The Annals of Statistics
1985, Vol. 13, No. 2, 727–742

EFFICIENCIES OF CHI-SQUARE AND LIKELIHOOD RATIO GOODNESS-OF-FIT TESTS

BY M. P. QUINE AND J. ROBINSON

University of Sydney

The classical problem of choice of number of classes in testing goodness of fit is considered for a class of alternatives, for the chi-square and likelihood ratio statistics. Pitman and Bahadur efficiencies are used to compare the two statistics and also to analyse the effect for each statistic of changing the number of classes for the case where the number of classes increases asymptotically with the number of observations. Overall, the results suggest that if the class of alternatives is suitably restricted the number of classes should not be very large.

1. Introduction. Although goodness-of-fit tests from grouped data constitute a classical problem in inference, the problem of choice of number of classes is still unresolved. For instance, the recommendations of Mann and Wald (1942), Kempthorne (1967) and Hamdan (1963) are radically different. In recent years there have been some attempts to illuminate the problem by focusing attention on a particular class of alternatives; see for example Holst (1972), Gvanceladze and Chibisov (1979).

Specifically, we consider the problem of testing the goodness of fit of a continuous distribution F_0 to a set of N observations grouped into n equal probability intervals. Our methods could also be used in a more general context so long as the n probabilities were of the same order, but the practical implications would be the same. Without loss of generality we suppose that $F_0'(x) = 1$, $0 < x < 1$. Since we are concerned with limit results, we take $n = n(N)$. We assume that $n(\cdot)$, taken as a function of the continuous variable x, is regularly varying, that is for some q, $n(ax)/n(x) \to a^q$ as $x \to \infty$, for all $a > 0$. We assume further that $n(N) \leq CN$ for some positive constant C, so that $0 \leq q \leq 1$. These assumptions include all cases of practical interest in which the number of classes increase (asymptotically) with N. In the case of Bahadur efficiency, we consider the null hypothesis

$$H_0: f(x) = 1$$

versus the fixed alternative

$$H_1: f(x) = 1 + g(x),$$

where $\int_0^1 g(x)\,dx = 0$ and $0 < \| g \|_5 = (\int_0^1 |g(x)|^5\,dx)^{1/5} < \infty$. In the case of Pitman efficiency, we consider the same null versus a sequence of alternatives.

Received August 1983; revised March 1984.
AMS 1982 *subject classifications*. Primary 62G20; secondary 60F05, 60F10.
Key words and phrases. Pitman efficiency, Bahadur efficiency, chi-square, likelihood ratio, goodness-of-fit, central limit theorem, large deviations.

That is, we suppose that for each integer $N \geq 1$, the alternative is

$$H_{1N}: f(x) = 1 + \nu(N)h_N(x),$$

where $\int_0^1 h_N(x)\,dx = 0$, $\| h_N - g \|_5 \to 0$ for some g with $\| g \|_5 < \infty$ and $\int_0^1 g(x)\,dx = 0$, and $\nu(N)$ is chosen so that the power for a test of size α has a limit in $(\alpha, 1)$.

Suppose $X_1, \cdots, X_{n(N)}$ are the numbers of observations in the intervals. We deal with the two test statistics

$$S_N = \sum_{k=1}^{n(N)} X_k^2 \quad \text{and} \quad T_N = \sum_{k=1}^{n(N)} X_k \log X_k$$

(taking $0 \log 0 = 0$). We compare both Pitman and Bahadur efficiencies for the following pairs, where $n' = n'(N)$, another choice of number of intervals, is also assumed to be regularly varying with index of regular variation $q' \in [0, 1]$:

(i) S_N and $S'_N = \sum_{k=1}^{n'(N)} X_k^2$
(ii) T_N and $T'_N = \sum_{k=1}^{n'(N)} X_k \log X_k$
(iii) S_N and T_N.

In the case of Pitman efficiency, we show in the next section that in cases (i) and (ii), for $n(N) \to \infty$, $N/n(N)$, $N/n'(N) \geq \varepsilon > 0$,

$$\text{PE}(S_N, S'_N) = \text{PE}(T_N, T'_N) = c^{1/(2-q)} \tag{1.1}$$

if $n'(N)/n(N) \to c \in (0, \infty)$, and

$$\text{PE}(S_N, S'_N) = \text{PE}(T_N, T'_N) = \infty \tag{1.2}$$

if $n'(N)/n(N) \to \infty$. This complements the results of Gvanceladze and Chibisov (1979) and provides a further contrast to the results of Mann and Wald (1942). We also show that in (iii), if $n(N)$ and $N/n(N) \to \infty$,

$$\text{PE}(S_N, T_N) = 1. \tag{1.3}$$

This extends the well-known result that (1.3) holds when $n(N)$ is constant and $N \to \infty$, and shows that the result $\text{PE}(S_N, T_N) > 1$ when $N/n(N) \to \lambda_0$, $0 < \lambda_0 < \infty$ (Holst, 1972) is an extreme case, although even here computations suggest that the maximum efficiency is very close to 1 (see Table 1).

Rather different results are given in Section 3 in the case of Bahadur efficiency.

TABLE 1
PE(S, T) *for* $N/n \to \lambda$

λ	**0.01**	**0.05**	**0.1**	**0.5**	**1.0**
PE(S, T)	1.001	1.006	1.013	1.058	1.102
λ	**1.5**	**2.0**	**2.37**	**2.5**	**3**
PE(S, T)	1.131	1.145	1.148	1.147	1.142
λ	**5**	**10**	**50**	**100**	**500**
PE(S, T)	1.095	1.038	1.007	1.003	1.001

We show that for $n(N) \to \infty$, $N/n(N)$, $N/n'(N) \geq \varepsilon > 0$,

$$\mathrm{BE}(S_N, S'_N) = c^{1/(2-q)} \tag{1.4}$$

if $n'(N)/n(N) \to c \in (0, \infty)$, and

$$\mathrm{BE}(S_N, S'_N) = \infty \tag{1.5}$$

if $n'(N)/n(N) \to \infty$. We further show that if $n(N)$, $n'(N)$, $N/n(N)$, $N/n'(N) \to \infty$,

$$\mathrm{BE}(T_N, T'_N) = 1. \tag{1.6}$$

For case (iii), we show that

$$\mathrm{BE}(S_N, T_N) = 0 \tag{1.7}$$

if $N/n(N) \geq \varepsilon > 0$ and $n(N) \to \infty$, which extends the result that for fixed n,

$$\mathrm{BE}(S_N, T_N) \leq 1 \tag{1.8}$$

with inequality "nearly everywhere" (see Hoeffding, 1965; Bahadur, 1971, pages 31–32). Results (1.6)–(1.7) do not require that $n(N)$ and $n'(N)$ be regularly varying.

It should be noted that the Pitman and Bahadur criteria yield a similar verdict on χ^2 tests based on different numbers of classes. However, in the case of the likelihood ratio test, our results demonstrate a reverse of the more commonly encountered situation in that the tests are equivalent in the sense of Bahadur, but not in the sense of Pitman. In the comparison of χ^2 and likelihood ratio tests, conflicting results are achieved by the two criteria in the case $N/n(N) \to \lambda_0 \in (0, \infty)$: χ^2 is superior in the Pitman sense but infinitely inferior in the Bahadur sense. In the case $n(N)$, $N/n(N) \to \infty$, (1.3) and (1.7) simply extend the known results for finite n, that is (1.3) and (1.8).

As well as details of (1.1)–(1.8), Sections 2 and 3 contain definitions of Pitman and Bahadur efficiencies. Proofs are given in Sections 4 and 5.

2. Pitman efficiency. Consider a test of the hypothesis H_0, using a test statistic U_N based on a sample of N, against the alternative hypothesis H_{1N} so chosen that the power of the size α test based on U_N under H_{1N} tends to β, $\alpha < \beta < 1$, as N tends to infinity. Let U'_N be another test statistic and let N' be a sequence such that the power of the size α test based on $U'_{N'}$ under H_{1N} also tends to β as $N' \to \infty$. Then if the limit of N'/N exists and is the same for all such sequences N', we call it the *Pitman efficiency* of $\{U_N\}$ with respect to $\{U'_N\}$ and write

$$\mathrm{PE}(U_N, U'_N) = \lim N'/N.$$

Specifically, we consider the alternatives given in Section 1 and show that under H_0 and H_{1N}, if U_N is one of S_N, T_N there are values (defined in Section 4) $E^a_{0N}U_N$, $V^a_{0N}U_N$, $E^a_{1N}U_N$, $V^a_{1N}U_N$ such that when $n(N) \to \infty$ and $N/n(N) \geq \varepsilon >$

0 for S_N or $N/n(N) \to \infty$ for T_N,

$$(U_N - E^a_{jN}U_N)/(V^a_{jN}U_N)^{1/2} \to_d N(0, 1) \tag{2.1}$$

for $j = 0, 1$ and under H_0 and H_{1N}, respectively. We choose $\nu(N)$ so that

$$(E^a_{1N}U_N - E^a_{0N}U_N)/(V^a_{0N}U_N)^{1/2} \to b \tag{2.2}$$

for some constant $b \in (0, \infty)$ and show that for this $\{\nu(N)\}$

$$V^a_{1N}U_N/V^a_{0N}U_N \to 1, \tag{2.3}$$

so that the power depends only on (2.2). Then for another statistic U'_N (S'_N, for example) we choose $\{N'\}$ so that (2.2) holds for the same constant b when U_N is replaced by $U'_{N'}$ and $n(N)$ is changed to $n(N')$ or $n'(N')$ as appropriate. We then look at the ratio N'/N.

Consider first S_N and S'_N. In Section 4 we use the results of Morris (1975) to show that if $n \to \infty$, $N/n \geq \varepsilon > 0$, then (2.1) holds for $U_N = S_N$ with

$$E^a_0 S_N = N + N^2/n, \tag{2.4}$$

$$E^a_1 S_N = N + N^2/n + n^{-1}N^2\nu^2 \| g \|^2_2(1 + o(1)) \tag{2.5}$$

and, taking

$$\nu^2 = dn^{1/2}/N, \tag{2.6}$$

$$V^a_j S_N = 2N^2/n(1 + o(1)), \qquad j = 0, 1, \tag{2.7}$$

where here and in the sequel we omit the index N from n, ν, h, E^a_j and V^a_j when no confusion might arise. This ensures that (2.3) holds and that

$$\begin{aligned}(E^a_1 S_N - E^a_0 S_N)/(V^a_0 S_N)^{1/2} \\ = 2^{-1/2}n^{-1/2}N\nu^2 \| g \|^2_2(1 + o(1)) \to 2^{-1/2}d \| g \|^2_2 \in (0, \infty).\end{aligned} \tag{2.8}$$

Since (2.4), (2.5) and (2.7) still hold when n, N and S_N are replaced by n', N' and $S'_{N'}$, ν^2 remaining unchanged as in (2.6), it can be seen from (2.8) that in order for the test based on $S'_{N'}$ to have the same asymptotic power under H_{1N} we must choose N' so that

$$n(N)/N^2 \sim n'(N')/N'^2. \tag{2.9}$$

To obtain (1.1) for χ^2, note that since $n'(N')/n(N')$ also tends to c, (2.9) gives

$$n(N)/N^2 \sim cn(N')/N'^2 \sim n(aN')/(aN')^2$$

where $a = c^{-1/(2-q)}$. Noting that $R(x) = n(x)/x^2$ is regularly varying with index of regular variation $q - 2 < 0$, the representation theorem (Seneta, 1976, page 2) implies

$$R(r_x x)/R(x) \to r^{q-2} \quad \text{as} \quad x \to \infty,$$

if $r_x \to r \in [0, \infty]$, $0 < r_x < \infty$. Let a subsequence of N'/N have a further subsequence for which $N'/N \to t \in [0, \infty]$. Then for the latter subsequence

$R(aN')/R(N) \to (at)^{q-2}$, and since from above this limit is 1, we have $t = 1/a$. Thus for the whole sequence $N'/N \to 1/a = c^{1/(2-q)}$, establishing (1.1). This argument can also be used to establish (1.2) for χ^2. Suppose that $n'(N)/n(N) \to \infty$; then from (2.9) for arbitrarily large C there exists N_C such that for $N' > N_C$

$$n(N)/N^2 > Cn(N')/N'^2 \sim n(AN')/(AN')^2,$$

where $A = C^{-1/(2-q)}$. So asymptotically, $N < AN'$, giving (1.2).

Gvanceladze and Chibisov (1979) have shown for a similar scheme to ours that the power of the χ^2 test (and the likelihood ratio test) tends to the significance level as $N \to \infty$ whenever $n \to \infty$ for $\nu = N^{-1/2}$. Thus in a sense (1.1) (and (1.2)) provide refinements of their results and lend further weight to the practical implication that the number of classes chosen should not be too great. As Gvanceladze and Chibisov (1979) mention, this recommendation runs counter to that of Mann and Wald (1942), who use a minimax argument to obtain the relation $n \sim cN^{2/5}$ (see also Kendall and Stuart, 1973). To illustrate the inevitability of such a conclusion when a sufficiently wide class of alternatives is used, consider alternatives

$$F_N(x) = x + \nu(N)H_N(x), \qquad \int_0^1 dH_N(x) = 0.$$

In order to obtain substantial power with all alternatives satisfying $\sup_x |H_N(x)| \geq \varepsilon > 0$, we need to take $n > C\nu^{-1}$ (these alternatives include the discrete uniform distribution with jumps of size $c\nu$ at intervals of length $c\nu$). By the kind of argument centering on (2.7), to maximise power, β, against alternatives subject to this constraint, we need to take n as small as possible, and to get $\beta \in (\alpha, 1)$ we need $n^{-1/2}N\nu^2 \sim C$, so that $N \sim Cn^{5/2}$.

We may cite the case study of a normal shift alternative by Hamdan (1963) (see also Dahiya and Gurland, 1973) who concluded on numerical grounds that small n gives higher power. More generally, consider

$$H_0\colon F(x) = F_0(x),$$

$$H_{1N}\colon F(x) = F_0(x - \delta_N) = F_0(x) - \delta_N f_0(x)(1 + o(1)).$$

When standardized to a uniform null, the density under H_{1N} becomes

$$1 - \delta_N f_0'(F_0^{-1}(x))/f_0(F_0^{-1}(x))(1 + o(1))$$

which is asymptotically of the same form as our scheme.

We now consider cases (ii) and (iii). In Section 4, we show that for n and $N/n \to \infty$, (2.1) holds with $U_N = T_N$ under H_0 and H_{1N}, respectively, that (2.3) holds and that

$$(2.10) \qquad (E_1^a T_N - E_0^a T_N)/(V_0^a T_N)^{1/2} = 2^{-1/2} n^{-1/2} N \nu^2 \| g \|_2^2 (1 + o(1));$$

so again we take ν as at (2.6). Thus (1.3) is proved. The proofs of (1.1) and (1.2) for likelihood ratio follow those for χ^2.

The case (iii) when λ is finite has been considered by Holst (1972) who showed that the efficiency was

$$[\operatorname{corr}\{(Y-\lambda)-Y,\ Y\log Y\}]^2$$

where Y is Poisson with parameter λ. We tabulate this in Table I for some values of λ to show that it is never far from the value 1, which is the limit as $\lambda \to \infty$. The moments in the correlation were obtained using the Poisson probabilities up to the 1000th term.

3. Bahadur efficiency. A rather different measure of efficiency has been described by Bahadur (1971). We need a slightly more general definition than his. Let P_0, E_0, V_0, P_1, E_1, V_1 denote probability, expectation and variance under H_0 and H_1 (as given in Section 1), respectively. Let U_N and U'_N be two statistics based on N observations as in Section 2, and

$$1 - F_N(u) = P_0(U_N \geq u), \quad 1 - F'_N(u) = P_0(U'_N \geq u)$$

and define the random variables (levels)

$$L_N = 1 - F_N(U_N), \quad L'_N = 1 - F'_N(U'_N).$$

Suppose that there are sequences ϕ_N and ϕ'_N such that

$$(U_N - E_0 U_N)/\phi_N \to_{P_1} b, \quad (U'_N - E_0 U'_N)/\phi'_N \to_{P_1} b',$$

and that there are sequences ψ_N, ψ'_N such that for $N \to \infty$,

$$-(N\psi_N)^{-1}\log P_0(U_N > E_0 U_N + \phi_N t) \to c(t),$$

$$-(N\psi'_N)^{-1}\log P_0(U'_N > E_0 U'_N + \phi'_N t) \to c'(t)$$

for each $t \in I$, an open set containing b and b', and c, c' are continuous functions on I. Arguing as in Bahadur (1971, Theorem 7.2)

$$-(N\psi_N)^{-1}\log L_N \to_{P_1} c(b), \quad -(N\psi'_N)^{-1}\log L'_N \to_{P_1} c'(b').$$

Let N' be a sequence such that the levels based on U_N and $U'_{N'}$ are asymptotically equivalent. Then, if $0 < c(b), c'(b') < \infty$,

$$N'/N \sim \psi_N c(b)/\psi'_{N'} c'(b').$$

Thus we define the Bahadur efficiency of U_N with respect to U'_N as

$$\operatorname{BE}(U_N, U'_N) = rc(b)/c'(b')$$

if $\psi_N/\psi'_{N'} \to r \in [0, \infty]$.

Using the notation $\lambda = N/n$,

$$E_1 S_N - E_0 S_N = n\lambda^2\delta(1 + o(1)) \tag{3.1}$$

and

$$V_1 S_N = (2n\lambda^2(1+\delta) + 4n\lambda^3\xi)(1 + o(1)), \tag{3.2}$$

where

$$\delta = \int_0^1 g^2(x)\,dx,$$

$$\xi = \int_0^1 (1 + g(x))\left[g(x) - \int_0^1 g^2(x)\,dx\right]^2 dx.$$

So we can see that

$$(S_N - E_0 S_N)/(n\lambda^2) \to_{P_1} \delta$$

since under H_1 the left-hand side has expectation which tends to δ and variance which tends to 0, as $N \to \infty$. In Section 5, we prove the following result.

THEOREM 1. *For* $\lambda \geq \varepsilon > 0$ *and* $n \to \infty$,

$$\lim_{N\to\infty} -\{\log P_0(S_N > n\lambda + n\lambda^2(1+\delta))\}/(\lambda n^{1/2}\log n) = \tfrac{1}{2}\delta^{1/2}.$$

Thus in the definition of Bahadur efficiency we can take

$$\phi_N = N^2/n(N),\quad \phi_N' = N^2/n'(N),\quad \psi_N = n(N)^{-1/2}\log n(N),$$
$$\psi_N' = n'(N)^{-1/2}\log n'(N),\quad c(t) = c'(t) = \tfrac{1}{2}t^{1/2},$$

so that

$$N'/N \sim \psi_N/\psi_{N'}' \sim (n'(N')/n(N))^{1/2}\log n(N)/\log n'(N')$$

(1.4) and (1.5) follow from this in the same way that (1.1) and (1.2) for χ^2 follow from (2.9), since $n(x)/\log n(x)$ is regularly varying with the same index as $n(x)$.

REMARK. This theorem extends a result of Hoeffding (1965) who shows that for n finite,

$$\begin{aligned}\lim_{N\to\infty} &-[\log P_O(S_N > n\lambda + n\lambda^2(1+\delta))]/(n\lambda)\\ &= n^{-1}(1 + (\delta(n-1))^{1/2})\log(1 + (\delta(n-1))^{1/2})\\ &\quad + (1 - n^{-1})(1 - (\delta/(n-1))^{1/2})\log(1 - (\delta/(n-1))^{1/2}).\end{aligned}$$

As $n \to \infty$, the right-hand side is $\tfrac{1}{2}(\delta/n)^{1/2}\log n(1 + o(1))$.

Again using the notation of the introduction, for $\lambda \to \infty$, $n \to \infty$

$$E_1 T_N - E_0 T_N = n\lambda\delta'(1 + o(1)) \tag{3.3}$$

and

$$V_1 T_N = n\lambda\xi'(1 + o(1)), \tag{3.4}$$

where

$$\delta' = \int_0^1 (1 + g(x))\log(1 + g(x))\,dx$$

and

$$\xi' = \int_0^1 (1 + g(x))\left[\log(1 + g(x)) - \int_0^1 (1 + g(y))\log(1 + g(y))\, dy\right]^2 dx.$$

Thus

$$(T_N - E_0 T_N)/N \to_{P_1} \delta'.$$

since under H_1 the left-hand side has expectation which tends to δ' and variance which tends to 0 as $N \to \infty$. Now we need the following result.

THEOREM 2. *For $n, \lambda \to \infty$,*

$$\lim_{N\to\infty}[-N^{-1}\log P_0(T_N > n(\lambda \log \lambda + \lambda\delta'))] = \delta'.$$

This theorem extends the result that for n finite

$$\lim_{N\to\infty}\{-N^{-1}\log P_0(T_N > E_1 T_N)\} = \delta_n,$$

where $\delta_n = n^{-1}\sum_k (1 + \bar{g}_k)\log(1 + \bar{g}_k)$ and $\bar{g}_k = n \int_{(k-1)/n}^{k/n} g(x)\, dx$ (Bahadur, 1971, page 32), in the sense that $\delta_n \to \delta'$ as $n \to \infty$.

For $\lambda \to \infty$, Theorem 2 implies (1.6) and Theorems 1 and 2 imply (1.7). If $\lambda \to \lambda_0 \in (0, \infty)$,

(3.5) $$\lim_{N\to\infty}\{-N^{-1}\log P_0(T_N > E_1 T_N)\} \ge -\log E\{\exp(t(Y \log Y - A))\},$$

where $A = \lim_{N\to\infty} n^{-1} E_1 T_N$ and Y is Poisson (λ_0), for any $t \in (0, 1/\lambda_0)$; so (1.7) holds in this case also.

4. Proof of results of Section 2. The statistics under consideration are special cases of $U_N = \sum_k c(X_k, k)$, where $X_1, \cdots, X_k$ are defined in the introduction. Proofs of limit results are based on the fact that U_N has the conditional distribution of $V_N = \sum_k c(Y_k, k)$, where $Y_1, \cdots, Y_n$ are independent Poisson variates with parameters $\lambda_1, \cdots, \lambda_n$, given that $\sum_k Y_k = N$, where

$$\lambda_k = N \int_{(k-1)/n}^{k/n} (1 + \nu h_N(x))\, dx = \lambda(1 + \nu \bar{h}_{Nk}),$$

$k = 1, \cdots, n$. Under H_0, $\lambda_k = \lambda$, $k = 1, \cdots, n$. Such central limit results are obtained in Morris (1975) and, together with rates of convergence, in Quine and Robinson (1984). If

$$\mu = E \sum_k c(Y_k, k),$$

$$\gamma = N^{-1} \sum_k \operatorname{cov}(c(Y_k, k), Y_k)$$

and

$$\sigma^2 = n^{-1} \sum_k \operatorname{var}(c(Y_k, k) - \gamma Y_k),$$

then, in the two cases considered here, it is shown in Theorems 5.1 and 5.2 of

Morris (1975) that

$$n^{-1/2}(U_N - \mu)/\sigma \to_d N(0, 1)$$

if

(4.1) $$\min_k \lambda_k \geq \varepsilon > 0 \quad \text{and} \quad N^{-1}\max_k \lambda_k = o(1)$$

and

(4.2) $$\max_k \operatorname{var}(c(Y_k, k) - \gamma Y_k)/n\sigma^2 = o(1).$$

These follow trivially under H_0 since $\lambda \geq \varepsilon > 0$ and $n \to \infty$. Under H_{1N} from the Hölder inequality,

$$\nu \mid \bar{h}_{Nk} \mid \leq \nu n^{1/5} \parallel h_N \parallel_5 = o(\lambda^{-1/2})$$

when $n \to \infty$, since $\parallel h_N - g \parallel_5 \to 0$ and $\parallel g \parallel_5 < \infty$, and from (2.5) $\nu = O(n^{-1/4}\lambda^{-1/2})$. Thus (4.1) holds since $\lambda \geq \varepsilon > 0$. To complete the proof of asymptotic normality it remains to verify (4.2) for S_N and T_N under H_{1N}, to which end we need the first two moments for these statistics. This will enable us at the same time to show that (2.3) holds for S_N and T_N and to obtain (2.8) and (2.10).

For S_N, $c(Y_k, k) = Y_k^2$, so

$$\mu = \textstyle\sum_k (\lambda_k + \lambda_k^2) = n\lambda + n\lambda^2 + \lambda^2\nu^2 \sum_k \bar{h}_{Nk}^2.$$

If we use the notation $f^{(n)}$ for the projection of f in L_2 onto the step functions on the intervals $((k-1)/n, k/n)$, $k = 1, \cdots, n$, then $n^{-1} \sum_k h_{Nk}^2 = \parallel h_N^{(n)} \parallel_2^2$. Further

$$\parallel h_N^{(n)} - g \parallel_2 \leq \parallel h_N^{(n)} - g^{(n)} \parallel_2 + \parallel g^{(n)} - g \parallel_2 \leq \parallel h_N - g \parallel_2 + \parallel g^{(n)} - g \parallel_2.$$

Now $\parallel h_N - g \parallel_2 \leq \parallel h_N - g \parallel_5 \to 0$. Also since the continuous functions are dense in L_2, given $\varepsilon > 0$, there is a continuous function f such that $\parallel f - g \parallel_2 < \varepsilon$. So

$$\parallel g^{(n)} - g \parallel_2 \leq \parallel g^{(n)} - f^{(n)} \parallel_2 + \parallel f^{(n)} - f \parallel_2 + \parallel f - g \parallel_2 \leq 3\varepsilon$$

for large enough n, because f, being continuous, is Riemann integrable. Thus $\parallel h_N^{(n)} - g \parallel_2 \to 0$ and so $n^{-1} \sum_k \bar{h}_{Nk}^2 \to \parallel g \parallel_2^2$. So if $0 < \parallel g \parallel_2^2$,

$$\mu = n\lambda + n\lambda^2 + n\lambda^2\nu^2 \parallel g \parallel_2^2 (1 + o(1)).$$

In the same way we obtain

$$\gamma = 1 + 2\lambda + 2\textstyle\sum_k (\lambda_k - \lambda)^2/N = 1 + 2\lambda + O(n^{-1/2})$$

and so

$$\operatorname{var}(Y_k^2 - \gamma Y_k) = 2\lambda_k^2 + \lambda_k(1 + 2\lambda_k - \gamma)^2 = 2\lambda^2(1 + o(1)),$$

and

$$\sigma^2 = 2n\lambda^2(1 + o(1)).$$

Thus (4.2) holds under H_{1N}. Further if we use the notation $E_0^a S_N$, $V_0^a S_N$ and $E_1^a S_N$, $V_1^a S_N$ for μ and σ^2 under H_0 and H_{1N}, then (2.1) and (2.3) hold with asymptotic moments given in (2.4), (2.5) and (2.7). Also (2.8) follows immediately.

For T_N we will restrict attention to the case $\lambda \to \infty$; the case $\lambda \to \lambda_0 \in (0, \infty)$ was given by Holst (1972). Here we use the inequality

$$\begin{aligned} 0 \le{} & (1+x)\log(1+x) \\ & - (x + \tfrac{1}{2}x^2 - \tfrac{1}{6}x^3 + \cdots - [2\ell(2\ell+1)]^{-1}x^{2\ell+1}) \le x^{2\ell+2}/(2\ell+1), \end{aligned} \tag{4.3}$$

$x > 1$, to write

$$\begin{aligned} Y_k \log Y_k ={} & \lambda_k \log \lambda_k + (Y_k - \lambda_k)(\log \lambda_k + 1) \\ & + \frac{(Y_k - \lambda_k)^2}{2\lambda_k} - \frac{(Y_k - \lambda_k)^3}{6\lambda_k^2} + \theta \frac{(Y_k - \lambda_k)^4}{3\lambda_k^3}, \end{aligned}$$

where $\theta = \theta(Y_k - \lambda_k) \in [0, 1]$. Then since $E(Y_k - \lambda_k)^j$ is a polynomial of degree $[j/2]$ in λ_k,

$$EY_k \log Y_k = \lambda_k \log \lambda_k + \tfrac{1}{2} + O(\lambda^{-1}), \quad \gamma = 1 + \log \lambda + O(\lambda^{-1})$$

and so

$$\begin{aligned} Y_k & \log Y_k - EY_k \log Y_k - \gamma(Y_k - \lambda_k) \\ & = (Y_k - \lambda_k)[\log(\lambda_k/\lambda) + O(\lambda^{-1})] \\ & \quad + \left[\frac{(Y_k - \lambda_k)^2}{2\lambda_k} - \tfrac{1}{2}\right] - \frac{(Y_k - \lambda_k)^3}{6\lambda_k^2} + \theta \frac{(Y_k - \lambda_k)^4}{3\lambda_k^3}. \end{aligned}$$

Thus

$$\begin{aligned} \operatorname{var}(Y_k \log Y_k - \gamma Y_k) & = \tfrac{1}{2} + \lambda_k[\log(\lambda_k/\lambda)]^2 + O(\log(\lambda_k/\lambda)) + O(\lambda^{-1}) \\ & = \tfrac{1}{2} + o(1), \end{aligned}$$

using (4.3) and the fact that $\nu |\bar{h}_{Nk}| = o(\lambda^{-1/2})$. Thus (4.2) holds under H_{1N} in this case. If we again use the notation $E_0^a T_N$, $V_0^a T_N$ and $E_1^a T_N$, $V_1^a T_N$ for μ and σ^2 under H_0 and H_{1N}, then (2.1) and (2.3) hold with these asymptotic moments, and to obtain (2.10) we need to consider the difference

$$\begin{aligned} E_1^a T_N - E_0^a T_N & = \textstyle\sum_k [E_1 Y_k \log Y_k - E_0 Y_k \log Y_k] \\ & = \textstyle\sum_k \sum_{j \ge 1} [j \log(j/\lambda_k) e^{-\lambda_k} \lambda_k^j - j \log(j/\lambda) e^{-\lambda} \lambda^j]/j! \\ & \quad + \textstyle\sum_k \lambda_k \log(\lambda_k/\lambda). \end{aligned} \tag{4.4}$$

Using a two-term Taylor expansion for the first term, Δ say, gives

$$\Delta = \tfrac{1}{2} \textstyle\sum_k (\lambda_k - \lambda)^2 [(d^2/d\lambda^2) \sum_{j \ge 1} j \log(j/\lambda) e^{-\lambda} \lambda^j / j!]_{\lambda = \lambda_k^*},$$

where $\lambda_k^* \in [\min(\lambda, \lambda_k), \max(\lambda, \lambda_k)]$. So

$$|\Delta| \le \tfrac{1}{2}\sum_k (\lambda_k - \lambda)^2\lambda_k^{*-2}$$
$$\cdot\, |EY_k^*[2(Y_k^* - \lambda_k^*) - 1 - (Y_k^* - \lambda_k^*)^2\log(Y_k^*/\lambda_k^*) + Y_k^*\log(Y_k^*/;_k^*)]|$$

where Y_k^* is Poisson with parameter λ_k^*. This gives

$$\Delta = O(\textstyle\sum_k (\lambda_k - \lambda)^2\lambda_k^{*-2}) = O(n^{1/2}\lambda^{-1}),$$

using (4.3) and the same arguments as above. Also the second term of (4.4) is

$$\textstyle\sum_k \lambda_k\log(\lambda_k/\lambda) = \tfrac{1}{2}n\lambda\nu^2\,\|g\|_2^2(1 + o(1))$$

from (4.3) again. So (2.10) follows.

5. Proofs of results of Section 3. First we need to obtain the expectations and variances of the statistics under H_0 and H_1. Formulae for moments of multinomials can be obtained from Kendall and Stuart (1969, pages 84 and 141). These can be used to show by direct calculation that

$$E\textstyle\sum_k X_k^2 = N + (1 - N^{-1})\sum_k \lambda_k^2$$
$$V\textstyle\sum_k X_k^2 = [2\sum_k \lambda_k^2 + 4\sum_k \lambda_k(\lambda_k - N^{-1}\sum_k \lambda_k^2)^2](1 + o(1))$$

and so to obtain (3.1) and (3.2). Also using these moments and the inequality (4.3), we can obtain

$$E\textstyle\sum_k X_k\log X_k = \sum_k \lambda_k\log\lambda_k + O(n)$$
$$V\textstyle\sum_k X_k\log X_k = [\tfrac{1}{2}n + \sum_k \lambda_k(\log\lambda_k - N^{-1}\sum_k \lambda_k\log\lambda_k)^2](1 + o(1)),$$

from which (3.3) and (3.4) follow.

To derive lower bounds on large deviation probabilities in both χ^2 and LR cases, we use a lemma which is related to results of Brockwell (1964) (see also Hald, 1981). For $0 < x < 1$, $0 < p < 1$, let

$$g(x, p) = x\log(x/p) + (1 - x)\log((1 - x)/(1 - p)).$$

LEMMA 1. *If Z is Binomial (N, p), $q = 1 - p$, $0 < p, u < 1$, then for Nu integral,*

$$P(Z \ge Nu) \ge P(Z = Nu) \ge 0\cdot 8(2\pi Nu(1 - u))^{-1/2}\exp\{-Ng(u, p)\}.$$

PROOF. The Lemma follows from three applications of Stirling's formula in the form

$$1 < n!(2\pi)^{-1/2}n^{-n-1/2}e^n < 1.087.$$

Chi-square. In the sequel C is a generic constant. We have

$$P(\textstyle\sum_k (X_k - \lambda)^2 \ge n\lambda + n\lambda^2\delta)$$
$$\ge P(\textstyle\sum_2^n ((X_k - \lambda)^2 - \lambda) \ge 0,\ (X_1 - \lambda)^2 - \lambda \ge n\lambda^2\delta)$$
$$\ge P(\textstyle\sum_2^n ((X_k - \lambda)^2 - \lambda) \ge 0 \mid X_1 = \ell)P(X_1 = \ell),$$

where $m^2 = n\lambda^2(\delta + 1/N)$ and $\ell = [\lambda + m] + 1$, $[x]$ denoting the integer part of x. From Lemma 1, for $n \to \infty$

$$P(X_1 = \ell) \geq C\ell^{-1/2}\exp(-Ng(\ell/N, n^{-1})) = \exp\{-\tfrac{1}{2}n^{1/2}\lambda\delta^{1/2}\log n(1 + o(1))\},$$

and if $X_2', \cdots, X_n'$ are mult$(N', (n-1)^{-1}, \cdots, (n-1)^{-1})$, $N' = N - \ell$,

$$\begin{aligned} P(\textstyle\sum_2^n ((X_k - \lambda)^2 - \lambda) \geq 0 \mid X_1 = \ell) \\ &= P(\textstyle\sum_2^n (X_k' - \lambda)^2 \geq (n-1)\lambda) \\ &\geq P(\textstyle\sum_2^n (X_k' - (N'/(n-1)))^2 \geq (n-1)\lambda) \\ &= P\left(\frac{\sum_2^n (X_k' - (N'/(n-1)))^2 - N'}{(2N'^2/(n-1))^{1/2}} \geq \frac{\ell - \lambda}{(2N'^2/(n-1))^{1/2}}\right). \end{aligned}$$

Now $N'^2/(n-1) = n\lambda^2(1 + o(1))$, so

$$\begin{aligned} P(\textstyle\sum_2^n ((X_k - \lambda)^2 - \lambda) \geq 0 \mid X_1 = \ell) \\ &\geq P\left(\frac{\sum_2^n (X_k' - (N'/(n-1)))^2 - N'}{(2N'^2/(n-1))^{1/2}} \geq \left(\frac{\delta}{2}\right)^{1/2} + o(1)\right) \\ &\geq \varepsilon > 0 \end{aligned}$$

from (2.1) with $U_N = S_N$ and $j = 0$. Thus

(5.1) $$P(\textstyle\sum_k (X_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \geq \exp\{-\tfrac{1}{2}n^{1/2}\lambda\delta^{1/2}\log n(1 + o(1))\}$$

for $n \to \infty$.

Likelihood ratio. We have

$$\begin{aligned} P(\textstyle\sum_k X_k\log X_k \geq n\lambda\log\lambda + n\lambda\delta') \\ &= P(\textstyle\sum_k X_k\lambda^{-1}\log(X_k\lambda^{-1}) \geq n\delta') \\ &\geq P(X_1\lambda^{-1}\log(X_1\lambda^{-1}) + (N - X_1)\lambda^{-1}\log((N - X_1)/(N - \lambda)) \geq n\delta'), \end{aligned}$$

since by Jensen's inequality

$$\begin{aligned} (n-1)^{-1} \textstyle\sum_{k=2}^n (X_k/\lambda)\log(X_k/\lambda) \\ &\geq (N - X_1)\lambda^{-1}(n-1)^{-1}\log((N - X_1)\lambda^{-1}(n-1)^{-1}). \end{aligned}$$

Thus

$$P(\textstyle\sum_k X_k\log X_k \geq n\lambda\log\lambda + n\lambda\delta') \geq P(g(X_1/N, 1/n) \geq \delta').$$

Now $g(x, p)$ is increasing for $p \leq x \leq 1$, so choosing $x_n > 1/n$ such that $g(x_n, 1/n) = \delta'$,

$$P(\textstyle\sum_k X_k\log X_k \geq n\lambda\log\lambda + n\lambda\delta') \geq P(X_1/N \geq x_n)$$

(5.2) $$\geq C(([Nx_n] + 1)(1 - ([Nx_n] + 1)/N))^{-1/2}\exp(-Ng(([Nx_n] + 1)/N, 1/n))$$

$$= \exp(-N\delta'(1 + o(1))).$$

Upper bound for χ^2. Let $m^2 = n\lambda^2\delta + \lambda$ as before. Then using a truncation argument of the sort introduced by Nagaev (1969)

$$(5.3)\qquad \begin{aligned} &P\{\textstyle\sum_k X_k^2 \geq n(\lambda + (1+\delta)\lambda^2)\} \\ &\qquad = P\{\textstyle\sum_k ((X_k - \lambda)^2 - \lambda) \geq n\lambda^2\delta\} \\ &\qquad \leq nP(X_1 - \lambda \geq m) + P(\textstyle\sum_k (X_k^* - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \end{aligned}$$

where $X_k^* - \lambda = (X_k - \lambda)I(X_k - \lambda < m)$. If Z is Binomial (N, p) then (Feller, 1968, page 151) for $k > Np$,

$$\begin{aligned} P(Z \geq k) &\leq k(1-p)(k - Np)^{-1}P(Z = k) \\ &\leq k(1-p)(k-Np)^{-1}(2\pi k(1 - k/N))^{-1/2}\exp\{-Ng(k/N, p)\} \end{aligned}$$

as in Lemma 1, so with $k = [m + \lambda]$, $p = n^{-1}$,

$$(5.4)\qquad \begin{aligned} P(X_1 - \lambda \geq m) &\leq C(m+\lambda)m^{-1}((m+\lambda)(1 - (m+\lambda)/N))^{-1/2} \\ &\qquad \cdot \exp\{-Ng([m+\lambda]/N, 1/n)\} \\ &= \exp\{-\tfrac{1}{2}m \log n(1 + o(1))\} \end{aligned}$$

as $n \to \infty$, $\lambda \geq \varepsilon > 0$.

Now $\mathscr{L}(X_1, \cdots, X_n) = \mathscr{L}(Y_1, \cdots, Y_n \mid \sum_k Y_k = n\lambda)$, where $\{Y_k\}$ are as in Section 4, so setting $Z_k - \lambda = (Y_k - \lambda)I(Y_k - \lambda < m)$,

$$(5.5)\qquad \begin{aligned} &P(\textstyle\sum_k (X_k^* - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \\ &\qquad = P(\textstyle\sum_k (Z_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta;\ \sum_k Y_k = n\lambda)/P(\sum_k Y_k = n\lambda) \\ &\qquad \leq P(\textstyle\sum_k (Z_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta)/P(\sum_k Y_k = n\lambda). \end{aligned}$$

For $t > 0$,

$$(5.6)\qquad \begin{aligned} &P(\textstyle\sum_k (Z_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \\ &\qquad \leq (E\exp(t(Z_1 - \lambda)^2))^n/\exp(t(n\lambda + n\lambda^2\delta)) \end{aligned}$$

and

$$(5.7)\qquad \begin{aligned} E\exp\{t(Z_1 - \lambda)^2\} &= \textstyle\sum_{0 \leq j < m+\lambda} \exp(t(j-\lambda)^2)p_j + \sum_{j \geq m+\lambda} p_j \\ &= S(0, m+\lambda) + R \end{aligned}$$

say, where $p_j = \lambda^j e^{-\lambda}/j!$, $j = 0, 1, \cdots$. It is easy to show (see the proof of Lemma 2 below) that

$$R \leq \exp(-\tfrac{1}{2}n^{1/2}\lambda\delta^{1/2}(1 + o(1))).$$

Next consider

$$(5.8)\qquad \begin{aligned} S(\lambda - (\theta/t)^{1/2}, \lambda + (\theta/t)^{1/2}) &\leq 1 + (1+\theta)\textstyle\sum_{|j-\lambda| \leq (\theta/t)^{1/2}} t(j-\lambda)^2 p_j \\ &\leq 1 + (1+\theta)\lambda t, \end{aligned}$$

since $e^x < 1 + (1+\theta)x$ for $0 < x < \theta < 1.5$. We will show below that the remainder of the sum in (5.7) is $o(\lambda^2 t)$ for a value of t chosen later. So

$$\begin{aligned} &P(\textstyle\sum_k (Z_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \\ (5.9) \qquad &< \{1 + (1+\theta)\lambda t(1 + o(\lambda))\}^n \exp(-t(n\lambda + n\lambda^2\delta)) \\ &< \exp\{-n\lambda t(\lambda\delta - \theta + o(\lambda))\}. \end{aligned}$$

Choosing $\theta < \min(\eta\lambda\delta, 1.5)$ for $\eta > 0$, it follows that this is less than $\exp\{-n\lambda^2 \cdot\delta(1-\eta)t(1+o(1))\}$. Taking $t = \varepsilon m^{-1}\log n$ for $0 < \varepsilon < \frac{1}{2}$, it follows that the left-hand side of (5.9) is less than $\exp\{-(1-\eta)\varepsilon n^{1/2}\lambda\delta^{1/2}\log n(1+o(1))\}$. This together with (5.3) and (5.4) yields

$$P(\textstyle\sum_k (X_k - \lambda)^2 \geq n\lambda + n\lambda^2\delta) \leq \exp\{-(1-\eta)\varepsilon n^{1/2}\lambda\delta^{1/2}\log n\,(1+o(1))\},$$

which taken in conjunction with (5.1) establishes Theorem 1, since $\eta > 0$ and $0 < \varepsilon < \frac{1}{2}$ are arbitrary.

It remains only to show that the terms of the sum in (5.7) not included in (5.8) are $o(n^{-1/2}\lambda\log n) = o(\lambda^2 t)$. The following lemma will be used in the sequel.

LEMMA 2. *If* $\lambda e^{tj_2}/j_1 < \frac{3}{4}$,

$$S(j_1, \lambda + j_2) < C\exp\{(j_1 - \lambda)(1 + tj_2) - j_1\log(j_1/\lambda)\}.$$

PROOF. The sum is dominated by

$$\lambda^{j_1}\exp\{-\lambda - \lambda tj_2 + tj_1j_2\}(j_1!)^{-1}\textstyle\sum_{j=0}^{\infty}(\lambda e^{tj_2}/j_1)^j$$

and the result follows by Stirling's formula.

Taking $j_1 = \frac{1}{2}m/\log n$ and $j_2 = m$ in Lemma 2, we obtain

$$\begin{aligned} &S(\tfrac{1}{2}m/\log n, m + \lambda) < C\exp\{\tfrac{1}{2}m(1 + \varepsilon\log n)/\log n - \tfrac{1}{4}m(1+o(1))\} \\ (5.10) \qquad &= C\exp\{\tfrac{1}{2}(\varepsilon - \tfrac{1}{2})m(1+o(1))\} \\ &= o(n^{-1/2}\lambda\log n) \end{aligned}$$

since $\varepsilon < \frac{1}{2}$. Now we need to consider two cases. First suppose $\lambda < \log n$, so that $(\theta/t)^{1/2} > \lambda$ for large n, and

$$\begin{aligned} &S(\lambda + (\theta/t)^{1/2}, \tfrac{1}{2}m/\log n) \\ (5.11) \qquad &< C\exp\{(\theta/t)^{1/2}(1 + \tfrac{1}{2}\varepsilon) - (\theta/t)^{1/2}\log(t^{-1/2}\lambda^{-1})(1+o(1))\} \\ &= o(n^{-1/2}\lambda\log n) \end{aligned}$$

on putting $j_2 = \frac{1}{2}m/\log n - \lambda$ and $j_1 = \lambda + (\theta/t)^{1/2}$ in Lemma 2. For $\lambda < \log n$, (5.9) follows from (5.8), (5.10) and (5.11). Now suppose $\lambda \geq \log n$. For λ large enough, we need to consider $S(0, \lambda - (\theta/t)^{1/2})$, $S(\lambda + (\theta/t)^{1/2}, 3\lambda)$ and $S(3\lambda, \frac{1}{2}m/\log n)$. From Lemma 2 with $j_2 = \frac{1}{2}m/\log n - \lambda$ and $j_1 = 3\lambda$,

$$\begin{aligned} &S(3\lambda, \tfrac{1}{2}m/\log n) < C\exp\{3\lambda(\tfrac{1}{2}\varepsilon + 1 - \log 3) - \lambda(1 + \tfrac{1}{2}\varepsilon)\} \\ (5.12) \qquad &< C\exp\{-3\lambda/4\} \\ &= o(n^{-1/2}\lambda\log n). \end{aligned}$$

Further if $1 \le j \le 3\lambda$, from the left-hand inequality of (4.3),

$$\exp\{t(j-\lambda)^2\}p_j < C\exp\{\lambda t(j-\lambda)^2/\lambda - (j-\lambda)^2/(6\lambda)\}$$
$$= C\exp\{-((j-\lambda)^2/(6\lambda))(1+o(1))\}$$

so

$$S(0, \lambda - (\theta/t)^{1/2}) < e^{-1/2\lambda} + C\textstyle\sum_{1\le j<\lambda-(\theta/t)^{1/2}} \exp\{-(j-\lambda)^2/(7\lambda)\}$$

$$(5.13)\qquad \begin{aligned} &< C\int_{-\infty}^{\lambda-(\theta/t)^{1/2}} \exp\{-(x-\lambda)^2/(7\lambda)\}\,dx \\ &= O(\lambda^{1/2}\exp\{-(\theta/7)(\lambda t)^{-1}\}) \\ &= o(n^{-1/2}\lambda\log n). \end{aligned}$$

Similarly,

$$(5.14)\qquad S(\lambda + (\theta/t)^{1/2}, 3\lambda) = o(n^{-1/2}\lambda\log n).$$

For $\lambda \ge \log n$, (5.9) follows from (5.8), (5.10), (5.12), (5.13) and (5.14).

Upper bound for likelihood ratio. Let $\{Y_k\}$ be as above. Then

$$P(\textstyle\sum_k X_k\log X_k \ge n(\lambda\log\lambda + \lambda\delta'))$$

$$(5.15)\qquad \begin{aligned} &= P(\textstyle\sum_k (X_k\log(X_k/\lambda) - (X_k-\lambda)) \ge n\lambda\delta') \\ &\le \frac{P(\sum_k (Y_k\log(Y_k/\lambda) - (Y_k-\lambda)) \ge n\lambda\delta')}{P(\sum_k Y_k = n\lambda)} \\ &\le C(n\lambda)^{1/2}P(\textstyle\sum_k (Y_k\log(Y_k/\lambda) - (Y_k-\lambda)) \ge n\lambda\delta') \\ &\le \frac{C(n\lambda)^{1/2}\prod_{k=1}^{n} E\exp\{t(Y_k\log(Y_k/\lambda) - (Y_k-\lambda))\}}{\exp\{tn\lambda\delta'\}}. \end{aligned}$$

Now

$$E\exp\{t(Y_k\log(Y_k/\lambda) - (Y_k-\lambda)\} = \textstyle\sum_{j=0}^{\infty} p_j\exp\{tj\log(j/\lambda) - t(j-\lambda)\}$$

with p_j as above. Using Stirling's formula and the inequality (4.3), we have for $j \ge 1$,

$$\exp\{tj\log(j/\lambda) - t(j-\lambda)\}p_j$$
$$\le C\exp\{-\tfrac{1}{2}(1-t)(j-\lambda)^2(1-\tfrac{1}{3}(j-\lambda)/\lambda)/\lambda - \tfrac{1}{2}\log j\}.$$

So

$$\textstyle\sum_{0\le j<3\lambda} \exp(t(j\log(j/\lambda) - (j-\lambda)))p_j$$
$$\le C\int_{-\infty}^{\infty} \exp(-\tfrac{1}{3}(1-t)(x-\lambda)^2/\lambda)\,dx \le C\lambda^{1/2}/(1-t)^{1/2} = C\lambda$$

for $t = 1 - \lambda^{-1}$. Also in this case

$$\sum_{j\geq 3\lambda} \exp(t(j\log(j/\lambda) - (j-\lambda)))p_j < C\sum_{j\geq 3\lambda} \exp((t-1)(j\log(j/\lambda) - (j-\lambda))) < C\sum_{j\geq 3\lambda} \exp(-\lambda^{-1}j(\log 3 - 1)) < C\lambda$$

and so from (5.15)

$$P(\textstyle\sum_k X_k \log X_k \geq n(\lambda\log\lambda + \lambda\delta')) < C\exp\{n(\log\lambda + C - \lambda\delta')(1 + o(1))\} = \exp\{-N\delta'(1 + o(1))\}.$$

This and (5.2) imply Theorem 2.

Finally, (3.9) is proved using the argument at (5.15) with $n(\lambda\log\lambda + \lambda\delta')$ replaced by $E_1^a T_N$.

Acknowledgement. We wish to thank a referee whose detailed comments finally dispelled a number of misconceptions and led to significant improvements.

REFERENCES

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia.

BROCKWELL, P. J. (1964). An asymptotic expansion for the tail of a binomial distribution. *J. Appl. Probab.* **1** 166–169.

DAHIYA, R. C. and GURLAND, J. (1973). How many classes in the Pearson chi-square test? *J. Amer. Statist. Assoc.* **68** 707–712.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* 1. Wiley, New York.

GVANCELADZE, L. G. and CHIBISOV, D. M. (1979). On tests of fit based on grouped data. In *Contributions to Statistics, J. Hájek Memorial Volume.* J. Jurecková, ed. 79–89, Academia, Prague.

HALD, A. (1981). *Statistical Theory of Sampling Inspection by Attributes.* Academic, London.

HAMDAN, M. A. (1963). The number and width of classes in the chi-square test. *J. Amer. Statist. Assoc.* **58** 678–689.

HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

HOLST, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* **59** 137–145.

KEMPTHORNE, O. (1967). The classical problem of inference—goodness of fit. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 235–249. Univ. California Press.

KENDALL, M. G. and STUART, A. (1969). *The Advanced Theory of Statistics. Vol. 1: Distribution Theory,* 3rd ed. Griffin, London.

KENDALL, M. G. and STUART, A. (1973). *The Advanced Theory of Statistics. Vol. 2: Inference and Relationship,* 3rd ed. Griffin, London.

MANN, H. B. and WALD, A. (1942). On the choice of the number of intervals in the application of the chi-square test. *Ann. Math. Statist.* **13** 306–317.

MORRIS, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188.

NAGAEV, A. V. (1969). Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I. *Theory Probab. Appl.* **14** 51–64.

QUINE, M. P. and ROBINSON, J. (1984). Normal approximations to sums of scores based on occupancy numbers. *Ann. Probab.* **13** 794–804.

SENETA, E. (1976). Regularly varying functions. *Lecture Notes in Math.* **508** Springer-Verlag, Berlin.

DEPT. OF MATHEMATICAL STATISTICS
UNIVERSITY OF SYDNEY
NSW 2006, AUSTRALIA