0}[\Delta_s(\theta_0) = c_s]\} - \{P_{\theta_s}[\Delta_s(\theta_0) > c_s] + \gamma_s P_{\theta_s}[\Delta_s(\theta_0) = c_s]\}| \\
\leq \quad & |P_{\theta_0}[\Delta_s(\theta_0) > c_s] - P_{\theta_s}[\Delta_s(\theta_0) > c_s]| + \gamma_s\,|P_{\theta_0}[\Delta_s(\theta_0) = c_s] - P_{\theta_s}[\Delta_s(\theta_0) = c_s]| \\
\leq \quad & 3\|P_{\theta_0} - P_{\theta_s}\|, \text{ by (4.17)} \\
\longrightarrow \quad & 0, \text{ by (4.18)},
\end{aligned}$$

so that

$$\mathcal{E}_{\theta_s}\varphi_s \to \alpha. \tag{4.19}$$

A similar calculation shows that the test ψ_s, which is most powerful, has also asymptotic power α; that is, $\mathcal{E}_{\theta_s}\psi_s \to \alpha$, so that

$$\mathcal{E}_{\theta_s}\varphi_s - \mathcal{E}_{\theta_s}\psi_s \to 0\ . \tag{4.20}$$

Relation (4.20) is same as the relation (4.13) which led us to a contradiction to (4.4). Hence (4.20) also implies that (4.4) cannot hold and hence the proof of the theorem is complete. ■

As it was pointed out in the course of the proof of Theorem 10, assumption (A5) was only used for the purpose of arriving at a contradiction to (4.4) in the case that the alternatives θ_n either do not approach θ_0 sufficiently fast or stay away from it, in the sense that $\{n^{\frac{1}{2}}(\theta_n - \theta_0)\}$ is unbounded. So, if we restrict our alternatives to those θ_n's such that θ_n tends to θ_0 at an appropriate rate, in the sense that $\{n^{\frac{1}{2}}(\theta_n - \theta_0)\}$ stays bounded (from above), then the sequence of tests $\{\varphi_n\}$ defined by (4.2) and (4.3) is asymptotically locally most powerful for testing $H_0 :\ \theta = \theta_0$ against $A :\ \theta \in \omega = \{\theta \in \Theta;\ \theta > \theta_0\}$.

When we want to test $H_0 :\ \theta = \theta_0$ against $A' :\ \theta \in \omega' = \{\theta \in \Theta;\ \theta < \theta_0\}$, a theorem analogous to Theorem 10 also holds true, the proof of which is essentially similar to the proof of Theorem 10. More precisely, we define the sequence of tests $\{\varphi'_n\}$ as

$$\varphi'_n = \varphi'_n(\Delta_n(\theta_0)) \quad = \quad \begin{cases} 1 & \text{if } \Delta_n(\theta_0) < c'_n \\ \gamma'_n & \text{if } \Delta_n(\theta_0) = c'_n \\ 0 & \text{otherwise,} \end{cases} \tag{4.21}$$

where the sequences $\{c'_n\}$ and $\{\gamma'_n\}$ are determined by the requirement

$$\mathcal{E}_{\theta_0}\varphi'_n = \alpha \text{ for all } n. \tag{4.22}$$

Then we have

Theorem 11. *Under assumptions* (A1–A4) *and* (A5$'$)*, the sequence of level-α tests* $\{\varphi'_n\}$ *defined by* (4.21) *and* (4.22) *is AUMP for testing* $H_0 :\ \theta = \theta_0$ *against* $A' :\ \theta \in \omega'$*.*

Under assumptions *(A1–A4)*, the sequence of tests $\{\varphi'_n\}$ defined by (4.21) and (4.22) is asymptotically locally most powerful for testing $H_0 :\ \theta = \theta_0$ against $A' :\ \theta \in \omega'$ when we restrict the alternatives θ'_n's in such a way that $\{n^{\frac{1}{2}}(\theta_n - \theta_0)\}$ stays bounded (from below).

Examples where Theorems 10 and 11 can be applied to derive an optimal test, have been given in Roussas (1972, 2008) for Markov processes.

When the alternatives are two-sided, AUMP tests do not exist as one would expect. However, one can construct tests which are asymptotically uniformly most powerful unbiased. Let us recall the relevant definitions.

Definition 4.2. For testing $H_0 :\ \theta = \theta_0$, $\theta_0 \in \mathbb{R}$, against $A :\ \theta \in \omega'' = \{\theta \in \Theta;\ \theta \neq \theta_0\}$ at asymptotic level of significance α, the sequence of tests $\{\varphi_n\}$ is said to be *asymptotically unbiased* if $\liminf[\inf(\mathcal{E}_\theta\varphi_n;\ \theta \in \omega'')] \geq \alpha$ and is said to be *AUMP unbiased* (AUMPU) of asymptotic level of significance α if it is asymptotically unbiased and of asymptotic level α and

$$\limsup[\sup(\mathcal{E}_\theta\omega_n - \mathcal{E}_\theta\varphi_n;\ \theta \in \omega'')] \leq 0$$

for any sequence of asymptotically unbiased and of asymptotic level of significance α tests $\{\omega_n\}$.

When the sample size n is fixed, the unbiasedness of a test requires that its power remains $\geq \alpha$. When n is allowed to increase indefinitely, it seems natural to replace the concept of unbiasedness of a test by asymptotic unbiasedness, as is done in Definition 4.2.

Consider the sequence of tests $\{\varphi''_n\}$ defined by:

$$\varphi''_n = \varphi''_n(\Delta_n(\theta_0)) \quad = \quad \begin{cases} 1 & \text{if } \Delta_n(\theta_0) < a_n \text{ or } \Delta_n(\theta_0) > b_n \quad (a_n < b_n) \\ 0 & \text{if } a_n \leq \Delta_n(\theta_0) < b_n, \end{cases} \tag{4.23}$$

where the sequences $\{a_n\}$ and $\{b_n\}$ are chosen, so that

$$a_n \to -\xi_{\frac{1}{2}\alpha} \text{ and } b_n \to \xi_{\frac{1}{2}\alpha}, \tag{4.24}$$

and ξ_p is the upper pth quantile of the $N(0, \bar{\Gamma}(\theta_0))$, and $\bar{\Gamma}(\theta_0)$ is as in assumption (A3)(iii) (with θ replaced by θ_0).

Then we have the following result.

Theorem 12. *Under assumptions* (A1–A5), (A5′)*, the sequence of tests* $\{\varphi_n''(\Delta_n(\theta_0))\}$ *defined by* (4.23) *and* (4.24) *is AUMPU of asymptotic level of significance* α *for testing* $H_0:\ \theta = \theta_0$ *against* $A:\ \theta \in \omega''$*, where* ω'' *is given in Definition 4.2.*

Proof. (Outline) The proof of the theorem is presented in the following three stages. We first show that $\{\varphi_n''(\Delta_n(\theta_0))\}$ is of asymptotic level of significance α. Next it is proved that $\{\varphi_n''(\Delta_n(\theta_0))\}$ satisfies the condition of asymptotic unbiasedness; this is done by contradiction. Finally, it is shown that $\{\varphi_n''(\Delta_n(\theta_0))\}$ is AUMP within the class of asymptotically unbiased and of asymptotic level α tests. The proof of this fact is also by contradiction and consists of two steps. First we take care of those parameter values θ_n for which $\{n^{\frac{1}{2}}(\theta_n - \theta_0)\}$ is unbounded. For those θ_n for which $\{n^{\frac{1}{2}}(\theta_n - \theta_0)\}$ remains bounded, we employ the exponential approximation discussed in Section 3 in order to replace the given family by an exponential family. For this latter family, we set up the appropriate UMPU test which exists and then, by utilizing available results, we return to the original family and show the required optimal character of the sequence of tests $\{\varphi_n''(\Delta_n(\theta_0))\}$. The details of the proof run along the same lines as those of the proof of Theorem 5.1 in Roussas (1972, 2008), pp.115–120, and are omitted. ■

5 Efficient Tests for Local Asymptotic Mixed Normal Experiments

It has now become apparent that LAN is a valuable tool in drawing statistical inference about the underlying parameter. However, it has also been noted that there exists many situations where LAN is not satisfied. What is happening instead is that, under suitable conditions, the log-likelihood ratio statistic and the other entities closely related to it converge to a Locally Asymptotically Mixed Normal (LAMN) distribution. (For details the interested readers are referred to Roussas and Bhattacharya (2009), Le Cam and Yang (2000) and Jeganathan(1995), Davies(1985)). Let us introduce the concept of LAMN experiments in brief, and for the general case $\Theta \in \mathbb{R}^k$, $k \geq 1$.

Let $\theta_n = \theta + \delta_n^{-1} h$, where $h \in \mathbb{R}^k$, $\delta_n^{k\times k}$ is a sequence of norming factors such that $\delta_n^{k\times k}$ is a positive definite (p.d.) matrix with $\|\delta_n^{-1}\| \to 0$ as $n \to \infty$. The norming constants δ_n may depend on θ but are independent of the observations. For all sufficiently large n, $\theta_n \in \Theta$ whenever $\theta \in \Theta$.

Then the sequence of log-likelihood ratios is defined as

$$\Lambda_n(X_1, X_2, \ldots, X_n; \theta, \theta_n) = \Lambda_n(\theta, \theta_n) = \log \frac{dP_{n,\theta_n}}{dP_{n,\theta}}. \tag{5.1}$$

Under a standard set of assumptions (see Roussas and Bhattacharya (2009)), it can be shown that there exists a sequence of k-dimensional vectors $\{\Delta_n(\theta) = \Delta_n\}$ and a sequence of $k \times k$ symmetric almost sure (a.s.) p.d. random matrices $\{T_n(\theta) = T_n\}$ such that the log-likelihood ratio statistic, as defined in (5.1), can be approximately written as a sum of two terms, a term $h'\Delta_n$, which is linear in the local parameter h, and a term $-\frac{1}{2}h'T_n h$, which is quadratic in h; that is,

$$\Lambda_n(\theta_n, \theta) - \left(h'\Delta_n - \frac{1}{2}h'T_n h\right) \to 0 \text{ in } P_{n,\theta}\text{-probability}. \tag{5.2}$$

Further,

$$\mathcal{L}[(\Delta_n, T_n)|P_{n,\theta}] \Longrightarrow \mathcal{L}(\Delta, T), \tag{5.3}$$

where T is an a.s. p.d. random matrix and Δ is such that the conditional distribution of Δ, given T, is $N_k(0, T)$.

If the matrices T_n in the quadratic term converge to a non-random matrix, then the sequence of log-likelihood ratios is Locally Asymptotically Normal (LAN). Then $\Delta \sim N_k(0, T)$ and T is a p.d. matrix and a non-random quantity. These cases have been covered in earlier sections.

Now, let the random vectors Δ_n in equation (5.2) be represented in the form

$$\Delta_n = T_n^{1/2} W_n, \tag{5.4}$$

such that

$$\mathcal{L}[(\Delta_n, T_n)|P_{n,\theta}] \Longrightarrow \mathcal{L}(\Delta, T), \quad \Delta = T^{1/2}W,$$

where T is an a.s. p.d. random matrix and $W \sim N_k(0, I)$ independent of T. Then the sequence of models or experiments is *Locally Asymptotically Mixture of Normals* (LAMN). Clearly, under the LAMN conditions, the distribution of $\Delta|T$ is $N_k(0, T)$, and $\mathcal{E}[(h'\Delta)^2|T] = h'Th$. Under the LAMN conditions, the limiting distribution of T does not depend on the local parameter h; that is, $\mathcal{L}(T_n|\theta_n = \theta + \delta_n^{-1}h)$ has a limit independent of h. Moreover, it can be shown that the sequences of probability measures$\{P_{n,\theta}\}$ and $\{P_{n,\theta_n}\}$ are contiguous for every $h \in \mathbb{R}^k$.

In order to arrive at the efficient tests we need the following results stated in the form of lemmas. Justifications of the results stated below can be found in Le Cam and Yang (2000), Jeganathan (1995), Basu and Bhattacharya (1988, 1990, 1992), Davies (1985).

Lemma 5.1. *If the sequence of experiments $\{E_n\}$ satisfies the LAMN conditions at $\theta_0 \in \Theta$, then for every $h \in \mathbb{R}^k$, we have*

$$\mathcal{L}\{[\Lambda(\theta_n, \theta_0), \Delta_n, T_n]|P_{n,\theta_0}\} \implies \mathcal{L}(h'T^{1/2}W - \frac{1}{2}h'Th, T^{1/2}W, T), \tag{5.5}$$

$$\mathcal{L}\{[\Lambda(\theta_n, \theta_0), \Delta_n, T_n]|P_{n,\theta_n}\} \implies \mathcal{L}(h'T^{1/2}W + \frac{1}{2}h'Th, T^{1/2}W + Th, T), \tag{5.6}$$

where W is a $k \times 1$ standard normal vector independent of T.

Lemma 5.2. *If the sequence of experiments $\{E_n\}$ satisfies the LAMN conditions at $\theta_0 \in \Theta$, then for every $h \in \mathbb{R}^k$, we have*

$$\mathcal{L}\left(T_n^{-1/2}\Delta_n | P_{n,\theta_0}\right) \implies \mathcal{L}(W), \tag{5.7}$$

$$\mathcal{L}(T_n^{-1/2}\Delta_n | P_{n,\theta_n}) \implies \mathcal{L}(W + T^{1/2}h). \tag{5.8}$$

Lemma 5.3. *If the sequence of experiments $\{E_n\}$ satisfies the LAMN conditions at $\theta_0 \in \Theta$, then for every $h \in \mathbb{R}^k$, we have the joint convergence of (Δ_n, T_n) as follows:*

$$\mathcal{L}\left[(\Delta_n, T_n) | P_{n,\theta_0}\right] \implies \mathcal{L}(\Delta, T), \text{ where } \Delta = T^{1/2}W, \tag{5.9}$$

$$\mathcal{L}\left[(\Delta_n, T_n) | P_{n,\theta_n}\right] \implies \mathcal{L}\left(\widetilde{\Delta}, T\right), \text{ where } \widetilde{\Delta} = T^{1/2}W + Th. \tag{5.10}$$

Exponential approximation result similar to Theorem 7 also holds for the LAMN models with respect to a certain truncated version of Δ_n. (See Theorem 9.4 in Roussas and Bhattacharya (2009)). However, the approximating family under LAMN framework no longer belongs to a standard exponential family, but to a curved exponential family, so defined by Efron (1975). Roughly speaking, an exponential family is curved when the dimensionality of the sufficient statistics for θ is larger than the dimensionality of θ. For example, the normal family $N(\theta, \theta^2)$, $\theta \in \mathbb{R}$, is a curved exponential family.

Let us consider the case where $k = 1$ and replace $\Delta_n(\theta_0)$, $T_n(\theta_0)$, $\Delta(\theta_0)$ and $T(\theta_0)$ by Δ_n, T_n, Δ and T, respectively. We can define the following sequence of test functions $\{\widetilde{\varphi}_n\}$ for testing $H_0: \ \theta = \theta_0$ against $A'': \ \theta = \theta_n = \theta_0 + \delta_n^{-1}h$:

$$\widetilde{\varphi}_n = \widetilde{\varphi}_n(\Delta_n(\theta_0)) = \begin{cases} 1 & \text{if } \Lambda(\theta_n, \theta_0) > c''_n \\ \gamma''_n & \text{if } \Lambda(\theta_n, \theta_0) = c''_n \\ 0 & \text{otherwise,} \end{cases} \tag{5.11}$$

where the sequences $\{c''_n\}$ and $\{\gamma''_n\}$ are determined by the requirement

$$\mathcal{E}_{\theta_0}\widetilde{\varphi}_n = \alpha \text{ for all } n. \tag{5.12}$$

Using (5.2) for $k = 1$, we can rewrite (5.11) as

$$\widetilde{\varphi}_n = \widetilde{\varphi}_n(\Delta_n(\theta_0)) = \begin{cases} 1 & \text{if } h\Delta_n - \frac{1}{2}h^2T_n > c''_n \\ \gamma''_n & \text{if } h\Delta_n - \frac{1}{2}h^2T_n = c''_n \\ 0 & \text{otherwise,} \end{cases} \tag{5.13}$$

where the sequences $\{c''_n\}$ and $\{\gamma''_n\}$ are determined by the requirement $\mathcal{E}_{\theta_0}\widetilde{\varphi}_n = \alpha$, for all n.

Thus, we have the following theorem:

Theorem 13. *Let the sequence of experiments $\{E_n\}$ satisfy the LAMN condition at $\theta_0 \in \Theta$ and let $\{\widetilde{\varphi}_n\}$ be a sequence of tests defined by (5.11) and (5.12). Then for every $h(\neq 0) \in \mathbb{R}$, the sequence of tests $\{\widetilde{\varphi}_n\}$ is asymptotically most powerful of size α, and the upper bound of the asymptotic power function $\beta_{\widetilde{\varphi}_n}(\theta_n)$ of the test at θ_n is given by*

$$\lim_{n \hookrightarrow \infty} \sup \beta_{\widetilde{\varphi}_n}(\theta_n) \leq P\left(hT^{1/2}W + \frac{h^2}{2}T > c''\right),$$

where W is a $N(0,1)$ variable independent of T, and c'' is such that

$$P_{\theta_0}\left(hT^{1/2}W - \frac{h^2}{2}T > c''\right) = \alpha.$$

Proof. It is clear that the sequence of tests $\{\widetilde{\varphi}_n\}$ defined by (5.11) and (5.12) is the Neyman-Pearson test and hence is most powerful for testing $H_0: \ \theta = \theta_0$ against $A'': \ \theta = \theta_n$. Now we have seen that under LAMN condition $\Lambda(\theta_n, \theta_0)$ and $h\Delta_n - \frac{1}{2}h^2 T_n$ are differentially (asymptotically) equivalent in the sense that the difference between $\Lambda(\theta_n, \theta_0)$ and $h\Delta_n - \frac{1}{2}h^2 T_n$ converges to zero in P_{n,θ_0}-probability. Since the probability measures P_{n,θ_0} and P_{n,θ_n} are contiguous, the difference between $\Lambda(\theta_n, \theta_0)$ and $h\Delta_n - \frac{1}{2}h^2 T_n$ also converges to zero in P_{n,θ_n}-probability. Thus, tests given by (5.11) and (5.13) are differentially (asymptotically) equivalent and hence the tests defined by (5.13) and (5.12) are also asymptotically most powerful. Now, let A be the class of all asymptotically size α tests $\widetilde{\varphi}'_n$ for which the power function $\beta_{\widetilde{\varphi}'_n}(\theta)$ is continuous in θ and $\alpha_n = \mathcal{E}_{\theta_0}\widetilde{\varphi}'_n \to \alpha$, where $0 < \alpha < 1$. Then we have

$$\begin{aligned}
\beta_{\widetilde{\varphi}'_n}(\theta_n) &\leq P_{\theta_n}(\Lambda(\theta_n, \theta_0) > c''_n) + \gamma''_n P_{\theta_n}(\Lambda(\theta_n, \theta_0) = c''_n) \\
&= P_{\theta_n}\left(h\Delta_n - \frac{h^2}{2}T_n > c''_n\right) + \gamma''_n P_{\theta_n}\left(h\Delta_n - \frac{h^2}{2}T_n = c''_n\right) \\
&\to P\left(hT^{1/2}W + h^2 T - \frac{h^2}{2}T > c''\right) + \gamma'' P\left(hT^{1/2}W + h^2 T - \frac{h^2}{2}T = c''\right) \\
&= P\left(hT^{1/2}W + \frac{h^2}{2}T > c''\right), \qquad (5.14)
\end{aligned}$$

using Lemma 3.1 and the fact that W is a $N(0,1)$ variable which is independent of T .

The constant c'' ($c''_n \to c''$ as $n \to \infty$) is such that

$$P_{\theta_0}\left(hT^{1/2}W - \frac{h^2}{2}T > c''\right) = \alpha. \qquad (5.15)$$

Since $\widetilde{\varphi}'_n \in A$ is arbitrary, we replace $\widetilde{\varphi}'_n$ by $\widetilde{\varphi}_n$ in the above proof and Theorem 13 follows. ■

It is usually difficult to find a test $\widetilde{\varphi}'_n$ for which the power function $\beta_{\widetilde{\varphi}'_n}(\theta)$ has the largest value for each θ compared to any other test of the same size $\alpha_n(\theta_0)$. However, it is possible to compare two tests, $\widetilde{\varphi}'_n$ and $\widetilde{\varphi}''_n$, of the same size $\alpha_n(\theta_0)$, by comparing their respective power functions $\beta^1_{\widetilde{\varphi}'_n}(\theta)$ and $\beta^2_{\widetilde{\varphi}''_n}(\theta)$, which satisfy one kind of inequality for some values of θ and the reverse inequality for the other values of θ in the admissible range of θ (see the plot of the power functions in Bhattacharya and Roussas(1999)); i.e.,

$$\beta^1_{\widetilde{\varphi}'_n}(\theta) \geq \beta^2_{\widetilde{\varphi}''_n}(\theta), \text{ for some values of } \theta,$$

and

$$\beta^1_{\widetilde{\varphi}'_n}(\theta) \leq \beta^2_{\widetilde{\varphi}''_n}(\theta), \text{ for the other values of } \theta \text{ in the admissible range of values of } \theta.$$

Such knowledge is very useful in comparing two tests of the same size, but in practice finding the power function for all admissible values of θ is extremely difficult. Thus for comparison purposes, we prefer to depend on a single numerical measure which is easy to compute, and the approach of comparison must be asymptotic in nature; i.e., such a measure must be based on large samples. A first order criterion of asymptotic efficiency of a test is defined as follows: a test is efficient when it maximizes the derivative of the power function $\beta_{\widetilde{\varphi}'_n}(\theta)$ at the point θ_0; i.e., $\beta'_{\widetilde{\varphi}'_n}(\theta_0)$ is maximum (Rao 1973). If an AUMP test does not exist, then we try to find tests which are best for alternatives close to the null hypothesis (close in the sense that they converge to the null hypothesis at a fixed rate) and hope that those tests will also perform well for distant alternatives. On the basis of this idea, locally efficient tests may be derived.

The criterion for local efficiency of a test is given by the Pitman power which is defined below.

Definition 5.1. The *Pitman power* $\beta_p(\theta_0; h, \alpha)$ for any test $\widetilde{\varphi}_n \in A$, A being the class of all asymptotically size-α tests for which the power function $\beta_{\widetilde{\varphi}_n}(\theta)$ is continuous in θ and $\alpha_n = \alpha_n(\theta_0) = E_{\theta_0}\widetilde{\varphi}_n \to \alpha$, where $0 < \alpha < 1$, is defined as

$$\beta_p(\theta_0; h, \alpha) = \lim_{n\to\infty} \beta_{\widetilde{\varphi}_n}(\theta_n), \text{ where } \theta_n = \theta_n(h) = \theta_0 + \delta_n^{-1}h \text{ and } \theta_n \to \theta_0, \tag{5.16}$$

when the limit exists.

Another criterion, known as local power criterion, is defined below.

Definition 5.2. The *local power* $\beta_l(\theta_0; \alpha)$ of any test $\widetilde{\varphi}_n \in A$ is defined as

$$\begin{aligned}\beta_l(\theta_0; \alpha) &= \lim_{n\to\infty} \left\{ \delta_n^{-1} \left[\frac{d\beta_{\widetilde{\varphi}_n}(\theta_n)}{d\theta}\right]\bigg|_{\theta=\theta_0} \right\} \\ &= \lim_{n\to\infty} [\delta_n^{-1}\beta'_{\widetilde{\varphi}_n}(\theta_0)], \text{ where } \beta'_{\widetilde{\varphi}_n}(\theta_0) = \left.\frac{d\beta_{\widetilde{\varphi}_n}(\theta_n)}{d\theta}\right|_{\theta=\theta_0}, \end{aligned}\tag{5.17}$$

when the limit exists.

It is clear from (5.14) that the test defined by (5.13) and (5.12) maximizes the limiting power function for a specified sequence of alternatives $\theta_n = \theta_n(h) = \theta_0 + \delta_n^{-1}h$ in the following sense:

Let the derivative of the power function at θ_0 exist and be non-zero. Thus

$$\beta'_{\widetilde{\varphi}_n}(\theta_n) = \lim_{\theta_n \to \theta_0} \frac{\beta_{\widetilde{\varphi}_n}(\theta_n) - \beta_{\widetilde{\varphi}_n}(\theta_0)}{\theta_n - \theta_0} \text{ exists and is non-zero.}$$

Hence as $\theta_n \to \theta_0$ at a specified rate, we have the following approximation (which also can be obtained via a Taylor expansion of $\beta_{\widetilde{\varphi}_n}(\theta_n)$ around θ_0 and with $\beta_{\widetilde{\varphi}_n}(\theta_0) = \alpha$)

$$\beta_{\widetilde{\varphi}_n}(\theta_n) \simeq (\theta_n - \theta_0)\beta'_{\widetilde{\varphi}_n}(\theta_n) + \alpha. \tag{5.18}$$

Now, for $h > 0$ (i.e., for $\theta_n - \theta_0 > 0$), $\beta_{\widetilde{\varphi}_n}(\theta_n)$ is maximized in the neighborhood of θ_0 if $\beta'_{\widetilde{\varphi}_n}(\theta_n)$ is maximized in the neighborhood of θ_0. The test defined by (5.13) and (5.12), being differentially (asymptotically) equivalent to the Neyman-Pearson most powerful test defined by (5.11) and (5.12), is most powerful for testing $H_0 : \ \theta = \theta_0$ against $A'' : \ \theta = \theta_n$, and hence $\beta'_{\widetilde{\varphi}_n}(\theta_n)$ becomes maximum; i.e.,

$$\beta'_{\widetilde{\varphi}_n}(\theta_n) \geq \beta'_{\widetilde{\varphi}'_n}(\theta_n), \text{ for any other test } \widetilde{\varphi}'_n \in A.$$

It is readily seen from the above discussion that the Pitman power is maximized for this test.

According to the local power criterion (5.17), we have

$$\begin{aligned}
\beta_l(\theta_0; \alpha) &= \lim_{n\to\infty} \left\{ \delta_n^{-1} \left[\frac{d\beta_{\widetilde{\varphi}_n}(\theta_n)}{d\theta} \right] \Bigg|_{\theta=\theta_0} \right\} \\
&= \lim_{n\to\infty} [\delta_n^{-1} \beta'_{\widetilde{\varphi}_n}(\theta_0)] \\
&= \lim_{n\to\infty} \left\{ \delta_n^{-1} \lim_{\theta_n \to \theta_0} \left[\frac{\beta_{\widetilde{\varphi}_n}(\theta_n) - \beta_{\widetilde{\varphi}_n}(\theta_0)}{\theta_n - \theta_0} \right] \right\} \\
&= \lim_{n\to\infty} h^{-1} [\beta_{\widetilde{\varphi}_n}(\theta_n) - \beta_{\widetilde{\varphi}_n}(\theta_0)] \\
&= \lim_{n\to\infty} h^{-1} \beta_{\widetilde{\varphi}_n}(\theta_n) - \frac{\alpha}{h}.
\end{aligned} \tag{5.19}$$

Again, (5.19) establishes that the test which will have maximum Pitman power will also have maximum local power for fixed h; i.e., for a chosen sequence of alternatives $\theta_n(h) = \theta_0 + \delta_n^{-1} h$.

It is to be noted that the power function of the sequence of tests defined by (5.13) and (5.12) is not free of h even in the limit, since the critical region of the tests $\widetilde{\varphi}_n$ defined by $R_n = h\Delta_n - \frac{1}{2}h^2 T_n > c''_n$ is not free of h even in the limit. This is due to the fact that the convergence $h\Delta_n - \frac{1}{2}h^2 T_n \Longrightarrow hT^{1/2}W + h^2 T - \frac{h^2}{2}T = hT^{1/2}W + \frac{h^2}{2}T$ under P_{θ_n} holds. As a result, the sequence of tests is not AUMP for different values of h, though it is asymptotically most powerful against the chosen sequence of alternatives $\theta_n(h) = \theta_0 + \delta_n^{-1} h$. Hence, under LAMN conditions, no AUMP test exists for testing $H_0 : \ \theta = \theta_0$ against $A'' : \ \theta = \theta_n = \theta_0 + \delta_n^{-1} h$. Bhattacharya and Roussas (1999) have discussed examples from the LAMN model, where the asymptotically efficient tests for the parameter of an autoregressive process of order one have been derived. Performances of different tests have been compared with respect to their Pitman power at different values of h. For better understanding of the complexities that may arise in testing $H_0: \ \theta = \theta_0$ against $A'': \ \theta = \theta_n = \theta_0 + \delta_n^{-1} h$, under LAMN conditions, the interested reader is referred to Section 6.8 in Le Cam and Yang (2000, pp.164–168).

References

[1] Basu, A.K. and Bhattacharya, D. (1988). Local asymptotic mixed normality of the log-likelihood based on stopping times. *Calcutta Statist. Assoc. Bull.*, **37**, 143–159.

[2] Basu, A.K., and Bhattacharya, D. (1990). Weak convergence of randomly stopped log-likelihood ratio statistic to mixed Gaussian process. *Calcutta Statist. Assoc.Bull.*, **39**, 137–149.

[3] Basu, A.K., and Bhattacharya, D. (1992). On the asymptotic non-null distribution of randomly stopped log-likelihood ratio statistic. *Calcutta Statist. Assoc.Bull.*, **42**, 255–260.

[4] Bhattacharya, D. and Roussas, George G. (1999). On asymptotically efficient tests for autoregressive processes. *Stochastic Modelling and Applications*, **2**, 17–30.

[5] Davies, R.B. (1985). Asymptotic inference when the amount of information is random. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, **II**, L. Le Cam, R.A. Olson eds., 841–864, Wodsworth, Monterery, California.

[6] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.*, **3**, 1189–1242.

[7] Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series model. *Econometric Theory*, **11**, 818–887.

[8] Johnson, R.A. and Roussas, George G. (1969). Asymptotically most powerful tests in Markov processes. *Ann. Math. Statist.*, **40**, 1207–1215.

[9] Johnson, R.A. and Roussas, George G. (1970). Asymptotically optimal tests in Markov processes. *Ann. Math. Statist.*, **41**, 918–938.

[10] Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. of California Publications in Statistics*, **3**, 37–98.

[11] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer series in Statistics. Springer–Verlag.

[12] Le Cam, L. and Yang, G.L. (2000). *Asymptotics in Statistics, Some Basic Concepts.* Lecture Notes in Statistics. Springer–Verlag.

[13] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications* (2nd edition), John Wiley, New York.

[14] Roussas, George G. (1972). *Contiguity of Probability Measures: Some applications in statistics*, Cambridge University Press, Cambridge.

[15] Roussas, George G. (1979). Asymptotic distribution of the log-likelihood function for stochastic processes. *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **47**, 31–46.

[16] Roussas, George G. (2008). *Contiguity of Probability Measures: Some applications in statistics*, Cambridge University Press, Cambridge.

[17] Roussas, George G. and Bhattacharya, D. (2007). Asymptotic expansion, exponential approximation and Hajek-Inagaki representation theorem under a general dependence set-up. *Proceedings of the 20th Panhellenic Statistical Conference*, Nikosia, Cyprus, 45–65 (Invited keynote lecture for the annual meeting).

[18] Roussas, George G. and Bhattacharya, D. (2009). Revisiting Local Asymptotic Normality (LAN) and passing on to Local Asymptotic Mixed Normality (LAMN) and Local Asymptotic Quadratic (LAQ) experiments. Technical Report, University of California, Davis, submitted for publication.

[19] Stamatelos, G.D. (1976). Asymptotic distribution of the log-likelihood function for stochastic processes: some examples. *Bulletin Société Mathématique Grèce*, 12, 92–116.

[20] Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Statist.*, **12**, 1–19.

[21] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54**, 426–482.