EFFICIENT ESTIMATION OF A SHIFT PARAMETER FROM GROUPED DATA[1]

By P. K. Bhattacharya

University of Arizona

1. Summary. Universally efficient procedures for testing and estimation problems have been briefly explored by Hájek [3] and Stein [7]. In this paper we consider two populations having frequency functions $f(x)$ and $f(x - \theta)$ where the common form f and the shift parameter θ are unknown. A method of estimating θ when one sample is reduced to a frequency distribution over a given set of class-intervals is suggested by the likelihood principle and the asymptotic efficiency of this estimator relative to the appropriate maximum likelihood estimator based on the complete data is found to be the ratio of the Fisher-information in a grouped observation to the Fisher-information in an ungrouped observation.

2. Introduction. $X_1, \cdots, X_m$ are independent random variables with common distribution function F and $Y_1, \cdots, Y_n$ are independent random variables with common distribution function G. There exists a θ such that $G(x) = F(x - \theta)$ for all x, the function F and the location parameter θ being unknown. Consider another probability model for $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ which is equivalent to the one mentioned above; there exists a distribution function H and constants θ_1 and θ_2 such that $F(x) = H(x - \theta_1)$, $G(x) = H(x - \theta_2)$. The parameter θ in the former model is the same as $\theta_2 - \theta_1$ in the latter.

Suppose the distribution function H in the second model is everywhere differentiable and let $h(x) = H'(x)$. If h were known, the maximum likelihood estimate T^*_{m+n} of $\theta_2 - \theta_1$ based on $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ is $T^*_{2n} - T^*_{1m}$ where T^*_{1m} and T^*_{2n} are such that

$$\textstyle\sum_{i=1}^m \log h(X_i - T^*_{1m}) \geqq \sum_{i=1}^m \log h(X_i - t) \quad \text{for all } t$$

and

$$\textstyle\sum_{i=1}^n \log h(Y_i - T^*_{2n}) \geqq \sum_{i=1}^n \log h(Y_i - t) \quad \text{for all } t.$$

It is known (Wald [8]) that under very general conditions T^*_{m+n} converges in probability to $\theta_2 - \theta_1$. It is also known (Cramér [2]) that under certain regularity conditions on h, the likelihood equations

$$(1) \quad \textstyle\sum_{i=1}^m [\partial \log h(X_i - \theta_1)/\partial\theta_1] = 0 \quad \text{and} \quad \sum_{i=1}^n [\partial \log h(Y_i - \theta_2)/\partial\theta_2] = 0$$

have sequences of roots $\{\hat{T}_{1m}\}$ and $\{\hat{T}_{2n}\}$ respectively which have the property that both $m^{\frac{1}{2}}(\hat{T}_{1m} - \theta_1)$ and $n^{\frac{1}{2}}(\hat{T}_{2n} - \theta_2)$ are asymptotically normal with mean

Received 21 March 1966; revised 29 June 1967.

[1] Research supported in part by NSF Grant No. GP-6082.

0 and variance $1/I$ as m and n tend to infinity where I is the Fisher-information for the problem, viz.

$$I = \int_{-\infty}^{\infty} \{h'(x)\}^2/h(x)\,dx = \int_{-\infty}^{\infty} \{f'(x)\}^2/f(x)\,dx \tag{2}$$

where $f(x) = h(x - \theta_1) = F'(x)$. Let $m + n = N$, $m/N = \lambda_N$ and suppose m and n tend to infinity in such a manner that λ_N tends to some λ in the open interval $(0, 1)$. Then $N^{\frac{1}{2}}\{(\hat{T}_{2n} - \hat{T}_{1m}) - (\theta_2 - \theta_1)\}$ is asymptotically normal with mean 0 and variance $1/\lambda(1 - \lambda)I$.

The maximum likelihood estimate $\hat{T}_{m+n} = \hat{T}_{2n} - \hat{T}_{1m}$ of $\theta = \theta_2 - \theta_1$ can also be visualized in a different way. Let $g(x) = h(x - \theta_2) = G'(x)$. Now

$$\begin{aligned}\textstyle\sum_{i=1}^{m} \log h(X_i - \hat{T}_{1m}) &= \textstyle\sum_{i=1}^{m} \log h(X_i - \hat{T}_{2n} + \hat{T}_{m+n}) \\ &= \textstyle\sum_{i=1}^{m} \log g_n(X_i \mid \hat{T}_{m+n})\end{aligned}$$

where $g_n(x \mid t) = h(x - \hat{T}_{2n} + t)$ can be considered an estimate of $g(x + t) = h(x - \theta_2 + t)$ which is the density function of the random variables $Y_1 - t, \cdots, Y_n - t$. Thus in the computation of the maximum likelihood estimate of θ, the role of Y-sample is to estimate the function $g_n(\ \mid t)$ in the above manner for different values of t and role of the X-sample is to find out that value of t for which $\sum_{i=1}^{m} \log g_n(X_i \mid t)$ is a maximum. This way of looking at the maximum likelihood estimate $\hat{T}_{m+n}$ suggests that even when the form of the density function h is unknown it may be possible to estimate the density function of the random variables $Y_1 - t, \cdots, Y_n - t$ by some other function $\hat{g}_n(\ \mid t)$ which makes no use of the form of h and can be constructed only from $Y_1 - t, \cdots, Y_n - t$. With such an estimate $\hat{g}_n(\ \mid t)$ we can then proceed formally as in the method of maximum likelihood by choosing that value of t as the estimate of θ at which $\sum_{i=1}^{m} \log \hat{g}_n(X_i \mid t)$ is a maximum.

We shall now proceed with an approximation to the procedure outlined above.

In what follows, we shall avoid unnecessary complications by writing $m = N\lambda$ and $n = N(1 - \lambda)$ instead of $m = N\lambda_N$ and $n = N(1 - \lambda_N)$ and carry out the analysis as if $N\lambda$ is an integer. It can be easily verified that since $\lambda_N \to \lambda$ as $N \to \infty$, this simplification will not affect our analysis in any way.

Choose k real numbers $a_1 < \cdots < a_k$ and let $\chi_1, \cdots, \chi_{k+1}$ denote the indicator functions of the intervals $C_1 = (-\infty, a_1]$, $C_j = (a_{j-1}, a_j]$, $j = 2, \cdots, k$, and $C_{k+1} = (a_k, \infty)$ respectively. Let

$$\nu_{jN} = \textstyle\sum_{i=1}^{N\lambda} \chi_j(X_i), \quad p_{jN} = \nu_{jN}/N\lambda, \quad q_{jN}(t) = \textstyle\sum_{i=1}^{N(1-\lambda)} \chi_j(Y_i - t)/N(1 - \lambda), \qquad j = 1, \cdots, k + 1.$$

For any given t, when we have reduced the data $X_1, \cdots, X_{N\lambda}, Y_1, \cdots, Y_{N(1-\lambda)}$ to $\chi_j(X_1), \cdots, \chi_j(X_{N\lambda}), \chi_j(Y_1 - t), \cdots, \chi_j(Y_{N(1-\lambda)} - t)$, $j = 1, \cdots, k + 1$, the analog of the frequency function $g(\ \mid t)$ of $Y_i - t$ becomes the multinomial probability distribution

$$q_j(t) = P[Y_i - t \in C_j], \qquad j = 1, \cdots, k + 1,$$

and the log likelihood of $\chi_j(X_i)$, $i = 1, \cdots, N\lambda$, $j = 1, \cdots, k+1$, assuming that they have come from a multinomial population with probabilities $q_j(t)$, $j = 1, \cdots, k+1$, is $\sum_{j=1}^{k+1} \nu_{jN} \log q_j(t)$. In this log likelihood, if each $q_j(t)$ is replaced by its estimate $q_{jN}(t)$, we get

$$Z_N(t) = \sum_{j=1}^{k+1} \nu_{jN} \log q_{jN}(t) = N\lambda \sum_{j=1}^{k+1} p_{jN} \log q_{jN}(t)$$

as the discrete analog of $\sum_{i=1}^{N\lambda} \log \hat{g}_N(X_i \mid t)$.

This approximation forces us to make another approximation. Since $Z_N(t)$ is a step-function for any given sample, the usual technique for deriving the asymptotic distribution of the value of t at which Z_N is maximized, fails. For this reason, we choose and fix $\delta > 0$ and a real number t_0 and fix our attention to the set $\Theta = \{t_0 + r\delta/N^{\frac{1}{2}}: r = 0, \pm 1, \pm 2, \cdots\}$ in search of an estimate of θ.

We now define a maximum empirical likelihood estimate of θ as a point $t_0 + r\delta/N^{\frac{1}{2}}$ such that

$$(3) \quad Z_N(t_0 + r\delta/N^{\frac{1}{2}}) > \max\{Z_N(t_0 + (r-1)\delta/N^{\frac{1}{2}}), Z_N(t_0 + (r+1)\delta/N^{\frac{1}{2}})\}.$$

In what follows, we shall refer to (3) as the empirical likelihood inequality. As there may be more than one solution for the likelihood equations (1), the empirical likelihood inequality may be satisfied for more than one point in Θ. It will be shown however (Theorem 1) that for arbitrary K, the probability that the empirical likelihood inequality is satisfied for more than one point in Θ lying in the interval $(\theta - K/N^{\frac{1}{2}}, \theta + K/N^{\frac{1}{2}})$, tends to zero as $N \to \infty$. We also note that for any given sample size N there may be sample points for which the empirical likelihood inequality has no solution. In such a case let us make the convention of treating t_0 as "a solution of the empirical likelihood inequality." With this convention we can now talk about a sequence of solutions of the empirical likelihood inequality. It will be shown (Theorem 2) that the empirical likelihood inequality has a sequence of solutions $\{T_N\}$ for which

$$\begin{aligned} \lim_{N\to\infty} P_\theta[T_N \leqq t_0 + r\delta/N^{\frac{1}{2}}] &= \lim_{N\to\infty} P_\theta[N^{\frac{1}{2}}(T_N - \theta) \leqq c + s\delta] \\ &= \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1-\lambda)I_k)^{\frac{1}{2}}) \end{aligned}$$

where Φ is the standard normal distribution function,

$$(4) \quad \begin{aligned} I_k = I_k(a_1, \cdots, a_k) &= \{f(a_1)\}^2/F(a_1) \\ &+ \textstyle\sum_{j=2}^{k} \{f(a_j) - f(a_{j-1})\}^2/\{F(a_j) - F(a_{j-1})\} \\ &+ \{f(a_k)\}^2/\{1 - F(a_k)\}, \end{aligned}$$

$c = N^{\frac{1}{2}}(t_0 - \theta) + r_0\delta$, $s = r - r_0$, r_0 being such that $t_0 + (r_0 - 1)\delta/N^{\frac{1}{2}} < \theta \leqq t_0 + r_0\delta/N^{\frac{1}{2}}$ (obviously $0 \leqq c < \delta$). In other words, the empirical likelihood inequality has a sequence of solutions $\{T_N\}$ which has the property that $N^{\frac{1}{2}}(T_N - \theta)$ is asymptotically roughly normal with mean 0 and variance $1/\lambda(1-\lambda)I_k$.

A comparison of I_k given by (4) with I given by (2) shows that though $I_k < I$,

$I_k \to I$ as $k \to \infty$, $a_1 \to -\infty$, $a_k \to \infty$ and $a_j - a_{j-1} \to 0$, $j = 2, \cdots, k$, under mild conditions (see Section 4 (ii)). Thus for all f, the maximum empirical likelihood estimate of θ behaves asymptotically more and more like the maximum likelihood estimate of θ corresponding to f as the class-intervals are made finer and finer.

3. Asymptotic properties of solutions of the empirical likelihood inequality. We shall assume throughout our analysis that F has positive probabilities over each of the intervals $C_1, \cdots, C_{k+1}$ and is continuously twice differentiable at $a_1, \cdots, a_{k+1}$.

The following notation will be used:

$$\alpha_j = P[X_1 \varepsilon C_j] = \int_{-\infty}^{\infty} \chi_j(x)\, dF(x), \qquad j = 1, \cdots, k+1,$$

$$\beta_j(t) = P_\theta[Y_1 - t \varepsilon C_j] = \int_{-\infty}^{\infty} \chi_j(x - t + \theta)\, dF(x), \qquad j = 1, \cdots, k+1,$$

$$\alpha_1' = F'(a_1),\ \alpha_j' = F'(a_j) - F'(a_{j-1}),\ j = 2, \cdots, k, \qquad \alpha_{k+1}' = -F'(a_k),$$

$$\alpha_1'' = F''(a_1),\ \alpha_j'' = F''(a_j) - F''(a_{j-1}),\ j = 2, \cdots, k,\ \alpha_{k+1}'' = -F''(a_k).$$

The $(k + 1)$ -dimensional column vectors whose jth coordinates are $\chi_j(X_i)$, $\chi_j(Y_i - t)$, p_{jN}, $q_{jN}(t)$, α_j, $\beta_j(t)$, α_j', α_j'', $\log (\beta_j(t)/\alpha_j) - 1$, $\alpha_j/\beta_j(t)$ and 1 will be denoted by ξ_i, $\eta_i(t)$, p_N, $q_N(t)$, α, $\beta(t)$, γ_1, γ_2, $\phi(t)$, $\psi(t)$ and e respectively. The corresponding row vectors will be denoted by the same symbols with primes. Obviously, $p_N = \sum_{i=1}^{N\lambda} \xi_i/N\lambda$ and $q_N(t) = \sum_{i=1}^{N(1-\lambda)} \eta_i(t)/N(1 - \lambda)$.

For every integer s, $(c + s\delta)/N^{\frac{1}{2}}$ will be denoted by $\delta_N(s)$.

A, $B(t)$, G_1 and G_2 are $(k + 1) \times (k + 1)$ diagonal matrices whose jth diagonal elements are α_j, $\beta_j(t)$, α_j' and α_j'' respectively. For integers $s_1 < s_2$, $D_{s_1 s_2}$ is the $k \times k$ diagonal matrix whose jth diagonal element is $F(a_j + \delta_N(s_2)) - F(a_j + \delta_N(s_1))$. E_1 and E_2 are $k \times k$ diagonal matrices whose jth diagonal elements are $F'(a_j)$ and $F''(a_j)$ respectively. $J_{s_1 s_2}$, $\tilde{J}_{s_1 s_2}$, H_1, $\tilde{H}_1$, H_2 and $\tilde{H}_2$ are $(k + 1) \times (k + 1)$ matrices defined as follows.

$$J_{s_1 s_2} = \underset{1 \quad\ \ k}{\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & D_{s_1 s_2} \end{array}\right)}\begin{array}{l} 1 \\ k \end{array}, \qquad \tilde{J}_{s_1 s_2} = \underset{k \quad\ \ 1}{\left(\begin{array}{c|c} 0 & 0 \\ \hline D_{s_1 s_2} & 0 \end{array}\right)}\begin{array}{l} 1 \\ k \end{array},$$

$$H_1 = \underset{1 \quad k}{\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & E_1 \end{array}\right)}\begin{array}{l} 1 \\ k \end{array}, \qquad \tilde{H}_1 = \underset{k \quad 1}{\left(\begin{array}{c|c} 0 & 0 \\ \hline E_1 & 0 \end{array}\right)}\begin{array}{l} 1 \\ k \end{array},$$

$$H_2 = \underset{1 \quad k}{\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & E_2 \end{array}\right)}\begin{array}{l} 1 \\ k \end{array}, \qquad \tilde{H}_2 = \underset{k \quad 1}{\left(\begin{array}{c|c} 0 & 0 \\ \hline E_2 & 0 \end{array}\right)}\begin{array}{l} 1 \\ k \end{array}.$$

It can be easily verified that the following relations hold among the vectors and matrices defined above.

LEMMA 1

(a) $e'\alpha = 1$, $Ae = \alpha$, $(A - \alpha\alpha')e = 0$ (*null vector*).

(b) $e'\gamma_1 = e'\gamma_2 = 0$.

(c) $G_1 e = \gamma_1$, $G_2 e = \gamma_2$.

(d) $\gamma_1' A^{-1}(A - \alpha\alpha')A^{-1}\gamma_1 = \gamma_1' A^{-1}\gamma_1 = I_k$.

(e) $(\tilde{H}_1 - H_1)e = (\tilde{H}_2 - H_2)e = 0$ (*null vector*).

(f) $e'(\tilde{H}_1 - H_1) = \gamma_1'$.

Let

$$U_{iN}(s) = \phi(\theta + \delta_N(s))'\xi_i, \qquad i = 1, \cdots, N\lambda,$$

$$V_{iN}(s) = \psi(\theta + \delta_N(s))'\eta_i(\theta + \delta_N(s)), \qquad i = 1, \cdots, N(1 - \lambda),$$

$$U_N(s) = \textstyle\sum_{i=1}^{N\lambda} U_{iN}(s)/\lambda = N\phi(\theta + \delta_N(s))'p_N,$$

$$V_N(s) = \textstyle\sum_{i=1}^{N(1-\lambda)} V_{iN}(s)/(1 - \lambda) = N\psi(\theta + \delta_N(s))'q_N(\theta + \delta_N(s)),$$

$$W_{1N}(s) = \{U_N(s) - U_N(s - 1)\} + \{V_N(s) - V_N(s - 1)\},$$

and $W_{2N}(s) = \{U_N(s) - U_N(s + 1)\} + \{V_N(s) - V_N(s + 1)\}$.

We shall first compute

$\lim_{N\to\infty} P_\theta[W_{1N}(s) \leqq 0, W_{2N}(s) \leqq 0]$ and $\lim_{N\to\infty} P_\theta[W_{1N}(s) > 0, W_{2N}(s) > 0]$ for a fixed integer s and will then show that the inequalities $W_{1N}(s) > 0$ and $W_{2N}(s) > 0$ together are asymptotically equivalent to the empirical likelihood inequality.

Let η_{is}, $i = 1, \cdots, N(1 - \lambda)$, be a $3(k + 1)$-dimensional random variable defined as

$$\eta_{is} = \begin{pmatrix} \eta_i(\theta + \delta_N(s - 1)) \\ \eta_i(\theta + \delta_N(s)) \\ \eta_i(\theta + \delta_N(s + 1)) \end{pmatrix},$$

and β_s a $3(k + 1)$-dimensional vector defined as

$$\beta_s = \begin{pmatrix} \beta(\theta + \delta_N(s - 1)) \\ \beta(\theta + \delta_N(s)) \\ \beta(\theta + \delta_N(s + 1)) \end{pmatrix}.$$

In what follows, all probabilities (P), expectations (E), variances (Var) and covariances (Cov) are computed when θ obtains.

LEMMA 2. $\xi_1, \cdots, \xi_{N\lambda}, \eta_{1s}, \cdots, \eta_{N(1-\lambda),s}$ *are independent random variables with*

(*a*) $E(\xi_i) = \alpha$, Cov $(\xi_i) = A - \alpha\alpha'$, $i = 1, \cdots, N\lambda$,

(b) $E(\eta_{is}) = \beta_s$,

$$\text{Cov}(\eta_{is}) = \begin{pmatrix} \Sigma_{s-1,s-1} & \Sigma_{s-1,s} & \Sigma_{s-1,s+1} \\ \Sigma_{s,s-1} & \Sigma_{ss} & \Sigma_{s,s+1} \\ \Sigma_{s+1,s-1} & \Sigma_{s+1,s} & \Sigma_{s+1,s+1} \end{pmatrix}, \qquad i = 1, \cdots, N(1 - \lambda),$$

where $\Sigma_{s_1 s_2} = B(\theta + \delta_N(s_1)) - \beta(\theta + \delta_N(s_1))\beta(\theta + \delta_N(s_1))'$ *if* $s_1 = s_2$

$= B(\theta + \delta_N(s_1)) - \beta(\theta + \delta_N(s_1))$

$\cdot\beta(\theta + \delta_N(s_2))' - J_{s_1 s_2} + \hat{J}_{s_1 s_2}$ *if* $s_1 < s_2$

for large N, *and* $\Sigma_{s_2 s_1} = \Sigma'_{s_1 s_2}$.

PROOF. (a) and the expectation part of (b) can be easily verified and we shall omit this part of the proof.

$$E[\chi_j(Y_i - \theta - \delta_N(s))\chi_{j'}(Y_i - \theta - \delta_N(s))]$$
$$= P[Y_i - \theta - \delta_N(s) \varepsilon C_j, Y_i - \theta - \delta_N(s) \varepsilon C_{j'}]$$
$$= \beta_j(\theta + \delta_N(s)) \quad \text{if } j = j',$$
$$= 0 \quad \text{if } j \neq j'.$$

Hence $\Sigma_{ss} = B(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s))\beta(\theta + \delta_N(s))'$.

Let $s_1 < s_2$.

$$E[\chi_j(Y_i - \theta - \delta_N(s_1))\cdot\chi_j(Y_i - \theta - \delta_N(s_2))]$$
$$= P[Y_i - \theta - \delta_N(s_1) \varepsilon C_j, Y_i - \theta - \delta_N(s_2) \varepsilon C_j]$$
$$= P[Y_i - \theta - \delta_N(s_1) \varepsilon C_j] \quad \text{for } j = 1$$
$$= P[Y_i - \theta - \delta_N(s_1) \varepsilon C_j] - P[a_{j-1} + \theta + \delta_N(s_1)$$
$$< Y_i \leqq a_{j-1} + \theta + \delta_N(s_2)] \quad \text{for } j = 2, \cdots, k+1,$$
$$= \beta_j(\theta + \delta_N(s_1)) \quad \text{for } j = 1$$
$$= \beta_j(\theta + \delta_N(s_1)) - \{F(a_{j-1} + \delta_N(s_2))$$
$$- F(a_{j-1} + \delta_N(s_1))\} \quad \text{for } j = 2, \cdots, k+1.$$

Now suppose N is so large that

$$\min_{j=2,\cdots,k} (a_j - a_{j-1}) > (s_2 - s_1)\delta/N^{\frac{1}{2}} = \delta_N(s_2) - \delta_N(s_1). \tag{5}$$

For $j \neq j'$,

$$E[\chi_j(Y_i - \theta - \delta_N(s_1))\cdot\chi_{j'}(Y_i - \theta - \delta_N(s_2))]$$
$$= P(Y_i - \theta - \delta_N(s_1) \varepsilon C_j, Y_i - \theta - \delta_N(s_2) \varepsilon C_{j'}]$$
$$= 0 \quad \text{for } j' \neq j - 1$$
$$= P[a_j + \theta + \delta_N(s_1) < Y_i \leqq a_j + \theta + \delta_N(s_2)] \quad \text{for } j' = j - 1$$

because (5) implies that

$$a_{j-1} + \theta + \delta_N(s_2) < a_j + \theta + \delta_N(s_1), \qquad j = 2, \cdots, k.$$

Thus,

$$\text{Cov}[\chi_j(Y_i - \theta - \delta_N(s_1)), \chi_j(Y_i - \theta - \delta_N(s_2))]$$
$$= \beta_j(\theta + \delta_N(s_1)) - \beta_j(\theta + \delta_N(s_1))\beta_j(\theta + \delta_N(s_2)) \quad \text{for } j = 1$$

$$= \beta_j(\theta + \delta_N(s_1)) - \beta_j(\theta + \delta_N(s_1))\beta_j(\theta + \delta_N(s_2)) - \{F(a_{j-1} + \delta_N(s_2)) - F(a_{j-1} + \delta_N(s_1))\} \quad \text{for } j = 2, \cdots, k+1.$$

These can be recognized as the diagonal elements of

$$B(\theta + \delta_N(s_1)) - \beta(\theta + \delta_N(s_1))\beta(\theta + \delta_N(s_2))' - J_{s_1s_2} + \tilde{J}_{s_1s_2} \tag{6}$$

as soon as we note that the first diagonal element of $J_{s_1s_2}$ and all the diagonal elements of $\tilde{J}_{s_1s_2}$ are zeros while for $j = 2, \cdots, k+1$, the jth diagonal element of $J_{s_1s_2}$ is $F(a_{j-1} + \delta_N(s_2)) - F(a_{j-1} + \delta_N(s_1))$.

Again, for $j = 2, \cdots, k+1$,

$$\text{Cov}\,[\chi_j(Y_i - \theta - \delta_N(s_1)), \chi_{j-1}(Y_i - \theta - \delta_N(s_2))]$$
$$= -\beta_j(\theta + \delta_N(s_1))\beta_{j-1}(\theta + \delta_N(s_2)) + \{F(a_j + \delta_N(s_2)) - F(a_j + \delta_N(s_1))\}.$$

These can be recognized as the immediate sub-diagonal elements of the matrix given in (6) as soon as we note that all the off-diagonal elements of $B(t)$ and $J_{s_1s_2}$ are zeros while the immediate sub-diagonal elements of $\tilde{J}_{s_1s_2}$ are $F(a_j + \delta_N(s_2)) - F(a_j + \delta_N(s_1))$, $j = 1, \cdots, k$.

Finally, for all other (j, j')

$$\text{Cov}\,[\chi_j(Y_i - \theta - \delta_N(s_1)), \chi_{j'}(Y_i - \theta - \delta_N(s_2))] = -\beta_j(\theta + \delta_N(s_1))\beta_{j'}(\theta + \delta_N(s_2)),$$

which can be recognized as the (j, j')th elements of the matrix given in (6) other than those on the diagonal or immediately below the diagonal because all such elements of $B(t)$, $J_{s_1s_2}$ and $\tilde{J}_{s_1s_2}$ are zeros. This completes the proof.

The asymptotic behavior of $W_{1N}(s)$ and $W_{2N}(s)$ depends on the values of ϕ and ψ at $\theta + \delta_N(r)$, $r = s-1$, s and $s+1$ and the expectations and covariances of the random vectors ξ_i and η_{is} given in Lemma 2. For this reason, we shall need the asymptotic expansions of the values of β, B, ϕ and ψ at $\theta + \delta_N(r)$, $r = s-1$, s and $s+1$ and of $J_{s_1s_2}$ and $\tilde{J}_{s_1s_2}$ for $s_1 < s_2$ taking values $s-1$, s and $s+1$ in powers of $N^{-\frac{1}{2}}$ as $N \to \infty$. These asymptotic expansions are given in the following lemma.

LEMMA 3. *For any fixed s, as* $N \to \infty$,

(a) $\beta(\theta + \delta_N(s)) = \alpha + \delta_N(s)\gamma_1 + \frac{1}{2}\delta_N^2(s)\gamma_2 + o(N^{-1})$,

(b) $B(\theta + \delta_N(s)) = A + \delta_N(s)G_1 + \frac{1}{2}\delta_N^2(s)G_2 + o(N^{-1})$,

(c) $\phi(\theta + \delta_N(s)) = -e + \delta_N(s)A^{-1}\gamma_1 + \frac{1}{2}\delta_N^2(s)(A^{-1}\gamma_2 - A^{-2}G_1\gamma_1) + o(N^{-1})$

(d) $\psi(\theta + \delta_N(s)) = e - \delta_N(s)A^{-1}\gamma_1 - \frac{1}{2}\delta_N^2(s)(A^{-1}\gamma_2 - 2A^{-2}G_1\gamma_1) + o(N^{-1})$,

(e) $J_{s_1s_2} = (s_2 - s_1)\delta N^{-\frac{1}{2}}H_1 + (s_2 - s_1)\delta\{c + \frac{1}{2}(s_1 + s_2)\delta\}N^{-1}H_2 + o(N^{-1})$ *for* $s_1 < s_2$,

(f) $\tilde{J}_{s_1s_2} = (s_2 - s_1)\delta N^{-\frac{1}{2}}\tilde{H}_1 + (s_2 - s_1)\delta\{c + \frac{1}{2}(s_1 + s_2)\delta\}N^{-1}\tilde{H}_2 + o(N^{-1})$ *for* $s_1 < s_2$.

PROOF. $\beta_j(\theta + \delta_N(s)) = \beta_j(\theta) + \delta_N(s)\beta_j'(\theta) + \frac{1}{2}\delta_N^2(s)\beta_j''(\theta_j)$, where θ_j lies between θ and $\theta + \delta_N(s)$. Since $\beta_j(\theta) = \alpha_j$, $\beta_j'(\theta) = \alpha_j'$, $\beta_1''(\theta_1) = F''(a_1')$, $\beta_j''(\theta_j) = F''(a_j') - F''(a_{j-1}')$, $j = 2, \cdots, k$, $\beta_{k+1}''(\theta_{k+1}) = -F''(a_k')$ where a_j' lies between a_j and $a_j + \delta_N(s)$, it follows from the continuity of F'' at $a_1, \cdots, a_k$ that as $N \to \infty$ (i.e., as $\delta_N(s) \to 0$),

$$\beta_j(\theta + \delta_N(s)) = \alpha_j + \delta_N(s)\alpha_j' + \tfrac{1}{2}\delta_N^2(s)\alpha_j'' + o(N^{-1}). \tag{7}$$

(a), (b), (c), and (d) follows from (7). Again,

$$F(a_j + \delta_N(s_2)) - F(a_j + \delta_N(s_1))$$
$$= \{\delta_N(s_2) - \delta_N(s_1)\}\, F'(a_j) + \tfrac{1}{2}\delta_N^2(s_2)F''(a_j') - \tfrac{1}{2}\delta_N^2(s_1)F''(a_j'')$$

where a_j' lies between a_j and $a_j + \delta_N(s_2)$ and a_j'' lies between a_j and $a_j + \delta_N(s_1)$. It now follows from the continuity of F'' at $a_1, \cdots, a_k$ that as $N \to \infty$,

$$\begin{aligned} &F(a_j + \delta_N(s_2)) - F(a_j + \delta_N(s_1)) \\ &= \{\delta_N(s_2) - \delta_N(s_1)\}F'(a_j) + \tfrac{1}{2}\{\delta_N^2(s_2) - \delta_N^2(s_1)\}F''(a_j) + o(N^{-1}) \\ &= (s_2 - s_1)\delta N^{-\frac{1}{2}}F'(a_j) + (s_2 - s_1)\delta\{c + \tfrac{1}{2}(s_1 + s_2)\delta\}N^{-1}F''(a_j) + o(N^{-1}). \end{aligned} \tag{8}$$

(e) and (f) follows from (8).

$W_{1N}(s)$ is obtained from sums of independent random variables $U_{iN}(s) - U_{iN}(s-1)$, $i = 1, \cdots, N\lambda$, $V_{iN}(s) - V_{iN}(s-1)$, $i = 1, \cdots, N(1-\lambda)$, and $W_{2N}(s)$ is obtained from sums of independent random variables $U_{iN}(s) - U_{iN}(s+1)$, $i = 1, \cdots, N\lambda$, $V_{iN}(s) - V_{iN}(s+1)$, $i = 1, \cdots, N(1-\lambda)$. Since we are now going to study the asymptotic properties of $W_{1N}(s)$ and $W_{1N}(s) + W_{2N}(s)$, we shall need some moments of the random variables mentioned above.

LEMMA 4. *For fixed s, as* $N \to \infty$,

(a) $\mu_{iN}(s) = E[U_{iN}(s) - U_{iN}(s+1)] = \delta\{c + (s + \frac{1}{2})\delta\}I_k/N + o(N^{-1})$,

$\mu_{iN}'(s) = E[V_{iN}(s) - V_{iN}(s+1)] = 0$;

(b) $\sigma_{iN}^2(s) = \operatorname{Var}[U_{iN}(s) - U_{iN}(s+1)] = \delta^2 I_k/N + o(N^{-1})$

$\sigma_{iN}'^2(s) = \operatorname{Var}[V_{iN}(s) - V_{iN}(s+1)] = \delta^2 I_k/N + o(N^{-1})$;

(c) $\tau_{iN}(s) = \operatorname{Cov}[U_{iN}(s) - U_{iN}(s-1), U_{iN}(s) - U_{iN}(s+1)]$

$= -\delta^2 I_k/N + o(N^{-1})$,

$\tau_{iN}'(s) = \operatorname{Cov}[V_{iN}(s) - V_{iN}(s-1), V_{iN}(s) - V_{iN}(s+1)]$

$= -\delta^2 I_k/N + o(N^{-1})$;

(d) $\rho_{iN}^3(s) = E\,|U_{iN}(s) - U_{iN}(s+1) - \mu_{iN}(s)|^3 \leqq K/N^{\frac{3}{2}}$,

$\rho_{iN}'^3(s) = E\,|V_{iN}(s) - V_{iN}(s+1)|^3 \leqq K'(s)/N^{\frac{3}{2}}$;

where $K = 2\delta^3(\gamma_1'A^{-2}\gamma_1)^{\frac{3}{2}}$ *and* $K'(s) = 2(|s| + 1)\delta^3(\gamma_1'A^{-2}\gamma_1)^{\frac{3}{2}}$.

PROOF. (a)

$$\mu_{iN}(s) = \{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1))\}'\alpha,$$

and

$$\mu'_{iN}(s) = \psi(\theta + \delta_N(s))'\beta(\theta + \delta_N(s)) - \psi(\theta + \delta_N(s + 1))'\beta(\theta + \delta_N(s + 1)) = 0$$

since for all t, $\psi(t)'\beta(t) = \sum_1^{k+1} \alpha_j = 1$. To complete the proof of this part, we expand $\phi(\theta + \delta_N(s))$ and $\phi(\theta + \delta_N(s + 1))$ as in Lemma 3 and simplify by using Lemma 1.

(b) and (c).

$$\sigma^2_{iN}(s) = \{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1))\}'(A - \alpha\alpha') \cdot\{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1))\}$$

and

$$\tau_{iN}(s) = \{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s - 1))\}'(A - \alpha\alpha') \cdot\{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1))\},$$

and each of these is brought to its desired form by an application of Lemma 3 followed by simplification with the use of Lemma 1. To prove the other parts, we write

$$\sigma'^2_{iN}(s) = \psi(\theta + \delta_N(s))'\Sigma_{ss}\psi(\theta + \delta_N(s)) + \psi(\theta + \delta_N(s + 1))' \cdot\Sigma_{s+1,s+1}\psi(\theta + \delta_N(s + 1) - 2\psi(\theta + \delta_N(s))'\Sigma_{s,s+1}\psi(\theta + \delta_N(s + 1)),$$

$$\tau'_{iN}(s) = \psi(\theta + \delta_N(s))'\Sigma_{ss}\psi(\theta + \delta_N(s)) - \psi(\theta + \delta_N(s))'\Sigma_{s,s+1}\psi(\theta + \delta_N(s + 1)) - \psi(\theta + \delta_N(s - 1))'\Sigma_{s-1,s}\psi(\theta + \delta_N(s)) + \psi(\theta + \delta_N(s - 1))' \cdot\Sigma_{s-1,s+1}\psi(\theta + \delta_N(s + 1)).$$

Substituting the expressions for the Σ-matrices given in Lemma 2, using the expansions of Lemma 3 and simplifying by repeated use of Lemma 1, we get

$$\psi(\theta + \delta_N(s_1))'\Sigma_{s_1s_2}\psi(\theta + \delta_N(s_2)) = \delta_N(s_1)\delta_N(s_2)I_k + o(N^{-1}).$$

$\sigma'^2_{iN}(s)$ and $\tau'_{iN}(s)$ are now easily brought to their desired forms.

(d) $$\rho^3_{iN}(s) \leqq E[|U_{iN}(s) - U_{iN}(s + 1)| + |\mu_{iN}(s)|]^3.$$

Now $U_{iN}(s) - U_{iN}(s + 1) = \{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1)\}'\xi_i$, and since ξ_i is a $(k + 1)$-dimensional random vector whose one coordinate is 1 and the other k coordinates are zeros,

$$|U_{iN}(s) - U_{iN}(s + 1)| \leqq [\{\phi(\theta + \delta_N(s)) - \phi(\theta + \delta_N(s + 1))\}'\{\phi(\theta + \delta_N(s) - \phi(\theta + \delta_N(s + 1)\}]^{\frac{1}{2}} = \delta(\gamma_1'A^{-2}\gamma_1/N)^{\frac{1}{2}} + o(N^{-\frac{1}{2}})$$

with probability 1. The desired inequality for $\rho_{iN}^3(s)$ now follows from the fact that $\mu_{iN}(s) = O(N^{-1})$. Again, $\eta_i(t)$ is bounded for all t, we have from Lemma 3,

$$\begin{aligned} V_{iN}(s) - V_{iN}(s+1) &= e'\{\eta_i(\theta + \delta_N(s)) - \eta_i(\theta + \delta_N(s+1))\} \\ &\quad + \delta_N(s+1)\gamma_1'A^{-1}\eta_i(\theta + \delta_N(s+1)) \\ &\quad - \delta_N(s)\gamma_1'A^{-1}\eta_i(\theta + \delta_N(s)) + o(N^{-\frac{1}{2}}). \end{aligned} \tag{9}$$

For each t, the vector $\eta_i(t)$ has 1 for one coordinate and 0 for the other coordinates. Hence $e'\eta_i(t) = 1$ for all t. Thus the first term on the right-hand side of (9) is 0. One can also verify that for each t,

$$|\gamma_1'A^{-1}\eta_i(t)| \leqq (\gamma_1'A^{-2}\gamma_1)^{\frac{1}{2}} \tag{10}$$

with probability 1. From (9) and (10) and because $\theta \leqq c < \delta$, the desired inequality for $\rho_{iN}'^3(s)$ follows.

The results of this lemma will now be used to study the convergence of $W_{1N}(s)$ in law and the convergence of $W_{1N}(s) + W_{2N}(s)$ in probability.

LEMMA 5. *For fixed s, $W_{1N}(s)$ is asymptotically normally distributed with mean $-\delta\{c + (s - \frac{1}{2})\delta\}I_k$ and variance $\delta^2 I_k/\lambda(1 - \lambda)$.*

PROOF. $W_{1N}(s)$ is the sum of independent random variables

$$\lambda^{-1}\{U_{iN}(s) - U_{iN}(s-1)\}, \quad i = 1, \cdots, N\lambda,$$
$$\text{and} \quad (1-\lambda)^{-1}\{V_{iN}(s) - V_{iN}(s-1)\}, \quad i = 1, \cdots, N(1-\lambda).$$

From Lemma 4, the sum of the expectations of these random variables is

$$\mu_N(s) = -\delta\{c + (s - \tfrac{1}{2})\delta\}I_k + o(1), \tag{11}$$

the sum of the variances of these random variables is

$$\sigma_N^2(s) = \delta^2 I_k/\lambda(1-\lambda) + o(1), \tag{12}$$

and the sum of the third absolute moments of these random variables is

$$\rho_N^3(s) \leqq N^{-\frac{1}{2}}[K/\lambda^2 + K'(s-1)/(1-\lambda)^2].$$

To show that the central limit theorem applies to $W_{1N}(s)$, we shall now verify Liapounoff's condition (see Cramér [2]). This verification is immediate as soon as we note that

$$\rho_N^3(s)/\sigma_N^3(s) \leqq 2N^{-\frac{1}{2}}\cdot[(1-\lambda)^2 K + \lambda^2 K'(s-1)]/\delta^3 I_k^{\frac{3}{2}}(\lambda(1-\lambda))^{\frac{1}{2}}$$

for sufficiently large N. Moreover, it follows from (11) and (12) that

$$\Phi((w - \mu_N(s))/\sigma_N(s)) = \phi((w + \delta\{c + (s - \tfrac{1}{2})\delta\}I_k)/\delta\cdot(I_k/\lambda(1-\lambda))^{\frac{1}{2}}) + o(1)$$

as $N \to \infty$, and that completes the proof.

LEMMA 6. *For fixed s, $W_{1N}(s) + W_{2N}(s)$ converges in probability to $\delta^2 I_k$ as $N \to \infty$.*

PROOF. We have from Lemma 4,

$$E[W_{1N}(s) + W_{2N}(s)] = -\delta\{c + (s - \tfrac{1}{2})\delta\}I_k + \delta\{c + (s + \tfrac{1}{2})\delta\}I_k + o(1)$$
$$= \delta^2 I_k + o(1),$$
$$\text{Var}\,[W_{1N}(s)] = \delta^2 I_k/\lambda(1 - \lambda) + o(1),$$
$$\text{Var}\,[W_{2N}(s)] = \delta^2 I_k/\lambda(1 - \lambda) + o(1),$$

and

$$\text{Cov}\,[W_{1N}(s), W_{2N}(s)] = -\delta^2 I_k/\lambda(1 - \lambda) + o(1),$$

as $N \to \infty$. It now follows from the Tchebychev-inequality that

$$W_{1N}(s) + W_{2N}(s) - E[W_{1N}(s) + W_{2N}(s)]$$

converges in probability to 0 as $N \to \infty$. But $E[W_{1N}(s) + W_{2N}(s)]$ converges to $\delta^2 I_k$ as $N \to \infty$, and that completes the proof.

Lemmas 5 and 6 suggest that for the purpose of computing the limiting probabilities for the events $\{W_{1N}(s) > a_1, W_{2N}(s) > a_2\}$ and $\{W_{1N}(s) \leqq a_1, W_{2N}(s) \leqq a_2\}$, we may replace $W_{2N}(s)$ by $\delta^2 I_k - W_{1N}(s)$ and then obtain the above limiting probabilities from the asymptotic distribution of $W_{1N}(s)$. The following lemma tells us that this procedure is in order.

LEMMA 7. *$\{X_n\}$ and $\{Y_n\}$ are sequences of random variables such that the distribution function of X_n converges to a distribution function F at all points of continuity of F and $X_n + Y_n$ converges in probability to a constant μ. If $a + b < \mu$ and if a and $\mu - b$ are points of continuity of F, then*

$$\lim_{n \to \infty} P[X_n \leqq a, Y_n \leqq b] = 0$$

and

$$\lim_{n \to \infty} P[X_n > a, Y_n > b] = F(\mu - b) - F(a).$$

PROOF. Choose arbitrary $\epsilon > 0$. Since

$$\lim_{n \to \infty} P[|X_n + Y_n - \mu| > \epsilon] = 0,$$

it follows that

$$\lim_{n \to \infty} P[X_n \leqq a, Y_n \leqq b] \leqq \lim_{n \to \infty} P[X_n \leqq a, X_n \geqq \mu - b - \epsilon].$$

The right-hand side of the above inequality becomes 0 if we choose $\epsilon < \mu - a - b$. To prove the second part we notice that

$$\lim_{n \to \infty} P[a < X_n < \mu - b - \epsilon] \leqq \lim_{n \to \infty} P[X_n > a, Y_n > b]$$
$$\leqq \lim_{n \to \infty} P[a < X_n < \mu - b + \epsilon]$$

for arbitrary choice of $\epsilon > 0$ and as $\epsilon \to 0$, both the extremes of the above inequality tend to $F(\mu - b) - F(a)$.

From Lemmas 5, 6, and 7 we now conclude,

LEMMA 8. *For fixed s and for $a_1 + a_2 < \delta^2 I_k$,*

$$\lim_{N \to \infty} P[W_{1N}(s) \leqq a_1, W_{2N}(s) \leqq a_2] = 0$$

and

$$\lim_{N\to\infty} P[W_{1N}(s) > a_1, W_{2N}(s) > a_2]$$
$$= \Phi([\delta^2 I_k - a_2 + \delta\{c + (s - \tfrac{1}{2})\delta\}I_k]/\delta\cdot(I_k/\lambda(1 - \lambda))^{\frac{1}{2}})$$
$$- \Phi([a_1 + \delta\{c + (s - \tfrac{1}{2})\delta\}I_k]/\delta\cdot(I_k/\lambda(1 - \lambda))^{\frac{1}{2}}).$$

We shall now find the stochastic orders of

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s - 1))\} - W_{1N}(s)$$

and

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s + 1))\} - W_{2N}(s)$$

as $N \to \infty$. For definitions of stochastic order relations o_p and O_p and for general theorems concerning the algebra of o_p and O_p, the reader is referred to Mann and Wald [5] and Pratt [6].

LEMMA 9.

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s - 1))\} = W_{1N}(s) + o_p(1)$$

and

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s + 1))\} = W_{2N}(s) + o_p(1),$$

as $N \to \infty$.

PROOF.

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s - 1))\}$$
$$= N \sum_{j=1}^{k+1} p_{jN}\{\log q_{jN}(\theta + \delta_N(s)) - \log p_{jN}\}$$
$$- N \sum_{j=1}^{k+1} p_{jN}\{\log q_{jN}(\theta + \delta_N(s - 1)) - \log p_{jN}\}.$$

Since the second partial derivatives of the function

$$f(u_1, \cdots, u_{k+1}, v_1, \cdots, v_{k+1}) = \sum_{j=1}^{k+1} u_j(\log v_j - \log u_j)$$

are continuous at $(\alpha_1, \cdots, \alpha_{k+1}, \beta_1(t), \cdots, \beta_{k+1}(t))$ and since $N^{\frac{1}{2}}(p_{jN} - \alpha_j) = O_p(1)$, $N^{\frac{1}{2}}(q_{jN}(t) - \beta_j(t)) = O_p(1)$, $j = 1, \cdots, k + 1$, we have the following stochastic Taylor expansion of $Nf(p_{1N}, \cdots, p_{k+1,N}, q_{1N}(t), \cdots, q_{k+1,N}(t))$ about $(\alpha_1, \cdots, \alpha_{k+1}, \beta_1(t), \cdots, \beta_{k+1}(t))$:

$$N \sum_{j=1}^{k+1} p_{jN}(\log q_{jN}(t) - \log p_{jN})$$
$$= N\phi(t)'p_N + N\psi(t)'q_N(t) - \tfrac{1}{2}\{N^{\frac{1}{2}}(p_N - \alpha)\}'A^{-1}\{N^{\frac{1}{2}}(p_N - \alpha)\}$$
$$- \tfrac{1}{2}\{N^{\frac{1}{2}}(q_N(t) - \beta(t))\}'AB(t)^{-2}\{N^{\frac{1}{2}}(q_N(t) - \beta(t))\}$$
$$+ N^{\frac{1}{2}}(p_N - \alpha)\}'B(t)^{-1}\{N^{\frac{1}{2}}(q_N(t) - \beta(t))\} + o_p(1).$$

Thus,

$$\lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s - 1))\}$$
$$= W_{1N}(s) - \tfrac{1}{2}R_{1N}(s) + R_{2N}(s) + o_p(1),$$

where

$$\begin{aligned} R_{1N}(s) = {} & \{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s)))\}'AB(\theta + \delta_N(s))^{-2} \\ & \cdot\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s)))\} \\ & - \{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1))) - \beta(\theta + \delta_N(s-1)))\}' \\ & \cdot AB(\theta + \delta_N(s-1))^{-2}\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) - \beta(\theta + \delta_N(s-1)))\} \end{aligned}$$

and

$$\begin{aligned} R_{2N}(s) = {} & \{N^{\frac{1}{2}}(p_N - \alpha)\}'B(\theta + \delta_N(s))^{-1}\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s)))\} \\ & - \{N^{\frac{1}{2}}(p_N - \alpha)\}'B(\theta + \delta_N(s-1))^{-1} \\ & \cdot\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) - \beta(\theta + \delta_N(s-1)))\}. \end{aligned}$$

We now note from Lemma 3 that for any fixed s,

$$B(\theta + \delta_N(s))^{-1} = A^{-1} + o(1) \quad \text{and} \quad AB(\theta + \delta_N(s))^{-2} = A^{-1} + o(1).$$

We also note from Lemmas 2 and 3 that for any fixed s, the covariance matrix of $N^{\frac{1}{2}}\{q_N(\theta + \delta_N(s)) - q_N(\theta + \delta_N(s-1))\}$, i.e., $\Sigma_{ss} - \Sigma_{s,s-1} - \Sigma_{s-1,s} + \Sigma_{s-1,s-1}$ is $o(1)$. Thus,

$$\begin{aligned} R_{1N}(s) = {} & \{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s)))\}'A^{-1} \\ & \cdot\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s)))\} \\ & - \{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) - \beta(\theta + \delta_N(s-1)))\}' \\ & \cdot A^{-1}\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) - \beta(\theta + \delta_N(s-1)))\} \\ & + O_p(1)\cdot o(1)\cdot O_p(1) \\ = {} & \{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s))) + N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) \\ & - \beta(\theta + \delta_N(s-1)))\}'A^{-1}\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s))) \\ & - N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) - \beta(\theta + \delta_N(s-1)))\} + o_p(1) \\ = {} & O_p(1)\cdot O(1)\cdot o_p(1) + o_p(1) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} R_{2N}(s) = {} & \{N^{\frac{1}{2}}(p_N - \alpha)\}'A^{-1} \\ & \cdot\{N^{\frac{1}{2}}(q_N(\theta + \delta_N(s)) - \beta(\theta + \delta_N(s))) - N^{\frac{1}{2}}(q_N(\theta + \delta_N(s-1)) \\ & - \beta(\theta + \delta_N(s-1)))\} + O_p(1)\cdot o(1)\cdot O_p(1) \\ = {} & O_p(1)\cdot O(1)\cdot o_p(1) + o_p(1) = o_p(1). \end{aligned}$$

This completes the proof of the first part of the lemma. The proof of the other part follows exactly in the same way.

We shall now prove the main theorems.

THEOREM 1. *For arbitrary K, the probability that the empirical likelihood inequality has more than one solution in the interval $(\theta - K/N^{\frac{1}{2}}, \theta + K/N^{\frac{1}{2}})$ tends to zero as $N \to \infty$.*

PROOF. There exists a finite number of points of Θ in the interval $(\theta - K/N^{\frac{1}{2}}, \theta + K/N^{\frac{1}{2}})$. In order that there exist two points $t_0 + r_1\delta/N^{\frac{1}{2}}$ and $t_0 + r_2\delta/N^{\frac{1}{2}}$ at whieh inequality (3) is satisfied, there must exist some r between r_1 ann r_2 for which

$$Z_N(t_0 + r\delta/N^{\frac{1}{2}}) \leqq Z_N(t_0 + (r-1)\delta/N^{\frac{1}{2}})$$

and

$$Z_N(t_0 + r\delta/N^{\frac{1}{2}}) \leqq Z_N(t_0 + (r+1)\delta/N^{\frac{1}{2}}),$$

i.e., there exists an integer s between $-(K+1)\delta$ and $K\delta$ such that

$$Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s-1)) \leqq 0$$

and

$$Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s+1)) \leqq 0.$$

Let $R_{1N}^*(s) = \lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s-1)\} - W_{1N}(s)$ and $R_{2N}^*(s) = \lambda^{-1}\{Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s+1))\} - W_{2N}(s)$. Then

$$\begin{aligned}
P[Z_N(\theta + \delta_N(s)) &- Z_N(\theta + \delta_N(s-1)) \leqq 0,\\
&Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s+1)) \leqq 0]\\
&= P[W_{1N}(s) + R_{1N}^*(s) \leqq 0, W_{2N}(s) + R_{2N}^*(s) \leqq 0]\\
&\leqq P[W_{1N}(s) \leqq \epsilon_1, W_{2N}(s) \leqq \epsilon_2] + P[R_{1N}^*(s) < -\epsilon_1] + P[R_{2N}^*(s) < -\epsilon_2]
\end{aligned}$$

for arbitrary $\epsilon_1 > 0$ and $\epsilon_2 > 0$. Since

$$\lim_{N\to\infty} P[|R_{1N}^*(s)| > \epsilon_1] = \lim_{N\to\infty} P[|R_{2N}^*(s)| > \epsilon_2] = 0$$

by Lemma 9,

$$\begin{aligned}
\lim_{N\to\infty} P[Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s-1)) \leqq 0,&\\
Z_N(\theta + \delta_N(s)) - Z_N(\theta + \delta_N(s+1)) \leqq 0]&\\
\leqq \lim_{N\to\infty} P[W_{1N}(s) \leqq \epsilon_1, W_{2N}(s) \leqq \epsilon_2].&
\end{aligned}$$

Choosing ϵ_1 and ϵ_2 such that $\epsilon_1 + \epsilon_2 < \delta^2 I_k$, the right-hand side of the above inequality is found to be zero by the first part of Lemma 8. Since this holds for each s between $-(K+1)\delta$ and $K\delta$, the desired result follows.

THEOREM 2. *The empirical likelihood inequality has a sequence of solutions $\{T_N\}$ for which*

$$\lim_{N\to\infty} P[N^{\frac{1}{2}}(T_N - \theta) \leqq c + s\delta] = \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1-\lambda)I_k)^{\frac{1}{2}}).$$

PROOF. For each sample of size N, we define a solution T_N of the empirical likelihood inequality as follows:

If (3) has no solution, then $T_N = t_0 = \theta + \delta_N(-r_0)$ by convention.

If (3) has at least one solution, then

$T_N = \theta + \delta_N(0)$ if (3) is satisfied at $\theta + \delta_N(0)$ and for positive integers s,

$T_N = \theta + \delta_N(s)$ if (3) is satisfied at $\theta + \delta_N(s)$ but not satisfied at any of the points $\theta + \delta_N(s')$, $s' = -s, \cdots, s - 1$,

$T_N = \theta + \delta_N(-s)$ if (3) is satisfied at $\theta + \delta_N(-s)$ but not satisfied at any of the points $\theta + \delta_N(s')$, $s' = -s + 1, \cdots, s - 1$.

We shall prove the theorem for the sequence $\{T_N\}$ defined above.

Let $A_N(s)$ be the set of all samples of size N for which inequality (3) is satisfied at $\theta + \delta_N(s)$ and $B_N(s)$ the set of all samples of size N for which $T_N = \theta + \delta_N(s)$. Then

$$P[N^{\frac{1}{2}}(T_N - \theta) \leqq c + s\delta] = P[\textstyle\bigcup_{r \leqq s} B_N(r)],$$

and we want to prove that this probability tends to $\Phi\{[c + (s + \frac{1}{2})\delta\} \cdot (\lambda(1 - \lambda)I_k)^{\frac{1}{2}})$ in the limit as $N \to \infty$.

We first note the following facts about $\{A_N(s)\}$ and $\{B_N(s)\}$:

(a) For any given $s_1 \neq s_2$, $\lim_{N\to\infty} P[A_N(s_1) \cap A_N(s_2)] = 0$.

(b) For any given s,

$$\lim_{N\to\infty} P[A_N(s)] = \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) - \Phi(\{c + (s - \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}).$$

(c) $\lim_{N\to\infty} P[\bigcup_{-\infty<s<\infty} A_N(s)] = 1$ or $\lim_{N\to\infty} P[\bigcap_{-\infty<s<\infty} A_N(s)'] = 0$ where $A_N(s)'$ denotes the complement of $A_N(s)$ in the sample space of size N.

(d) For any $s_1 \neq s_2$, $B_N(s_1)$ and $B_N(s_2)$ are disjoint.

(e) For each N, $\bigcup_{-\infty<s<\infty} B_N(s)$ is the entire sample space of size N.

(a) is merely a restatement of Theorem 1 while (d) and (e) follow from the fact that T_N is well-defined over the entire sample space of size N and the only values that T_N can take are $\theta + \delta_N(s)$, s integer. To prove (b), we note that for any given s, by the arguments used in proving Theorem 1, $\lim_{N\to\infty} P[A_N(s)]$ lies between $\lim_{N\to\infty} P[W_{1N}(s) > \epsilon_1, W_{2N}(s) > \epsilon_2]$ and $\lim_{N\to\infty} P[W_{1N}(s) > -\epsilon_1, W_{2N}(s) > -\epsilon_2]$ for arbitrary $\epsilon_1 > 0$ and $\epsilon_2 > 0$. Choosing ϵ_1 and ϵ_2 such that $\epsilon_1 + \epsilon_2 < \delta^2 I_k$, applying the second part of Lemma 8 and allowing ϵ_1 and ϵ_2 to tend to zero, both of these tend to the expression in the right-hand side of (b). To prove (c), for arbitrary $\epsilon > 0$ choose s_1 and s_2 such that

$$\Phi(\{c + (s_1 - \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) < \epsilon/2$$

and

$$\Phi(\{c + (s_2 + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) > 1 - \epsilon/2.$$

Then

$$\begin{aligned}
\lim_{N\to\infty} P[\textstyle\bigcup_{-\infty<s<\infty} A_N(s)] &\geqq \lim_{N\to\infty} P[\textstyle\bigcup_{s_1 \leqq s \leqq s_2} A_N(s)] \\
&= \textstyle\sum_{s=s_1}^{s_2} \lim_{N\to\infty} P[A_N(s)] \\
&= \Phi(\{c + (s_2 + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) - \Phi(\{c + (s_1 - \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\
&> 1 - \epsilon
\end{aligned}$$

by (a) and (b). Since the above holds for arbitrary $\epsilon > 0$, (c) follows.

We shall now express the sets $B_N(s)$ in terms of the sets $A_N(s)$.

$$\begin{aligned} B_N(0) &= A_N(0) \text{ if } r_0 \neq 0 \\ &= A_N(0) \cup [\cap_{-\infty<r<\infty} A_N(r)'] \text{ if } r_0 = 0. \end{aligned}$$

For $s = 1, 2, \cdots$,

$$\begin{aligned} B_N(s) &= A_N(s) \cap \{\cap_{-s \leq r \leq s-1} A_N(r)'\} \text{ if } s \neq -r_0 \\ &= [A_N(s) \cap \{\cap_{-s \leq r \leq s-1} A_N(r)'\}] \cup [\cap_{-\infty<r<\infty} A_N(r)'] \text{ if } s = -r_0, \end{aligned}$$

and

$$\begin{aligned} B_N(-s) &= A_N(-s) \cap \{\cap_{-s+1 \leq r \leq s-1} A_N(r)'\} \quad \text{if } s \neq r_0 \\ &= [A_N(-s) \cap \{\cap_{-s+1 \leq r \leq s-1} A_N(r)'\}] \cup [\cap_{-\infty<r<\infty} A_N(r)'] \quad \text{if } s = r_0. \end{aligned}$$

It now follows from (a) and (c) that

(f) For any given s, $\lim_{N\to\infty} P[B_N(s)] = \lim_{N\to\infty} P[A_N(s)]$.

We now conclude from (b) and (f) that

(g) For any given s,

$$\begin{aligned} \lim_{N\to\infty} P[B_N(s)] = \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\ - \Phi(\{c + (s - \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}). \end{aligned}$$

By (d) we have for any given s,

$$P[N^{\frac{1}{2}}(T_N - \theta) \leq c + s\delta] = P[\textstyle\bigcup_{-\infty<r\leq s} B_N(r)] = \sum_{r=-\infty}^{s} P[B_N(r)].$$

For arbitrary $\epsilon > 0$, let us choose s_1 and s_2 again as we did for proving (c). We consider ϵ so small that $s_1 < s < s_2$. Then

$$\begin{aligned} \lim_{N\to\infty} \textstyle\sum_{r=-\infty}^{s} P[B_N(r)] &\geq \lim_{N\to\infty} \textstyle\sum_{r=s_1}^{s} P[B_N(r)] \\ &= \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) - \Phi(\{c + (s_1 - \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\ &> \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) - \epsilon \end{aligned}$$

by (g). Again,

$$\begin{aligned} &\lim_{N\to\infty} \textstyle\sum_{r=-\infty}^{s} P[B_N(r)] \\ &= \lim_{N\to\infty} \textstyle\sum_{r=-\infty}^{s_2} P[B_N(r)] - \sum_{r=s+1}^{s_2} \lim_{N\to\infty} P[B_N(r)] \\ &= \lim_{N\to\infty} P[\textstyle\bigcup_{-\infty<r\leq s_2} B_N(r)] - \Phi(\{c + (s_2 + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\ &\quad + \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\ &\leq 1 - \Phi(\{c + (s_2 + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) + \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) \\ &< \Phi(\{c + (s + \tfrac{1}{2})\delta\}(\lambda(1 - \lambda)I_k)^{\frac{1}{2}}) + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary in the above two inequalities, the theorem is proved.

4. Miscellaneous remarks.

(i) *Choice of* t_0. Any consistent estimate of θ with asymptotic variance of the order of N^{-1}, e.g., the median of $X_i - Y_{i'}$, $i = 1, \cdots, m$, $i' = 1, \cdots, n$, or median of X's $-$ median of Y's (which is easier to compute) will be a reasonable choice for t_0. Such a choice of t_0 has the advantage that in case of more than one solution of the empirical likelihood inequality, the one lying close to t_0 can be picked out as the estimate with the property proved in Theorem 2.

(ii) *Convergence of* I_k to I. Suppose $f(x) \to 0$ as $x \to \pm\infty$, $\log f(x)$ is continuously differentiable everywhere and the Fisher-information I given by (2) exists. Then $I_k(a_1, \cdots, a_k)$ converges to I as $k \to \infty$, $a_1 \to -\infty$, $a_j - a_{j-1} \to 0$, $j = 2, \cdots, k$, and $a_k \to \infty$. To prove this, let

$$f_j(x) = f(x)/\alpha_j, \qquad x \varepsilon C_j,$$
$$= 0 \qquad \text{elsewhere}, \quad j = 1, \cdots, k+1,$$

and let E_j and Var_j denote respectively the expectation and the variance with respect to the probability density function f_j. Then

$$E_j[f'(X)/f(X)] = \alpha_j'/\alpha_j \quad \text{and} \quad E_j[\{f'(X)/f(X)\}^2] = \alpha_j^{-1} \int_{C_j} \{f'(x)\}^2/f(x)\,dx,$$

$j = 1, \cdots, k+1$, since $f(x) \to 0$ as $x \to \pm\infty$. Hence

$$\begin{aligned} I - I_k(a_1, \cdots, a_k) &= \textstyle\sum_{j=1}^{k+1} \int_{C_j} \{f'(x)\}^2/f(x)\,dx - \sum_{j=1}^{k+1} \alpha_j'^2/\alpha_j \\ &= \textstyle\sum_{j=1}^{k+1} \alpha_j \,\text{Var}_j\,[f'(X)/f(X)] \\ &\leqq \alpha_1 E_1[\{f'(X)/f(X)\}^2] + \alpha_{k+1}E_{k+1}[\{f'(X)/f(X)\}^2] \\ &\quad + \textstyle\sum_{j=2}^{k} \alpha_j \,\text{Var}_j\,[f'(X)/f(X)] \\ &= \textstyle\int_{C_1 \cup C_{k+1}} \{f'(x)\}^2/f(x)\,dx + \sum_{j=2}^{k} \alpha_j\,\text{Var}_j\,[f'(X)/f(X)]. \end{aligned}$$

Choose and fix $\epsilon > 0$. Since I exists, we can find $a < b$ such that

$$\textstyle\int_{-\infty}^{a} \{f'(x)\}^2/f(x)\,dx + \int_{b}^{\infty} \{f'(x)\}^2/f(x)\,dx < \frac{1}{2}\epsilon.$$

Since $f'(x)/f(x)$ is continuous everywhere, it is uniformly continuous on $[a, b]$. Hence, we can find $\delta > 0$ such that

$$|f'(x_1)/f(x_1) - f'(x_2)/f(x_2)| < (\tfrac{1}{2}\epsilon)^{\frac{1}{2}} \tag{13}$$

whenever $x_1, x_2 \,\varepsilon\, [a, b]$ and $|x_1 - x_2| < \delta$. Now let $k > 1 + (b-a)/\delta$, $a_1 = a$, $a_k = b$ and $a_j = a + (j-1)(b-a)/(k-1)$, $j = 2, \cdots, k-1$. For such $a_1, \cdots, a_k$,

$$\textstyle\int_{C_1 \cup C_{k+1}} \{f'(x)\}^2/f(x)\,dx < \frac{1}{2}\epsilon,$$

and since $P[M_1 \leqq Y \leqq M_2] = 1$ implies $\text{Var}\,(Y) \leqq (M_2 - M_1)^2$, (13) implies $\text{Var}_j\,[f'(X)/f(X)] \leqq \frac{1}{2}\epsilon$, $j = 2, \cdots, k$. Hence

$$I - I_k(a_1, \cdots, a_k) < \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon \textstyle\sum_{j=2}^{k} \alpha_j \leqq \epsilon.$$

To complete the proof we note that the preceding arguments imply that if $I - I_k(a_1, \cdots, a_k) < \epsilon$ and if $a_1' < \cdots < a_m'$ are such that $\{a_1, \cdots, a_k\}$ is a subset of $\{a_1', \cdots, a_m'\}$ then $I - I_m(a_1', \cdots, a_m') < \epsilon$.

(iii) *Choice of class-intervals.* It should be our aim to choose the class-intervals in such a manner that I_k/I is sufficiently close to 1 irrespective of the form of f. In the next paragraph we shall give a general method of constructing consistent estimates $\hat{I}_k$ of I_k for a given choice of class-intervals. As a rough and ready method for choosing class-intervals we make them finer and finer till the relative increase in estimated I_k is sufficiently small. However, in order to make our estimate of θ robust in the sense of high asymptotic efficiency for all f, we need a consistent estimate $\hat{I}$ of I. Then we can make the class-intervals sufficiently fine so that $\hat{I}_k/\hat{I}$ attains a desired value. Bhattacharya [1] has given a method of obtaining consistent estimates of I under certain conditions which assumes a knowledge about the rate at which the tails of f approach 0.

(iv) *Consistent estimation of* I_k. If $f(a_j)$ and $F(a_j)$, $j = 1, \cdots, k$, are replaced by any consistent estimators of these quantities in the expression of I_k given in (4), we will get a consistent estimator of I_k.

(v) *Large sample tests for hypotheses about* θ. Suppose we want to test the null hypothesis $H_0 : \theta = \theta_0$ against alternatives on both sides. We know (Theorems 1 and 2) that under H_0, for arbitrary K, with probability tending to 1 as $N \to \infty$, there is a unique solution T_N of the empirical likelihood inequality in the interval $(\theta_0 - K/N^{\frac{1}{2}}, \theta_0 + K/N^{\frac{1}{2}})$ and that $N^{\frac{1}{2}}(T_N - \theta_0)$ is asymptotically roughly normal with mean 0 and variance $1/\lambda(1 - \lambda)I_k$. Using a consistent estimate $\hat{I}_k$ of I_k as indicated in the above paragraph, we now define the statistic

$$\tau_N = (N\lambda(1 - \lambda)\hat{I}_k)^{\frac{1}{2}}(T_N - \theta_0)$$

which is asymptotically roughly standard normal under H_0. Thus when $\delta > 0$ is chosen sufficiently small, τ_N can be used as a test criterion for testing H_0 and the critical region $|\tau_N| \geqq \Phi^{-1}(1 - \frac{1}{2}\alpha)$ will have an approximate level of significance α when N is large. It can be easily seen that for any f, the asymptotic efficiency of this test relative to the corresponding maximum likelihood test (in the sense of Pitman) is I_k/I. The test criterion τ_N can also be used for testing the null hypothesis $H_0 : \theta \leqq \theta_0$ (or $\theta \geqq \theta_0$) against alternatives $H_1 : \theta > \theta_0$ (or $\theta < \theta_0$).

REFERENCES

[1] BHATTACHARYA, P. K. (1967). Estimation of a probability density function and its derivatives. Submitted to *Sankhyā*.

[2] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[3] HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

[4] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[5] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[6] PRATT, J. W. (1959). On a general concept of 'in probability'. *Ann. Math. Statist.* **30** 549–558.

[7] STEIN, C. (1956). Efficient non-parametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–195. Univ. of California Press.

[8] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.