Ann. Inst. Statist. Math.
Vol. 43, No. 2, 245–260 (1991)

A GEOMETRIC LOOK AT NUISANCE PARAMETER EFFECT OF LOCAL POWERS IN TESTING HYPOTHESIS

SHINTO EGUCHI

Department of Mathematics, Shimane University, Matue 690, Japan

(Received October 24, 1989; revised February 22, 1990)

Abstract. This paper is concerned with the theory of testing hypothesis with composite null hypothesis or with nuisance parameters. The asymptotic behaviour of the likelihood ratio and the associated test statistics are investigated. Under a class of local alternatives with local orthogonality relative to the nuisance parameter vector, a unique decomposition of local power is presented. The decomposition consists of two parts; one is the influence of nuisance parameters and the other is the power corresponding to the simple case where the nuisance parameters are known. The decomposition formula is applied to some examples, including the gamma, Weibull and location-scale family.

Key words and phrases: Curvature tensor, composite null hypothesis, likelihood ratio, local power, local unbias, nuisance parameters, variance coefficient.

1. Introduction

We are concerned with testing hypothesis based on a random sample from a distribution P_ω depending on a parameter vector ω, which varies in Ω. Let ω be partitioned into two vectors ξ and θ of dimension p and q, respectively, so that $\Omega = \Xi \times \Theta$. The problem of testing hypothesis H: $\theta = \theta_0$ with a nuisance vector ξ is considered. Thus H constitutes the subplane $\Xi \times \{\theta_0\}$ in Ω and is referred to as the composite null hypothesis (see Cox and Hinkley (1974)). Under a sequence of alternatives K_n: $\theta = \theta_0 + \epsilon/\sqrt{n}$ with arbitrarily fixed ξ, the likelihood ratio or the associated test statistics asymptotically has a noncentral chi-square distribution with q degrees of freedom and a noncentrality parameter $\nu = \epsilon^T I(\xi, \theta_0)\epsilon$, where $I(\xi, \theta_0)$ denotes the Fisher information matrix of θ, evaluated at (ξ, θ_0). The higher-order asymptotic powers are investigated by Hayakawa (1975) (see Harris and Peers (1980) for the efficient scores statistics and also Peers (1971) for a simple null hypothesis case).

In the present paper we deal with the more general problem of testing the hypothesis H: $\omega \in \Omega_0$ against K: $\omega \notin \Omega_0$. Here Ω_0 is a p-dimensional subsurface of Ω, which is practically introduced by either a parametrized case or a constrained case as in Section 2. The general form can be locally reduced to the partitioned form by a one-to-one transformation ϕ of $\mathcal{N}_0$ into $\Xi \times \Theta$ such that $\phi(\Omega_0) =$

$\Xi \times \{\theta_0\}$, where $\mathcal{N}_0$ denotes a tubular neighbourhood around Ω_0. The domain $\mathcal{N}_0$ of ϕ cannot generally be extended to the full space Ω. However the local powers can be given by application of Hayakawa's formula to the reduced form. Such transformations ϕ constitute a subclass of transformations on Ω. Here an embarrassing aspect arises; the form of null hypothesis and the corresponding test statistics are independent of the choice of ϕ; nevertheless, this is not so of local powers.

The objective of the paper is to give intrinsic characteristics of the test procedures in the local powers. For this we modify the sequence K_n into K_n^* such that the Kullback-Leibler divergence from K_n^* to H is asymptotically the noncentrality parameter ν to $o(n^{-1/2})$. Furthermore we give some insight into the behaviour of K_n^* in the large sample size n. In effect we assume the local orthogonality of ξ with θ in the sense that the mixed part of the information matrix of ξ relative to θ is zero when evaluated at θ_0 (see Cox and Reid (1987) for statistical interpretation). We present a decomposition formula of local powers under the adjusted sequence K_n^* of alternatives. The decomposition consists of the sum of the two parts; one is the local power for the case with known nuisance parameters and the other part represents the effect of nuisance parameters. The latter part is expressed by the imbedding curvature tensors.

McCullagh and Cox (1986) have discussed the Bartlett adjustment for chi-square approximation from the invariance viewpoint (see also Voss (1989) for further discussion). Kumon and Amari (1983) have given an invariant representation of local powers for the simple case of one-parameter curved exponential model in terms of modifying the range of the signed root squares of the test statistics. In a further development of this result, Amari (1985) has also pointed out that the imbedding curvature plays a fundamental role for expressing the effect of nuisance parameters, of which intuition coincides with our result from another approach (see also Kumon and Amari (1988)).

The present paper adopts the differential geometric framework (see Amari (1985) and Eguchi (1985) for detailed discussion). In Section 2 we investigate the effect of the composite null hypothesis. First two practical cases of the composite null hypothesis are introduced as are the corresponding partitioned forms. The simple case where the nuisance parameters are known is reviewed for the comparison with the composite case. We present a decomposition formula of the local powers in the composite null hypothesis case. Section 3 introduces the applications of the decomposition formula to some examples. Section 4 discusses about local unbiasedness.

2. Effect of nuisance parameters

Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution having density $f(x \mid \omega)$, where the space of possible values of ω is assumed to be an open subset of R^{p+q}, say Ω. Let Ω_0 be a p-dimensional submanifold of Ω. We consider a problem of testing hypothesis H: $\omega \in \Omega_0$ against K: $\omega \notin \Omega_0$. We wish to look at the effect of the composite null hypothesis from a geometric point of view.

2.1 *The forms of hypotheses*

For the test problem, the following three types of test procedures based on the log-likelihood function,

$$\bar{l}(\omega) = \frac{1}{n}\sum_{j=1}^{n} \log f(x_j \mid \omega)$$

are available:

the likelihood ratio test statistic, $S_1 = 2n\{\bar{l}(\hat{\omega}) - \bar{l}(\hat{\omega}_0)\}$,
the efficient scores test statistic, $S_2 = n\bar{e}(\hat{\omega}_0)^T I_\omega(\hat{\omega}_0)^{-1}\bar{e}(\hat{\omega}_0)$ and
the MLE test statistic, $S_3 = n(\hat{\omega} - \hat{\omega}_0)^T I_\omega(\hat{\omega})(\hat{\omega} - \hat{\omega}_0)$

with $\bar{e}(\omega) = (\partial/\partial\omega)\bar{l}(\omega)$ and the information matrix $I_\omega(\omega_1)$ of ω when evaluated at ω_1. Here $\hat{\omega}$ is the MLE of ω by maximization over Ω and $\hat{\omega}_0$ is the restricted MLE to Ω_0.

All the test statistics S_i's enjoy invariance under one-to-one transformations of the sample space. Furthermore, consider a transformation ϕ of ω into ω^*, so that the form of hypothesis is rewritten as H: $\omega^* \in \phi(\Omega_0)$ and K: $\omega^* \notin \phi(\Omega_0)$. Then both of the statistics S_1 and S_2 keep invariant under such a transformation ϕ. Generally S_3 is slightly changed to

$$S_3^* = n(\phi(\hat{\omega}) - \phi(\hat{\omega}_0))^T I_{\omega^*}(\phi(\hat{\omega}))(\phi(\hat{\omega}) - \phi(\hat{\omega}_0)),$$

where $I_{\omega^*}(\phi(\hat{\omega}))$ denotes the information matrix of ω^*, evaluated at $\phi(\hat{\omega})$, i.e.

$$I_{\omega^*}(\phi(\hat{\omega})) = \left(\frac{\partial\phi}{\partial\omega}(\hat{\omega})^T\right)^{-1} I_\omega(\hat{\omega}) \left(\frac{\partial\phi}{\partial\omega}(\hat{\omega})\right)^{-1}.$$

In a subsequent discussion, the difference between S_3 and S_3^* will be expressed as only the terms coming from the effect of parameters in the local powers.

Henceforth we introduce some notations of differential geometry for the parameter vector $(\xi,\, \theta)$ which leads to the reduced form of the hypothesis. The information metric g, the skewness tensor T and the exponential connection $\Gamma^{(e)}$ have the components with respect to $(\xi,\, \theta)$-coordinates

$$(2.1) \qquad g_{\alpha\beta} = E(e_\alpha e_\beta), \qquad T_{\alpha\beta\gamma} = E(e_\alpha e_\beta e_\gamma) \qquad \text{and} \qquad \Gamma^{(e)}_{\alpha\beta,\gamma} = E(\partial_\alpha e_\beta e_\gamma)$$

respectively, for $1 \le \alpha,\, \beta,\, \gamma \le p+q$, where $e_\alpha = (\partial/\partial\zeta^\alpha)\log f(\cdot \mid \zeta)$ with $\zeta = (\xi,\, \theta)$. Furthermore we use the letters $a, b, c, \ldots$ for the indices of ξ-part and $i, j, k, \ldots$ for those of θ-part. We assume the local orthogonality of ξ to θ, which is defined by

$$(g_{\xi\theta})_0 = ((g_{ai})_0)_{a=1,\ldots,p,\, i=1,\ldots,q} = 0,$$

where the subscript 0 denotes the evaluation at $(\xi,\, \theta_0)$. Hereafter the subscript is omitted throughout the paper.

The condition implies that the θ-subcoordinate are orthogonally transverse to Ω_0. Hence the components of the embedding curvature tensor $H^{(m)}$ of Ω_0 with respect to $\Gamma^{(m)}$, or the m-curvature are written as

$$H^{(m)}_{ab,i} = -\Gamma^{(e)}_{ai,b}, \tag{2.2}$$

where $\Gamma^{(m)}$ denotes the mixture connection, which is conjugate to $\Gamma^{(e)}$. Similarly the e-curvature $H^{(e)}$ has the components

$$H^{(e)}_{ab,i} = H^{(m)}_{ab,i} - T_{abi}.$$

Define a line $C_\epsilon = \{(\xi, \theta(t));\ t \in (-1, 1)\}$ by $\theta(t) = \theta_0 + t\epsilon$ with the direction vector ϵ, so that C_ϵ goes through the local alternative K_n at $t = n^{-1/2}$ and traverses Ω_0 at $t = 0$. For a subsequent discussion we introduce the following geometric quantities associated with C_ϵ;

$$\begin{aligned} &\nu = g_{ij}\epsilon^i\epsilon^j, \quad \mu = T_{ijk}\epsilon^i\epsilon^j\epsilon^k, \quad \mu_1 = T_{ijk}g^{jk}\epsilon^i, \\ &\Lambda_1 = \Gamma^{(e)}_{ij,k}\epsilon^i\epsilon^j\epsilon^k, \quad \Lambda_2 = -(2\Gamma^{(e)}_{ij,k} + \Gamma^{(e)}_{ki,j})g^{jk}\epsilon^j, \\ &h^{(m)} = H^{(m)}_{ab,i}g^{ab}\epsilon^i \quad \text{and} \quad h^{(e)} = h^{(m)} - T_{abi}g^{ab}\epsilon^i. \end{aligned} \tag{2.3}$$

when evaluated at (ξ, θ_0). Here and hereafter we use the summation convention. We note that ν, μ, μ_1, $h^{(m)}$ and $h^{(e)}$ are invariant under transformations of (ξ, θ) with local orthogonality. The term ν is the square length of the tangent vector of C_ϵ at θ_0. The coefficients μ and μ_1 come from the skewness along the θ-parameter curves to be tested in the parametric model. The terms $h^{(e)}$ and $h^{(m)}$ are led to by the embedding curvatures $H^{(e)}$ and $H^{(m)}$ of the null hypothesis or the nuisance parameters. On the other hand, Λ_1 and Λ_2 are parametrization-dependent. We shall show that only the terms express the local powers in the test procedures S_i's.

2.2 *Simple null hypothesis case*

We pay attention to the case where the parameter vector ξ is known. Then the corresponding statistics are given by

$$\begin{aligned} S_{10} &= S_{10}(\xi) = 2n\{\bar{l}(\xi, \tilde{\theta}) - \bar{l}(\xi, \theta_0)\}, \\ S_{20} &= S_{20}(\xi) = n\bar{e}_1(\xi, \theta_0)\bar{e}_j(\xi, \theta_0)g^{ij}(\xi, \theta_0), \\ S_{30} &= S_{30}(\xi) = n(\tilde{\theta} - \theta_0)^i(\tilde{\theta} - \theta_0)^j g_{ij}(\xi, \tilde{\theta}), \end{aligned} \tag{2.4}$$

where $\tilde{\theta}$ is the MLE of θ in the case where ξ is known. Thus the comparison between S_i's and S_{i0}'s will illustrate what influence ξ exerts on the performance of test procedures.

We note the parametrization-invariance of S_{10} and S_{20}. Of course both of these two statistics have invariant distributions under the null hypothesis but this is not the case for the asymptotic powers under the local alternatives K_n. This is caused by the choice of $\{K_n\}$. For discussion consistent with the null case we wish to obtain the invariant expression of local power. Hence we introduce an adjusted local alternative K_n^*: $\theta = \theta_0 + \epsilon/\sqrt{n} + \Delta(\epsilon)/n$ with the i-th component

$$\Delta^i(\epsilon) = -\frac{1}{2}\Gamma_{jk}^{(\alpha)i}(\xi,\, \theta_0)\epsilon^j\epsilon^k.$$

with $\alpha = -1/3$, where $\Gamma_{jk}^{(\alpha)i}$ denote the components of the α-connection $\Gamma^{(\alpha)} \equiv (1-\alpha)\Gamma^{(m)}/2 + (1+\alpha)\Gamma^{(e)}/2$. We have the following theorem from the geometric viewpoint.

THEOREM 2.1. *Under the adjusted local alternatives K_n^*, the asymptotic densities of the statistics S_1, S_2 and S_3 are given, to $o(1/\sqrt{n})$, by*

$$(2.5)\quad g_{10}(s) = f_q(s \mid \nu) + \frac{\mu}{6\sqrt{n}}\{f_q(s \mid \nu) - 2f_{q+2}(s \mid \nu) + f_{q+4}(s \mid \nu)\},$$

$$(2.6)\quad g_{20}(s) = f_q(s \mid \nu) + \frac{1}{6\sqrt{n}}[\mu\{f_q(s \mid \nu) - 2f_{q+2}(s \mid \nu) + f_{q+6}(s \mid \nu)\} - 3\mu_1\{f_{q+2}(s \mid \nu) - f_{q+4}(s \mid \nu)\}]$$

and

$$(2.7)\quad \begin{aligned} g_{30}(s) = f_q(s \mid \nu) + \frac{1}{\sqrt{n}}[&\mu\{f_q(s \mid \nu) - 2f_{q+2}(s \mid \nu) + f_{q+6}(s \mid \nu)\} \\ &- \mu_1\{f_{q+2}(s \mid \nu) - f_{q+4}(s \mid \nu)\} \\ &- 3\Lambda_1\{f_{q+4}(s \mid \nu) - f_{q+6}(s \mid \nu)\} \\ &+ \Lambda_2\{f_{q+2}(s \mid \nu) - f_{q+4}(s \mid \nu)\}], \end{aligned}$$

where $f_q(s \mid \nu)$ denotes the noncentral chi-square density function with q degrees of freedom and a noncentral parameter ν.

The proof will be outlined in the Appendix.

Thus the invariances of ν, μ and μ_1 lead to the invariant densities of g_{10} and g_{20}. The noncentral parameter ν is the square length of the tangent vector of C_ϵ at θ_0.

Alternatively, the statistic S_{30} does not have such an invariant property. Thus the terms Λ_1 and Λ_2 in g_{03} tell us the dependence of the statistic S_{30}. Let ϕ_0 be a transformation of θ into θ^* and let S_{30}^* be the corresponding statistic to S_{30} via ϕ_0. We see that the density of S_{03}^* is the right-hand side of (2.7) with coefficients

$$\Lambda_\kappa^* = \Lambda_\kappa + \frac{\partial\phi_0^m}{\partial\theta^i\partial\theta^j}\frac{\partial\phi_0^l}{\partial\theta^k}g_{ml}\epsilon^i\epsilon^j\epsilon^k$$

in place of Λ_κ for $\kappa = 1, 2$, noting the transformation rule of $\{\Gamma_{ij,k}^{(e)}\}$ as coefficients of affine connection. Specifically we choose ϕ_0 as

$$\phi_0^i(\theta) = (\theta - \theta)^i - \frac{1}{2}\Gamma_{jk}^{(e)i}(\theta_0)(\theta - \theta_0)^j(\theta - \theta_0)^k,$$

so that the terms Λ_1^* and Λ_2^* vanish and μ and μ_1 keep invariant in (2.7).

The Kullback-Leibler divergence from K to H is defined by

$$\rho(K, H) = \int \{\log f(y_n \mid \omega) - \log f(y_n \mid \omega_0)\} f(y_n \mid \omega) dm(y_n)$$

with $y_n = (x_1, \ldots, x_n)$ and the dominating measure m, where the parameters ω and ω_0 designate K and H, respectively. Then we see that the original local alternatives K_n satisfy

$$\rho(K_n, H) = \nu + O(n^{-1/2}),$$

and further the adjusted alternatives K_n^* satisfy

$$\rho(K_n^*, H) = \nu + o(n^{-1/2}),$$

which implies that the ρ-sphere centered at K_n^* is approximated by the ellipsoid $\{(\xi, \theta_0 + \epsilon);\ \nu \le c\}$ in Θ up to $o(n^{-1/2})$. Furthermore we have another viewpoint. Define a curve $C_\epsilon^* = \{(\xi, \theta(t));\ t \in (-1, 1)\}$ by $\theta(t) = \theta_0 + \epsilon t + \Delta(\epsilon)t^2$, so that C_ϵ^* goes through K_n^* when $t = n^{-1/2}$ and traverses the subsurface Ω_0 at $t = 0$. Thus we observe that the curve satisfies

$$\ddot{\theta}(0)^i + \Gamma_{jk}^{(\alpha)i}(\xi, \theta_0)\dot{\theta}^j(0)\dot{\theta}^k(0) = 0 \qquad (i = 1, \ldots, q),$$

which implies that C_ϵ^* is orthogonally $\Gamma^{(\alpha)}$-transverse to Ω_0 (see Eguchi (1983) for the relation of such a transversality with the estimation theory).

2.3 *Decomposition of local powers*

We now return to the general form of the composite null hypothesis H: $\omega \in \Omega_0$, which reduces to H: $\theta = \theta_0$ with nuisance parameters ξ. We assume the local orthogonality condition as introduced in Subsection 2.2. Under the adjusted local alternatives $\{K_n^*\}$ as defined in Subsection 2.2, we have the following asymptotic result, of which proof will be given in the Appendix.

THEOREM 2.2. *The asymptotic densities $g_i(s)$ of S_i $(i = 1, 2, 3)$ under the adjusted local alternatives K_n^* are decomposed into*

$$g_1(s) = g_{10}(s) - \frac{h^{(m)}}{\sqrt{n}}\{f_q(s \mid \nu) - f_{q+2}(s \mid \nu)\},$$

$$g_2(s) = g_{20}(s) - \frac{h^{(m)}}{\sqrt{n}}\{f_q(s \mid \nu) - f_{q+2}(s \mid \nu)\},$$

$$g_3(s) = g_{30}(s) + \frac{h^{(e)}}{\sqrt{n}}\{f_q(s \mid \nu) - f_{q+2}(s \mid \nu)\},$$

where g_{i0}'s are defined in Theorem 2.1.

From Theorem 2.2 it follows that the local powers are expressed as

$$\begin{aligned}&\Pr(S_i \ge s \mid K_n^*)\\ &\quad = \Pr(S_{i0} \le s \mid K_n^*) + \frac{1}{\sqrt{n}}h^{(m)}\{\Pr(\chi_{q,\nu}^2 \ge s) - \Pr(\chi_{q+2,\nu} \ge s)\},\end{aligned}$$

for $i = 1, 2$, where $\chi^2_{q,\nu}$ denotes the chi-square random variable with q degrees of freedom and noncentrality parameter ν. Thus the effect of nuisance parameters appears only as $h^{(m)}$ or $h^{(e)}$ in the local powers.

In practical situations, the subspace Ω_0 specifying the composite null hypothesis is given in either of two forms. One is of parametrized form, $\Omega_0 = k(\Xi) = \{k(\xi);\ \xi \in \Xi\}$ with an open subset Ξ of R^p, where $k(\xi)$ is a non-singular mapping from Ξ to Ω in the sense that the Jacobian matrix

$$J(\xi) = \left(J^\alpha_a(\xi) = \frac{\partial k^\alpha}{\partial \xi^a}(\xi)\right)_{\alpha=1,\dots,p+q,\ a=1,\dots,p}$$

is of rank p. Thus each component of ξ constitutes nuisance parameters in the testing problem. Alternatively the other is of constrained form

$$\Omega_0 = h^{-1}(\theta_0) = \{\omega \in \Omega_0;\ h(\omega) = \theta_0\}, \tag{2.8}$$

where h is a mapping of Ω into Θ with the Jacobian matrix of rank q. In this case the null hypothesis is H: $h(\omega) = \theta_0$. In such situations, the unrestricted MLE $\hat{\omega}$ of ω has often a simple form; the restricted MLE $\hat{\omega}_0$ is intractable and needs to be solved by some iterative methods. For example, assume that the underlying density $f(x \mid \beta)$ belongs to a regular exponential family of order $p+q$, $f(x \mid \beta) = \exp[b(x) + x^T\beta - \psi(\beta)]$ with expectation parameter vector $\omega = E_\beta X$. Let a null hypothesis H be of the form (2.8). Then the MLE test S_3 has a more tractable version $\tilde{S}_3 = h(\hat{\omega})V(\hat{\omega})h(\hat{\omega})^T$ depending only on $\hat{\omega}$, where $\hat{\omega} = \sum_{\alpha=1}^n x_\alpha/n$ and

$$V(\hat{\omega}) = \left(\frac{\partial h}{\partial \omega}(\hat{\omega})I_\omega(\hat{\omega})\frac{\partial h}{\partial \omega}(\hat{\omega})^T\right)^{-1}.$$

The simple version $\tilde{S}_3$ has the same asymptotic behaviour as S_3, of which proof will be given at Remark 1 in the Appendix.

We return to the parametrized form $\Omega_0 = \{k(\xi);\ \xi \in \Xi\}$. Let a $(p+q) \times q$ matrix $J^\perp(\xi)$ satisfy the orthogonality condition

$$J(\xi)^T I_\omega(k(\xi))J^\perp(\xi) = 0$$

with $p \times q$ zero matrix 0. The mapping of (ξ, θ) into ω,

$$\phi(\xi, \theta) = k(\xi) + J^\perp(\xi)(\theta - \theta_0)$$

leads to the direct application of Theorem 2.2; in $g_i(s)$'s

$$\nu = \delta^\alpha\delta^\beta g_{\alpha\beta}, \qquad \mu = \delta^\alpha\delta^\beta\delta^\gamma T_{\alpha\beta\gamma},$$
$$\mu_1 = \delta^\alpha T_{\alpha\beta\gamma}g^{\alpha\beta} - \delta^\alpha J^\beta_a J^\gamma_b T_{\alpha\beta\gamma}g^{ab},$$
$$h^{(m)} = \delta^\alpha H^{(m)}_{ab,\alpha}g^{ab} \quad \text{and} \quad h^{(e)} = h^{(m)} - \delta^\alpha J^\beta_a J^\gamma_b T_{\alpha\beta\gamma}g^{ab},$$

where $\delta^\alpha = J^{\perp\alpha}_i\epsilon^i$ and g^{ab} is the inverse element of $J^\alpha_a J^\beta_b g_{\alpha\beta}$ with $J^\alpha_a = \partial k^\alpha(\xi)/\partial\xi^a$ and the (α, i)-element $J^{\perp\alpha}_i$ of $J^\perp(\xi)$.

For the constrained form $\Omega_0 = \{\omega \in \Omega;\ h(\omega) = \theta_0\}$, we can express $g_i(s)$'s in terms of $h(\omega)$. Let $B_i^\alpha = g_{ij} g^{\alpha\beta} B_\beta^j$, where $B_\beta^j = \partial h^j / \partial \omega^\beta$ and let g_{ij} be the inverse element of $g^{ij} = B_\alpha^j B_\beta^i g^{\alpha\beta}$. Then in $g_i(s)$'s we have

$$\nu = \delta^\alpha \delta^\beta g_{\alpha\beta}, \quad \mu = \delta^\alpha \delta^\beta \delta^\gamma T_{\alpha\beta\gamma}, \quad \mu_1 = \delta^\alpha B_i^\beta B_j^\gamma T_{\alpha\beta\gamma} g^{ij},$$
$$h^{(m)} = \delta^\alpha (B_j^\beta B_k^\gamma g^{jk} - g^{\beta\gamma}) \Gamma^{(e)}_{\beta\gamma,\alpha} \quad \text{and}$$
$$h^{(e)} = h^{(m)} - \delta^\alpha (B_j^\beta B_k^\gamma g^{jk} - g^{\beta\gamma}) \Gamma^{(e)}_{\beta\gamma,\alpha}$$

where $\delta^\alpha = B_i^\alpha \epsilon^i$.

We introduce both of the two forms for the following example, which is related to the ABO-blood system.

Example 1. Let p, q and r be frequency parameters of the alleles A, B and O, respectively, with $r = 1 - p - q$. One considers the problem of testing whether a population is subject to a Hardy-Weinberg equilibrium. That is the cell parameters $(\omega_1, \omega_2, \omega_3, \omega_4)$ with four phenotypes A, B, AB and O are written as, respectively, $(p^2 + 2pr, q^2 + 2qr, 2pq, r^2) = \omega(p, q)$, say. Thus the nuisance parameter vector (p, q) designates the null hypothesis. The MLE of (p, q) under the null hypothesis is not known to be in a closed form, so that the three statistics S_i's have no closed form. Alternatively let $h(\omega) = \sqrt{(\omega_1 + \omega_4)} + \sqrt{(\omega_2 + \omega_4)} - \sqrt{\omega_4} - 1$, so that $h[\omega(p, q)] = 0$ for any p and q. Hence the hypothesis is rewritten as H: $h(\omega) = 0$ in the constrained form. Thus the simple version of S_3 is given by $\tilde{S}_3 = h(\hat{\omega})^2 / v(\hat{\omega})$ with the vector $\hat{\omega}$ of observed frequencies, where $v(\omega) = H_1^2/\omega_1 + H_2^2/\omega_2 + H_3^2/\omega_3 + H_4^2/\omega_4$. Here

$$H_1 = 1/2\sqrt{\omega_4} - 1/2\sqrt{(\omega_2 + \omega_4)}, \qquad H_2 = 1/2\sqrt{\omega_4} - 1/2\sqrt{(\omega_1 + \omega_4)},$$
$$H_3 = 1/2\sqrt{\omega_4} - 1/2\sqrt{(\omega_1 + \omega_4)} - 1/2\sqrt{(\omega_2 + \omega_4)} \quad \text{and}$$
$$H_4 = H_1 + H_2 + H_3.$$

Let a $(p + q) \times q$ matrix $J^\perp(\xi)$ satisfy the orthogonality condition

$$J(\xi)^T I_\omega(k(\xi)) J^\perp(\xi) = O \tag{2.9}$$

with $p \times q$ zero matrix O.

The following example relates to the functional relation of a sample in a location-scale model.

Example 2 (location-scale model). Let $f(x)$ be a probability density function on R such that $f(x) = f(-x)$ and $f > 0$ on R. In the location-scale model $\{\sigma^{-1} f((x - \mu)/\sigma);\ \sigma > 0, \mu \in R\}$, we wish to test the hypothesis for which $\omega = (\mu, \sigma)$ belongs to $\{k(\xi) = (\mu(\xi), \sigma(\xi));\ \xi \in \Xi\}$ with an open interval Ξ. The typical example is then the case of a known variation coefficient, say c, or

$k(\xi) = (ce^{\xi}, e^{\xi})$ for $\xi \in R$. Finney (1976) expands the model further in a regression situation. We here assume that the following moments exist;

$$a_1 = \int l'^2(t)f(t)dt, \qquad a_2 = \int t^2 l'^2(t)f(t)dt,$$
$$b_1 = \int tl'(t)l''(t)f(t)dt, \qquad b_2 = \int t^3 l'(t)l''(t)f(t)dt,$$

with $l(t) = \log f(t)$. For example, if f is the standard normal density, $f(t) = \exp(-t^2/2)/\sqrt{(2\pi)}$, then $(a_1, a_2, b_1, b_2) = (1, 3, 1, 3)$. We locally parametrize the alternative hypothesis around the null hypothesis by

$$\omega(\xi, \theta) = (\mu(\xi) + a_2\sigma'(\xi)\theta, \sigma(\xi) - a_1\mu'(\xi))$$

for $\theta \in (-\epsilon, \epsilon)$ with a sufficiently small $\epsilon > 0$. Note that $\omega(\xi, 0) = k(\xi)$ and that ξ is orthogonal to θ when $\theta = 0$.

Similarly, the constrained form leads to a one-to-one transformation ϕ of ω into (ξ, θ), defined by $(\xi, \theta) = \phi(\omega) = (h^*(\omega), h(\omega))$. Here $h^*(\omega) = J^*(\omega)\omega$ with a $(p+q) \times p$ matrix $J^*(\omega)$ satisfying the condition

$$\frac{\partial h}{\partial \omega}(\omega)^T I_\omega(\omega) J^*(\omega) = 0,$$

so that θ and ξ are orthogonal at θ_0.

3. Some examples

We apply the formula for the effect of nuisance parameters to some examples. First we review a testing problem concerning covariance in the m-variate normal distribution with mean vector ξ and covariance Σ.

Example 3. The null hypothesis considered is H: $\Sigma = I$, where ξ forms the vector of nuisance parameters orthogonal to Σ in the global sense. Sugiura (1973) has given the asymptotic distribution of the modified likelihood ratio statistic S_1 under a sequence of local alternatives K_n: $\Sigma = I + \epsilon/\sqrt{n}$ with symmetric matrix ϵ, see Hayakawa (1975) for the exact likelihood ratio test and Nagao (1974) for the modified scores statistic S_2. In our formulation, the adjusted local alternatives are given by

$$K_n^*: \Sigma = I + \epsilon/\sqrt{n} + \mathrm{tr}(\epsilon^2)I/6n.$$

Under K_n^* the asymptotic densities of S_1 and S_2 are then

$$(3.1)\quad g_1(s) = \left[f_q(s \mid \nu) + \frac{\mu}{6\sqrt{n}}\{f_q(s \mid \nu) - 2f_{q+2}(s \mid \nu) + f_{q+4}(s \mid \nu)\}\right] - \frac{1}{2\sqrt{n}}\operatorname{tr}(\epsilon)\{f_q(s \mid \nu) - f_{q+2}(s \mid \nu)\},$$

$$(3.2)\quad g_2(s) = \left[f_q(s \mid \nu) + \frac{1}{6\sqrt{n}}\{\mu f_q(s \mid \nu) - (3\mu_1 + 2\mu)f_{q+2}(s \mid \nu) + 3\mu_1 f_{q+4}(s \mid \nu) + \mu f_{q+6}(s \mid \nu)\}\right] - \frac{1}{2\sqrt{n}}\operatorname{tr}(\epsilon)\{f_q(s \mid \nu) - f_{q+2}(s \mid \nu)\},$$

respectively with $q = m(m+1)/2$, $\nu = \operatorname{tr}(\epsilon^2)/4$, $\mu = \operatorname{tr}(\epsilon^3)$ and $\mu_1 = m\operatorname{tr}(\epsilon)$. From Theorem 2.2, we can interpret that the bracket of the right-hand side of (3.1) or (3.2) is the asymptotic distribution under the simple case with known ξ and that the other term expresses the influence of the nuisance parameter vector ξ.

We next consider the problem of testing exponentiality.

Example 4. The density of gamma distribution takes the form

$$f(x \mid \xi, \theta) = \left(\frac{\xi}{\theta}\right)^{-\theta} x^{\theta-1}\exp\left(-\frac{\theta}{\xi}x\right) \Big/ \Gamma(\theta),$$

where ξ is orthogonal to θ in the global sense (see Cox and Reid (1987)). Consider the problem of testing the hypothesis $\theta = 1$, which designates the family of exponential distributions, against $\theta \neq 1$. The adjusted local alternatives are given by $\theta = 1 + \epsilon/\sqrt{n} + c\epsilon^2/n$ with $c = (1 + \psi'''(1))/3(\psi''(1) - 1)$, where $\psi(\theta) = \log\Gamma(\theta)$. The skewness tensor and embedding curvature tensor of the family of gamma distributions are given as follows;

$$(g_{\theta\theta}, g_{\theta\xi}, g_{\xi\xi}) = \left(\psi''(\theta) - \frac{1}{\theta}, 0, \frac{\theta}{\xi^2}\right),$$

$$(H_{\xi\xi,\theta}, T_{\theta\theta\theta}) = \left(\frac{1}{\xi^2}, \frac{1}{\theta^2} + \psi'''(\theta)\right),$$

from which it follows that the asymptotic densities of S_1 are

$$g_1(s) = f_1(s \mid \nu) + \frac{1}{\sqrt{n}}[(1 - \psi'''(1))\epsilon^3\{f_1(s \mid \nu) - 2f_3(s \mid \nu) + f_5(s \mid \nu)\} + \epsilon\{f_1(s \mid \nu) - f_3(s \mid \nu)\}].$$

with $\nu = \epsilon^2(\psi''(1) - 1)$.

The following example is not of (curved) exponential family.

Example 5. Let X be a random variable with the Weibull distribution. Cox and Reid (1987) give the following expression of density

$$f(x \mid \xi, \theta) = c\frac{\theta}{\xi}\left(\frac{x}{\xi}\right)^{\theta-1} \exp\left\{-c\left(\frac{x}{\xi}\right)^{\theta}\right\}$$

by the global orthogonal system of parameters ξ and θ, where $c = \exp\{\Gamma'(2)\}$. We are concerned with a test for the null hypothesis $\theta = 1$ against alternatives $\theta \neq 1$ with the nuisance parameter ξ. The Kullback-Leibler divergence takes the form

$$\begin{aligned}\rho(\xi, \theta;\ \xi_1, \theta_1) &= \int_0^\infty f(x \mid \xi, \theta)\{\log f(x \mid \xi, \theta) - \log f(x \mid \xi, \theta)\}dx \\ &= \log(\theta/\theta_1) - \theta_1 \log(\xi/\xi_1) + (1 - \theta_1/\theta)(1 - \Gamma'(2)) \\ &\quad + \Gamma(1 + \theta_1/\theta)\exp\{\Gamma'(2)(1 - \theta_1/\theta)\} - 1,\end{aligned}$$

since the changed variable $Y \equiv c(X/\xi)^\theta$ has a standard exponential distribution: $\Pr(Y \leq y) = 1 - e^{-y}$. We apply the expression to the formula given by Eguchi (1983):

$$\begin{aligned} g_{\xi\xi} &= \left(-\frac{\partial^2}{\partial\xi^2}\rho\right)_{\theta_1=\theta,\xi_1=\xi} = \theta/\xi^2, \qquad g_{\xi\theta} = \left(-\frac{\partial^2}{\partial\theta\partial\xi}\rho\right)_{\theta_1=\theta,\xi_1=\xi} = 0, \\ g_{\theta\theta} &= \left(-\frac{\partial^2}{\partial\theta^2}\rho\right)_{\theta_1=\theta,\xi_1=\xi} = (1 + \Gamma''(2) - \Gamma'(2)^2)/\theta^2, \\ H_{\xi\xi,\theta} &= \left(\frac{\partial^3}{\partial\theta\partial\xi\partial\xi_1}\rho\right)_{\theta_1=\theta,\xi_1=\xi} = \frac{2\theta}{\xi^2}. \end{aligned}$$

Consequently we conclude that the part of the influence of the nuisance parameter vector is given by

$$h^{(m)} = H_{\xi\xi,\theta}(g_{\xi\xi})^{-1}\epsilon = 2\epsilon.$$

We note that the local power is free from ξ, which is a very special case in addition to Examples 1 and 2.

In the following example a global orthogonal vector relative to the parameter to be tested is not known. Hence we introduce local orthogonal vectors of parameters.

Example 6. Let X have a bivariate normal distribution with mean vector 0 and unknown covariance matrix

$$\begin{pmatrix} \sigma_1^2 & , & \sigma_1\sigma_2\rho \\ * & , & \sigma_2^2 \end{pmatrix}.$$

We wish to test for the hypothesis $\rho = \rho_0$ with a fixed value ρ_0, where (σ_1, σ_2) is a vector of nuisance parameters. Transform (ξ_1, ξ_2) into (σ_1, σ_2) as

$$\sigma_i = \xi_i\left\{1 + \frac{\rho_0(\rho - \rho_0)}{2(1 - \rho_0^2)}\right\}$$

for $i = 1, 2$, so that (ξ_1, ξ_2) is orthogonal to ρ, evaluated at ρ_0. The geometric characteristics are given by

$$(g^{\xi_1\xi_1}, g^{\xi_1\xi_2}, g^{\xi_2\xi_2}) = \frac{1}{4}(\sigma_1^2(2-\rho^2), \sigma_1\sigma_2\rho, \sigma_2^2(2-\rho^2)).$$

$$(H^{(m)}_{\xi_1\xi_1\rho}, H^{(m)}_{\xi_1\xi_2\rho} = H^{(m)}_{\xi_2\xi_1\rho}, H^{(m)}_{\xi_2\xi_2\rho}) = \frac{\rho}{(1-\rho)^2}\left(\frac{1+\rho^2}{\sigma_1^2}, \frac{\rho^2}{\sigma_1\sigma_2}, \frac{1+\rho^2}{\sigma_2^2}\right),$$

which yields that the corresponding coefficient to the influence of nuisance parameters is given by

$$h^{(m)} = \sum_{i,j} H^{(m)}_{\xi_i\xi_j\rho} g^{\xi_i\xi_j\epsilon} = \rho_0(2+\rho_0^2)/2(1-\rho_0^2)$$

when evaluated at (ξ, ρ_0). In this way the influence term vanishes at $\rho_0 = 0$ for any ξ.

We finally return to Example 2.

Example 2 (continued). Under the condition for moments we see that the coefficient of nuisance parameter effect is expressed as

$$h^{(m)}(\xi) = \left[\frac{\mu'(\xi)}{\sigma(\xi)}\{(a_1a_2 + a_1 + b_2 - 2b_1)\sigma'^2(\xi) + (a_1b_1 + a_1^2)\mu'^2(\xi)\}\right.$$
$$\left. + a_1a_2\{\mu''(\xi)\sigma'(\xi) - \mu'(\xi)\sigma''(\xi)\}\right] \Big/ (a_2\sigma'^2(\xi) + \mu'^2(\xi)),$$

which is reduced to $h^{(m)}(\xi) = 2c$ for the case of a known variation coefficient c under the normal family.

4. Discussion

We discuss the local unbiasedness of the test statistics. According to Peers' investigation, the likelihood ratio statistics S_{10} is locally unbiased but so are not the other two statistics S_{20} and S_{30}. We can modify both S_2 and S_{30} into being locally unbiased in the following way;

$$S_{02}^{\dagger} = S_{02} - \tilde{\mu}_1 \quad \text{and} \quad S_{30}^{\dagger} = S_{30} - \tilde{\Lambda}_2, \quad \text{where}$$
$$\tilde{\mu}_1 = T_{ijk}g^{jk}\bar{e}^i \quad \text{and} \quad \tilde{\Lambda}_2 = \kappa_{ijk}g^{jk}\bar{e}^i,$$

with $\kappa_{ijk} = E\partial_i\partial_j e_k$. Thus the modified statistics $S_{20}^{\dagger}$ and $S_{30}^{\dagger}$ have the common asymptotic density

$$g_0^{\dagger}(s) = f_q(s \mid \nu) + \frac{\mu}{6\sqrt{n}}\{f_q(s \mid \nu) - 2f_{q+2}(s \mid \nu) + f_{q+6}(s \mid \nu)\},$$

so that the linear term of ϵ in the local power of $S_2^\dagger$ and $S_3^\dagger$ vanishes. This implies the local unbiasedness of $S_2^\dagger$ and $S_3^\dagger$. Of course there exists a variety of modifications; for example,

$$S_{20}^{\dagger\dagger} = S_{20} - \tilde{\tau}_1 \quad \text{and} \quad S_{30}^{\dagger\dagger} = S_{30} - \tilde{\kappa},$$

where

$$\tilde{\tau}_1 = \frac{1}{3}(T_{ijk}\bar{e}^i\bar{e}^j\bar{e}^k)_0 \quad \text{and} \quad \tilde{\kappa} = (\kappa_{ijk}\bar{e}^i\bar{e}^j\bar{e}^k)_0$$

for which asymptotic densities are common and coincide with that of S_{10} defined in (2.4).

We next consider the composite null hypothesis case. As Hayakawa has suggested, all the S_i's have no local unbiasedness. Similarly, S_i's can be modified into being locally unbiased;

$$S_1^\dagger = S_1 + \hat{h}^{(m)}, \qquad S_2^\dagger = S_2 + \hat{h}^{(m)} - \hat{\mu}_1, \qquad S_3 = S_3^\dagger - \hat{h}^{(e)} - \hat{\Lambda}_2,$$

where the symbol "^" denotes the evaluation at the MLE $(\hat{\xi}_0, \theta_0)$ in the null hypothesis. Thus $S_i^\dagger$'s are found to be locally unbiased since the asymptotic densities are reduced to that of S_{10}.

All the modifications discussed here are supported only in the asymptotic sense. In effect the modified statistics do not satisfy the nested condition (cf. Section 4.2 in Cox and Hinkley (1974) though this aspect may be asymptotically negligible).

Acknowledgements

I am grateful to the referee for valuable and kind suggestions, which lead to more comprehensive expressions of local powers. Also I wish to thank Professors Y. Fujikoshi, K. Maekawa and R. Nishii of Hiroshima University for helpful suggestions and comments on the earlier version.

Appendix

We present a simple outline of proofs of Theorems 2.1 and 2.2 (see Eguchi (1987) for detailed proofs). The derivations used here are almost based on the same way by Hayakawa (1975) except for adding to a geometric interpretation. We use the following notation:

$$\bar{e}_\alpha = \bar{e}_\alpha(\xi, \theta) = \frac{1}{n}\partial_\alpha l(\xi, \theta), \qquad \kappa_{\alpha\beta\gamma} = E\partial_\alpha\partial_\beta\bar{e}_\gamma,$$
$$\bar{A}_{\alpha\beta} = \partial_\alpha\bar{e}_\beta + g_{\alpha\beta} - \Gamma_{\alpha\beta}^{(e)\gamma}\bar{e}_\gamma,$$

where $\partial_\alpha = \partial/\partial(\xi, \theta)^\alpha$ and $l(\xi, \theta)$ denotes the log-likelihood. We note that $\{\bar{e}_\alpha\}_{1\le\alpha\le p+q}$ and $\{\bar{A}_{\alpha\beta}\}_{1\le\alpha,\beta\le p+q}$ are exactly uncorrelated. Here we use the letters $\alpha, \beta, \gamma, \ldots$ as the indices of the full coordinates (ξ, θ), so that symbolically $\alpha = (a, i)$.

LEMMA A.1. *Under a sequence of alternatives* K_n: $\theta = \theta_0 + \epsilon/\sqrt{n}$, *the test statistics* S_{i0}, $i = 1, 2, 3$ *are expanded as*

$$S_{10} = n\left\{ D^2 + \bar{A}_{ij}(\bar{e}^i\bar{e}^j - \bar{\epsilon}^i\bar{\epsilon}^j) - \frac{1}{3}T_{ijk}\bar{e}^i\bar{e}^j\bar{e}^k \right.$$
$$\left. - \Gamma^{(e)}_{kj,i}\bar{e}^i\bar{\epsilon}^j\bar{\epsilon}^k + \frac{1}{3}\kappa_{ijk}\bar{\epsilon}^i\bar{\epsilon}^j\bar{\epsilon}^k \right\};$$
$$S_{20} = n\{ D^2 + 2\bar{A}_{ij}(\bar{e} - \bar{\epsilon})^i\bar{e}^j + T_{ijk}\bar{e}^i\bar{e}^j\bar{\epsilon}^k$$
$$+ (T - \Gamma^{(e)})_{kj,i}\bar{e}^i\bar{e}^j\bar{\epsilon}^k - \Gamma^{(e)}_{ijk}\bar{\epsilon}^i\bar{\epsilon}^j\bar{\epsilon}^k\};$$
$$S_{30} = n\{ D^2 + 2\bar{A}_{ij}(\bar{e} - \bar{\epsilon})^i\bar{e}^j + \Gamma^{(e)}_{ij,k}\bar{e}^i\bar{e}^j\bar{e}^k$$
$$+ \kappa_{ijk}\bar{e}^i\bar{e}^j\bar{\epsilon}^k + (T + 2\Gamma^{(e)})_{ij,k}\bar{e}^i\bar{\epsilon}^j\bar{\epsilon}^k\},$$

where $\bar{\epsilon} = \epsilon/\sqrt{n}$, $D^2 = (\bar{e} - \bar{\epsilon})^i(\bar{e} - \bar{\epsilon})^j g_{ij}$.

The proof follows from a straightforward but complicated routine using the Taylor theorem by neglecting the terms of $o_P(1/\sqrt{n})$.

By correcting the expressions in Lemma A.1 under the adjusted local alternatives K_n^*, it follows that the moment functions are given by

$$E[\exp\{tS_{10}\} \mid K_n^*] = (1-2t)^{-q/2}\exp\{u\nu\} \times \left(1 + \frac{2}{3\sqrt{n}}u^2\mu\right),$$
$$E[\exp\{tS_{20}\} \mid K_n^*] = (1-2t)^{-q/2}\exp\{u\nu\}$$
$$\times \left[1 + \frac{1}{\sqrt{n}}\left\{(2u^2 + u)\mu_1 + \left(\frac{4}{3}u^2 + 2u^2 + \frac{1}{3}u\right)\mu\right\}\right],$$
$$E[\exp\{tS_{30}\} \mid K_n^*] = (1-2t)^{-q/2}\exp\{u\nu\}$$
$$\times \left[1 + \frac{1}{\sqrt{n}}\left\{\left(\frac{1}{3}\mu + \mu_2 - \mu_3\right)u \right.\right.$$
$$\left.\left. + (2\mu + 4\mu_2 - 2\mu_3)u^2 + \left(\frac{4}{3}\mu + 4\mu_2\right)u^3\right\}\right],$$

where $u = t/(1-2t)$ and ν, μ, μ_1, μ_2 and μ_3 are defined in (2.3). Consequently the inversion formula leads to (2.5), (2.6) and (2.8).

Next we give the sketch of the proof of Theorem 2.2. The following relation will be helpful in giving the asymptotic powers of S_i's.

Let $\hat{\xi}_0$ be the MLE of ξ under the null hypothesis, or under known θ_0 and let $\hat{\xi}$ be the ξ-part in the simultaneous MLE of (ξ, θ). Then under the local alternatives K_n: $\theta = \theta_0 + \bar{\epsilon}$ with $\bar{\epsilon} = \epsilon/\sqrt{n}$ the difference $\hat{\xi} - \hat{\xi}_0$ is expanded to be of order 2 in terms $(\bar{e}^\alpha, \bar{A}^{\alpha\beta}, \bar{\epsilon}^i)$. This is easily seen from the expansion of the estimating equations for $\hat{\xi}_0$, $\bar{e}_a(\hat{\xi}_0, \theta_0) = 0$ $(a = 1, \ldots, p)$.

The test statistics S_i's for the composite null hypothesis have the following relation with the corresponding statistics for the simple null hypothesis by letting nuisance parameters be known.

LEMMA A.2. *Under the adjusted local alternatives* K_n^*: $\theta = \theta_0 + \bar{\epsilon} + \Delta(\bar{\epsilon})$ *with* $\bar{\epsilon} = \epsilon/\sqrt{n}$, *it holds that*

$$S_1 = S_{10} + n\{2\bar{A}_{ai}\bar{e}^i\bar{e}^a + \Gamma^{(e)}_{ia,j}\bar{e}^a(\bar{e}^i\bar{e}^j - 2\bar{e}^i\bar{\epsilon}^j + \bar{\epsilon}^i\bar{\epsilon}^j) - H^{(m)}_{ab,i}(\bar{e}+\bar{\epsilon})^i\bar{e}^a\bar{e}^b\}$$

$$S_2 = S_{20} + 2n\{T_{iaj}\bar{e}^a\bar{e}^j + (\Gamma^{(e)}_{ia,j} - \Gamma^{(e)}_{ij,a})\bar{\epsilon}^j\bar{e}^a - \frac{1}{2}H^{(m)}_{ab,i}\bar{e}^a\bar{e}^b\}(\bar{e}-\bar{\epsilon})^i$$

$$S_3 = S_{30} + 2n\{\bar{A}_{ai}(\bar{e}+\bar{\epsilon})^i\bar{e}^a + \partial_a g_{ij}\bar{e}^a(\bar{e}+\bar{\epsilon})^i(\bar{e}+\bar{\epsilon})^j - (T+\Gamma^{(e)})_{ia,j}\bar{\epsilon}^j\bar{e}^a(\bar{e}+\bar{\epsilon})^i + \frac{1}{2}H^{(e)}_{ab,i}(\bar{e}+\bar{\epsilon})^i\bar{e}^a\bar{e}^b\}.$$

The proof is here omitted (see Eguchi (1987)).

By a similar argument to the simple null hypothesis case in Subsection 2.2, we obtain that

$$E(e^{tS_1} \mid K_n^*) = E(e^{tS_{10}} \mid K_n^*)\left(1 - \frac{uh^{(m)}}{\sqrt{n}}\right),$$

$$E(e^{tS_2} \mid K_n^*) = E(e^{tS_{20}} \mid K_n^*)\left(1 - \frac{uh^{(m)}}{\sqrt{n}}\right),$$

$$E(e^{tS_3} \mid K_n^*) = E(e^{tS_{30}} \mid K_n^*)\left\{1 - \frac{uh^{(e)}}{\sqrt{n}}\right\},$$

to order $o(n^{-1/2})$ because of Lemmas A.1 and A.2, where $u = t/(1-2t)$ and h and τ are defined in (2.3). Hence, the inversion formula leads to the asymptotic densities of S_i's. The proof is complete.

Remark 1. The simple versions $\tilde{S}_3$ as discussed in Subsection 2.1 are asymptotically equivalent to S_3, noting that $\tilde{S}_3$ is exactly equal to S_3^* and that $S_3 - S_3^* = O(\|(\bar{A}_{\alpha\beta}, \bar{e}_\alpha, \bar{\epsilon}_i)\|^4)$.

REFERENCES

Amari, S. (1985). *Differential-Geometrical Methods in Statistics*, Springer, New York.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimator in a curved exponential family, *Ann. Statist.*, **11**, 793–803.

Eguchi, S. (1985). A differential geometric approach to statistical inference on the basis of contrast functionals, *Hiroshima Math. J.*, **15**, 341–391.

Eguchi, S. (1987). A geometric look at local power comparison of test statistics, Statistical Research Group of Hiroshima Univ., TR 202.

Finney, D. J. (1976). Radioligand assay, *Biometrics*, **32**, 721–740.

Harris, P. and Peers, H. W. (1980). The local power of efficient scores test statistics, *Biometrika*, **67**, 525–529.

Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative, *Biometrika*, **58**, 577–587.

Kumon, M. and Amari, S. (1983). Geometrical theory of higher-order asymptotics of test, interval estimator, conditional inference, *Proc. Roy. Soc. London Ser. A*, **387**, 429–458.

Kumon, M. and Amari, S. (1988). Differential geometry of testing hypothesis—A higher order asymptotic theory in multi-parameter curved exponential family—, *J. Fac. Engrg. Univ. Tokyo*, **36**.

McCullagh, P. and Cox, D. R. (1986). Invariants and likelihood ratio statistics, *Ann. Statist.*, **14**, 1419–1430.

Nagao, H. (1974). Asymptotic nonnull distributions of certain test criteria for a covariance matrix, *J. Multivariate Anal.*, **4**, 409–418.

Peers, H. W. (1971). Likelihood ratio and associated test criteria, *Biometrika*, **58**, 577–587.

Sugiura, N. (1973). Asymptotic non-null distribution of the likelihood ratio criteria for covariance matrix under local alternatives, *Ann. Statist.*, **1**, 718–728.

Voss, P. W. (1989). Fundamental equations for statistical submanifolds with applications to the bartlett corrections, *Ann. Inst. Statist. Math.*, **41**, 429–450.