ASYMPTOTIC EXPANSIONS OF THE LIKELIHOOD RATIO TEST STATISTIC WITH ORDERED HYPOTHESES

By Yazhen Wang *
University of Missouri-Columbia, U. S. A.
Sankhya A 57 (1995), pp. 410-423

SUMMARY. Likelihood ratio tests are used to test ordered hypotheses involving the parameters of k independent samples from an exponential family. For the test of constancy of the parameters versus an ordered alternative, the likelihood ratio test statistic is asymptotically distributed as a mixture of k chi-squared distributions under the null hypothesis. This paper derives an asymptotic expansion of the null distribution of the likelihood ratio test statistic in powers of $n^{-1/2}$, where n is the average sample size. In particular, it is shown that the null distribution of the likelihood ratio test statistic is generally order $n^{-1/2}$ away from the limiting distribution. For the equal sample size case, however, all odd powers of $n^{-1/2}$ vanish, and, similar to the Bartlett adjustment, the likelihood ratio test statistic can be adjusted through multiplication by scale factors such that the null distribution of the adjusted likelihood ratio test statistic is n^{-2} away from the limiting distribution.

1. Introduction

For $i = 1, \cdots, k$, suppose that $\{X_{ij}, 1 \leq j \leq n_i\}$ is a random sample of size n_i from a population with the density $f(x;\theta_i)$, and also suppose that the k samples are independent. Set $n = \sum_{i=1}^k n_i/k$, the average sample size, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$. With $H_0 : \theta_1 = \cdots = \theta_k$, and $H_1 : \theta_1 \leq \cdots \leq \theta_k$, the likelihood ratio test statistic for H_0 versus $H_1 - H_0$ is

$$\Lambda = 2log\{L(\hat{\boldsymbol{\theta}}^*)/L(\hat{\boldsymbol{\theta}}_0)\},$$

where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$, $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}_1^*, \cdots, \hat{\theta}_k^*)$, and $\hat{\boldsymbol{\theta}}_0 = (\hat{\theta}_0, \cdots, \hat{\theta}_0)$ are the unrestricted MLE of $\boldsymbol{\theta}$, the MLE of $\boldsymbol{\theta}$ under H_1, and the MLE of $\boldsymbol{\theta}$ under H_0, respectively. All

*Research supported in part by the Research Board Grant in the University of Missouri.
AMS 1980 subject classifications. Primary 62F05; secondary 62F30.
Key Words and Phrases. Asymptotic expansion, Bartlett adjustment, chi-bar-squared distribution, Edgeworth expansion, isotonic estimation, likelihood ratio statistic, ordered hypothesis.

these quantities, of course, depend on n_i but we suppress this dependence to ease the notation. Under H_0, the limiting distribution of the Λ is a mixture of k chi-squared distributions. Specifically, for $\boldsymbol{\theta} \in H_0$,

$$P_{\boldsymbol{\theta}}(\Lambda \leq x) = P(1, k, \mathbf{w}) + \sum_{m=2}^{k} P(m, k, \mathbf{w}) \int_0^x \chi^2_{m-1}(t)dt + o(1), \tag{1.1}$$

as $n_i \to \infty$, where $\chi^2_{m-1}(t)$ is the density of a chi-squared distribution with $m-1$ degrees of freedom for $m \geq 2$, $P(m, k, \mathbf{w})$ is a probability function which does not depend on the density f, see Robertson et al. (1988) or Section 2. This paper derives an asymptotic expansion of the null distribution of the likelihood ratio test statistic, Λ, in powers of $n^{-1/2}$. Because the ordered alternative destroys the symmetry, unlike the expansions in Box (1954) and Chandra and Ghosh (1979) for the null distribution of the likelihood ratio test statistic for the hypothesis with an unrestricted alternative, odd powers of $n^{-1/2}$ will appear in the asymptotic expansion of the null distribution of the Λ, and thus the null distribution of the Λ is generally order $n^{-1/2}$ away from its limiting distribution. For the equal sample size case, however, all odd powers of $n^{-1/2}$ vanish, and thus the null distribution of the Λ is generally order n^{-1} away from its limiting distribution. Moreover, the coefficient of n^{-1} is a linear function, and thus, similar to the Bartlett adjustment, an adjustment through multiplication by scale factors will be made for the Λ such that the null distribution of the adjusted likelihood ratio test statistic is n^{-2} away from the limiting distribution of the Λ.

Section 2 introduces notation and reviews results about the likelihood ratio test for an ordered alternative. Section 3 gives an asymptotic expansion of the null distribution of the Λ. Section 4 gives an asymptotic expansion in the equal sample size case. A Bartlett-type adjustment for the Λ is also considered in Section 4.

Notation. Throughout this paper, denote by $\chi^2_m(t)$ the density of a chi-squared distribution with m degrees of freedom. Convention, denote by χ^2_0 the delta function, or equivalently, the formal density of the random variable which is identically zero. Under this convention, the sum of the first two terms in the right hand side of (1.1) can be written as $\sum_{m=1}^{k} P(m, k, \mathbf{w}) \int_0^x \chi^2_{m-1}(t)dt$.

2. Preliminaries

This section introduces notation and reviews results about tests with an ordered alternative. The material can be found in Chapter 1, Chapter 2 and Section 4.1 of Robertson et al. (1988). Some can also be found in Barlow et al. (1972).

Suppose that an exponential family of distributions has the following probability densities with respect to a σ-finite measure

$$f(x; \theta) = exp\{p(\theta)K(x) + S(x) + q(\theta)\}, \tag{2.1}$$

where $\theta \in \Theta$, an open interval.

Assumption I:

(a) Both p and q have continuous second derivatives on Θ.

(b) $p'(\theta) > 0$ for all $\theta \in \Theta$.

(c) $q'(\theta) = -\theta p'(\theta)$ for all $\theta \in \Theta$.

The assumption I(c), which is needed in order to apply isotonic regression theory, appears rather strong, but many of the commonly studied exponential families satisfy these condition, see Robertson et al. (1988, p.33-38). By properties of the exponential family, $E[K(X)] = \theta$, and $\mathrm{Var}[K(X)] = [p'(\theta)]^{-1}$.

Let X_{ij} be independent with the density of X_{ij} given by $f(x;\theta_i)$, $i = 1, \cdots, k$, $j = 1, \cdots, n_i$, $n = \sum_{i=1}^k n_i/k$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$. Let $H_0 : \theta_1 = \cdots = \theta_k$, $H_1 : \theta_1 \le \cdots \le \theta_k$. Denote by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$, $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}_1^*, \cdots, \hat{\theta}_k^*)$, and $\hat{\boldsymbol{\theta}}_0 = (\hat{\theta}_0, \cdots, \hat{\theta}_0)$ the unrestricted MLE of $\boldsymbol{\theta}$, the MLE of $\boldsymbol{\theta}$ under H_1, and the MLE of $\boldsymbol{\theta}$ under H_0, respectively, then

$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$, $\hat{\theta}_i = \sum_{j=1}^{n_i} K(X_{ij})/n_i$,

$\hat{\theta}_0 = \sum_{i=1}^k w_i \hat{\theta}_i / \sum_{i=1}^k w_i$, and

$\hat{\boldsymbol{\theta}}^* = P_{\mathbf{w}}(\hat{\boldsymbol{\theta}}|\mathcal{J})$,

where $\mathbf{w} = (w_1, \cdots, w_k)$, $w_i = n_i/n$, $P_{\mathbf{w}}(\hat{\boldsymbol{\theta}}|\mathcal{J})$ is the projection of $\hat{\boldsymbol{\theta}}$ on the σ-lattice, $\mathcal{J}$, the cone of monotone function on $\{1, 2, \cdots, k\}$, or equivalently, $\hat{\boldsymbol{\theta}}^*$ is the isotonic regression of $\hat{\boldsymbol{\theta}}$ with weights $\mathbf{w}$, see Theorem 1.5.2 on p.34 of Robertson et al. (1988). Then the likelihood ratio test statistic for H_0 vs $H_1 - H_0$ is

$$\Lambda = 2\sum_{i=1}^k n_i \hat{\theta}_i [p(\hat{\theta}_i^*) - p(\hat{\theta}_0)] + 2\sum_{i=1}^k n_i [q(\hat{\theta}_i^*) - q(\hat{\theta}_0)].$$

The following characterization of $\hat{\boldsymbol{\theta}}^*$ will be used in the subsequent sections. Given a sequence $y_1, \cdots, y_k$ and a subset B of $\{1, \cdots, k\}$, denote by $Av(B, y, \mathbf{w})$ the weighted average $\sum_{i\in B} w_i y_i / \sum_{i\in B} w_i$. When $w_1 = \cdots = w_k$, $\mathbf{w}$ will be suppressed and $Av(B, y, \mathbf{w})$ will be written as $Av(B, y)$. A subset L of $\{1, 2, \cdots, k\}$ is said to be a lower layer, if $i \in L$ implies all $j \in L$ for $j \le i$.

Given a partition, $B_1, \cdots, B_m$, of $\{1, 2, \cdots, k\}$, if $\hat{\boldsymbol{\theta}}^*$ is constant on each B_ℓ and the value of $\hat{\boldsymbol{\theta}}^*$ on B_ℓ increases with ℓ, we call $B_1, \cdots, B_m$ the ordered level sets of $\hat{\boldsymbol{\theta}}^*$ and m the number of level sets in $\hat{\boldsymbol{\theta}}^*$.

For lower layers $L_0, L_1, \cdots, L_m$, with $\emptyset = L_0 \subset L_1 \subset \cdots \subset L_m = \{1, 2, \cdots, k\}$, set $B_i = L_i - L_{i-1}$. By Lemma A on p.70 of Robertson et al. (1988), the ordered level sets of $\hat{\boldsymbol{\theta}}^*$ are $B_1, \cdots, B_m$ if and only if

$$Av(B_1, \hat{\boldsymbol{\theta}}, \mathbf{w}) < Av(B_2, \hat{\boldsymbol{\theta}}, \mathbf{w}) < \cdots < Av(B_m, \hat{\boldsymbol{\theta}}, \mathbf{w}), \tag{2.2}$$

and for $L_{i-1} \subset L \subset L_i$ with $L - L_i \ne \emptyset$, $L - L_{i-1} \ne \emptyset$,

$$Av(B_i, \hat{\boldsymbol{\theta}}, \mathbf{w}) \le Av(L - L_{i-1}, \hat{\boldsymbol{\theta}}, \mathbf{w}), \quad \text{for } i = 1, 2, \cdots, k. \tag{2.3}$$

Define $G(B_1, \cdots, B_m)$ to be the event on which (2.2) and (2.3) are satisfied. Let $W_\ell = \sum_{i\in B_\ell} w_i$, $U_\ell = Av(B_\ell, \hat{\boldsymbol{\theta}}^*, \mathbf{w}) = \sum_{i\in B_\ell} w_i\hat{\theta}_i/W_\ell$, $U = \sum_{\ell=1}^m W_\ell U_\ell / \sum_{\ell=1}^m W_\ell = \hat{\theta}_0$, then on the event $G(B_1, \cdots, B_m)$, $\hat{\theta}_i^* = U_\ell$ for $i \in B_\ell$. Hence,

$$
\begin{aligned}
\Lambda &= 2\sum_{\ell=1}^{m}\sum_{i\in B_\ell} n_i\{\hat{\theta}_i[p(U_\ell) - p(U)] + [q(U_\ell) - q(U)]\} \\
&= 2n\sum_{\ell=1}^{m} W_\ell\{U_\ell[p(U_\ell) - p(U)] + [q(U_\ell) - q(U)]\}. \qquad (2.4)
\end{aligned}
$$

Denote by M the number of level sets in $\hat{\boldsymbol{\theta}}^*$. Under the H_0, the level probability

$$P(M = m) = \sum_{B_1,\cdots,B_m} P(G(B_1, \cdots, B_m)), \qquad (2.5)$$

where the summation is taken over all possible m ordered level sets. If $\hat{\theta}_1, \cdots, \hat{\theta}_k$ are i.i.d. with a continuous distribution function, then the probability function of M, denote by $P(m, k)$, has the probability generating function $P_k(s) = s(s+1)\cdots(s+k-1)/k!$ which is free from the distribution $f(x, \theta)$, see Theorem 2.4.2 on p.81 of Robertson et al. (1988). When $f(x, \theta)$ is the density of a normal distribution with mean θ and variance 1, denote by $P(m, k, \mathbf{w})$ the level probabilities $P(M = m)$ under H_0, which is independent of θ, with weights $\mathbf{w} = (w_1, \cdots, w_k)$, $w_i = n_i/n$, see Robertson et al. (1988, p.74-77). Note that $P(m, k, \mathbf{w}) = P(m, k)$ if $w_1 = \cdots = w_k$.

Finally,

$$P(\Lambda \le x) = \sum_{m=1}^{k}\sum_{B_1,\cdots,B_m} P(\Lambda \le x, G(B_1, \cdots, B_m)) = \sum_{m=1}^{k} P(\Lambda \le x, M = m). \qquad (2.6)$$

If $f(x, \theta)$ is the density of a normal distribution with mean θ and variance 1, from (2.4) and (2.6), one can derive that the distribution of Λ is a mixture of k chi-squared distributions with the distribution function $\sum_{m=1}^k P(m, k, \mathbf{w}) \int_0^x \chi^2_{m-1}(u)du$, which is called a chi-bar-squared distribution, see Theorem 2.3.1 and its Corollary on p.69 of Robertson et al. (1988).

Generally, for $0 < a_i < k$, $\sum_{i=1}^k a_i = k$, if $n = \sum_{i=1}^k n_i/k \to \infty$ and $n_i/n \to a_i$, then $\hat{\theta}_1, \cdots, \hat{\theta}_k$ are asymptotic normal. The operation $P_{\mathbf{w}}(\mathbf{x}|\mathcal{J})$ is continuous in $(\mathbf{x}, \mathbf{w})$, so $\hat{\boldsymbol{\theta}}^*$ depends on $\hat{\boldsymbol{\theta}}$ continuously. Therefore, under H_0, Λ is asymptotically distributed as a chi-bar-squared distribution with the distribution function $\sum_{m=1}^k P(m, k, \mathbf{w}) \int_0^x \chi^2_{m-1}(u)du$, where $\mathbf{w} = (w_1, \cdots, w_k)$, $w_i = a_i$. See Theorem 4.1.1 on p.164 of Robertson et al. (1988).

3. Asymptotic expansions

Suppose that $w_i = n_i/n = a_i$, $i = 1, \cdots, k$, are fixed constants. Also suppose that $p(\theta)$ and $q(\theta)$ have continuous derivatives up to the sth order. Under H_0, expanding

$p(\cdot)$ and $q(\cdot)$ about the common value $\theta_1 = \cdots = \theta_k$ using Taylor's theorem,

$$\Lambda = 2\sum_{i=1}^{k} n_i\{\sum_{j=1}^{s} \frac{\hat{\theta}_i p^{(j)}(\theta_1) + q^{(j)}(\theta_1)}{j!}[(\hat{\theta}_i^* - \theta_1)^j - (\hat{\theta}_0 - \theta_1)^j]\} + o_p(n^{-s/2+1}). \quad (3.1)$$

Recalling the assumption I(c) that $q'(\theta) = -\theta p'(\theta)$, taking jth order derivative and then adding $\hat{\theta}_i p^{(j)}(\theta_1)$ on both sides of the equation, one obtains

$$\hat{\theta}_i p^{(j)}(\theta_1) + q^{(j)}(\theta_1) = (\hat{\theta}_i - \theta_1)p^{(j)}(\theta_1) - (j-1)p^{(j-1)}(\theta_1). \quad (3.2)$$

However,

$$\sum_{i=1}^{k} w_i(\hat{\theta}_i - \theta_1)(\hat{\theta}_0 - \theta_1)^j = \sum_{i=1}^{k} w_i(\hat{\theta}_0 - \theta_1)^{j+1}, \quad (3.3)$$

and note that $\{\hat{\theta}_i^* - \theta_1 : 1 \le i \le k\}$ is the isotonic regression of $\{\hat{\theta}_i - \theta_1 : 1 \le i \le k\}$ with weights $\mathbf{w}$, so, by Theorem 1.3.6 on p.21 of Robertson et al. (1988),

$$\sum_{i=1}^{k} w_i(\hat{\theta}_i - \theta_1)(\hat{\theta}_i^* - \theta_1)^j = \sum_{i=1}^{k} w_i(\hat{\theta}_i^* - \theta_1)^{j+1}. \quad (3.4)$$

Substituting (3.2) into (3.1) and applying (3.3) and (3.4), one gets

$$\Lambda = 2n\sum_{j=2}^{s} \frac{p^{(j-1)}(\theta_1)}{j(j-1)!}\sum_{i=1}^{k} w_i\{(\hat{\theta}_i^* - \theta_1)^j - (\hat{\theta}_0 - \theta_1)^j\} + o_p(n^{-s/2+1}). \quad (3.5)$$

Let $\mathbf{Z} = (Z_1, \cdots, Z_k)$, $Z_i = (np'(\theta_1))^{1/2}(\hat{\theta}_i - \theta_1)$, $\mathbf{Z}^* = (Z_1^*, \cdots, Z_k^*)$ the isotonic regression of $(Z_1, \cdots, Z_k)$ with weights $\mathbf{w}$, $Z_0 = \sum_{i=1}^k w_i Z_i / \sum_{i=1}^k w_i$, and

$$d_j = \frac{2p^{(j+1)}(\theta_1)(p'(\theta_1))^{-j/2-1}}{(j+2)(j+1)!}.$$

Then, substituting d_j, Z_i and Z_0 into (3.5), one gets that under H_0,

$$\Lambda = \sum_{j=0}^{s-2} d_j n^{-j/2}\sum_{i=1}^{k} w_i\{(Z_i^*)^{j+2} - (Z_0)^{j+2}\} + o_p(n^{-s/2+1}). \quad (3.6)$$

Suppose that the Cramer's condition

$$\limsup_{|t|\to\infty} |E_f\{exp(itK(X))\}| < 1, \quad (3.7)$$

is satisfied, then $\mathbf{Z}$ has a probability density p which admits an Edgeworth expansion of the form

$$p(\mathbf{z}) = \sum_{j=0}^{s-2} n^{-j/2}p_j(\mathbf{z})\phi(\mathbf{z}/\sqrt{\mathbf{w}}) + o(n^{-s/2+1}), \quad (3.8)$$

where $\mathbf{z}/\sqrt{\mathbf{w}} = (z_1/\sqrt{w_1}, \cdots, z_k/\sqrt{w_k})$, ϕ is the density of the standard normal distribution, $p_0(\mathbf{z}) = 1$, and p_j is an odd polynomial for odd j and an even polynomial for even j. See Barndorff-Nielsen & Cox (1979,1984).

By (3.6) and (3.8), the following asymptotic expansion, proved in the appendix, can be derived.

Theorem 3.1 : *Suppose that Cramer's condition (3.7) is satisfied and $p(\theta)$ and $q(\theta)$ have continuous derivatives up to the sth order, then under H_0,*

$$P(\Lambda \leq x) = \sum_{m=1}^{k} P(m, k, \mathbf{w}) \int_0^x \sum_{j=0}^{s-2} n^{-j/2} b_{mj}\, \chi^2_{m-1+j}(u) du + o(n^{-s/2+1}), \quad (3.9)$$

where the b_{mj} are constants, $b_{m0} = 1$, and $P(m, k, \mathbf{w})$ is the level probability defined in Section 2.

Remark 3.1: For an unrestricted alternative, the asymptotic expansion of the likelihood ratio test statistic in Theorem 1 and Remark 2.2 of Chandra and Ghosh (1979) is in powers of n^{-1} even for the non i.i.d. sample. Because the ordered alternative destroys the symmetry, odd powers of $n^{-1/2}$ do appear in the expansion (3.9) in the case of unequal sample sizes, which is verified by the following example.

Example 3.1: $k = 2$, $n_1 = 2n$, $n_2 = n$. Suppose that X_{ij} follows the exponential distribution

$$f(x, \theta_i) = exp\{-x/\theta_i - \log \theta_i\},$$

then $\hat{\theta}_i = \bar{X}_i$, $i = 1, 2$, $\hat{\theta}_0 = (2\bar{X}_1 + \bar{X}_2)/3$, and

$$\Lambda = -2 \sum_{i=1}^{2} n_i log(\hat{\theta}_i^*/\hat{\theta}_0) = -2 \sum_{i=1}^{2} n_i log(\hat{\theta}_i/\hat{\theta}_0) 1(\hat{\theta}_1 \leq \hat{\theta}_2),$$

where $1(A)$ is the indicate function of A. Let $Z = 2\bar{X}_1/(2\bar{X}_1 + \bar{X}_2)$. Then

$$\Lambda = -2n\{2 log(Z) + \log(1 - Z) - log(4/27)\} 1(Z \leq 2/3),$$

or equivalently,

$$(4/27)e^{-\Lambda/2n} = Z^2(1 - Z) 1(Z \leq 2/3).$$

Solving the cubic equation, one gets

$$\begin{aligned} Z &= 1/3 - 2/3\, cos\{cos^{-1}(1 - 2e^{-\Lambda/2n})/3 + \pi/3\} \\ &= 2/3 + (2/3\sqrt{6})\, \Lambda^{1/2} n^{-1/2} - (2/27)\, \Lambda n^{-1} + O(n^{-3/2}), \end{aligned} \quad (3.10)$$

and

$$\frac{dZ}{d\Lambda} = -\frac{e^{-\Lambda/4n}}{9} \Lambda^{-1/2} n^{-1/2} \{\frac{\sqrt{6}}{2} - \frac{1}{3} \Lambda^{1/2} n^{-1/2} - \frac{25\sqrt{6}}{144} \Lambda\, n^{-1} + O(n^{-3/2})\}. \quad (3.11)$$

However, $n_i \bar{X}_i$ follows a gamma distribution $\Gamma(n_i, 1/\theta_i)$. Then, under H_0, Z has the beta distribution $B(2n, n)$. By (3.10) and (3.11), the conditional density of Λ given $Z \leq 2/3$ is equal to

$$\frac{\Gamma(3n)}{\Gamma(2n)\Gamma(n)} (4/27)^{n-1}\, e^{-\frac{\lambda}{2}}\, e^{\lambda/2n}\, Z(\lambda)\, \frac{dZ}{d\lambda}$$
$$= (2\pi)^{-1/2}\, \lambda^{-1/2}\, e^{-\frac{\lambda}{2}}\{1 + \frac{1}{3\sqrt{6}}\, \lambda^{1/2}\, n^{-1/2} - \frac{23}{72}\, \lambda\, n^{-1} + O(n^{-3/2})\},$$

and $\Lambda = 0$ if $Z > 2/3$. One can see that the term $n^{-1/2}$ does appear in the expansion of the null distribution of Λ.

Remark 3.2: The assumption that n_i/n are equal to the fixed constants a_i, $i = 1, \cdots, k$, can be relaxed to $n_i = a_i n + a_{i1} n^{1/2} + \cdots$.

Remark 3.3: The expansion in Theorem 3.1 can be generalized to the following independent samples with nonidentical distributions: X_{ij} has the density $f(x; \theta_i, \tau_i)$, where

$$f(x; \theta, \tau) = exp\{p_1(\theta) p_2(\tau) K(x; \tau) + S(x; \tau) + q(\theta, \tau)\},$$

$\tau \in \Xi$, $\theta \in \Theta$, an open interval.

4. Asymptotic expansions in the equal sample size case

Suppose that $n_1 = \cdots = n_k = n$, then $w_1 = \cdots = w_k = 1$, and thus $\mathbf{w}$ will be suppressed in the notation. In the equal sample size case, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$, $\hat{\theta}_i = \sum_{j=1}^{n} K(X_{ij})/n$, $\hat{\theta}_0 = \sum_{i=1}^{k} \hat{\theta}_i/(nk)$, and $\hat{\boldsymbol{\theta}}^*$ is the isotonic regression of $\hat{\boldsymbol{\theta}}$ with equal weights.

Theorem 4.1 : *Suppose that the conditions of Theorem 3.1 are satisfied, and $n_1 = \cdots = n_k = n$, then under H_0,*

$$P(\Lambda \leq x) = \sum_{m=1}^{k} P(m,k) \int_0^x \chi^2_{m-1}(u)\{1 + \frac{b_m(u - m + 1)}{2n} + \sum_{j=2}^{[s/2]-1} n^{-j} \varphi_{mj}(u)\} du$$
$$+ o(n^{-s/2+1}),$$

where the φ_{mj} are polynomials, and

$$b_m = \{\sum_{\mathbf{r} \in \mathcal{R}_m} c_{\mathbf{r},m} / \prod_{i=1}^{k} r_i!\, i^{r_i}\}/P(m,k). \tag{4.1}$$

Remark 4.1: For the unequal sample size case, odd powers appear in the expansion because the ordered alternative destroys the symmetry. However, the asymmetry caused by the ordered alternative is compensated by the equal sample size, and then all odd powers of $n^{-1/2}$ vanish in the expansion of the null distribution of likelihood ratio test statistic in the equal sample size case. Moreover, the coefficient of the term n^{-1} is a linear function. These will be clearly shown in the following example.

Example 4.1: $k = 2$, $n_1 = n_2 = n$. Suppose that X_{ij} follows an exponential distribution,

$$f(x, \theta_i) = exp\{-x/\theta_i - \log \theta_i\},$$

then $\hat{\theta}_i = \bar{X}_i$, $i = 1, 2$, $\hat{\theta}_0 = (\bar{X}_1 + \bar{X}_2)/2$, and

$$\Lambda = -2n \sum_{i=1}^{2} log(\hat{\theta}_i^*/\hat{\theta}_0) = -2n \sum_{i=1}^{2} log(\hat{\theta}_i/\hat{\theta}_0) 1(\hat{\theta}_1 \leq \hat{\theta}_2).$$

Let $Z = \bar{X}_1/(\bar{X}_1 + \bar{X}_2)$. Then

$$\Lambda = -2n\{log(Z) + \log(1 - Z) + log\, 4\} 1(2Z \leq 1),$$

and thus $Z = \frac{1}{2}(1 - \{1 - \exp(-\frac{\Lambda}{2n})\}^{1/2})$. However, $n\bar{X}_i \sim \Gamma(n, 1/\theta_i)$. Then, under H_0, Z has the beta distribution $B(n, n)$. The conditional density of Λ given $Z \leq 1/2$ whose probability is one half is equal to

$$\frac{\Gamma(2n)}{n\,[\Gamma(n)]^2\, 2^{2n+1}}\, e^{-\frac{\lambda}{2}}(1 - e^{-\frac{\lambda}{2n}})^{-1/2} = (2\pi)^{-1/2} \lambda^{-1/2}\, e^{-\frac{\lambda}{2}}(1 + \frac{\lambda - 1}{8n} + O(n^{-2})),$$

and $\Lambda = 0$ if $Z > 1/2$. This shows that the odd power terms $n^{-1/2}$ and $n^{-3/2}$ vanish, and the coefficient of the term n^{-1} is a linear function.

Proof of Theorem 4.1: Theorem 4.1 can be directly proved by the method in Section 3 and saddle-point approximations. However, here we will give a simple proof by applying Theorem 1.2 in Boswell and Brunk (1969).

Let $\mathcal{R}_m = \{\mathbf{r} = (r_1, \cdots, r_k) : r_1, \cdots, r_k$ are nonnegative integers, $\sum_{i=1}^{k} i\, r_i = k$, and $\sum_{i=1}^{k} r_i = m\}$. For $\mathbf{r} \in \mathcal{R}_m$, define $\mathcal{A}^{\mathbf{r}} = \{\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m) : \alpha_1, \cdots, \alpha_m$ are positive integers; and exactly r_i of the components of $\boldsymbol{\alpha}$ are equal to $i\}$.

For $\mathbf{r} \in \mathcal{R}_m$, $\boldsymbol{\alpha} \in \mathcal{A}^{\mathbf{r}}$, set $\beta_j = \sum_{i=1}^{j} \alpha_i$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)$, $Y_j = \sum_{i=\beta_{j-1}+1}^{\beta_j} \hat{\theta}_i$, $\bar{Y}_j = Y_j/(n\alpha_j)$, for $j = 1, \cdots, m$, $\bar{Y}_0 = \sum_{j=1}^{m} \alpha_j \bar{Y}_j / \sum_{j=1}^{m} \alpha_j$, and

$$\Lambda(\boldsymbol{\alpha}) = 2n \sum_{j=1}^{m} \alpha_j Y_j [p(Y_j) - p(Y_0)] + 2 \sum_{j=1}^{m} \alpha_j [q(Y_j) - q(Y_0)].$$

By Theorem 1.2 or Equation (2.8) in Boswell and Brunk (1969), one has that under H_0,

$$P(\Lambda \leq x) = \sum_{m=1}^{k} \sum_{\mathbf{r} \in \mathcal{R}_m} \{P(\Lambda(\boldsymbol{\alpha}) \leq x) / \prod_{i=1}^{k} r_i!\, i^{r_i}\}, \tag{4.2}$$

where $\boldsymbol{\alpha}$ is selected arbitrarily from $\mathcal{A}^{\mathbf{r}}$ for each $\mathbf{r} \in \mathcal{R}_m$.

Suppose, for a moment, that $\theta_1 = \cdots = \theta_{\beta_1}, \cdots, \theta_{\beta_{m-1}+1} = \cdots = \theta_{\beta_m}$ (Recall that $\beta_m = k$) and denote the m common values of θ_i by $\xi_1, \cdots, \xi_m$. Treating $Y_1, \cdots, Y_m$ as independent observations, one can see that $\Lambda(\boldsymbol{\alpha})$ is the likelihood ratio test statistic for testing $\xi_1 = \cdots = \xi_m$ against that at least two of the $\xi_1, \cdots, \xi_m$ differ. Applying Theorem 2 in Chandra and Ghosh (1979), the second paragraph on page 1072 of Bickel and Ghosh (1990), and theory of Bartlett adjustments, e.g. Barndorff-Nielsen and Cox (1979, 1984), Barndorff-Nielsen and Hall (1988), Lawley (1956), one has that under H_0,

$$P(\Lambda(\boldsymbol{\alpha}) \leq y) = \int_0^y \chi^2_{m-1}(u)\{1 + \frac{c_{\mathbf{r},m}\,(u-m+1)}{2n} + \sum_{j=2}^{[s/2]-1} n^{-j}\varphi_{\mathbf{r}mj}(u)\}du$$
$$+\, o(n^{-s/2+1}), \tag{4.3}$$

where the $\varphi_{\mathbf{r}mj}$ are polynomials, and the $c_{\mathbf{r},m}$ are constants satisfying

$$E\{\Lambda(\boldsymbol{\alpha})\} = (m-1)(1 + c_{\mathbf{r},m}n^{-1}) + O(n^{-2}). \tag{4.4}$$

Recall that $P(m,k)$ has the probability generating function $P_k(s) = s(s+1)\cdots(s+k-1)/k!$ which is the generating function of the absolute values of the Stirling's Number of the First Kind, and then

$$P(m,k) = \sum_{r \in \mathcal{R}_m} 1/\prod_{i=1}^{k} r_i!\, i^{r_i}, \tag{4.5}$$

see Abramowitz and Stegun (1965, p.824) and Boswell and Brunk (1969, p.373). Substituting (4.3) into (4.2) and applying (4.5), one can easily prove the theorem. □

Corollary 4.1 : *Suppose that the Cramer's condition (3.7) is satisfied, and $n_1 = \cdots = n_k = n$, and $p(\theta)$ and $q(\theta)$ have continuous derivatives up to the sixth order. Then under H_0,*

$$P(\Lambda \leq x) = \sum_{m=1}^{k} P(m,k)\int_0^x \chi^2_{m-1}(u)\{1 + \frac{b_m(u-m+1)}{2\,n}\}du + O(n^{-2}).$$

Set

$$\Lambda' = \sum_{m=1}^{k} \Lambda\, 1(M=m)/(1 + b_m n^{-1}), \tag{4.6}$$

where the b_m are defined in (4.1).

The following theorem shows that, like the Bartlett adjustment, after the adjustment (4.6) the null distribution of the likelihood ratio statistic Λ is order n^{-2} away from the chi-bar-squared distribution.

Theorem 4.2 : *Under the conditions of Corollary 4.1,*

$$P(\Lambda' \leq x) = \sum_{m=1}^{k} P(m,k)\int_0^x \chi^2_{m-1}(t)dt + O(n^{-2}).$$

Proof: By Theorem 2.1 in Boswell and Brunk (1969) and the arguments to prove Theorem 4.1, one can easily prove the theorem. □

Remark 4.2: By (1.1), the limit distribution of the likelihood ratio test statistic Λ under H_0 is a mixture of k chi-squared distributions with mixture probabilities $P(m,k)$. Under H_0, it can be shown that the conditional distribution Λ given $M = m$ is asymptotically distributed as a chi-squared distribution with $m-1$ degrees of freedom, see the proofs of Theorem 2.3.1 and Theorem 4.1.1 of Robertson et al. (1988); so theory of Bartlett adjustments implies that the null distribution of Λ will be order n^{-2} away from its limit distribution after Λ is divided by $1 + b_{m,k}(n-1)^{-1}$ conditioning on $M = m$. Therefore, Theorem 1 essentially follows from Bartlett adjustment phenomenon and symmetry, and the adjustment (4.6) is just a mixture of Bartlett adjustments.

Remark 4.3: From (4.4) to (4.1), one can see that the constants $1 + c_{r,m,k}(n-1)^{-1}$ in (4.4) are Bartlett adjustment constants, see Barndorff-Nielsen and Cox (1979,1984), Barndorff-Nielsen and Hall (1988), Bickel and Ghosh (1990), Hayakawa (1977) and Lawley (1956); so the adjustment constants $1 + b_{m,k}(n-1)^{-1}$ in (4.6) are the weighted average of Bartlett adjustment constants.

Remark 4.4: The problem of testing for trend in gamma variables may arise in several different ways. For example, testing exponential means and normal variances. For this problem the constants $b_1, \cdots, b_k$ do not depend on the unknown parameters $\boldsymbol{\theta}$. Wang (1993) gave an explicit formula for computing b_m and tabulated the values of b_m for $k \leq 10$.

Acknowledgments. I would like to express my thanks to professor Peter J. Bickel for his encouragements and professor Farroll T. Wright for his comments and correction of the language. Also I would like to thank a referee for many helpful suggestions.

Appendix

By (1.1), the limit distribution of the likelihood ratio test statistic, Λ, under H_0 is a mixture of k chi-squared distributions with mixture probabilities $P(m,k)$, so the terms corresponding to $j = 0$ in the expansion (3.9) is true. Hence, we will focus on derivation of the terms with the orders $n^{-j/2}$, $j \geq 1$. In the following proofs, a, b, c and d are constants which do not depend on n.

Lemma A.1 *:*

$$P(G(B_1, \cdots, B_m)) = \sum_{j=0}^{s-2} n^{-j/2} b_{mj1} + o(n^{-s/2+1}), \tag{A.1}$$

and

$$P(M = m) = P(m,k,w)(1 + \sum_{j=0}^{s-2} n^{-j/2} b_{mj2}) + o(n^{-s/2+1}), \tag{A.2}$$

where $G(B_1, \cdots, B_m)$ is the event on which (2.2) and (2.3) are satisfied, M is the number of level sets in $\hat{\boldsymbol{\theta}}^*$ defined in Section 2.

PROOF. By the Edgeworth expansion (3.8). $\square$

LEMMA A.2 *:*

$$P(\Lambda \leq x, G(B_1, \cdots, B_m)) = \int_0^x \{\chi^2_{m-1}(t) + \sum_{j=1}^{s-2} n^{-j/2} b_{mj3} \chi^2_{m-1+j}(t)\}dt + o(n^{-s/2+1}). \tag{A.3}$$

PROOF. The proof of this lemma is similar to the proofs of Theorems 1 and 2 in Chandra and Ghosh (1979) and the Main Theorem in Chandra (1985), so I sketch the proof only.

Define a transformation T_1 as follows:

a) First send $(z_1, z_2, \cdots, z_k)$ to $(u_1, \cdots, u_m, v_1, \cdots, v_{k-m})$, where $\mathbf{u} = (u_1, \cdots, u_m)$ and $\mathbf{v} = (v_1, \cdots, v_{k-m})$ are defined as follows: For $\ell = 1, \cdots, m$, let Q_ℓ be an orthogonal matrix with last row $\{w_j^{1/2}/(\sum_{j \in B_\ell} w_j)^{1/2} : j \in B_\ell\}$. Then $(u_1, \cdots, u_m, v_1, \cdots, v_{k-m})$ is equal to $\{Q_\ell \{w_j^{1/2} z_j : j \in B_\ell\}, \ell = 1, \cdots, m\}$ in the way that
$$u_\ell = \sum_{j \in B_\ell} w_j z_j / (\sum_{j \in B_\ell} w_j)^{1/2}, \quad \ell = 1, \cdots, m,$$
and $v_1, \cdots, v_{k-m}$ are the rest of $Q_\ell \{w_j^{1/2} z_j : j \in B_\ell\}$.

b) Second, send $(u_1, u_2, \cdots, u_m)$ to $(r, \beta_1, \cdots, \beta_{m-1})$ by means of a polar transformation and have the v_i unaltered, e.g.
$$u_1 = r \prod_{i=1}^{m-1} cos\beta_i, \quad u_j = r sin\beta_{m-j+1} \prod_{i=1}^{m-j} cos\beta_i, \quad \text{for } 2 \leq j \leq m-1,$$
where $r > 0$, $-\pi/2 < \beta_i \leq \pi/2$, $i = 1, \cdots, m-2$, $0 < \beta_{m-1} \leq 2\pi$.

The Jacobian of the transformation T_1 is $r^{m-1} J(\boldsymbol{\beta})$, where

$$J(\boldsymbol{\beta}) = c\,(cos\beta_1)^{m-2}(cos\beta_2)^{m-3} \cdots (cos\beta_{m-2}),$$

$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{m-1})$, and c is a positive constant depending on k and W_ℓ only. Denote by $R(r, \boldsymbol{\beta}, \mathbf{v})$ an expression of the form

$$r^{d_0} \prod_{i=1}^{m} (u_i/r)^{d_{1i}} \, [\prod_{i=1}^{k-m} v^{d_{2i}}],$$

where the d_{ij} are integers, $d_0 = \sum_{i=1}^m d_{1i}$, and $R_{ij}(r, \boldsymbol{\beta}, \mathbf{v})$ is a finite sum of constant multiples of terms of the form $R(r, \boldsymbol{\beta}, \mathbf{v})$ with the property that the powers of r are even numbers for even j.

Under a normal underlying distribution, r, $\boldsymbol{\beta}$, and $\mathbf{v}$ are independent. By (3.8), the joint distribution of r, $\boldsymbol{\beta}$ and $\mathbf{v}$ has the form

$$exp(-r^2/2)\psi(\boldsymbol{\beta},\mathbf{v})(1+\sum_{j=1}^{s-2} n^{-j/2}R_{1j}(r,\boldsymbol{\beta},\mathbf{v}))+o(n^{-s/2+1}),$$

where $\psi(\boldsymbol{\beta},\mathbf{v})$ is a function of $\boldsymbol{\beta}$ and $\mathbf{v}$ only. Let $R(r)=r^{m-1}exp(-r^2/2)$, $1(G(B_1,\cdots,B_m))$ be the indicator of the set $G(B_1,\cdots,B_m)$. One has

$$P(\Lambda\le x, G(B_1,\cdots,B_m))=\int_{\Lambda\le x} R(r)\psi(\boldsymbol{\beta},\mathbf{v})\{1+\sum_{j=1}^{s-2} n^{-j/2}R_{1j}(r,\boldsymbol{\beta},\mathbf{v})\}$$
$$J(\boldsymbol{\beta})1(G(B_1,\cdots,B_m))dr\,d\boldsymbol{\beta}\,d\mathbf{v}+o(n^{-s/2+1}). \tag{A.4}$$

The following argument is similar to those for proving (2.6) on p.30 of Chandra and Ghosh (1979) and Lemma 2.1 on p.106-108 of Chandra (1985). Define a transformation T_2 sending $(r,\boldsymbol{\beta},\mathbf{v})$ to $(r',\boldsymbol{\beta},\mathbf{v})$, where $r'=\Lambda^{1/2}$. By (3.6), one gets

$$(r')^2=r^2(1+\sum_{j=1}^{s-2} n^{-j/2}R_{2j}(r,\boldsymbol{\beta}))+o(n^{-s/2+1}),$$

$$r'=r(1+\sum_{j=1}^{s-2} n^{-j/2}R_{3j}(r,\boldsymbol{\beta}))+o(n^{-s/2+1}),$$

$$r=r'(1+\sum_{j=1}^{s-2} n^{-j/2}R_{4j}(r',\boldsymbol{\beta}))+o(n^{-s/2+1}),$$

and the Jacobian of T_2 is equal to

$$\frac{dr}{dr'}=1+\sum_{j=1}^{s-2} n^{-j/2}R_{5j}(r',\boldsymbol{\beta})+o(n^{-s/2+1}).$$

Because (2.2) and (2.3) are determined by $\boldsymbol{\beta}$ and $\mathbf{v}$, respectively, the event $G(B_1,\cdots,B_m)$ depends on $\boldsymbol{\beta}$ and $\mathbf{v}$ only which are independent of r under a normal underlying distribution. Applying the transformation T_2 in the integral in the right hand side of (A.4), one gets

$$P(\Lambda\le x, G(B_1,\cdots,B_m))=\int_{(r')^2\le x}(r')^{m-1}exp(-(r')^2/2)\psi(\boldsymbol{\beta},\mathbf{v})\{1+$$
$$\sum_{j=1}^{s-2} n^{-j/2}R_{6j}(r',\boldsymbol{\beta},\mathbf{v})\}J(\boldsymbol{\beta})1(G(B_1,\cdots,B_m))\,dr'\,d\boldsymbol{\beta}\,d\mathbf{v}+o(n^{-s/2+1}). \tag{A.5}$$

One now integrates with respect to $(\boldsymbol{\beta},\mathbf{v})$ and obtains (A.3). $\square$

Proof of Theorem 3.1: By (2.4), Lemma A.1 and Lemma A.2,

$$
\begin{aligned}
P(\Lambda \le x) &= \sum_{m=1}^{k} \sum_{B_1,\cdots,B_m} P(\Lambda \le x, G(B_1,\cdots,B_m)) \\
&= \sum_{m=1}^{k} \sum_{B_1,\cdots,B_m} \int_0^x \{\chi^2_{m-1}(t) + \sum_{j=1}^{s-2} n^{-j/2} b_{mj3} \chi^2_{m-1+j}(t)\} dt + o(n^{-s/2+1}) \\
&= \sum_{m=1}^{k} P(m,k,\mathbf{w}) \int_0^x \sum_{j=0}^{s-2} n^{-j/2} b_{mj} \chi^2_{m-1+j}(t) dt + o(n^{-s/2+1}). \quad \square
\end{aligned}
$$

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D.(1972): Statistical Inference Under Order Restrictions. New York: Wiley.

Barndorff-Nielsen, O. E. and Cox, D. R.(1979): Edgeworth and saddle-point approximations with statistical applications. J. Roy. Statist. Soc. Ser. B 41 279-312.

——— (1984): Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. J. Roy. Statist. Soc. Ser. B 46 483-495.

Barndorff-Nielsen, O.E and Hall, P.(1988): On the level-error after Bartlett adjustment of the likelihood ratio statistic. Biometrika 75 374-378.

Bhattacharya, R. N. and Ghosh, J. K.(1978): On the validity of the formal Edgeworth expansion. Ann. Statist. 6 434-451.

Bickel, P. J. and Ghosh, J. K.(1990): A decomposition for the likelihood ratio statistic and the Bartlett correction – A Bayesian argument. Ann. Statist. 18 1070-1090.

Boswell, M. T. and Brunk, H. D.(1969): Distribution of likelihood ratio in testing against trend. Ann. Math. Statist. 40 371-380.

Box, G. E. P.(1949). A general distribution theory for a class of likelihood criteria. Biometrika 36 317-346.

Chandra, T.(1985): Asymptotic expansions of perturbed chi-square variables. Sankhyā Ser. A 47 100-110.

Chandra, T. and Ghosh, J. K.(1979): Valid asymptotic expansion for the likelihood ratio statistic and other perturbed χ^2 variables. Sankhyā Ser. A 41 22-47.

Chandra, T. and Mukerjee, R. (1991): Bartlett-type modification for Rao's efficient score statistic. J. Multivariate Anal. 36 103-112.

Ghosh, J. K. and Mukerjee, R. (1991): Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. J. Multivariate Anal. 38 385-393.

Hayakawa, T.(1977): The likelihood ratio criterion and the asymptotic expansion of its distribution. Ann. Inst. Statist. Math. 29 359-378.

LAWLEY, D. N.(1956): A general method for approximating to the distribution of the likelihood ratio criteria. Biometrika 43 295-303.

MUKERJEE, R. and CHANDRA, K. (1991): Bartlett-type adjustment for the conditional likelihood ratio statistic of Cox and Reid. Biometrika 78, 365-372.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L.(1988): Order Restricted Statistical Inferences. New York: Wiley.

WANG, Y. (1994): A Bartlett adjustment for the likelihood ratio statistic with an ordered alternative. Statis. and Probab. Lett., **20** 347-352.

DEPARTMENT OF STATISTICS
222 MATHEMATICAL SCIENCES BUILDING
UNIVERSITY OF MISSOURI-COLUMBIA
COLUMBIA, MISSOURI 65211, U. S. A.