Statistical Analysis of Diffusion Tensors in Diffusion-weighted Magnetic Resonance Image Data (Technical Details) *

Hongtu Zhu, Heping Zhang, Joseph G. Ibrahim, and Bradley S. Peterson

Abstract

In this technical report, we give detailed information about how to establish asymptotic theory for one-step weighted least-squares estimates of tensors, estimated eigenvalues and eigenvectors, and pseudo-likelihood ratio statistics. We establish the strong convergence rate and asymptotic normality for the one-step weighted least-squares estimates of tensors. We derive the first-order and second-order expansions of the eigenvalues and eigenvectors of the estimated diffusion tensors. We also derive the asymptotic distributions of pseudo-likelihood ratio statistics under the null hypotheses to classify tensor morphologies.

---

*H. Zhu is Associate Professor of Biostatistics (E-mail: hzhu@bios.unc.edu), Department of Biostatistics and Biomedical Research Imaging Center, and J. G. Ibrahim is Alumni Distinguished Professor of Biostatistics (E-mail: ibrahim@bios.unc.edu), Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599-7420. H. Zhang is Professor of Biostatistics (E-mail: heping.zhang@yale.edu), Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034. B. Peterson is Professor of Psychiatry (E-mail: petersob@childpsych.columbia.edu), Department of Psychiatry, Columbia University Medical Center and the New York State Psychiatric Institute. We thank the Editor, the Associate Editor, and two anonymous referees for valuable suggestions, which greatly helped to improve our presentation. Thanks to Dr. Jason Royal for his invaluable editorial assistance. This work was supported in part by NSF grant SES-06-43663 to Dr. Zhu, NIDA grants DA016750 and DA017713 to Dr. Zhang, NIDA grant DA017820 and NIMH grants MH068318 and K02-74677 to Dr. Peterson, NIH grants GM 70335 and CA 74015 to Dr. Ibrahim, as well as by the Suzanne Crosby Murphy Endowment at Columbia University Medical Center, and by the Thomas D. Klingenstein and Nancy D. Perlman Family Fund.

1 Assumptions

The following assumptions are needed to facilitate the technical details, although they are not the weakest possible conditions.

(C1) The errors η_i are independent and $\sup_i E\eta_i^2 < \infty$.

(C2) $\lambda_{\min}(A_n) \to \infty$.

(C3) θ_* is an interior point of Θ and $\sup_i b_i < \infty$;

(C4) $\lim_{C\to\infty} \sup_i E[\eta_i^2 \mathbf{1}\{|\eta_i| > C\}] = 0$ and $\inf_i E[\eta_i^2] > 0$, where $\mathbf{1}(\cdot)$ denotes the indicator function.

(C5) $\max_{1\le i\le n} \mathbf{z}_i^T (A_n)^{-1}\mathbf{z}_i \to 0$ as $n \to \infty$.

(C6) $\sup_i E[\eta_i^4] < \infty$.

(C7) $\sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^T$ is always positive definite for $n \ge 7$, and the distribution of $(\log S_1, \cdots, \log S_n)$ is absolutely continuous with respect to $n-$dimensional Lebesgue measure.

(C8) The three eigenvalues of $\widehat{\mathbf{D}}$ are distinct with probability one.

(C9) $\sqrt{n}\text{vecs}(\widehat{\mathbf{D}} - \mathbf{D})$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ_D.

(C10) $\mathbf{Q}_n$ converges to a matrix $\mathbf{Q}$, which satisfies $0 < \lambda_{\min}(\mathbf{Q}) \le \lambda_{\max}(\mathbf{Q}) < \infty$, where $\mathbf{Q}_n = G_{n,*}^{1/2} B_{n,*}^{-1} G_{n,*}^{1/2}$ and $\lambda_{\max}(\mathbf{Q})$ denotes the maximum eigenvalue of $\mathbf{Q}$.

Comments. Conditions (C1)-(C2) are sufficient and necessary conditions for $\hat{\theta}_{LS}$ to be strongly consistency (Lai, Robbins, and Wei 1979; Chen, Hu, and Ying 1999). Condition (C3) is a natural condition for diffusion tensor imaging, because diffusion tensor is associated with the covariance matrix of a diffusion process and b_i, the b factor, usually range from 0 to 3,000 s/mm^2 (Kingsley 2006 a, b, c). Conditions (C4)-(C6) are standard conditions to establish the asymptotic normality of $\hat{\theta}_{LS}$ for a linear heteroskedastic model (Eicker 1963; White 1980). Condition (C7) is similar to the condition used for sample covariance matrix in Okamoto (1973). Conditions (C1)-(C7) are sufficient conditions for Conditions (C8) and (C9). Condition (C10) is required to ensure the asymptotic distributions of $PLRT(i)$.

2 Proof of Theorem 1

Lemma 1. Assume that

(D1) $\sum_{i=1}^{n} \mathbf{z}_{n,i}\mathbf{z}_{n,i}^T = \mathbf{I}_7$, where $\mathbf{z}_{n,i}$ is a 7×1 vector;

(D2) $|f_{n,i}(\theta)| \leq C_1||\theta||$ and $|f_{n,i}(\theta_1) - f_{n,i}(\theta_2)| \leq C_2||\theta_1 - \theta_2||$ for all $i = 1, \cdots, n$, where C_1 and C_2 are constants;

(D3) $\{\epsilon_i : i = 1, \cdots, n\}$ is a sequence of independent random variables satisfying $E\epsilon_i = 0$ and $\sup_i E|\epsilon_i| < \infty$;

(D4) $\lim_{n\to\infty} \sup_i \mathbf{z}_{n,i}^T\mathbf{z}_{n,i} = 0$ and $\sup_i E\epsilon_i^2 < \infty$.

We have the following results.

(i) If assumptions (D1)-(D3) are true, then $\sup_{||\theta||\leq M} |\mathbf{e}_k^T W_n(\theta)\mathbf{e}_l| \to 0$ in probability for all $k, l = 1, \cdots, 7$, where $W_n(\theta) = \sum_{i=1}^{n}(\mathbf{z}_{n,i}\mathbf{z}_{n,i}^T)f_{n,i}(a_n\theta)\epsilon_i$, $\lim_{n\to\infty} a_n = 0$ and $\mathbf{e}_k$ is a 7×1 vector with the $k-$th component as one and zero otherwise;

(ii) If assumptions (D1)-(D4) are true, then $\sup_{||\theta||\leq M} |\mathbf{e}_k^T W_n'(\theta)\mathbf{e}_l| \to 0$ in probability for all $k, l = 1, \cdots, 7$, where $W_n'(\theta) = \sum_{i=1}^{n}(\mathbf{z}_{n,i}\mathbf{z}_{n,i}^T)f_{n,i}(\theta)\epsilon_i$.

Proof of Lemma 1: The proof consists of two steps as follows:

Step 1: $W_n(\theta)$ (or $W_n'(\theta)$) converges to zero in probability for each $||\theta|| \leq M$;

Step 2: $\{W_n(\theta) : n \geq 1\}$ (or $\{W_n'(\theta) : n \geq 1\}$) is stochastically equicontinuous on $||\theta|| \leq M$.

We prove Steps 1 and 2 for $W_n(\theta)$ as follows. For Step 1, let $B_n = \sum_{i=1}^{n} |(\mathbf{e}_k^T\mathbf{z}_{n,i}\mathbf{z}_{n,i}^T\mathbf{e}_l)||\epsilon_i|$. We have

$$EB_n \leq (\sup_i E|\epsilon_i|)\sum_{i=1}^{n} |\mathbf{e}_k^T\mathbf{z}_{n,i}\mathbf{z}_{n,i}^T\mathbf{e}_l| \leq 7(\sup_i E|\epsilon_i|)$$

and $\sup_{n\geq 1} EB_n < \infty$, which yields $B_n = O_p(1)$. Thus,

$$|(\mathbf{e}_k^T W_n(\theta)\mathbf{e}_l| \leq a_nC_1MB_n = a_nO_p(1) = o_p(1)$$

holds for any $||\theta|| \leq M$. For Step 2, because $|W_n(\theta) - W_n(\theta')| \leq C_2a_n||\theta - \theta'||B_n$ and $B_n = O_p(1)$, Lemma 1 (a) of Andrews (1992) yields the statement in Step 2.

We prove Steps 1 and 2 for $W_n'(\theta)$ as follows. To check Step 1, we apply the weak law of large numbers (Theorem 1 in Chow and Teicher 1988; p.338). We check the following three conditions stated as follows:

$$\text{(i)} \quad \sum_{i=1}^{n} P\{|X_{n,i}| \geq \epsilon\} \to 0 \ \text{ for } \ \epsilon > 0;$$

$$\text{(ii)} \quad \sum_{i=1}^{n} \mathrm{Var}(X_{n,i}^2 \mathbf{1}\{|X_{n,i}| < 1\}) \to 0; \quad \text{(iii)} \quad \sum_{i=1}^{n} E(X_{n,i} \mathbf{1}\{|X_{n,i}| < 1\}) \to 0;$$

where $X_{n,i} = (\mathbf{e}_k^T \mathbf{z}_{n,i} \mathbf{z}_{n,i}^T \mathbf{e}_l) f_{n,i}(\theta)\epsilon_i$. Using the Mapkob and Cauchy-Schwartz inequality, we can prove condition (i) by noting that

$$\sum_{i=1}^{n} P\{|X_{n,i}| \geq \epsilon\} \leq \epsilon^{-2} \sum_{i=1}^{n} E|X_{n,i}|^2 \leq \epsilon^{-2} C_1 M \sum_{i=1}^{n} (\mathbf{z}_{n,i}^T \mathbf{z}_{n,i})^2 E|\epsilon_i|^2 \leq C \sup_i \mathbf{z}_{n,i}^T \mathbf{z}_{n,i},$$

where C and C_1 are constants. Condition (ii) can be proved by noting that

$$\sum_{i=1}^{n} \mathrm{Var}(X_{n,i}^2 \mathbf{1}\{|X_{n,i}| < 1\}) \leq \sum_{i=1}^{n} E|X_{n,i}|^2 \leq C \sup_i \mathbf{z}_{n,i}^T \mathbf{z}_{n,i}.$$

To check condition (iii), we note that $|\sum_{i=1}^{n} E(X_{n,i} \mathbf{1}\{|X_{n,i}| < 1\})|$ can be bounded by

$$\begin{aligned} \left| \sum_{i=1}^{n} E(X_{n,i} \mathbf{1}\{|X_{n,i}| \geq 1\}) \right| &\leq \sum_{i=1}^{n} \mathbf{z}_{n,i}^T \mathbf{z}_{n,i} E\left[|\epsilon_i| \mathbf{1}\{(\max_j \mathbf{z}_{n,j}^T \mathbf{z}_{n,j}) |\epsilon_i| \geq 1\} \right] \\ \leq \quad \sum_{i=1}^{n} \mathbf{z}_{n,i}^T \mathbf{z}_{n,i} \sqrt{E|\epsilon_i|^2} &\sqrt{P\{(\max_j \mathbf{z}_{n,j}^T \mathbf{z}_{n,j}) |\epsilon_i| \geq 1\}} \leq (\max_j \mathbf{z}_{n,j}^T \mathbf{z}_{n,j}) C, \end{aligned}$$

because $E(X_{n,i}) = 0$. To check Step 2, we can show that $\{W_n'(\theta) : n \geq 1\}$ is stochastically equicontinuous on $||\theta|| \leq M$ by following the same reasoning for $W_n(\theta)$. This completes the proof of Lemma 1.

Proof of Theorem 1: To prove Theorem 1 (a), we consider two different cases for $\widehat{\theta}^{(0)}$ in $\Theta_{*,\delta'}$: a fixed $\widehat{\theta}^{(0)}$ and a random $\widehat{\theta}^{(0)}$.

We prove Theorem 1 (a) for the fixed $\widehat{\theta}^{(0)} \in \Theta_{*,\delta'}$ as follows. It follows from condition (C3) that

$$0 < m = \inf_{i \geq 1, \theta \in \Theta_{*,\delta'}} \exp(2\mathbf{z}_i^T \theta) \leq \sup_{i \geq 1, \theta \in \Theta_{*,\delta'}} \exp(2\mathbf{z}_i^T \theta) = M < \infty,$$

where $\Theta_{*,\delta'} = \{\theta : ||\theta - \theta_*|| \leq \delta'\}$ for any $\delta' > 0$. Thus, for all $\theta \in \Theta_{*,\delta'}$, we have $mA_n \leq B_n(\theta) \leq MA_n$ and

$$m\lambda_{\min}(A_n) \leq \lambda_{\min}(B_n(\theta)) \leq M\lambda_{\min}(A_n). \tag{1}$$

By using (C1) and (C2), we can use equation (1) and Theorem 1 of Lai et al. (1979) to infer, for any $\delta > 0$ and $\theta \in \Theta_{*,\delta'}$, we have $\left(\sum_{i=1}^n \exp(2\mathbf{z}_i^T\theta)\mathbf{z}_i\mathbf{z}_i^T\right)^{-1} \sum_{i=1} \exp(2\mathbf{z}_i^T\theta)\mathbf{z}_i\eta_i = o\left(\left\{[\log \lambda_{\min}(A_n)]^{1+\delta}/\lambda_{\min}(A_n)\right\}^{1/2}\right)$ almost surely. See also Theorem 1 of Chen et al. (1999).

For the random $\widehat{\theta}^{(0)}$, such as $\widehat{\theta}_{LS}$, we prove Theorem 1 (a) by showing that

$$\sup_{\theta \in \Theta_{*,\delta'}} || [B_n(\theta)]^{-1} \sum_{i=1}^n \exp(2\mathbf{z}_i^T\theta)\mathbf{z}_i\eta_i|| = o\left(\left\{[\log \lambda_{\min}(A_n)]^{1+\delta}/\lambda_{\min}(A_n)\right\}^{1/2}\right)$$

holds almost surely. It follows from equation (1) that it is sufficient to show that

$$\sup_{\theta \in \Theta_{*,\delta'}} ||A_n^{-1} \sum_{i=1}^n \exp(2\mathbf{z}_i^T\theta)\mathbf{z}_i\eta_i|| = o\left(\left\{[\log \lambda_{\min}(A_n)]^{1+\delta}/\lambda_{\min}(A_n)\right\}^{1/2}\right), \quad a.s. \tag{2}$$

To prove (2), we mainly generalize the methods used in Lai et al. (1979), who proved strong consistency of $\widehat{\theta}_{LS}$. Note that an extra term $\exp(2\mathbf{z}_i^T\theta)$ appears with each ϵ_i. The proof consists of three steps. We first show that

$$\sum_{i=1}^\infty c_i \exp(2\mathbf{z}_i^T\theta)\eta_i \text{ converges a.s. for all sequences } \{c_i\} \text{ such that } \sum_{i=1}^n c_i^2 < \infty. \tag{3}$$

Second, we apply the same techniques used in Lai et al. (1979) to prove a general version of Theorem 2 in Lai et al. (1979), in which we replace ϵ_i by $\exp(2\mathbf{z}_i^T\theta)\eta_i$ for all i. Finally, we apply a Chung-style uniform law of large numbers in Zaman (1989) to prove (2).

To avoid replicating the proof in Lai et al. (1979), we only show (3) as follows. Let $S_n(\theta) = \sum_{i=1}^n c_i \exp(2\mathbf{z}_i^T\theta)\eta_i$, $||\cdot||$ is the common L_2 norm, $||\cdot||_0$ is the supremum norm: $||f||_0 = \sup_{\theta \in \Theta_{*,\delta'}} |f(\theta)|$, and $||\cdot||_L$ is the Lipschitz norm given by

$$||f||_L = \sup_{\theta \in \Theta_{*,\delta'}} |f(\theta)| + \sup_{\theta,\theta' \in \Theta_{*,\delta'}:\theta \neq \theta'} |f(\theta) - f(\theta')|||\theta - \theta'||^{-1}.$$

We first show that the series of random function $S_n(\theta)$ converges uniformly in quadratic mean. Under assumption (C1), we use the type 2 inequality (Araujo and Gine 1980; Zaman 1989) to conclude that for any m, n,

$$E||S_n(\theta) - S_m(\theta)||_0^2 \leq C \sum_{i=m}^n E||c_i \exp(2\mathbf{z}_i^T\theta)\eta_i||_L^2 \leq CM^2 \sup_i (E\eta_i^2) \sum_{i=m}^n c_i^2,$$

where C is a constant. Thus, the series $\{S_n(\theta) : n \geq 1\}$ is Cauchy in quadratic mean and is convergent in quadratic mean. We can apply the Ito-Nisio lemma (Ito and Nisio 1968) to conclude that $S_n(\theta)$ converges uniformly to a limit function $s(\theta) \in \{f(\theta)|f : \Theta_{*,\delta'} \to (-\infty, +\infty)$ and $||f||_0 < \infty\}$ almost surely. This completes the proof of (3).

To prove Theorem 1 (b), we first prove that

$$[G_n(\widehat{\theta}^{(k)})]^{-1/2} B_n(\widehat{\theta}^{(k)})(\widehat{\theta}^{(k)} - \theta_*) = [G_n(\theta_*)]^{-1/2} \sum_{i=1}^{n} \omega_{i,*} \mathbf{z}_i \eta_i [1 + o_p(1)], \tag{4}$$

and then we apply the Lindeberg-Feller Theorem to prove that $[G_n(\theta_*)]^{-1/2} \sum_{i=1}^n \omega_{i,*} \mathbf{z}_i \eta_i$ converges to $N(0, \sigma^2 \mathbf{I}_7)$, where $\omega_{i,*} = \exp(\mathbf{z}_i^T \theta_*)$.

Because $[G_n(\widehat{\theta}^{(k)})]^{-1/2} B_n(\widehat{\theta}^{(k)})(\widehat{\theta}^{(k)} - \theta_*)$ can be written as

$$[G_n(\widehat{\theta}^{(k)})]^{-1/2} B_n(\widehat{\theta}^{(k)}) [B_n(\widehat{\theta}^{(k-1)})]^{-1} \sum_{i=1}^{n} \omega_i^{(k-1)} \mathbf{z}_i \eta_i,$$

we can prove (4) by using the following steps:

$$||[G_n(\widehat{\theta}^{(k)})]^{-1} G_n(\theta_*) - \mathbf{I}_7|| + ||B_n(\widehat{\theta}^{(k)})[B_n(\widehat{\theta}^{(k-1)})]^{-1} - \mathbf{I}_7|| \to 0 \quad \text{a.s.}, \tag{5}$$

$$[G_n(\theta_*)]^{-1/2} \sum_{i=1}^{n} \omega_i^{(k-1)} \mathbf{z}_i \eta_i = [G_n(\theta_*)]^{-1/2} \sum_{i=1}^{n} \omega_{i,*} \mathbf{z}_i \eta_i [1 + o_p(1)]. \tag{6}$$

It follows from Theorem 1 (a) and conditions (C2) and (C3) that (5) is true. Furthermore, by using (C3) and Theorem 1 (a), we have that $\sup_i |\omega_i^{(k-1)} - \omega_{i,*}|$ converges to zero almost surely. Thus, (6) is proved, and so is (4). It follows from conditions (C1)-(C5) and the Lindeberg-Feller Theorem that $[G_n(\theta_*)]^{-1/2} \sum_{i=1}^n \omega_{i,*} \mathbf{z}_i \eta_i$ converges to $N(\mathbf{0}, \mathbf{I}_7)$ in distribution.

To prove Theorem 1 (c), let $T_n(\theta) = [G_n(\theta_*)]^{-1/2} F_n(\theta) [G_n(\theta_*)]^{-1/2} - \mathbf{I}_7$. We note that $T_n(\widehat{\theta}^{(k)})$ can be rewritten as

$$[G_n(\theta_*)]^{-1/2} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^T \exp(4\mathbf{z}_i^T \theta_*) E\eta_i^2 [\exp(4\mathbf{z}_i^T \Delta^{(k)}) \mathbf{e}_i(\widehat{\theta}^{(k)})^2 / E\eta_i^2 - 1][G_n(\theta_*)]^{-1/2},$$

where $\Delta^{(k)} = \widehat{\theta}^{(k)} - \theta_*$. Now $\mathbf{e}_i(\widehat{\theta}^{(k)})^2 = \eta_i^2 - 2\mathbf{z}_i^T \Delta^{(k)} \eta_i + (\mathbf{z}_i^T \Delta^{(k)})^2$, $T_n(\widehat{\theta}^{(k)})$ can be written as the sum of term (I), term (II), and term (III), where

$$\text{term (I)} = \sum_{i=1}^{n} \mathbf{z}_{n,i} \mathbf{z}_{n,i}^T \exp(4\mathbf{z}_i^T \Delta^{(k)}) [\eta_i^2 - E\eta_i^2] / E\eta_i^2,$$

$$\text{term (II)} = \sum_{i=1}^{n} \mathbf{z}_{n,i} \mathbf{z}_{n,i}^T \exp(4\mathbf{z}_i^T \Delta^{(k)}) [-2\mathbf{z}_i^T \Delta^{(k)} \eta_i] / E\eta_i^2,$$

$$\text{term (III)} = \sum_{i=1}^{n} \mathbf{z}_{n,i}\mathbf{z}_{n,i}^T\{\exp(4\mathbf{z}_i^T\Delta^{(k)})(\mathbf{z}_i^T\Delta^{(k)})^2/E\eta_i^2 + \exp(4\mathbf{z}_i^T\Delta^{(k)}) - 1\},$$

and $\mathbf{z}_{n,i} = [G_n(\theta_*)]^{-1/2}\mathbf{z}_i \exp(2\mathbf{z}_i^T\theta_*)\sqrt{E\eta_i^2}$. Because

$$\sup_i |\exp(4\mathbf{z}_i^T\Delta^{(k)})(\mathbf{z}_i^T\Delta^{(k)})^2/E\eta_i^2 + \exp(4\mathbf{z}_i^T\Delta^{(k)}) - 1| \leq C_3||\Delta^{(k)}||,$$

term (III) converges to zero almost surely. Applying Lemma 1 leads to the result that every element of terms (I) and (II) converges to zero in probability.

3 Proof of Theorem 2

Proof of Theorem 2: We prove Theorem 2 (a) for $\widehat{\theta}_{LS}$ as follows. The estimated eigenvalues $\{m_1, m_2, m_3\}$ are the roots of

$$g(m) = |\widehat{\mathbf{D}} - m\mathbf{I}_3| = m^3 - m^2 I_1(\widehat{\mathbf{D}}) + mI_2(\widehat{\mathbf{D}}) - I_3(\widehat{\mathbf{D}}) = 0,$$

where $I_1(\widehat{\mathbf{D}}) = \text{trace}[\widehat{\mathbf{D}}]$, $I_3(\widehat{\mathbf{D}}) = |\widehat{\mathbf{D}}|$, and

$$I_2(\widehat{\mathbf{D}}) = \widehat{D}_{11}\widehat{D}_{22} + \widehat{D}_{11}\widehat{D}_{33} + \widehat{D}_{22}\widehat{D}_{33} - (\widehat{D}_{12}^2 + \widehat{D}_{13}^2 + \widehat{D}_{23}^2).$$

Let $d(\widehat{\mathbf{D}})$ be the discriminant of the polynomial $g(m)$. We know (Okamoto, 1973) that

$$\text{the eigenvalues of } \widehat{\mathbf{D}} \text{ are distinct if and only if } d(\widehat{\mathbf{D}}) \neq 0.$$

Thus, it suffices to prove that $d(\widehat{\mathbf{D}}) \neq 0$ holds with probability one. Because $\widehat{\theta}_{LS}$ is a linear combination of $\log S_i$ and $d(\widehat{\mathbf{D}})$ is a polynomial in the elements of $\widehat{\mathbf{D}}$, $d(\widehat{\mathbf{D}})$ is a polynomial in the elements of $\{\log S_1, \cdots, \log S_n\}$, denoted as $f(\log S_1, \cdots, \log S_n)$. Using the lemma in Okamoto (1973), we only need to show that $f(\log S_1, \cdots, \log S_n)$ is not identically zero. If we set $\log S_i = \mathbf{z}_i^T\theta_{0*}$, in which θ_{0*} corresponds to a diffusion tensor with three distinct eigenvalues, then $f(\mathbf{z}_1^T\theta_{0*}, \cdots, \mathbf{z}_n^T\theta_{0*})$ is not equal to zero. This proves Theorem 2 (a).

We prove Theorem 2 (b) in two steps. In Step 1, we consider any fixed $\widehat{\theta}^{(0)} \in \Theta_{*,\delta'}$. Thus,

$$\widehat{\theta}^{(1)} = [\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i^T \exp(2\mathbf{z}_i^T\widehat{\theta}^{(0)})]^{-1} \sum_{i=1}^{n} \exp(2\mathbf{z}_i^T\widehat{\theta}^{(0)})\mathbf{z}_i \log S_i$$

is a polynomial function of $\{\log S_1, \cdots, \log S_n\}$. Similar to the argument for Theorem 2 (a), we can use the lemma in Okamoto (1973) to complete the proof of Theorem 2 (b).

In Step 2, we consider any random $\widehat{\theta}^{(0)} \in \Theta_{*,\delta'}$. Note that $\widehat{\theta}^{(1)} = [B_n(\widehat{\theta}_{LS})]^{-1} \sum_{i=1}^n \exp(2\mathbf{z}_i^T \widehat{\theta}_{LS})\mathbf{z}_i \log S_i$. Let $\widehat{\mathbf{D}}^{(1)}$ be the diffusion tensor of $\widehat{\theta}^{(1)}$ and let $d(\widehat{\mathbf{D}}^{(1)})$ be the discriminant of the polynomial $g(m) = |m\mathbf{I}_3 - \widehat{\mathbf{D}}^{(1)}| = 0$. Thus, by using Fubini's Theorem, we have $P(\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}|\hat{\theta}_{LS} \in \Theta_{*,\delta'})$ can be written as

$$\int P(\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}|\widehat{\theta}_{LS} = \theta, \hat{\theta}_{LS} \in \Theta_{*,\delta'})p(\widehat{\theta}_{LS} = \theta|\hat{\theta}_{LS} \in \Theta_{*,\delta'})d\theta, \tag{7}$$

where $\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}$ denotes the event $d(\widehat{\mathbf{D}}^{(1)}) = 0$, $p(\widehat{\theta}_{LS} = \theta|\hat{\theta}_{LS} \in \Theta_{*,\delta'})$ is the conditional density function of $\widehat{\theta}_{LS}$ given $\hat{\theta}_{LS} \in \Theta_{*,\delta'}$, and $P(\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}|\widehat{\theta}_{LS} = \theta, \hat{\theta}_{LS} \in \Theta_{*,\delta'})$ is the conditional probability of $\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}$ given $\widehat{\theta}_{LS} = \theta$ and $\hat{\theta}_{LS} \in \Theta_{*,\delta'}$. We note that $\widehat{\theta}_{LS}$ is a linear combination of $\{\log S_i : i = 1, \cdots, n\}$ and $\widehat{\theta}^{(1)}$ given $\widehat{\theta}_{LS} = \theta$ is a linear function of $\{\log S_i : i = 1, \cdots, n\}$. It follows from Okamoto's (1973) lemma that $P(\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}|\widehat{\theta}_{LS} = \theta, \hat{\theta}_{LS} \in \Theta_{*,\delta'}) = 0$ holds for almost every $\theta \in \Theta_{*,\delta'}$. Thus, $P(\{d(\widehat{\mathbf{D}}^{(1)}) = 0\}|\hat{\theta}_{LS} \in \Theta_{*,\delta'}) = 0$.

4 Proof of Theorem 3

Proof of Theorem 3. For an isotropic tensor, we have $\Lambda = \lambda\mathbf{I}_3$, $\Gamma = \mathbf{I}_3$, and $\mathbf{C}_n^T = \mathbf{E}$. Recall that $\mathbf{T}_n = \lambda\mathbf{I}_3 + n^{-1/2}\mathbf{U}_n$, we have

$$\mathbf{T}_n = \mathbf{C}_n^T\mathbf{M}\mathbf{C}_n = \mathbf{C}_n^T(\lambda\mathbf{I}_3 + n^{-1/2}\mathbf{H}_n)\mathbf{C}_n = \lambda\mathbf{I}_3 + n^{-1/2}\mathbf{C}_n^T\mathbf{H}_n\mathbf{C}_n,$$

where $\mathbf{H}_n = \sqrt{n}(\mathbf{M} - \lambda\mathbf{I}_3)$. Thus, $\mathbf{U}_n = \mathbf{C}_n^T\mathbf{H}_n\mathbf{C}_n$, and $\mathbf{C}_n$ and $\mathbf{H}_n$ are uniquely defined as continuous functions of $\mathbf{U}_n$ with the proper ordering except on a set of probability 0. Using a theorem due to Rubin (Anderson, 2003; Theorem 13.5.3), we can infer that the limiting distribution of $\mathbf{H}_n$ and $\mathbf{C}_n$ is determined by $\mathbf{C}^T\mathbf{H}\mathbf{C} = \mathbf{U}$ and the distribution of $\mathbf{U}$, in which $\mathbf{H} = \text{diag}(h_1, h_2, h_3)$ and $\mathbf{C} = (c_{ij})$ satisfy $h_1 > h_2 > h_3$, $c_{ii} > 0$ for $i = 1, 2, 3$, and $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{3\times 3}$. Note that the density of the distribution of $\mathbf{U}$ is proportional to $|\Sigma_\mathbf{U}|^{-1/2}\exp\{-\frac{1}{2}\text{vecs}(\mathbf{U})^T\Sigma_\mathbf{U}^{-1}\text{vecs}(\mathbf{U})\}$. Using a result due to Hsu, P.L. (Deemer and Olkin, 1951), we can obtain that the Jacobian of the transformation from $\mathbf{U}$ to $\mathbf{H}$ and $\mathbf{C}$ is proportional to $(h_1 - h_2)(h_2 - h_3)(h_1 - h_3)$. Thus, combining the above two results, we can obtain the joint density of $\mathbf{H}$ and $\mathbf{C}$ as given in Theorem 3.

Because $h_1 \geq h_2 \geq h_3$ are three eigenvalues of $\mathbf{U}$, $-h_3 \geq -h_2 \geq -h_1$ are the corresponding eigenvalues of $-\mathbf{U}$. Moreover, since $\mathbf{U}$ and $-\mathbf{U}$ follow the same distribution, h_2 and $-h_2$ follow the same distribution. Thus, $E(h_2) = E(-h_2)$, which yields that $E(h_2) = 0$. Similarly, we can show that $E(h_1 + h_2 + h_3) = 0$. We can use the explicit form of $p(\mathbf{h}, \mathbf{C})$ to infer that $E(h_1 - h_2) > 0$ and $E(h_2 - h_3) > 0$. Finally, we get $E(h_1) > E(h_2) = 0 > E(h_3)$.

5 Proof of Theorem 4

Proof of Theorem 4. We have $\mathbf{T}_n = \Lambda + n^{-1/2}\mathbf{U}_n = \mathbf{C}_n^T \mathbf{M} \mathbf{C}_n = \mathbf{C}_n^T(\Lambda + n^{-1/2}\mathbf{H}_n)\mathbf{C}_n$. Using a matrix representation, we get

$$\begin{pmatrix} \lambda_1 \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0}^T & \lambda_3 \end{pmatrix} + n^{-1/2} \begin{pmatrix} \mathbf{U}_{n,11} & \mathbf{U}_{n,12} \\ \mathbf{U}_{n,21} & \mathbf{U}_{n,22} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{n,11}^T & n^{-1/2}\mathbf{F}_{n,21}^T \\ n^{-1/2}\mathbf{F}_{n,12}^T & \mathbf{C}_{n,22} \end{pmatrix} \times$$
$$\begin{pmatrix} \lambda_1 \mathbf{I}_2 + n^{-1/2}\mathbf{H}_{n,1} & \mathbf{0} \\ \mathbf{0}^T & \lambda_3 + n^{-1/2} h_{n,3} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{n,11} & n^{-1/2}\mathbf{F}_{n,12} \\ n^{-1/2}\mathbf{F}_{n,21} & \mathbf{C}_{n,22} \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_1 \mathbf{C}_{n,11}^T \mathbf{C}_{n,11} & \mathbf{0} \\ \mathbf{0}^T & \lambda_3 \mathbf{C}_{n,22}^2 \end{pmatrix} + n^{-1/2} \times$$
$$\begin{pmatrix} \mathbf{C}_{n,11}^T \mathbf{H}_{n,1} \mathbf{C}_{n,11} & \lambda_1 \mathbf{C}_{n,11}^T \mathbf{F}_{n,12} + \lambda_3 \mathbf{F}_{n,21}^T \mathbf{C}_{n,22} \\ \lambda_1 \mathbf{F}_{n,12}^T \mathbf{C}_{n,11} + \lambda_3 \mathbf{C}_{n,22} \mathbf{F}_{n,21} & \mathbf{C}_{n,22}^2 h_{n,3} \end{pmatrix} + n^{-1}\mathbf{M}_n,$$

where $\mathbf{F}_{n,12} = \sqrt{n}\mathbf{C}_{n,12}$, $\mathbf{F}_{n,21} = \sqrt{n}\mathbf{C}_{n,21}$, and $\mathbf{M}_n$ is given by

$$\begin{pmatrix} \mathbf{M}_{n,11} & \mathbf{M}_{n,12} \\ \mathbf{M}_{n,21} & \mathbf{M}_{n,22} \end{pmatrix} = \begin{pmatrix} (\lambda_3 + n^{-1/2}h_{n,3})\mathbf{F}_{n,21}^T \mathbf{F}_{n,21} & \mathbf{C}_{n,11}^T \mathbf{H}_{n,1} \mathbf{F}_{n,12} + h_{n,3} \mathbf{C}_{n,22} \mathbf{F}_{n,21}^T \\ \mathbf{F}_{n,12}^T \mathbf{H}_{n,1} \mathbf{C}_{n,11} + h_{n,3} \mathbf{C}_{n,22} \mathbf{F}_{n,21} & \lambda_1 \mathbf{F}_{n,12}^T \mathbf{F}_{n,12} + n^{-1/2} \mathbf{F}_{n,12}^T \mathbf{H}_{n,1} \mathbf{F}_{n,12} \end{pmatrix}.$$

Because $\mathbf{C}_n^T\mathbf{C}_n = \mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_3$, we know that

$$\begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{n,11}^T\mathbf{C}_{n,11} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_{n,22}^2 \end{pmatrix} + n^{-1/2} \times$$
$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{n,11}^T\mathbf{F}_{n,12} + \mathbf{F}_{n,21}^T\mathbf{C}_{n,22} \\ \mathbf{F}_{n,12}^T\mathbf{C}_{n,11} + \mathbf{C}_{n,22}\mathbf{F}_{n,21} & 0 \end{pmatrix} + n^{-1} \begin{pmatrix} \mathbf{F}_{n,21}^T\mathbf{F}_{n,21} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{F}_{n,12}^T\mathbf{F}_{n,12} \end{pmatrix}.$$

This gives

$$\mathbf{C}_{n,11}^T\mathbf{C}_{n,11} = \mathbf{I}_2 - n^{-1}\mathbf{F}_{n,21}^T\mathbf{F}_{n,21}, \quad \mathbf{C}_{n,11}\mathbf{C}_{n,11}^T = \mathbf{I}_2 - n^{-1}\mathbf{F}_{n,12}\mathbf{F}_{n,12}^T,$$

$$\mathbf{C}_{n,11}^T\mathbf{F}_{n,12} + \mathbf{F}_{n,21}^T\mathbf{C}_{n,22} = \mathbf{0}, \quad \text{and} \quad \mathbf{C}_{n,22}^2 = 1 - n^{-1}\mathbf{F}_{n,12}^T\mathbf{F}_{n,12}.$$

Combining the above results, we get

$$\mathbf{U}_{n,11} = \mathbf{C}_{n,11}^T\mathbf{H}_{n,1}\mathbf{C}_{n,11} + n^{-1/2}(\mathbf{M}_{n,11} - \lambda_1\mathbf{F}_{n,21}^T\mathbf{F}_{n,21}),$$

$$\mathbf{U}_{n,12} = (\lambda_1 - \lambda_3)\mathbf{C}_{n,11}^T\mathbf{F}_{n,12} + n^{-1/2}\mathbf{M}_{n,12},$$

and $\mathbf{U}_{n,22} = \mathbf{C}_{n,22}^2 h_{n,3} + n^{-1/2}(\mathbf{M}_{n,22} - \lambda_3\mathbf{F}_{n,12}^T\mathbf{F}_{n,12})$. Furthermore, by following the same reasoning in Theorem 13.5.1 of Anderson (2003) and Anderson (1963), it follows that

$$\begin{aligned}
&\mathbf{C}_{n,11}^T\mathbf{H}_{n,1}\mathbf{C}_{n,11} = \mathbf{U}_{n,11} + n^{-1/2}(\lambda_1 - \lambda_3)\mathbf{F}_{n,21}^T\mathbf{F}_{n,21} + o_p(n^{-1/2}),\\
&h_{n,3} = \mathbf{U}_{n,22} - (\lambda_1 - \lambda_3)n^{-1/2}\mathbf{F}_{n,12}^T\mathbf{F}_{n,12} + o_p(n^{-1/2}),\\
&\mathbf{C}_{n,22}^2 = 1 - n^{-1}\mathbf{F}_{n,12}^T\mathbf{F}_{n,12} + o_p(n^{-1}),\\
&\mathbf{C}_{n,11}^T\mathbf{F}_{n,12} = -\mathbf{F}_{n,21}^T\mathbf{C}_{n,22} = -\mathbf{F}_{n,21}^T + o_p(n^{-1/2}), \text{ and}\\
&\mathbf{C}_{n,11}^T\mathbf{F}_{n,12} = O_p(1) = (\mathbf{U}_{n,12} - n^{-1/2}\mathbf{M}_{n,12})/(\lambda_1 - \lambda_3) + o_p(n^{-1/2}).
\end{aligned} \tag{8}$$

The above results lead to Theorem 4 (a), (b), and (c). By using the transformation given by $\tilde{h}_1 = -h_2$ and $\tilde{h}_2 = -h_1$, we can prove that $E(h_1 + h_2) = 0$ and $E(h_1) > 0 > E(h_2)$.

By using $\mathbf{C}_n^T = \Gamma^T\mathbf{E}$, we obtain $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \Gamma\mathbf{C}_n^T$, which leads to

$$(\mathbf{e}_1, \mathbf{e}_2) = (\mathbf{v}_1, \mathbf{v}_2)\mathbf{C}_{n,11}^T + n^{-1/2}\mathbf{v}_3\mathbf{U}_{n,12}^T\mathbf{C}_{n,11}^T/(\lambda_1 - \lambda_3) + o_p(n^{-1/2}). \tag{9}$$

Furthermore, by using (8), we have

$$
\begin{aligned}
\sqrt{n}(\mathbf{e}_3 - \mathbf{v}_3) &= (\mathbf{v}_1, \mathbf{v}_2)\mathbf{F}_{n,21}^T - 0.5\mathbf{v}_3 n^{-1/2}\mathbf{F}_{n,12}^T\mathbf{F}_{n,12} + o_p(n^{-1/2}) \\
&= -(\mathbf{v}_1, \mathbf{v}_2)[\mathbf{I}_2 - n^{-1/2}(\mathbf{C}_{n,11}^T\mathbf{H}_{n,1}\mathbf{C}_{n,11} - \mathbf{U}_{n,22}\mathbf{I}_2)(\lambda_1 - \lambda_3)^{-1}]\mathbf{U}_{n,12}(\lambda_1 - \lambda_3)^{-1} \\
&- 0.5\mathbf{v}_3 n^{-1/2}\mathbf{U}_{n,21}\mathbf{C}_{n,11}^T\mathbf{C}_{n,11}\mathbf{U}_{n,12}/(\lambda_1 - \lambda_3)^2 + o_p(n^{-1/2}).
\end{aligned}
$$

Using (9) and the results in Theorem 4 (a), (b), and (c), the proof of Theorem 4 (d) immediately follows.

Proof of Corollaries 1 and 2. The technical arguments are similar to the proof of Theorem 4 and thus the details are omitted for brevity.

6 Proof of Theorem 5

Proof of Theorem 5: The key idea in deriving the asymptotic distributions of $PLRT(i)$ is as follows. After some algebraic and probabilistic manipulations, we get

$$
\begin{aligned}
\ell_n(\theta|\widehat{\theta}_{LS}) - \ell_n(\theta_*|\widehat{\theta}_{LS}) &= \widehat{\theta}^{(1)T}B_n(\widehat{\theta}_{LS})\widehat{\theta}^{(1)} - (\theta - \widehat{\theta}^{(1)})^T B_n(\widehat{\theta}_{LS})(\theta - \widehat{\theta}^{(1)}) \qquad (10) \\
&= \mathbf{Z}_n^T\mathbf{Q}_n\mathbf{Z}_n - [\mathbf{K}_n(\theta - \theta_*) - \mathbf{Z}_n]^T\mathbf{Q}_n[\mathbf{K}_n(\theta - \theta_*) - \mathbf{Z}_n][1 + o_p(1)],
\end{aligned}
$$

where $\mathbf{K}_n = G_{n,*}^{-1/2}B_{n,*}$, $\mathbf{Q}_n = G_{n,*}^{1/2}B_{n,*}^{-1}G_{n,*}^{1/2}$, and $\mathbf{Z}_n = C_{n,*}^{-1/2}\sum_{i=1}^n \mathbf{z}_i\eta_i\omega_{i,*}$, in which $G_{n,*} = G_n(\theta_*)$, $B_{n,*} = B_n(\theta_*)$, and $\omega_{i,*} = \exp(2\mathbf{z}_i^T\theta_*)$. Thus, we establish a quadratic expansion of $\ell_n(\theta|\widehat{\theta}_{LS})$ in θ about θ_*. Finally, we apply the asymptotic results in Andrews (2001) and Zhu and Zhang (2006) to deriving the limiting distributions of $PLRT(i)$.

It follows from Theorem 1 of Andrews (2001) that

$$
\max_{\theta\in\Theta(j)} \ell_n(\theta|\widehat{\theta}_{LS}) = \ell_n(\theta_*|\widehat{\theta}_{LS}) + \mathbf{Z}_n^T\mathbf{Q}_n\mathbf{Z}_n - \max_{\omega\in\Omega(j)}[\omega - \mathbf{Z}_n]^T\mathbf{Q}_n[\omega - \mathbf{Z}_n][1 + o_p(1)], \qquad (11)
$$

in which $\{\mathbf{K}_n(\theta-\theta_*)/b_n : n \geq 1\}$ locally approximates a cone $\Omega(j)$, where $\mathbf{K}_n = [G_n(\theta_*)]^{-1/2}B_n(\theta_*)$, $b_n \to \infty$, and $b_n \leq C\lambda_{\min}(\mathbf{K}_n) \leq C\sqrt{\lambda_{\min}(A_n)}$. The parameter spaces $\Theta(i)$ can be, respectively, written as

$$\Theta(1) = \{(\log S_0, \lambda): \ \log S_0 \in R, \ \ \mathbf{D} = \lambda \mathbf{I}_3 \geq 0\},$$
$$\Theta(2) = \{(\log S_0, a, b, c, d): \ \log S_0 \in R, \ \ \mathbf{D} = a^2\mathbf{I}_3 - a^2 \sin^2(b)\mathbf{v}\mathbf{v}^T\}, \ \ \text{and} \tag{12}$$
$$\Theta(3) = \{(\log S_0, a, b, c, d): \ \log S_0 \in R, \ \ \mathbf{D} = a^2 \sin^2(b)\mathbf{I}_3 + a^2[1 - \sin^2(b)]\mathbf{v}\mathbf{v}^T\},$$

in which $\mathbf{v} = (\cos(d), \cos(c)\sin(d), \sin(c)\sin(d))^T$. Therefore, we have

$$PLRT(j) = \max_{\omega \in \Omega(j)} [\omega - \mathbf{Z}_n]^T \mathbf{Q}_n [\omega - \mathbf{Z}_n] + o_p(1).$$

To derive the asymptotic distribution of $PLRT(j)$, we only need to study the geometric structure of $\Omega(j)$ for $j = 1, 2, 3$.

For $\theta \in \Theta(1)$, $\theta = \mathbf{G}_1\xi$, where

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_1^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Thus, the isotropic hypotheses can be written as $H_0^{(1)}: \theta = \mathbf{G}_1\xi, \ \xi \in R^2$ versus $H_1^{(1)}: \theta \in \Theta$. Because $\mathbf{D}$ has the form $\lambda\mathbf{I}_3$ with $\lambda > 0$ under $H_0^{(1)}$, we can get $\Omega(1) = \{\omega : \omega = \mathbf{G}_1\xi, \xi \in R^2\}$. Finally, we obtain that $PLRT(1)$ converges to $X(1) = \mathbf{Z}^T[\mathbf{Q} - \mathbf{Q}\mathbf{G}_1^T(\mathbf{G}_1\mathbf{Q}\mathbf{G}_1^T)^{-1}\mathbf{G}_1\mathbf{Q}]\mathbf{Z}$ in distribution.

For $\theta \in \Theta(2)$, we consider two different cases of $\mathbf{D} = \lambda_1\mathbf{I}_3 - (\lambda_1 - \lambda_3)\mathbf{v}_3\mathbf{v}_3^T$: $\lambda_1 - \lambda_3 > 0$ and $\lambda_1 = \lambda_3$. If $\lambda_1 > \lambda_3$, we define

$$\xi^T = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5) = (S_0, \lambda_1, \sqrt{\lambda_1 - \lambda_3}\mathbf{v}_3^T)$$

and ξ_* is the true value under the null hypothesis $H_0^{(2)}$. Thus, θ can be written as a function of ξ as follow:

$$\theta(\xi) = (\xi_1, \xi_2 - \xi_3^2, -\xi_3\xi_4, -\xi_5\xi_3, \xi_2 - \xi_4^2, -\xi_4\xi_5, \xi_2 - \xi_5^2)^T.$$

Differentiating θ with respect to ξ, we can prove that the rank of $\partial\theta(\xi)/\partial\xi$ evaluating at ξ_* is 5, because $\xi_{3,*}^2 + \xi_{4,*}^2 + \xi_{5,*}^2 = \lambda_1 - \lambda_3 > 0$. Let $\mathbf{G}_2 = \partial\theta(\xi_*)/\partial\xi$, we get that $\Omega(2) = \{\omega :$

$\omega = \mathbf{G}_2^T\xi, \xi \in R^5\}$ and $PLRT(2)$ converges to $X(2) = \mathbf{Z}^T[\mathbf{Q} - \mathbf{Q}\mathbf{G}_2^T(\mathbf{G}_2\mathbf{Q}\mathbf{G}_2^T)^{-1}\mathbf{G}_2\mathbf{Q}]\mathbf{Z}$ in distribution.

When $\lambda_1 = \lambda_3 > 0$, we cannot use the previous method since the rank of $\mathbf{G}_2$ is not full rank. Thus, we introduce a new parametrization $\xi = (S_0, \lambda_1, \lambda_1 - \lambda_3)^T = (\xi_1, \xi_2, \xi_3)^T$. Therefore, the diffusion tensor can be written as $\mathbf{D} = \xi_2\mathbf{I}_3 - \xi_3\mathbf{e}\mathbf{e}^T$, where $\mathbf{e} = (e_1, e_2, e_3)^T$ and $\mathbf{e}\mathbf{e}^T = 1$. Hence, θ can be written as

$$\theta(\xi|\mathbf{e}) = (\xi_1, \xi_2 - \xi_3 e_1^2, -\xi_3 e_1 e_2, -\xi_3 e_1 e_3, \xi_2 - \xi_3 e_2^2, -\xi_3 e_2 e_3, \xi_2 - \xi_3 e_3^2)^T.$$

Let $\Theta(2|\mathbf{e}) = \{\theta : \theta = \theta(\xi|\mathbf{e})\}$. Differentiating $\theta(\xi|\mathbf{e})$ with respect to ξ for any given $\mathbf{e}$, it follows that the rank of $\mathbf{G}_3(\mathbf{e}) = \partial\theta(\xi|\mathbf{e})/\partial\xi$ is 3 and $\Theta(2|\mathbf{e})$ can be approximated by $\Omega(2|\mathbf{e}) = \{\omega : \omega = \mathbf{G}_3(\mathbf{e})^T\xi, \xi_3 \in [0, \infty)\}$. Finally, we have

$$\begin{aligned} PLRT(2) &= \sup_{\mathbf{e}:\mathbf{e}^T\mathbf{e}=1,\, \omega\in\Omega(2|\mathbf{e})} [\mathbf{Z}_n - \omega]^T\mathbf{Q}_n[\mathbf{Z}_n - \omega] + o_p(1) \\ &\to^L \sup_{\mathbf{e}:\mathbf{e}^T\mathbf{e}=1} \mathbf{Z}^T\{\mathbf{Q} - \mathbf{Q}\mathbf{G}_3(\mathbf{e})^T[\mathbf{G}_3(\mathbf{e})\mathbf{Q}\mathbf{G}_3(\mathbf{e})^T]^{-1}\mathbf{G}_3(\mathbf{e})\mathbf{Q}\}\mathbf{Z}. \end{aligned}$$

Similar to $PLRT(2)$, we can establish the asymptotic distribution of $PLRT(3)$.

7 Approximating $X(i)$

Because similar procedure can be developed for $X(2)$ and $X(3)$, we only give a procedure for approximating $X(1)$ as follows. First, $X(1)$ can be written as $\mathbf{Z}^T\Sigma(1)\mathbf{Z}$ and $\Sigma(1) = \mathbf{Q} - \mathbf{Q}\mathbf{G}_1^T(\mathbf{G}_1\mathbf{Q}\mathbf{G}_1^T)^{-1}\mathbf{G}_1\mathbf{Q}$, in which $\mathbf{Q}$ is the limit of $\mathbf{Q}_n$, $\mathbf{G}_1$ is a matrix defined in the proof of Theorem 5, and $\mathbf{Z}$ is a multivariate Gaussian random vector that has mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_7$. Second, we can construct a consistent estimate of $\Sigma(1)$, $\widehat{\Sigma}(1) = \widehat{\mathbf{Q}} - \widehat{\mathbf{Q}}\mathbf{G}_1^T(\mathbf{G}_1\widehat{\mathbf{Q}}\mathbf{G}_1^T)^{-1}\mathbf{G}_1\widehat{\mathbf{Q}}$, where $\widehat{\mathbf{Q}} = F_n(\widehat{\theta}^{(1)})^{1/2}B_n(\widehat{\theta}^{(1)})^{-1}F_n(\widehat{\theta}^{(1)})^{1/2}$. Third, we can approximate $X(1)$ by a scaled χ^2 distribution $c_1\chi^2(\nu_1)$, where ν_1 is the degree of freedom (Chou et al. 1991). Fourth, we use the moment matching technique to match the mean and variance of $c_1\chi^2(\nu_1)$ with those of $X(1)$ in order to estimate c_1 and ν_1. Finally, we have $c_1 = \sum_{i=1}^6 \gamma_i^2/\sum_{i=1}^6 \gamma_i$ and $\nu_1 = (\sum_{i=1}^6 \gamma_i)^2/\sum_{i=1}^6 \gamma_i^2$, where γ_i are eigenvalues of $\widehat{\Sigma}(1)$.

REFERENCES

Anderson, T. W. (1963), "Asymptotic Theory for Principal Component Analysis," *Annals of Mathematical Statistics*, 34, 122-148.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis* (3rd ed.), Wiley Series in Probability and Statistics.

Andrews, D. W. K. (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

Andrews, D. W. K. (2001), "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 685-734.

Araujo, A., and Gine, E. (1980), *The Central Limit Theorem for Real- and Banach-valued Random Variables*, New York: Wiley.

Chen, K., Hu, I., and Ying, Z. (1999), "Strong Consistency of Maximum Quasi-likelihood Estimators in Generalized Linear Models with Fixed and Adaptive Designs," *Annals of Statistics*, 27, 1155-1163.

Chou, C. P., Bentler, P. M., and Satorra, A. (1991), "Scaled Test Statistics and Robust Standard Errors for Non-normal Data in Covariance Structure Analysis: a Monte Carlo Study," *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.

Chow, Y. S., and Teicher, H. (1978), *Probability Theory: Independence, Interchangeability, Martingales.* New York: Springer.

Deemer, W. L., and Olkin, I. (1951), "The Jacobians of Certain Matrix Transformation Useful in Multivariate Analysis, Based on Lectures by P.L.Hsu," *Biometrika*, 38, 345-367.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Ann. Math. Statist.* **34**, 447-456.

Ito, K., and Nisio, M. (1968), "On the Convergence of Sums of Independent Banach Space Valued Random Variables," *Osaka Mathematical Journal*, 5, 35-48.

Kingsley, P. B. (2006a), "Introduction to Diffusion Tensor Imaging Mathematics: Part I. Tensors, Rotations, and Eigenvectors," *Concepts in Magnetic Resonance Part A*, 28A, 101-122.

Kingsley, P. B. (2006b), “Introduction to Diffusion Tensor Imaging Mathematics: Part II. Anisotropy, Diffusion-weighting Factors, and Gradient Encoding Schemes,” *Concepts in Magnetic Resonance Part A*, 28A, 123-154.

Kingsley, P. B. (2006c), “Introduction to Diffusion Tensor Imaging Mathematics: Part III. Tensor Calculation, Noise, Simulations, and Optimization,” *Concepts in Magnetic Resonance Part A*, 28A, 155-179.

Lai, T. L., Robbins, H., and Wei, C. Z. (1979), “Strong Consistency of Least Squares Estimates in Multiple Regression II,” *Journal of Multivariate Analysis*, 9, 343-361.

Okamoto, M. (1973), “Distinctness of the Eigenvalues of a Quadratic Form in a Multivariate Sample,” *Annals of Statistics*, 1, 763-765.

White, H. L. (1980), “A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, 48, 817-838.

Zaman, A. (1989), “Consistency via Type 2 Inequalities: a Generalization of Wu’s Theorem,” *Econometric Theory*, 5, 272-286.

Zhu, H., and Zhang, H. (2006), “Asymptotics for Estimation and Testing Under Loss of Identifiability,” *Journal of Multivariate Analysis*, 97, 19-45.