The Annals of Statistics
1985, Vol. 13, No. 4, 1554–1580

ON MODERATE AND LARGE DEVIATIONS IN MULTINOMIAL DISTRIBUTIONS

BY WILBERT C. M. KALLENBERG

Free University

In this paper moderate and large deviation theorems are presented for the likelihood ratio statistic and Pearson's chi squared statistic in multinomial distributions. Let k be the number of parameters and n the number of observations. Moderate and large deviation theorems are available in the literature only if k is kept fixed when $n \to \infty$. Although here attention is focussed on $k = k(n) \to \infty$ as $n \to \infty$, explicit inequalities are obtained for both k and n fixed. These inequalities imply results for the whole scope of moderate and large deviations both for fixed k and for $k(n) \to \infty$ as $n \to \infty$. It turns out that the χ^2 approximation continues to hold in some sense, even if $k \to \infty$. The results are applied in studying the influence of the choice of the number of classes on the power in goodness-of-fit tests, including a comparison of Pearson's chi squared test and the likelihood ratio test. Also the question of combining cells in a contingency table is discussed.

1. Introduction. Let the random k-dimensional vector Y_n have a k-dimensional multinomial distribution with parameters n and $p = (p_1, \ldots, p_k)$, i.e.,

$$P_p(Y_n = y) = \frac{n!}{y_1! \cdots y_k!} \prod_{j=1}^{k} p_j^{y_j}, \tag{1.1}$$

where $y = (y_1, \ldots, y_k)$ has nonnegative integer components with sum n, and where p is any point in the simplex

$$S_k = \left\{(z_1, \ldots, z_k)\colon \sum_{j=1}^{k} z_j = 1,\ z_j \geq 0 \text{ for } j = 1, \ldots, k\right\}. \tag{1.2}$$

Hoeffding (1965a) has studied asymptotic properties of tests of simple and composite hypotheses concerning p, when the size α_n of the test tends to zero as $n \to \infty$ and k is kept fixed. The likelihood ratio (LR) statistic and Pearson's chi squared (χ^2) statistic are well known test statistics for this kind of testing problems. It is shown by Tumanyan (1954), Steck (1957), and Morris (1975) that under some conditions these statistics are asymptotically normal as $k \to \infty$ (and $n \to \infty$). However, to attack testing problems with level α_n tending to zero and $k = k(n) \to \infty$ as $n \to \infty$, these results are inadequate, since Hoeffding's paper deals with fixed k, and on the other hand, asymptotic normality can only be

Received February 1984; revised April 1985.
AMS 1980 *subject classifications.* Primary 60F10; secondary 62E20, 62E15, 62F20.
Key words and phrases. Moderate and large deviations in multinomial distributions, likelihood ratio statistic for the multinomial distribution, multinomial distribution, Pearson's chi squared statistic, contingency table.

applied if α is kept fixed. If the level α_n tends to zero, the theory of moderate and large deviations is needed.

It is the purpose of this paper to present moderate and large deviation theorems for the LR statistic, especially if $k = k(n) \to \infty$ as $n \to \infty$. The whole range of probabilities tending to zero very slowly up to the exponentially fast rate of convergence to zero is considered. By the relation between the LR statistic and Pearson's χ^2 statistic, results are also obtained for the latter one.

In fact, the asymptotics are derived from quite explicit inequalities with both n and k fixed. Although in this paper attention is focussed on $k \to \infty$, these inequalities also imply some improvements for fixed k. A more detailed comparison with Hoeffding's results is given in Section 2.

The χ^2 statistic and LR statistic for the multinomial distribution can be used to test goodness-of-fit. One of the major problems in such a case is the choice of k, the number of classes.

To attack this problem several asymptotic ($n \to \infty$) approaches can be made. In Kallenberg et al. (1985) a local approach associated with a fixed α is employed. Here we discuss the problem from a nonlocal point of view. There is a fairly good agreement between local and nonlocal theory. However, for some situations the conclusions are different. Neither of the asymptotic theories is definite. A lot of Monte Carlo studies shows that local theory is well reflected for moderate sample size in many examples. Here some Monte Carlo results are presented in situations where local and nonlocal theory lead to opposite conclusions. As is seen from these examples, nonlocal theory may be a better predictor of power behaviour.

As a second application χ^2 and LR tests are compared. For $k \to \infty$ the LR test is far more efficient in the sense of Bahadur than the χ^2 test. Moreover, for $k \to \infty$ the LR test attains the optimal Bahadur slope even within the class of tests based on nondiscretized observations! This surprising first-order optimality property asks for second-order investigations. The present large deviation results enable us to show that for the LR test based on discretized observations the well-known phenomena of first-order efficiency implies second-order efficiency does *not* come true! This relativizes the first-order optimality.

The third application concerns the question of combining cells in a large contingency table when testing independence. Collapsing cells such that the number of categories does not tend to infinity leads to a loss of efficiency in the sense of Bahadur.

Theorems on the multinomial distribution with $k = k(n) \to \infty$ as $n \to \infty$ differ in two ways from those with k fixed: (i) the dimension of the parameter space tends to infinity and so the number of parameters is growing with n; (ii) parameter points are not supposed to stay away from the boundary of the parameter space; for instance, $\min\{p_i\colon 1 \le i \le k\} \le k^{-1} \to 0$ as $k \to \infty$. For both reasons standard moderate and large deviation theory cannot be applied.

Both aspects are of interest. A growing number of parameters opens the possibility for closer approximations; if one uses a discretization, one can apply with a larger number of observations a finer discretization. The second aspect leads to investigations, also for fixed k, where small values of the parameters are involved, thus extending Hoeffding's results. Our proofs are based on the following approach. First, general inequalities are derived for the binomial distribution, i.e., $k = 2$ (Proposition 2.2). Then the proof is by induction on k, using the fact that the conditional distribution of Y_n given one component is again multinomial, but with k replaced by $k - 1$. Further the ideas of exponential centering (saddle-point method) are employed.

The results are precisely formulated in Section 2, applications are presented in Section 3, while proofs are given in Section 4.

2. Preliminaries and results. A crucial role in large and moderate deviation theory is played by the Kullback–Leibler information number. In the present case it is defined by

$$I_k(q, p) = \sum_{i=1}^{k} q_i \log(q_i/p_i), \qquad p, q \in S_k, \tag{2.1}$$

with the convention $r \log(r/s) = 0$ if $r = 0$. It is useful to think of it as a kind of "statistical distance" between the distribution of Y_1 under q and the distribution of Y_1 under p, although symmetry and the triangle inequality generally do not hold.

The Kullback–Leibler information number appears also in the form of the LR statistic: the LR test based on Y_n for the simple hypothesis $\{p\}$ against the alternative hypothesis $S_k - \{p\}$ rejects the null hypothesis for large values of $I_k(\overline{Y}_n, p)$, possibly with randomization on the set where the statistic assumes its critical value. [We denote by $\overline{Y}_n$ the vector $(\overline{Y}_{n1}, \dots, \overline{Y}_{nk}) = (n^{-1}Y_{n1}, \dots, n^{-1}Y_{nk})$.]

Therefore the LR statistic plays a natural role in the development of moderate and large deviation theorems for multinomial distributions. It is well known that both the LR statistic and Pearson's χ^2 statistic have an asymptotic χ^2 distribution with $(k - 1)$ degrees of freedom when k is kept fixed and $n \to \infty$. One might ask whether or not this χ^2 approximation continues to hold in the tails of the distribution, maybe even if $k \to \infty$. So it is of interest to derive an expression for probabilities of moderate and large deviations for the χ^2 distribution itself. This can be done by direct calculation, even if $k \to \infty$. Let χ_k^2 have a χ^2 distribution with k degrees of freedom. If $2nd_n \geq k$ then

$$\log \Pr\left(\chi_k^2 \geq 2nd_n\right) = -nd_n + \tfrac{1}{2}(k-2)\log(2end_n/k) + O(\log k) \tag{2.2}$$

as $n \to \infty$, irrespective of whether $k \to \infty$ or not.

By Theorem 3 of Hoeffding (1965b) it follows that (2.2) indeed holds if χ_{k-1}^2 is replaced by $2nI_k(\overline{Y}_n, p)$ provided that (i) k is fixed and (ii) $d_n < -\log(1 - \min\{p_i: 1 \leq i \leq k\}) - \beta$ for some $\beta > 0$. The last condition serves to avoid difficulties that may arise at the boundary of the parameter space (cf. Section 1 and the discussion below Proposition 2.2). Since $\min\{p_i: 1 \leq i \leq k\} \to 0$ if $k \to \infty$, it is clear that both conditions do not hold if $k \to \infty$. However, it is yet possible to derive a moderate and large deviation theorem for the LR statistic when $k \to \infty$, which has the same flavour as (2.2) (cf. also Corollary 2.5). In other words, the χ^2 approximation continues to hold in some sense, even if $k \to \infty$ and moderate and large deviation probabilities are under consideration.

THEOREM 2.1. *If $p_i \geq a/k$ $(i = 1,\dots,k)$ for some $a > 0$, and if $0 < d_n < 0.15$ and $2nd_n \geq k$, then for all $k = 2,3,\dots$, $n = 1,2,\dots$*

$$P_p\left(I_k(\overline{Y}_n, p) \geq d_n\right) \leq 2\left(\frac{2end_n}{k}\right)^{(k-2)/2} e^{-nd_n}\left(1 + 400a^{-1/2}d_n^{1/2}\right)^{2(k-2)} \cdot \prod_{i=3}^{k}\left(1 + \frac{10}{i}\right), \tag{2.3}$$

and if moreover $d_n < (a/125)^2$, then for all $k = 2,3,\dots$, $n = 1,2,\dots$

$$P_p\left(I_k(\overline{Y}_n, p) \geq d_n\right) \geq ca\left(\frac{2end_n}{k}\right)^{(k-3)/2} e^{-nd_n}\left(1 - 125a^{-1}d_n^{1/2}\right)^{k-2}\prod_{i=3}^{k}\left(1 - \frac{2}{i}\right) \tag{2.4}$$

for some constant $c > 0$.

It has to be noted that the bounds are quite explicit [the constant $c > 0$ in (2.4) goes back to the binomial case (cf. Proposition 2.2)]. Replacing $a^{-1/2}$ in (2.3) by $[k\min\{p_i\colon 1 \leq i \leq k\}]^{-1/2}$, there is no condition on p for (2.3) to hold (if $p_i = 0$ for some i we can return to a lower-dimensional multinomial distribution). This is useful when testing a composite null hypothesis (cf. Remark 2.1). Further, note that the condition

$$p_i \geq a/k, \qquad i = 1,\dots,k \quad \text{for some } a > 0 \tag{2.5}$$

stated in Theorem 2.1 and also in the following theorem and corollaries, does *not* imply that all p_is are of the same order of magnitude. For instance, if $p_i = (2k)^{-1}$ $(i = 1,\dots,k-1)$ and $p_k = \frac{1}{2} + (2k)^{-1}$, then (2.5) holds with $a = \frac{1}{2}$, while $p_k/p_1 \to \infty$ as $k \to \infty$. On the other hand, the bounds are not intended as direct numerical approximations for the involved probabilities, in the same way as Berry–Esseen bounds do not claim to yield sharp numerical bounds. They are useful because they have the right order of magnitude.

It is even possible to derive an upper bound which does not depend on p. Following the same method as in Kallenberg [(1978), Section 3.7], one can show that if $0 < d_n \leq 0.1$ and $nd_n \geq 1$, then for all $k \geq 2$

$$P_p\left(I_k(\overline{Y}_n, p) \geq d_n\right) \leq 4^{k-1}(nd_n)^{(k-2)/2}e^{-nd_n}. \tag{2.6}$$

The upper bound in (2.6), however, does not agree with (2.2) if $k \to \infty$.

The power of nd_n in the lower bound (2.4) equals $\frac{1}{2}(k-3)$. This corresponds to Hoeffding's (1965b) Theorem 3. The power of nd_n in the upper bound is $\frac{1}{2}$ higher. This can be explained by considering the special case $k = 2$—the binomial distribution.

PROPOSITION 2.2. *For all $n = 1,2,\dots$, $d_n > 0$ and all $p \in S_2$ we have*

$$P_p\left(I_2(\overline{Y}_n, p) \geq d_n\right) \leq 2e^{-nd_n}. \tag{2.7}$$

For each $\varepsilon > 0$ *there exists a constant* $c = c(\varepsilon) > 0$ *such that*

$$P_p(I_2(\bar{Y}_n, p) \geq d_n) \geq cp_1(1 - p_1)e^{-nd_n}(nd_n)^{-1/2} \tag{2.8}$$

for all $p \in S_2$, $n \in \mathbb{N}$, $\varepsilon n^{-1} \leq d_n \leq \log 2 - \varepsilon$.

We make some remarks on (2.7) and (2.8). First, we give a simple example showing that for a particular choice of p and d_n the upper bound in (2.7) is attained for each n.

EXAMPLE 2.1. Let $p = (\frac{1}{2}, \frac{1}{2})$ and $d_n = \log 2$, then $P_p(I_2(\bar{Y}_n, p) \geq d_n) = P_p(\bar{Y}_n = (1,0)) + P_p(\bar{Y}_n = (0,1)) = 2 \cdot (\frac{1}{2})^n = 2 \cdot \exp(-nd_n)$.

One might guess that the upper bound in (2.7) is not sharp for moderate deviations ($d_n \to 0$). The next example shows that the order of magnitude in (2.7) is the right one.

EXAMPLE 2.2. Let $p = p(n) = (p_1(n), p_2(n))$ with $\lim_{n\to\infty} p_1(n) = 0$, and let $d_n = I_2((0,1), p(n)) = -\log(1 - p_1(n))$, then $\lim_{n\to\infty} d_n = 0$ and $P_p(I_2(\bar{Y}_n, p) \geq d_n) > P_p(\bar{Y}_n = (0,1)) = \{1 - p_1(n)\}^n = \exp(-nd_n)$.

Hoeffding's condition $d_n < -\log(1 - \min\{p_i: 1 \leq i \leq k\}) - \beta$ for some $\beta > 0$ ensures that no difficulties arise with respect to the boundary of the parameter space. There may occur two types of troubles: the first one concerns the fact that the boundary of the critical region of the LR test overlaps the boundary of the parameter space. Examples 2.1 and 2.2 are typical for this situation. The price we have to pay for considering also this type of critical regions is the missing factor $(nd_n)^{-1/2}$ in (2.7) as compared with Hoeffding's result. The second difficulty which may arise, concerns the fact that p itself may be very near to the boundary of the parameter space.

In (2.8) we see which price we have to pay for considering also this type of parameter points—a factor $p_1(1 - p_1)$ is inserted at the right-hand side of (2.8). The following example illustrates the dependence on p_1.

EXAMPLE 2.3. Let $n = 1$, $p_1 < e^{-1}$, and $d_n = \frac{1}{2}$, then $P_p(I_2(\bar{Y}_n, p) \geq d_n) = P_p(Y_1 = (1,0)) = p_1$.

One might guess that for moderate deviations ($d_n \to 0$) the lower bound can be expressed as a function of nd_n alone, independent of p_1. The next example shows that this is not true (cf. Remark 4.1).

EXAMPLE 2.4. Let $p = p(n) = (p_1(n), p_2(n))$. Define $u_n = (a_n + \eta_n)n^{-1}$ with $a_n \in \mathbb{N}$, $a_n n^{-1} p_1^{-1}(n) \to \infty$, $a_n n^{-1} \log\{a_n n^{-1} p_1^{-1}(n)\} \to 0$, and $\eta_n \log\{a_n n^{-1} p_1^{-1}(n)\} \to 0$ as $n \to \infty$. Denoting $d_n = I_2((u_n, 1 - u_n), p(n))$, we have $\lim_{n\to\infty} d_n = 0$. Moreover, $I_2((0,1), p(n)) < d_n$ for sufficiently large n and hence, writing $t_n = (a_n + 1)n^{-1}$, $d_n^* = I_2((t_n, 1 - t_n), p(n))$ we have in view of (2.7)

$$P_p(I_2(\overline{Y}_n, p) \geq d_n) = P_p(\overline{Y}_{n1} \geq t_n) = P_p(I_2(\overline{Y}_n, p) \geq d_n^*) \leq 2e^{-nd_n^*}.$$

Choosing for instance $p_1(n) = n^{-1}$ and $a_n \in \mathbb{N}$, $a_n \to \infty$, $a_n n^{-1} \log a_n \to 0$ as $n \to \infty$, or $a_n = 1$, and $p_1(n) > n^{-2}$ such that $np_1(n) \to 0$ as $n \to \infty$, it is easily seen that the above conditions are satisfied and

$$e^{-nd_n^*} e^{nd_n} (nd_n)^{1/2} \to 0 \quad \text{as } n \to \infty,$$

indicating that we need some extra factor depending on p_1 in (2.8) to obtain *uniformity* in p.

In order to obtain inequalities for all $p \in S_2$, even if p is close to the boundary of the parameter space (and we need that), we have to pay for this uniformity in p here. But as is seen from Theorem 2.1 if we deal with dimension k, we are missing only *one* factor $(nd_n)^{-1/2}$ and *one* factor a [corresponding to $p_1(1 - p_1)$] and not such a factor to the power $k - 1$. Moreover, if k is fixed and Hoeffding's condition holds, then it is easy to show by our method of proof that in (2.7) and hence in (2.3) an upper bound can be obtained with the power of nd_n one-half lower.

For Pearson's chi squared statistic

$$Q_k^2(\overline{Y}_n, p) = \sum_{i=1}^{k} (\overline{Y}_{ni} - p_i)^2 / p_i,$$

we have the following result.

THEOREM 2.3. *If $p_i \geq a/k$ $(i = 1, \dots, k)$ for some $a > 0$ and if $0 < d_n(1 + u_n + v_n) < 0.15$ and $2nd_n(1 - u_n) \geq k$ with $u_n = \frac{1}{3}(2d_n a^{-1} k)^{1/2}$ and $v_n = \frac{4}{3} d_n a^{-1} k$, then for all $k = 2, 3, \dots$ and $n = 1, 2, \dots$ we have*

$$P_p(Q_k^2(\overline{Y}_n, p) \geq 2d_n)$$

$$\leq 2\left(\frac{2end_n}{k}\right)^{(k-2)/2} \exp\left[-nd_n + u_n\left(nd_n - \frac{k-2}{2}\right)\right] \cdot \left(1 + 400a^{-1/2}d_n^{1/2}\right)^{2(k-2)} \prod_{i=3}^{k}\left(1 + \frac{10}{i}\right), \tag{2.9}$$

and if moreover $8d_n < (a/125)^2$, then for all $k = 2, 3, \dots$, $n = 1, 2, \dots$

$$P_p(Q_k^2(\overline{Y}_n, p) \geq 2d_n) \geq ca\left(\frac{2end_n}{k}\right)^{(k-3)/2} \cdot \exp\left[-nd_n - (u_n + v_n)\left(nd_n - \frac{k-3}{2}\right)\right] \cdot \left(1 - 125a^{-1}(8d_n)^{1/2}\right)^{k-2} \exp\left[-\left(\frac{k-3}{4}\right)49u_n^2\right] \prod_{i=3}^{k}\left(1 - \frac{2}{i}\right) \tag{2.10}$$

for some constant $c > 0$.

Theorem 2.3 is of interest if $d_n k$ is small. This corresponds to the fact that the distribution of Pearson's χ^2 statistic and the LR statistic looks similar as long as small and moderate deviations are considered. However, they behave differently for large deviations. When k is fixed this is shown by Hoeffding [see, for instance, Hoeffding (1965b), page 218]; the case $k \to \infty$ is extensively discussed in Section 3.

Next we present three corollaries, which make applications easier. We start with a Chernoff-type large deviation result, i.e., exponentially fast convergence to zero.

COROLLARY 2.4. (Chernoff-type large deviation; $d_n = d$ fixed). *Let $n \to \infty$ and $k = k(n) = o(n)$ as $n \to \infty$, and let $p_i \geq a/k$ $(i = 1, \dots, k)$ for some $a > 0$. If $d_n = d < (a/125)^2$, then*

$$\lim_{n \to \infty} -n^{-1} \log P_p(I_k(\overline{Y}_n, p) \geq d) = d. \tag{2.11}$$

REMARK 2.1. The right-hand sides of (2.3) and (2.4) do not depend on p (only on a). Therefore, if $k = o(n)$ as $n \to \infty$ and $d_n = d < (a/125)^2$, then

$$\lim_{n \to \infty} -n^{-1} \log \sup_{p \in \mathscr{P}_0(a)} P_p(I_k(\overline{Y}_n, p) \geq d) = d,$$

where

$$\mathscr{P}_0(a) = \{p\colon p_i \geq a/k,\ i = 1, \dots, k\}.$$

This result is useful in testing a composite null hypothesis (cf. Section 3). Note that the situation where k is fixed and $p \in \operatorname{int} S_k$ is fixed, is covered by the above corollary. It may be used e.g., in the computation of Bahadur efficiency (cf. Section 3).

Now we turn to the case $d_n \to 0$. The following corollary covers the whole scope of moderate and large deviations, which do not lead to exponentially fast convergence to zero. Since $2nI_k(\overline{Y}_n, p)$ is asymptotically χ^2_{k-1}-distributed as $n \to \infty$ for fixed k and since the χ^2_k distribution is asymptotically normal as $k \to \infty$ (after suitable standardization), it may be expected that if $k \to \infty$ and the deviation is not too large, normal tail behaviour appears. That this is true, may be seen in (2.13). On the other hand, if we are further in the tail the typical large deviation feature appears (domination of the Kullback–Leibler number d_n [cf. (2.15)]).

COROLLARY 2.5. (Moderate and large deviation; $d_n \to 0$). *Let $n \to \infty$ and $k = k(n) = o(n)$ as $n \to \infty$, and let $p_i \geq a/k$ $(i = 1, \dots, k)$ for some $a > 0$. If $\lim_{n \to \infty} d_n = 0$ and $2nd_n \geq k$ then*

$$\log P_p(I_k(\overline{Y}_n, p) \geq d_n) = \left(\frac{k-3}{2}\right) \log\left(\frac{2end_n}{k}\right) - nd_n + O\left(\log(nd_n) + kd_n^{1/2}\right). \tag{2.12}$$

In particular, if $x_n \to \infty$, $x_n = o(\sqrt{k})$ *then*

$$\log P_p\big(2nI_k(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}x_n^2 + O\big(x_n^3 k^{-1/2} + \log k + k^{3/2}n^{-1/2}\big); \tag{2.13}$$

if $x_n = c\sqrt{k/2}$ *for some* $c > 0$ *then*

$$\log P_p\big(2nI_k(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}k\{c - \log(1 + c)\} + O\big(\log k + k^{3/2}n^{-1/2}\big); \tag{2.14}$$

if $x_n/\sqrt{k} \to \infty$ *and* $x_n k^{1/2} n^{-1} \to 0$ *then*

$$\log P_p\big(2nI_k(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}x_n\sqrt{2k}\,(1 + o(1)). \tag{2.15}$$

For Pearson's χ^2 statistic we obtain

COROLLARY 2.6. *Let* $n \to \infty$ *and* $k = k(n) = o(n^{1/2})$ *as* $n \to \infty$, *and let* $p_i \geq a/k$ $(i = 1, \dots, k)$ *for some* $a > 0$. *If* $\lim_{n\to\infty} kd_n = 0$ *and* $2nd_n\{1 - \frac{1}{3}(2d_n a^{-1}k)^{1/2}\} \geq k$ *then*

$$\begin{aligned}\log P_p\big(Q_k^2(\bar{Y}_n, p) \geq 2d_n\big) &= \left(\frac{k-3}{2}\right)\log\left(\frac{2end_n}{k}\right) - nd_n \\ &\quad + O\big(\log(nd_n) + kd_n^{1/2} + (kd_n)^{1/2}(2nd_n - k)\big).\end{aligned} \tag{2.16}$$

In particular, if $x_n \to \infty$, $x_n = o(\sqrt{k})$ *and* $n^{1/2}x_n k^{-3/2} \to \infty$ *then*

$$\log P\big(nQ_k^2(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}x_n^2 + O\big(x_n^3 k^{-1/2} + \log k + x_n k^{3/2} n^{-1/2}\big); \tag{2.17}$$

if $x_n = c\sqrt{k/2}$ *for some* $c > 0$ *then*

$$\log P_p\big(nQ_k^2(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}k\{c - \log(1 + c)\} + O\big(\log k + k^2 n^{-1/2}\big); \tag{2.18}$$

if $x_n/\sqrt{k} \to \infty$ *and* $x_n k^{3/2}/n \to 0$ *then*

$$\log P_p\big(nQ_k^2(\bar{Y}_n, p) \geq k + x_n\sqrt{2k}\big) = -\tfrac{1}{2}x_n\sqrt{2k}\,(1 + o(1)). \tag{2.19}$$

REMARK 2.2. It is well known that Hoeffding's result for the multinomial distribution can be generalized to k-parameter exponential families with fixed k [cf. Efron and Truax (1968), Kallenberg (1981)]. One may ask whether Theorem 2.1 can be generalized to exponential families. It might be of interest to consider such exponential families with dimension $k = k(n)$ growing with n. For instance, if we have a sufficiently smooth density $f(x; \theta)$, we may write

$$\begin{aligned}f(x;\theta) &= \exp\{\log f(x;\theta)\} \\ &\approx \exp\left\{\log f(x;\theta_0) + \sum_{j=1}^{k} \frac{(\theta-\theta_0)^j}{j!}\frac{\partial^j}{\partial\theta^j}\log f(x;\theta)\bigg|_{\theta=\theta_0}\right\},\end{aligned}$$

thus obtaining a curved exponential family as approximation to the density itself. By choosing $k = k(n)$ larger a better approximation may be expected.

Whether Theorem 2.1 holds for general exponential families is an open question. If $Y_1, \dots, Y_n$ are i.i.d. normal $N_k(\mu; I)$ with $\mu \in \mathbb{R}^k$ and I the $k \times k$ identity matrix, then the Kullback–Leibler information number $I_k(x, y)$ equals $\frac{1}{2}$ times the square of the Euclidean distance $\|x - y\|$ and hence

$$\begin{aligned} P_\mu\big(I_k(\overline{Y}_n, \mu) \geq d_n\big) &= P_\mu\big(\tfrac{1}{2}\|\overline{Y}_n - \mu\|^2 \geq d_n\big) \\ &= P_0\big(n\|\overline{Y}_n\|^2 \geq 2nd_n\big) \\ &= \Pr\big(\chi_k^2 \geq 2nd_n\big) \end{aligned}$$

and so we can apply (2.2).

3. Applications. In this section some statistical applications of the theorems of Section 2 are presented. Especially, we discuss the choice of the number of classes in goodness of fit tests, Bahadur efficiency and deficiency of LR and χ^2 tests, and the question of combining cells in a large contingency table.

Let $Z_1, \dots, Z_n$ be i.i.d. real-valued random variables with an absolutely continuous distribution. To test the simple hypothesis H_0 that the Z_is have given density h, we consider the classical Pearson χ^2 test. Let the range of the Z_is be divided into k disjoint intervals $A_1, \dots, A_k$, let $Y_{ni} = \#\{Z_j \in A_i\}$, and put $p_i = P_0(Z_1 \in A_i)$, $i = 1, \dots, k$, where P_0 denotes the distribution under H_0. Then H_0 is rejected for large values of the test statistic

$$Q_k^2(\overline{Y}_n, P) = \sum_{i=1}^{k} (\overline{Y}_{ni} - p_i)^2/p_i. \tag{3.1}$$

In applications of this test one of the major problems is the choice of the intervals $A_1, \dots, A_k$. We investigate the effect on the power of the χ^2 test of letting $k = k(n)$ tend to infinity (as $n \to \infty$) under the restrictions

$$\begin{aligned} &k = o(n^{1/2}) \quad \text{as } n \to \infty, \\ &p_i \geq a/k, \qquad i = 1, \dots, k \quad \text{for some } 0 < a < \infty, \end{aligned} \tag{3.2}$$

and

$$\max_{1 \leq i \leq k} p_i \to 0 \quad \text{as } k \to \infty.$$

To study the power for a given alternative density g consider the contamination family

$$g_\theta = (1 - \theta)h + \theta g, \qquad 0 \leq \theta \leq 1. \tag{3.3}$$

In line with applications, we assume that the support of P_0 is a possibly infinite interval S; i.e., $h > 0$ a.e. (λ) on S and $h = 0$ outside S where λ is the Lebesgue measure. It is also assumed that $g = 0$ outside S; this involves no loss of generality since goodness-of-fit tests are always performed conditional on the event that no observations lie outside S. The general idea is that the power of a test at g, i.e., at $\theta = 1$, for moderate sample sizes will be reflected in the asymptotic power of the test for $n \to \infty$ and $\theta = \theta_n \to 0$. This approach has proved to be very effective in parametric models. Complementary to Kallenberg et al. (1985), where strictly local theory ($n\theta_n^2$ bounded away from 0 and ∞) is developed, here we consider nonlocal alternatives for which $\lim_{n\to\infty} n\theta_n^2 = \infty$. For such nonlocal alternatives and a fixed significance level α, many tests will have limiting power one and a problem of comparison arises. To overcome this difficulty levels α_n are considered tending to zero at a rate such that

$$(3.4)\qquad (-\log \alpha_n)/n\theta_n^2 \text{ is bounded away from 0 and } \infty.$$

Note that this extends the strictly local case where $n\theta_n^2$ is bounded away from 0 and ∞ and $\alpha_n = \alpha$ is fixed. In its most extreme form, when θ_n is fixed, this approach is at the basis of the concept of Bahadur efficiency; in that case α_n tends to zero exponentially fast. Here we take an intermediate position assuming $\limsup_{n\to\infty} n\theta_n^4 < \infty$.

For smaller α_ns than given by (3.4), the limiting power of the χ^2 tests will be zero, for larger α_ns it will be one for all choices of $k = k(n)$ not increasing too fast. Considering a fixed sequence $\{\theta_n\}$, there is still a whole range of sequences $\{\alpha_n\}$ satisfying (3.4). By an appropriate choice of k, we may try to maximize the range of levels α_n for which a limiting power one is achieved.

Denoting by $\beta_\alpha(\theta; Q_k^2)$ the power of the size-α χ^2 test at θ, and writing

$$(3.5)\qquad \Delta_k = \Delta_k(A_1,\dots,A_k) = \sum_{i=1}^{k} p_i^{-1}\left\{\int_{A_i} f\,dH\right\}^2 \quad \text{with } f = g/h - 1,$$

the following proposition describes the influence of the choice of $k = k(n)$ on the power of the χ^2 tests.

PROPOSITION 3.1. *Let $n \to \infty$, $k = k(n) = o(n^{1/2})$ (k may remain bounded) and let $\min_{1\le i\le k} kp_i \ge a$ for some $0 < a < \infty$. Let $\{\theta_n\}$ satisfy both $\lim_{n\to\infty} n\theta_n^2 = \infty$ and $\limsup_{n\to\infty} n\theta_n^4 < \infty$ and let $\log \alpha_n = -bn\theta_n^2$ for some $0 < b < \infty$.*

(i) *If $k(n) = o(n\theta_n^2)$ then*

$$\lim_{n\to\infty} \beta_{\alpha_n}\left(\theta_n; Q_{k(n)}^2\right) = \begin{cases} 1 & \text{if } \lim_{n\to\infty} \Delta_{k(n)} > 2b, \\ 0 & \text{if } \lim_{n\to\infty} \Delta_{k(n)} < 2b. \end{cases}$$

(ii) *If $k(n) = dn\theta_n^2$, $0 < d < \infty$ then*

$$\lim_{n\to\infty} \beta_{\alpha_n}\left(\theta_n; Q_{k(n)}^2\right) = \begin{cases} 1 & \text{if } \int f^2\,dH > cd, \\ 0 & \text{if } \int f^2\,dH < cd, \end{cases}$$

where $c > 0$ is determined by $2b = d\{c - \log(1 + c)\}$.

The proof of Proposition 3.1 is in Section 4. Here we comment on its implications and relationships with the conclusions of the local theory developed in Kallenberg et al. (1985).

First of all it has to be noted that increasing k has two opposite effects:

—an increase of the local noncentralities $n\theta_n^2\Delta_k$ since Δ_k tends to become larger as k increases, which has a positive effect on the power;

—an increase of the variance of nQ_k^2, which has a negative effect on the power.

Strictly local theory ($n\theta_n^2$ bounded) leads to the conclusion that an improvement of the power by sending k to infinity is only obtained if the increase of Δ_k with k is quite strong to beat the negative effect of an increasing variance. It is shown in Kallenberg et al. (1985) that $k \to \infty$ improves the asymptotic local power of the χ^2 test if $\Delta_k/\sqrt{k} \to \infty$ and pushes it down if $\Delta_k/\sqrt{k} \to 0$.

For nonlocal alternatives the positive effect dominates the negative one if k increases slowly (cf. the proof of Proposition 3.1). Noting that $\lim_{k\to\infty}\Delta k = \int f^2\,dH \le \infty$ [cf. Lemma A in Kallenberg et al. (1985)], Proposition 3.1 leads to the following conclusions:

(a) if Δ_k grows fast, i.e., $\Delta_k/\sqrt{k} \to \infty$, both local and nonlocal theories state that $k \to \infty$ improves the asymptotic power;

(b) if $\Delta_k \to \infty$, i.e., $\int f^2\,dH = \infty$, but $\Delta_k/\sqrt{k} \to 0$ local theory suggests to take k bounded while by nonlocal theory $k \to \infty$ is preferred;

(c) if Δ_k converges to a finite limit, i.e., $\int f^2\,dH < \infty$, local theory prefers a bounded k; nonlocal theory then states that a slow increase of $k(n)$ to infinity is better than a fixed k [see (i)], while a fast increase of $k(n)$ to infinity is worse than a slowly increasing $k(n)$ [see (ii), cd is growing with d and $\lim_{d\to\infty} cd = \infty$]; moreover, if Δ_k does not increase much, the range of levels α_n for which the limiting power can be improved by a slow increase of k, is small; so in this case both theories suggest to take k not too large.

Here we have considered a fixed sequence $\{\theta_n\}$ and variable sequence $\{\alpha_n\}$. In view of (3.4), Proposition 3.1 can equivalently be interpreted in terms of a fixed sequence $\{\alpha_n\}$ and a whole range of alternatives θ_n.

REMARK 3.1. Nonlocal theory suggests taking $k = k(n) \to \infty$ at not too fast a rate. This is in sharp contrast to the criterion $\Delta_k/k \to \infty$ derived from Shirahata (1976) which is also based on $\alpha \to 0$. Note that Shirahata's criterion is never satisfied if $\min_{1\le i\le k} kp_i > a$ for some $0 < a < \infty$ since in that case $\Delta_k = o(k)$.

In two examples we confront the conclusions of the local and nonlocal theory with some Monte Carlo results for sample size $n = 50$.

EXAMPLE 3.1. Let h be the standard normal density ϕ and $g = (1/\sqrt{3})\phi(x/\sqrt{3})$ i.e., a normal distribution with variance 3.

EXAMPLE 3.2. Let h be the uniform distribution on $(0,1)$ and g the Beta $B(0.4, 0.4)$ distribution.

In both examples we take equiprobable classes under the null hypothesis. By routine calculations we have $\lim_{k\to\infty}\Delta_k = \infty$ and $\lim_{k\to\infty}\Delta_k/\sqrt{k} = 0$. Hence in

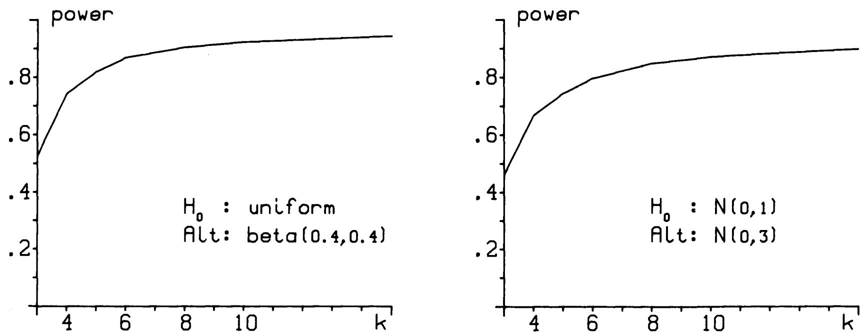

FIG. 1. *Power of χ^2 tests estimated by Monte Carlo from 10,000 samples. $n = 50$, $\alpha = 0.05$.*

both examples local theory suggests a bounded k, while nonlocal theory indicates $k \to \infty$ as the beter choice. The Monte Carlo results presented in Figure 1 indicate that nonlocal theory gives a better explanation of the power behaviour for these examples.

Proposition 3.1. deals with moderate deviations of the χ^2 statistic. Next we present an application where large deviations of both the χ^2 statistic and the LR test are applied. It is well known that the LR test beats the χ^2 test asymptotically at fixed alternatives. For instance, in the sense that for fixed k the Bahadur efficiency of the χ^2 test relative to the LR test is less than or equal to 1 and equals 1 only on a small set of alternatives [Bahadur (1971), pages 31–32]. The difference between the Bahadur slopes generally increases with k. Under strong moment conditions, Quine and Robinson (1985) show that if $k(n) = O(n) \to \infty$, the Bahadur efficiency of the χ^2 test relative to the LR test indeed equals 0. Their moment conditions can be weakened to $\int f^2\, dH < \infty$ (cf. Lemma 4.1). Here we discuss Bahadur efficiency in light of the theorems of Section 2. Let $k(n) = o(n) \to \infty$ and consider a fixed θ in (3.3) such that $\int g_\theta \log(g_\theta/h)\, dx < (125)^{-2}$. Further assume $\int f^2\, dH < \infty$ and take $p_i = k^{-1}$ $(i = 1, \dots, k)$. It now easily follows from Corollary 2.4 and Bahadur's (1971) Theorem 7.2 that the slope of the LR test based on the discretized observations $Y_{n1}, \dots, Y_{nk(n)}$ has exact Bahadur slope $2\int g_\theta \log(g_\theta/h)\, dx$. By Theorem 7.5 of Bahadur (1971) this implies the remarkable fact that the LR test based on $Y_{n1}, \dots, Y_{nk(n)}$ is Bahadur efficient at g_θ *even within the class of all tests based on the original observations* $Z_1, \dots, Z_n$! For instance, if $Z_1, \dots, Z_n$ are normally $N(\mu, 1)$-distributed and we are testing H_0: $\mu = 0$, the LR test based on the discretized observations $Y_{n1}, \dots, Y_{nk(n)}$ is Bahadur efficient if $k(n) = O(n) \to \infty$! This surprising first-order optimality property demands second-order investigations. In regular cases, a test which is Bahadur efficient has Bahadur deficiency of order $O(\log n)$—first-order efficiency implies second-order efficiency [cf. Kallenberg (1983)]. To obtain the Bahadur slope a first-order large deviation result as in Quine and Robinson (1985) suffices. To study Bahadur deficiency second-order large deviation theorems are needed. By Theorem 2.1 we have

$$-n^{-1}\log P_p\big(I_k(\overline{Y}_n, p) \ge d\big) = d - \frac{1}{2}\frac{k}{n}\log\frac{n}{k}(1 + o(1)). \tag{3.6}$$

Application of Theorem 5.2 in Morris (1975) yields

$$\{I_k(\overline{Y}_n, p) - I_k(p_\theta, p) + O(kn^{-1})\}\sigma_\theta^{-1}n^{1/2} \to_{D_{g_\theta}} N(0,1), \tag{3.7}$$

where $p_\theta = (p_{\theta 1}, \dots, p_{\theta k})$ with $p_{\theta i} = P_\theta(Z_1 \in A_i)$ $(i = 1, \dots, k)$ and where

$$\sigma_\theta^2 = \int g_\theta\left\{\log(g_\theta/h) - \int g_\theta \log(g_\theta/h)\right\}^2 dx.$$

Note that

$$I_k(p_\theta, p) \to \int g_\theta \log(g_\theta/h)\, dx \quad \text{as } k \to \infty. \tag{3.8}$$

It is seen from (3.6) and (3.7) that the Bahadur deficiency of the LR test based on the discretized observations depends on the rate of convergence of $k = k(n)$ to infinity in two ways—both by the convergence of the "Riemann–Stieltjes sum" $I_k(p_\theta, p)$ in (3.8) to the Kullback–Leibler information number $\int g_\theta \log(g_\theta/h)\, dx$, and by the term $\frac{1}{2}kn^{-1}\log(n/k)$ in (3.6). So the well known phenomena of first-order (Bahadur) efficiency implies second-order (Bahadur) efficiency does *not* hold here! The relativity of the earlier obtained first-order Bahadur optimality is clear from the lack of second-order efficiency.

Now we return to the comparison of LR and χ^2 tests. Choose $\tilde{y}_{n1}, \dots, \tilde{y}_{nk} \in \mathbb{N}$ with $\sum_{i=1}^k \tilde{y}_{ni} = n$ and $-1 \le \tilde{y}_{n1} - nk^{-1}\{1 + (k-1)^{1/2}d^{1/2}\} \le 1$, $-1 \le \tilde{y}_{ni} - nk^{-1}\{1 - (k-1)^{-1/2}d^{1/2}\} \le 1$ $(i = 2, \dots, k)$ [cf. Hoeffding (1965a), page 389]. Define $y_{n1} = \tilde{y}_{n1} + k - 1$, $y_{ni} = \tilde{y}_{ni} - 1$ $(i = 2, \dots, k)$, then $\sum_{i=1}^k y_{ni} = n$ and

$$\begin{aligned}\sum_{i=1}^k y_{ni}^2 &\ge \left[nk^{-1}\{1 + (k-1)^{1/2}d^{1/2}\} + k - 2\right]^2 \\ &\quad + (k-1)\left[nk^{-1}\{1 - (k-1)^{-1/2}d^{1/2}\} - 2\right]^2 \\ &\ge n^2k^{-1}(d+1)\end{aligned}$$

for n (and hence k) sufficiently large; therefore

$$Q_k^2(\bar{y}_n, p) = \sum_{i=1}^k \frac{(n^{-1}y_{ni} - 1/k)^2}{1/k} = kn^{-2}\sum_{i=1}^k y_{ni}^2 - 1/k \ge d$$

for n sufficiently large, and thus, applying Stirling's formula,

$$\begin{aligned}&-n^{-1}\log P_p\left(Q_k^2(\overline{Y}_n, p) \ge d\right) \\ &\quad \le -n^{-1}\log P_p(Y_1 = y_{n1}, \dots, Y_k = y_{nk}) = o(1)\end{aligned} \tag{3.9}$$

as $n \to \infty$. Since

$$Q_k^2(\overline{Y}_n, p) \to_{P_{g_\theta}} \int (g_\theta/h - 1)^2 h\, dx = \theta^2 \int f^2\, dH < \infty, \tag{3.10}$$

the exact Bahadur slope at g_θ of the χ^2 test equals 0, which implies that the Bahadur efficiency at g_θ of the χ^2 test relative to the LR test equals 0.

REMARK 3.2. In case $\int f^2\,dH < \infty$ the following curious situation for Pearson's χ^2 test may occur:

—for fixed α and bounded $n\theta_n^2$ local theory states that a bounded k is preferable [Proposition 4.1 in Kallenberg et al. (1985)];

—for $\alpha_n \to 0$, $n\theta_n^2 \to \infty$, and $n\theta_n^4$ bounded such that (3.4) holds, there may be a small range of θ_ns for which $k \to \infty$ at not too fast a rate has to be preferred (Proposition 3.1);

—for $\alpha_n \to 0$ at an exponential rate and a fixed θ again bounded k is preferable [(3.9) and (3.10) imply a Bahadur slope equal to 0 if $k \to \infty$, while for a fixed k a positive Bahadur slope is obtained].

As a third application we investigate the question of combining cells in a large contingency table. Let $N = (N_{11}, \dots, N_{1c}, N_{21}, \dots, N_{2c}, \dots, N_{r1}, \dots, N_{rc}) = (Y_{n1}, \dots, Y_{nk})$ with $k = rc$ be multinomial distributed with parameters n and $p = (p_{11}, \dots, p_{1c}, p_{21}, \dots, p_{2c}, \dots, p_{r1}, \dots, p_{rc})$. Write $p_{i.} = \sum_j p_{ij}$ and $p_{.j} = \sum_i p_{ij}$. We only consider tables with

$$\inf_i p_{i.} \geq a/r \quad \text{and} \quad \inf_j p_{.j} \geq a/c \quad \text{for some fixed } a \in (0, \infty). \tag{3.11}$$

Consider the null hypothesis of independence

$$H_0\colon p_{ij} = p_{i.}p_{.j}, \qquad i = 1, \dots, r;\ j = 1, \dots, c. \tag{3.12}$$

Define for $q = (q_{11}, \dots, q_{rc})$

$$I_k(q, H_0) = \inf_{p \in H_0} I_k(q, p) = \inf_{p \in H_0} \sum_{i,j} q_{ij}\{\log q_{ij}/(p_{i.}p_{.j})\},$$

then

$$I_k(q, H_0) = I_k(q, q^0) \quad \text{where } q_{ij}^0 = q_{i.}q_{.j}, \qquad i = 1, \dots, r;\ j = 1, \dots, c.$$

Take a sequence of alternatives $\{q_{ij}\}$ satisfying

$$\sum_{i,j} q_{ij} \log\{q_{ij}/(q_{i.}q_{.j})\} \to \gamma \in \left(0, a(125)^{-2}\right) \quad \text{if } r, c \to \infty. \tag{3.13}$$

(This corresponds to a "fixed" alternative; for instance, if

$$q_{ij} = \int_{A_{ij}} g_\theta(x, y)\,dx\,dy$$

with

$$g_\theta(x, y) = (1 - \theta)\int h(x, y)\,dy \int h(x, y)\,dx + \theta g(x, y)$$

for some densities h and g, and suitable classes A_{ij} in $\mathbb{R}^2$, then (3.13) holds in general with

$$\gamma = \int g_\theta(x, y) \log\left\{ g_\theta(x, y) \Big/ \left(\int g_\theta(x, y)\,dy \int g_\theta(x, y)\,dx \right) \right\} dx\,dy.$$

By Remark 2.1 we have for each $0 < d < a(125)^{-2}$

$$n^{-1}\log \sup_{p \in H_0} P_p(I_k(\overline{Y}_n, H_0) \geq d) \leq n^{-1}\log \sup_{p \in H_0} P_p(I_k(\overline{Y}_n, p) \geq d) \to -d. \tag{3.14}$$

Let $r = r(n)$, $c = c(n) \to \infty$ as $n \to \infty$. Assume

$$I_k(\overline{Y}_n, H_0) \to_{P_q} \lim_{r, c \to \infty} I_k(q, q^0) = \gamma. \tag{3.15}$$

In view of (the proofs of) Theorem 7.2 and 7.5 in Bahadur (1971), (3.14) and (3.15) imply that the exact Bahadur slope of the LR test equals 2γ. So one might say that again the LR test is Bahadur efficient. Applying the inequality

$$q_{i1}\log\frac{q_{i1}}{q_{i.}q_{.1}} + q_{i2}\log\frac{q_{i2}}{q_{i.}q_{.2}} \geq (q_{i1} + q_{i2})\log\frac{q_{i1} + q_{i2}}{q_{i.}(q_{.1} + q_{.2})},$$

it is easily seen that for bounded $r = r(n)$ and/or $c = c(n)$ the Bahadur slope is in general less than 2γ. Therefore from a nonlocal point of view combining cells in such a way that $r(n)$ and/or $c(n)$ remains bounded results in a loss of (Bahadur) efficiency.

4. Proofs. The main part of this section is devoted to the proof of (2.3). Before doing that we first prove Proposition 2.2.

PROOF OF PROPOSITION 2.2. The inequality (2.7) is implied by the more general lemma 2.3.1 of Kallenberg (1978) and the remark following it. Inequality (2.7) is also a special case of Theorem 1 in Fu (1983). It remains to prove (2.8). Without loss of generality assume $0 < p_1 \leq \frac{1}{2}$. Define $u_n > p_1$ by $I_2((u_n, 1 - u_n), p) = d_n$; then we have

$$\begin{aligned} P_p(I_2(\overline{Y}_n, p) \geq d_n) &\geq P_p(\overline{Y}_n \geq u_n) \\ &= \sum_{j \geq nu_n} \binom{n}{j} p_1^j (1 - p_1)^{n-j} e^{-nd_n} \left(\frac{u_n}{p_1}\right)^{nu_n} \left(\frac{1 - u_n}{1 - p_1}\right)^{n - nu_n} \\ &= \sum_{j \geq nu_n} \binom{n}{j} u_n^j (1 - u_n)^{n-j} \left(\frac{p_1(1 - u_n)}{u_n(1 - p_1)}\right)^{j - nu_n} e^{-nd_n}. \end{aligned} \tag{4.1}$$

Hence, with nt_n the smallest integer greater than or equal to nu_n,

$$P_p(I_2(\overline{Y}_n, p) \geq d_n) \geq \binom{n}{nt_n} u_n^{nt_n}(1 - u_n)^{n - nt_n} \frac{p_1(1 - u_n)e^{-nd_n}}{u_n(1 - p_1)}. \tag{4.2}$$

Using Stirling's formula it is seen that the right-hand side of (4.2) is greater than or equal to $c_1 p_1 e^{-nd_n}(nt_n)^{-1/2}$ for some constant $c_1 > 0$. If $u_n/d_n \to c_2 \in [0, \infty)$ the result is established. So from now on assume $u_n/d_n \to \infty$. Denote by ϕ the standard normal density and let $v_n = \{nu_n(1 - u_n)\}^{1/2}$. Returning to (4.1) we have by the Berry–Esseen theorem

$$
\begin{aligned}
&P_p(I_2(\bar{Y}_n, p) \ge d_n)\\
&\quad \ge \sum_{0 \le i \le 0.1 v_n - 1} \sum_{(j - nu_n)/v_n \in ((10i/v_n), [10(i+1)]/v_n)} \binom{n}{j} u_n^j (1 - u_n)^{n-j}\\
&\qquad \cdot \left(\frac{p_1(1 - u_n)}{u_n(1 - p_1)} \right)^{10(i+1)} e^{-nd_n}\\
&\quad \ge \sum_{0 \le i \le 0.1 v_n - 1} \left(\frac{p_1(1 - u_n)}{u_n(1 - p_1)} \right)^{10(i+1)} e^{-nd_n} \left[\frac{10\phi(1)}{v_n} - \frac{u_n^2 + (1 - u_n)^2}{v_n} \right]\\
&\quad \ge \sum_{0 \le i \le 0.1 v_n - 1} \left(\frac{p_1(1 - u_n)}{u_n(1 - p_1)} \right)^{10(i+1)} e^{-nd_n} \frac{1}{u_n^{1/2} n^{1/2}}.
\end{aligned} \tag{4.3}
$$

Since $u_n/d_n \to \infty$ implies $p_1/u_n \to 1$ and $(u_n - p_1)^2/\{2p_1 d_n(1 - p_1)\} \to 1$ the right-hand side of (4.3) is greater than or equal to $c_3(nd_n)^{-1/2}$ for some $c_3 > 0$. This completes the proof. □

REMARK 4.1. From the above proof we see that we do not need the extra factor $p_1(1 - p_1)$ in (2.8) if $u_n/d_n \to \infty$. If u_n/d_n is bounded and $nu_n \in \mathbb{N}$, we can estimate the sum in (4.1) by the first term, i.e., the term with $j - nu_n = 0$, implying that also in this case the extra factor $p_1(1 - p_1)$ in (2.8) can be omitted. However, in the remaining case, u_n/d_n is bounded and $nu_n \notin \mathbb{N}$, we may need some extra factor depending on p_1 in (2.8) as is seen form Example 2.4.

PROOF OF THEOREM 2.1. Since the method of the proof of (2.4) is similar to that of (2.3) and since the proof is rather long, we only will give the proof of (2.3) (cf. also Remark 4.2).

The proof is by induction on k. For $k = 2$ the inequality (2.3) reduces to (2.7) in Proposition 2.2 which already has been proved. Suppose that (2.3) is true for $k \ge 2$, and let $Y_n = (Y_{n1}, \dots, Y_{nk+1})$ have a $(k + 1)$-dimensional multinomial distribution. Without loss of generality, assume that $a(k + 1) \le p_1 \le 1/(k + 1)$. Define

$$
\begin{aligned}
I^*(r, s) &= r \log\left(\frac{r}{s} \right) + (1 - r)\log\left\{ \frac{1 - r}{1 - s} \right\}\\
&= I_2((r, 1 - r), (s, 1 - s)) \qquad (r, 1 - r), (s, 1 - s) \in S_2,\\
J_n &= \left\{ 0 \le j \le n \colon I^*\left(\frac{j}{n}, p_1 \right) < d_n - \frac{k}{2n} \right\}
\end{aligned}
$$

and

$$
\tilde{p}_{i-1} = \frac{p_i}{1 - p_1}, \qquad i = 2, \dots, k + 1.
$$

Note that $I^*(1, p_1) = -\log p_1 \ge \log(k + 1) > d_n$ and hence $n \notin J_n$. By Lemma

2.3.1. of Kallenberg (1978) and the remark following it we have

$$P_p\left(I^*(\overline{Y}_{n1}, p) \geq d_n - \frac{k}{2n}\right) \leq 2e^{-nd_n+k/2}.$$

Therefore

$$\begin{aligned}
&P_p(I_{k+1}(\overline{Y}_n, p) \geq d_n) \\
&\quad\leq P_p\left(I^*(\overline{Y}_{n1}, p_1) \geq d_n - \frac{k}{2n}\right) \\
&\qquad + \sum_{j\in J_n} P_p\left(\sum_{i=2}^{k+1} \overline{Y}_{ni}\log\left(\frac{\overline{Y}_{ni}}{p_i}\right) + \frac{j}{n}\log\left\{\frac{j/n}{p_1}\right\} \geq d_n,\ Y_{n1} = j\right) \\
&\quad\leq 2e^{-nd_n+k/2} + \sum_{j\in J_n} P_p(Y_{n1} = j) \qquad (4.4)\\
&\qquad \cdot P_p\left(\sum_{i=2}^{k+1} \frac{n}{n-j}\overline{Y}_{ni}\log\left\{\frac{n}{n-j}\frac{\overline{Y}_{ni}}{\tilde{p}_{i-1}}\right\} \geq \frac{n}{n-j}\left\{d_n - I^*\left(\frac{j}{n}, p_1\right)\right\}\middle| Y_{n1} = j\right) \\
&\quad= 2e^{-nd_n+k/2} + \sum_{j\in J_n} P_p(Y_{n1} = j) \\
&\qquad \cdot P_p\left(\sum_{i=1}^{k} \overline{Z}_{n-ji}\log\left\{\frac{\overline{Z}_{n-ji}}{\tilde{p}_i}\right\} \geq \frac{n}{n-j}\left\{d_n - I^*\left(\frac{j}{n}, p_1\right)\right\}\right),
\end{aligned}$$

where $\overline{Z}_{n-ji} = Z_{n-ji}/(n-j)$ and $Z_{n-j} = (Z_{n-j1}, \ldots, Z_{n-jk})$ has a k-dimensional multinomial distribution with parameters $n-j$ and $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_k)$. Now the induction hypothesis can be applied, since for $j \in J_n$

$$(n-j)\frac{n}{n-j}\left\{d_n - I^*\left(\frac{j}{n}, p_1\right)\right\} > k/2.$$

Because $\tilde{p}_{i-1} = p_i/(1-p_1) \geq a(k+1)^{-1}/(1-a(k+1)^{-1}) = \tilde{a}/k$ with $\tilde{a} = a/(1+k^{-1}(1-a))$, application of the induction hypothesis yields

$$\begin{aligned}
&P_p\left(\sum_{i=1}^{k} \overline{Z}_{n-ji}\log\left\{\frac{\overline{Z}_{n-ji}}{\tilde{p}_i}\right\} \geq \frac{n}{n-j}\left\{d_n - I^*\left(\frac{j}{n}, p_1\right)\right\}\right) \\
&\quad\leq 2\left(\frac{2en\{d_n - I^*(j/n, p_1)\}}{k}\right)^{(k-2)/2} e^{-n\{d_n - I^*(j/n, p_1)\}} \qquad (4.5)\\
&\qquad \cdot\left[1 + 400a^{-1/2}\{1 + k^{-1}(1-a)\}^{1/2} d_n^{1/2}\right]^{2(k-2)} \prod_{i=3}^{k}\left(1 + \frac{10}{i}\right).
\end{aligned}$$

For $x > 0$ and $0 < a \leq 1$

$$\left[\frac{1 + x\{1 + k^{-1}(1-a)\}^{1/2}}{1+x}\right]^{2(k-2)} \leq 1 + x,$$

implying that the right-hand side of (4.5) is less than or equal to

$$2\left(\frac{2end_n}{k}\right)^{(k-2)/2} e^{-n\{d_n - I^*(j/n, p_1)\}}\left(1 + 400a^{-1/2}d_n^{1/2}\right)^{2(k-2)+1}$$

(4.6)

$$\cdot\left\{\prod_{i=3}^{k}\left(1 + \frac{10}{i}\right)\right\}\left\{1 - \frac{I^*(j/n, p_1)}{d_n}\right\}^{(k-2)/2}.$$

Next we investigate $P_p(Y_{n1} = j)$. Here we use the method of exponential centering. Let X_{jn} have a binomial distribution with parameters n and j/n; then we have

$$P_p(Y_{n1} = j) = \Pr(X_{jn} = j)\exp\{-nI^*(j/n, p_1)\}. \tag{4.7}$$

(Note that $EX_{jn} = j$: centering!) The inequality

$$(2\pi)^{1/2}n^{n+1/2}e^{-n} < n! < (2\pi)^{1/2}n^{n+1/2}e^{-n}\left(1 + (4n)^{-1}\right)$$

yields the upper bound

$$\Pr(X_{jn} = j) \le \frac{1 + (4n)^{-1}}{\sqrt{2\pi}}\left\{\frac{n}{j(n-j)}\right\}^{1/2}, \qquad i \le j \le n - 1. \tag{4.8}$$

Now define

$$a_n^- = \sup\left\{0 \le j \le np_1 \colon I^*\left(\frac{j}{n}, p_1\right) \ge d_n - \frac{k}{2n}\right\},$$

$$a_n^+ = \inf\left\{np_1 \le j \le n \colon I^*\left(\frac{j}{n}, p_1\right) \ge d_n - \frac{k}{2n}\right\}$$

with the convention $a_n^- = 0$ if $I^*(0, p_1) = -\log(1 - p_1) < d_n - k/(2n)$. In view of (4.5), (4.6), (4.7), and (4.8) it holds that

$$\sum_{j \in J_n} P_p(Y_{n1} = j)P_p\left(\sum_{i=1}^{k} \bar{Z}_{n-ji}\log\left\{\frac{\bar{Z}_{n-ji}}{\tilde{p}_i}\right\} \ge \frac{n}{n-j}\left\{d_n - I^*\left(\frac{j}{n}, p_1\right)\right\}\right)$$

$$\le \sum_{j=a_n^-+1}^{a_n^+-1} 2\left(\frac{2end_n}{k}\right)^{(k-2)/2}$$

$$\cdot e^{-nd_n}\left(1 + 400a^{-1/2}d_n^{1/2}\right)^{2k-3}\left\{\prod_{i=3}^{k}\left(1 + \frac{10}{i}\right)\right\}$$

(4.9)

$$\cdot\left(1 + (4n)^{-1}\right)(2\pi)^{-1/2}\left\{1 - \frac{I^*(j/n, p_1)}{d_n}\right\}^{(k-2)/2}\left\{\frac{n}{j(n-j)}\right\}^{1/2}$$

$$+ \left[\max\left(0, d_n + \log(1 - p_1) - \frac{k}{2n}\right)\right]$$

$$\cdot 2\left[\frac{2en\{d_n + \log(1 - p_1)\}}{k}\right]^{(k-2)/2} e^{-nd_n}$$

$$\cdot\left(1 + 400a^{-1/2}d_n^{1/2}\right)^{2k-3}\left\{\prod_{i=3}^{k}\left(1 + \frac{10}{i}\right)\right\}.$$

The term $2\exp\{-nd_n + k/2\}$ in (4.4) will be estimated later on. Our next aim is to derive an upper bound for the first term in the right-hand side of (4.9). As a result we then obtain inequality (4.21).

Since $(1-x)\log(1-x) \geq -x$ for $0 < x < 1$, we have for $0 < p_1 z/(1-p_1) < 1$

$$\begin{aligned} & I^*(p_1 + p_1 z, p_1) \\ & \quad = p_1(1+z)\log(1+z) + (1 - p_1 - p_1 z)\log\left(1 - \frac{p_1 z}{1-p_1}\right) \\ & \quad > p_1(1+z)\log(1+z) - p_1 z \\ & \quad = p_1\{(1+z)\log(1+z) - z\}. \end{aligned} \tag{4.10}$$

Furthermore,

$$I^*(p_1 - p_1 z, p_1) > I^*(p_1 + p_1 z, p_1) \tag{4.11}$$

for $z > 0$ and $0 < p_1 < \frac{1}{2}$. Let $a(j) = |j/n - p_1| p_1^{-1}$, then we have,

$$\begin{aligned} & \sum_{j=a_n^- + 1}^{a_n^+ - 1} \left\{1 - \frac{I^*(j/n, p_1)}{d_n}\right\}^{(k-2)/2} \left\{\frac{n}{j(n-j)}\right\}^{1/2} \\ & \quad \leq \left\{1 - \frac{a_n^+ - 1}{n}\right\}^{-1/2} \\ & \quad \cdot \sum_{j=a_n^- + 1}^{a_n^+ - 1} \left[1 - \frac{p_1}{d_n}\{(1 + a(j)) \cdot \log(1 + a(j)) - a(j)\}\right]^{(k-2)/2} j^{-1/2}. \end{aligned} \tag{4.12}$$

For $j > 0.9 p_1 n$ the terms of the last series are decreasing or increasing—decreasing (in j). Therefore we can estimate the sum by

$$\begin{aligned} & \int_{j=b_n}^{a_n^+ - 1} \left[1 - \frac{p_1}{d_n}\{(1 + a(j)) \right. \\ & \quad \left. \cdot \log(1 + a(j)) - a(j)\}\right]^{(k-2)/2} j^{-1/2}\, dj + (0.9 n p_1)^{-1/2} \\ & \quad + \sum_{j=a_n^- + 1}^{0.9 p_1 n} \left[1 - \frac{p_1}{d_n}\{(1 + 0.1)\log(1 + 0.1) - 0.1\}\right]^{(k-2)/2} j^{-1/2}, \end{aligned} \tag{4.13}$$

where $b_n = \max(a_n^- + 1, 0.9 p_1 n)$.

First consider

$$\int_{np_1}^{a_n^+ - 1} \left[1 - \frac{p_1}{d_n}\{(1 + a(j)) \cdot \log(1 + a(j)) - a(j)\}\right]^{(k-2)/2} j^{-1/2}\, dj. \tag{4.14}$$

Substitute $z = (j/n - p_1)p_1^{-1}$, resulting in

$$\int_0^{\{(a_n^+ - 1)n^{-1} - p_1\}p_1^{-1}} \left[1 - \frac{p_1}{d_n}\{(1+z)\log(1+z) - z\}\right]^{(k-2)/2} \cdot (np_1)^{1/2}(1+z)^{-1/2}\,dz. \tag{4.15}$$

Next substitute $y^2 = (1+z)\log(1+z) - z$. Note that $2y\,dy = \log(1+z)\,dz$ and

$$2y \le 2^{1/2}(1+z)^{1/2}\log(1+z).$$

Hence the integral (4.14) is less than or equal to [cf. (4.15)]

$$\int_0^{(d_n/p_1)^{1/2}} \left\{1 - \frac{p_1}{d_n}y^2\right\}^{(k-2)/2} (2np_1)^{1/2}\,dy = (2nd_n)^{1/2}\int_0^1 (1-t^2)^{(k-2)/2}\,dt. \tag{4.16}$$

Let $k \ge 5$. Substituting $u = (1+t)/2$ we arrive at

$$\begin{aligned}
&(2nd_n)^{1/2}\int_{1/2}^1 u^{k/2-1}(1-u)^{k/2-1}2^{k-1}\,du \\
&= (2nd_n)^{1/2}2^{k-2}\frac{\Gamma(k/2)\Gamma(k/2)}{\Gamma(k)} \\
&\le (2nd_n)^{1/2}2^{k-2}\sqrt{2\pi}\,2^{-(k-1)}\left(1 - \frac{1}{k-1}\right)^{k-1} e(k-1)^{-1/2}\left(1 + \frac{1}{2k-4}\right)^2 \\
&\le (2nd_n)^{1/2}2^{-1}\sqrt{2\pi}\,(k+1)^{-1/2}\left(1 + \frac{3}{k+1}\right).
\end{aligned}$$

By direct computation of $\int_0^1(1-t^2)^{(k-2)/2}\,dt$ for $k = 2, 3, 4$ it follows that for all $k \ge 2$ the integral (4.14) is less than or equal to

$$(2nd_n)^{1/2}2^{-1}\sqrt{2\pi}\,(k+1)^{-1/2}\left(1 + \frac{3}{k+1}\right). \tag{4.17}$$

Next consider

$$\int_{b_n}^{np_1} \left[1 - \frac{p_1}{d_n}\{(1 + a(j))\cdot\log(1 + a(j)) - a(j)\}\right]^{(k-2)/2} j^{-1/2}\,dj. \tag{4.18}$$

Substitute $z = (p_1 - j/n)p_1^{-1}$, resulting in

$$\int_0^{(p_1 - b_n n^{-1})p_1^{-1}} \left[1 - \frac{p_1}{d_n}\{(1+z)\log(1+z) - z\}\right]^{(k-2)/2} \cdot (np_1)^{-1/2}(1-z)^{-1/2}\,dz. \tag{4.19}$$

Next substitute $y^2 = (1+z)\log(1+z) - z$. Note that $2y\,dy = \log(1+z)\,dz$ and for $0 \le z \le 0.1$

$$2y \le \left\{2^{1/2}(1-z)^{1/2}\log(1+z)\right\}(1 + 2^{1/2}y).$$

Hence the integral (4.18) is less than or equal to [cf. (4.19)]

$$\int_0^{(d_n/p_1)^{1/2}} \left\{1 - \frac{p_1}{d_n} y^2\right\}^{(k-2)/2} (2np_1)^{1/2}(1 + 2^{1/2}y)\, dy$$

$$(4.20)\qquad \begin{aligned} &= (2nd_n)^{1/2} \int_0^1 (1 - t^2)^{(k-2)/2} \left\{1 + \left(\frac{2d_n}{p_1}\right)^{1/2} t\right\} dt \\ &= (2nd_n)^{1/2} \int_0^1 (1 - t^2)^{(k-2)/2}\, dt + (2nd_n)^{1/2} k^{-1} \left(\frac{2d_n}{p_1}\right)^{1/2} \\ &\le (2nd_n)^{1/2} 2^{-1} \sqrt{2\pi}\,(k+1)^{-1/2} \left(1 + \frac{3}{k+1}\right) \\ &\quad + (2nd_n)^{1/2} (k+1)^{-1/2} (2a^{-1}d_n)^{1/2} (1 + k^{-1}) \end{aligned}$$

[cf. (4.16) and (4.17)].

In view of (4.4), (4.9), (4.12), (4.13), (4.17), and (4.20) we obtain

$$(4.21)\qquad \begin{aligned} &P_p(I_{k+1}(\bar{Y}_n, p) \ge d_n) \\ &\le 2e^{-nd_n + k/2} + \left\{\max\left(0, d_n + \log(1 - p_1) - \frac{k}{2n}\right)\right\} \\ &\quad \cdot 2\left[\frac{2en\{d_n + \log(1 - p_1)\}}{k}\right]^{(k-2)/2} \\ &\quad \cdot e^{-nd_n}(1 + 400a^{-1/2}d_n^{1/2})^{2k-3} \prod_{i=3}^{k}\left(1 + \frac{10}{i}\right) \\ &\quad + 2\left(\frac{2end_n}{k}\right)^{(k-2)/2} e^{-nd_n}(1 + 400a^{-1/2}d_n^{1/2})^{2k-3} \\ &\quad \cdot \left\{\prod_{i=3}^{k}\left(1 + \frac{10}{i}\right)\right\}\{1 + (4n)^{-1}\}(2\pi)^{-1/2}\left\{1 - \frac{a_n^+ - 1}{n}\right\}^{-1/2} \\ &\quad \cdot \Bigg[(2nd_n)^{1/2}\sqrt{2\pi}\,(k+1)^{-1/2}\left(1 + \frac{3}{k+1}\right) + (2nd_n)^{1/2}(k+1)^{-1/2} \\ &\quad \cdot (2a^{-1}d_n)^{1/2}(1 + k^{-1}) + (0.9np_1)^{-1/2} \\ &\quad + \sum_{j=a_n^- + 1}^{0.9p_1 n}\left[1 - \frac{p_1}{d_n}\{1.1\log 1.1 - 0.1\}\right]^{(k-2)/2} j^{-1/2}\Bigg]. \end{aligned}$$

Now we investigate some of the terms in the right-hand side of (4.21) a little bit closer. The obtained results will then be used to establish (4.28). Since the right-hand side of (2.3) is greater than 1 if $2nd_n < k + (3k)^{1/2}$ (and $k \ge 3$), we may assume in the sequel that

$$2nd_n \ge k + 1 + \{3(k+1)\}^{1/2}, \qquad k \ge 2.$$

Hence

$$2e^{-nd_n+k/2} \le \frac{3.5}{k+1} 2\left(\frac{2end_n}{k+1}\right)^{(k+1-2)/2} e^{-nd_n} \prod_{i=3}^{k}\left(1+\frac{10}{i}\right). \tag{4.22}$$

Because $(1 - Ax^{-1})^{(k-2)/2}x^{-1/2}$ is maximal on $x > A > 0$ if $x = (k-1)A$, and $-\log(1-p_1) > a/(k+1)$, we have

$$\max\left(0, d_n + \log(1-p_1) - \frac{k}{2n}\right)\left\{1 + \frac{\log(1-p_1)}{d_n}\right\}^{(k-2)/2} d_n^{-1/2}$$
$$\le \left\{1 - (k-1)^{-1}\right\}^{(k-2)/2} a^{-1/2}\left\{\frac{(k+1)}{(k-1)}\right\}^{1/2};$$

moreover,

$$\left(\frac{2end_n}{k+1}\right)^{-1/2}\left(\frac{k+1}{k}\right)^{(k-2)/2}\left(1 - \frac{1}{k-1}\right)^{(k-2)/2}\left(\frac{k+1}{k-1}\right)^{1/2} \le 1,$$

implying that the second term on the right-hand side of (4.21) is less than or equal to

$$2\left(\frac{2end_n}{k+1}\right)^{(k+1-2)/2} e^{-nd_n}\left(\frac{dn}{a}\right)^{1/2}\left(1 + 400\left(\frac{dn}{a}\right)^{1/2}\right)^{2k-3} \prod_{i=3}^{k}\left(1+\frac{10}{i}\right). \tag{4.23}$$

We proceed by deriving an upper bound for a_n^+. If $p_1 \ge d_n$ then, by (4.10), $I^*(p_1 + p_1(e-1), p_1) \ge d_n$ and hence $a_n^+ \le np_1 e$. If $p_1 < d_n$ then, again by (4.10),

$$I^*(p_1 + (e-1)d_n, p_1) > p_1\left\{\left(1 + (e-1)\frac{d_n}{p_1}\right)\log\left(1 + (e-1)\frac{d_n}{p_1}\right) - (e-1)\frac{d_n}{p_1}\right\}$$
$$> p_1\frac{d_n}{p_1} = d_n,$$

since $\{1 + (e-1)x\}\log(1 + (e-1)x) > ex$ for $x > 1$. So in this case $a_n^+ \le np_1 + nd_n(e-1)$. Because $(1-x)^{-1/2} \le 1 + x$ for $0 < x < (\sqrt{5} - 1)/2$, we have

$$\left\{1 - \frac{a_n^+ - 1}{n}\right\}^{-1/2} \le 1 + \frac{a_n^+ - - 1}{n} \le 1 + \frac{e}{k+1} + (e-1)d_n. \tag{4.24}$$

The next term is $(0.9np_1)^{-1/2}$; we have the following inequalities

$$(0.9np_1)^{-1/2} \le (0.9)^{-1/2}n^{-1/2}a^{-1/2}(k+1)^{1/2}$$
$$\le (0.9)^{-1/2}a^{-1/2}d_n^{1/2}(2nd_n)^{1/2}(k+1)^{-1/2}2^{1/2}. \tag{4.25}$$

For $0 < x < c^{-1}$ it holds that

$$\left(\frac{k-2}{2}\right)\log(1-cx) + \log x \le \left(\frac{k-2}{2}\right)\log\left(1 - \frac{2}{k}\right) + \log\frac{2}{ck}.$$

Writing $c = 1.1 \log 1.1 - 0.1$ this implies

$$\sum_{j=a_n^-+1}^{0.9p_1 n}\left[1-\frac{p_1}{d_n}\{1.1\log 1.1-0.1\}\right]^{(k-2)/2} j^{-1/2}$$

$$\leq \left(1-\frac{2}{k}\right)^{(k-2)/2}\frac{2}{ck}\frac{d_n}{p_1}\sum_{j=1}^{0.9p_1 n} j^{-1/2}$$

$$\leq \left(1-\frac{2}{k}\right)^{(k-2)/2}\frac{2}{ck}\frac{d_n}{p_1}2(0.9p_1 n)^{1/2} \tag{4.26}$$

$$\leq (2nd_n)^{1/2}(k+1)^{-1/2}a^{-1/2}d_n^{1/2} \cdot \left(1-\frac{2}{k}\right)^{(k-2)/2} 2^{3/2}c^{-1}(0.9)^{1/2}\left(\frac{k+1}{k}\right)$$

$$\leq (2nd_n)^{1/2}(k+1)^{-1/2}a^{-1/2}d_n^{1/2}\begin{cases}426 & \text{if } k \geq 3,\\ 829 & \text{if } k = 2.\end{cases}$$

Finally note that $I^*(0.61, 1/3) > 0.15$ and hence $I^*(0.61, p_1) > d_n$, implying $a_n^+ \leq 0.61n$ and thus

$$\left[1-\frac{a_n^+-1}{n}\right]^{-1/2} \leq 1.61. \tag{4.27}$$

Combining (4.21), (4.22), (4.23), (4.24), (4.25), (4.26), and (4.27) yields

$$P_p(I_{k+1}(\overline{Y}_n, p) \geq d_n)$$

$$\leq 2\left(\frac{2end_n}{k+1}\right)^{(k+1-2)/2} e^{-nd_n}\left(1+400a^{-1/2}d_n^{1/2}\right)^{2k-3}\left\{\prod_{i=3}^{k}\left(1+\frac{10}{i}\right)\right\}$$

$$\cdot\left[\frac{3.5}{k+1}+a^{-1/2}d_n^{1/2}+\left(\frac{2end_n}{k+1}\right)^{-1/2}\left(\frac{k+1}{k}\right)^{(k-2)/2}\{1+(4n)^{-1}\}\right.$$

$$\cdot(2\pi)^{-1/2}\min\left(1+\frac{e}{k+1}+(e-1)d_n, 1.61\right)$$

$$\cdot\left[(2nd_n)^{1/2}\sqrt{2\pi}(k+1)^{-1/2}\left(1+\frac{3}{k+1}\right)\right. \tag{4.28}$$

$$+(2nd_n)^{1/2}(k+1)^{-1/2}(2a^{-1}d_n)^{1/2}(1+k^{-1})$$

$$+(2nd_n)^{1/2}(k+1)^{-1/2}(0.9)^{-1/2}a^{-1/2}d_n^{1/2}2^{1/2}$$

$$\left.\left.+(2nd_n)^{1/2}(k+1)^{-1/2}a^{-1/2}d_n^{1/2}\begin{cases}426 & \text{if } k \geq 3\\ 829 & \text{if } k = 2\end{cases}\right]\right]$$

$$\leq 2\left(\frac{2end_n}{k+1}\right)^{(k+1-2)/2} e^{-nd_n}\left(1+400a^{-1/2}d_n^{1/2}\right)^{2k-3}\left\{\prod_{i=3}^{k}\left(1+\frac{10}{i}\right)\right\}$$

$$\cdot\left[\frac{3.5}{k+1}+a^{-1/2}d_n^{1/2}+e^{-1/2}\left(\frac{k+1}{k}\right)^{(k-2)/2}\left(1+\frac{d_n}{12}\right)\right.$$

$$\cdot\min\left(1+\frac{e}{k+1}+(e-1)d_n, 1.61\right)$$

$$\left.\cdot\left[1+\frac{3}{k+1}+a^{-1/2}d_n^{1/2}\begin{cases}172 & \text{if } k\geq 3\\ 333 & \text{if } k=2\end{cases}\right]\right]$$

$$\leq 2\left(\frac{2end_n}{k+1}\right)^{(k+1-2)/2} e^{-nd_n}\left(1+400a^{-1/2}d_n^{1/2}\right)^{2k-3}\prod_{i=3}^{k}\left(1+\frac{10}{i}\right)$$

$$\cdot\left(1+\frac{10}{k+1}\right)\left(1+400a^{-1/2}d_n^{1/2}\right),$$

which completes the proof of (2.3). □

REMARK 4.2. To prove (2.4) it is easier to condition on that component Y_{nj} of $(Y_{n1},\dots,Y_{nn})$ for which $p_j=\max\{p_i\colon 1\leq i\leq k+1\}$; then we have

$$\frac{p_i}{1-p_j}\geq\frac{a/(k+1)}{1-(k+1)^{-1}}=\frac{a}{k},\qquad i=1,\dots,k+1.$$

Theorem 2.3 is proved as follows: first the statistics Q^2 and I are related; application of Theorem 2.1 then yields the results.

PROOF OF THEOREM 2.3. Since $x\log(x/p_i)\geq x-p_i+\frac{1}{2}(x-p_i)^2p_i^{-1}-\frac{1}{6}(x-p_i)^3/p_i^2$ for all $x\geq 0$ and $i=1,\dots,k$, it holds that for all $y\in S_k$

$$I_k(y,p)=\sum_{i=1}^{k} y_i\log\left(\frac{y_i}{p_i}\right)$$

$$\geq\frac{1}{2}Q_k^2(y,p)-\frac{1}{6}\sum_{i=1}^{k}(y_i-p_i)^3p_i^{-2}.$$

If $Q_k^2(y,p)=2d_n$ then $(y_i-p_i)^2\leq 2p_id_n$ for all $i=1,\dots,k$; hence

$$\frac{1}{6}\sum_{i=1}^{k}|y_i-p_i|^3p_i^{-2}\leq\frac{1}{6}\max_{1\leq i\leq k}\frac{|y_i-p_i|}{p_i}Q_k^2(Y,p)\leq\frac{1}{6}\max_{1\leq i\leq k}\left(\frac{2d_n}{p_i}\right)^{1/2}2d_n$$

$$\leq u_nd_n.$$

Noting that $\min\{I_k(y,p)\colon Q_k^2(y,p)\geq 2d_n\}=\min\{I_k(y,p)\colon Q_k^2(y,p)=2d_n\}$ it follows that

$$\{y\colon Q_k^2(y,p)\geq 2d_n\}\subset\{y\colon I_k(y,p)\geq d_n(1-u_n)\}.$$

Because $\log(1-u_n)\leq -u_n$ and $1+400a^{-1/2}d_n^{1/2}(1-u_n)^{1/2}\leq 1+400a^{-1/2}d_n^{1/2}$, application of (2.3) yields (2.8).

Since $\log(1+x) \le x - \frac{1}{2}x^2 + \frac{1}{3}x^3$ for $-1 < x < \infty$, it holds that for all $y \in S_k$

$$\begin{aligned} I_k(y,p) &= \sum_{i=1}^{k} y_i \log\left(1 + \frac{y_i - p_i}{p_i}\right) \\ &\le \sum_{i=1}^{k} y_i \left\{ \frac{y_i - p_i}{p_i} - \frac{1}{2}\left(\frac{y_i - p_i}{p_i}\right)^2 + \frac{1}{3}\left(\frac{y_i - p_i}{p_i}\right)^3 \right\} \\ &= \frac{1}{2}Q_k^2(y,p) - \frac{1}{6}\sum_{i=1}^{k}(y_i - p_i)^3 p_i^{-2} + \frac{1}{3}\sum_{i=1}^{k}(y_i - p_i)^4 p_i^{-3}. \end{aligned}$$

In a similar way as above application of (2.4) yields (2.9). □

The proofs of Corollary 2.4, 2.5, and 2.6 are straightforward and therefore omitted.

PROOF OF PROPOSITION 3.1. By straightforward calculus it is seen that for $k = k(n) = O(n\theta_n^2)$ and $n\theta_n^4$ bounded the expectation and variance of nQ_k^2 satisfy

$$E_{\theta_n} nQ_k^2 \sim k - 1 + n\theta_n^2 \Delta_k$$

and

$$\operatorname{var}_{\theta_n}(nQ_k^2) \sim 2k - 2 + 4n\theta_n^2\Delta_k\left\{1 + \theta_n \Delta_k^{-1}\sum_i p_i^{-2}\left(\int_{A_i} f\,dH\right)^3 - \theta_n^2\Delta_k\right\}$$

as $n \to \infty$. (As usual $a_n \sim b_n$ stands for $\lim_{n\to\infty} a_n/b_n = 1$.) Since

$$\left|\Delta_k^{-1}\sum_i p_i^{-2}\left(\int_{A_i} f\,dH\right)^3\right| \le \frac{k}{a}\Delta_k^{-1}\sum_i p_i^{-1}\left(\int_{A_i} f\,dH\right)^2 = k/a,$$

it is easily seen that $(n\theta_n^2\Delta_k)^{-2}\operatorname{var}_{\theta_n}(nQ_k^2) \to 0$ as $n \to \infty$ if $k = O(n\theta_n^2)$. Let $k + x_n\sqrt{2k}$ be the critical value of the size-α_n χ^2 test, when we use $nQ_k^2(\overline{Y}_n, p)$ as test statistic.

If $k = o(n\theta_n^2)$, application of (2.19) yields $x_n\sqrt{2k} \sim 2bn\theta_n^2$. Since to first order $x_n\sqrt{2k}$ does not depend on k the effect of an "increasing variance" is negligible under the restriction $k = o(n\theta_n^2)$. By Chebyshev's inequality, if $\lim_{n\to\infty}\Delta_{k(n)} > 2b$,

$$\begin{aligned} P_{\theta_n}\left(nQ_k^2(\overline{Y}_n, p) \le k + x_n\sqrt{2k}\right) &\le \operatorname{var}_{\theta_n}(nQ_k^2)/\left\{k + x_n\sqrt{2k} - E_{\theta_n}nQ_k^2\right\}^2 \\ &\sim \operatorname{var}_{\theta_n}(nQ_k^2)/\left\{(2b - \Delta_k)n\theta_n^2\right\}^2 \to 0 \end{aligned}$$

as $n \to \infty$. Hence $\beta_{\alpha_n}(\theta_n; nQ_{k(n)}^2) \to 1$ if $\lim_{n\to\infty}\Delta_{k(n)} > 2b$. The other part of (i) may be proved similarly.

If $k = dn\theta_n^2$, application of (2.18) yields $x_n\sqrt{2k} \sim ck$ and since $\lim_{k\to\infty}\Delta_k = \int f^2\,dH \le \infty$, another application of Chebyshev's inequality completes the proof of part (ii). Note that $x_n\sqrt{2k} \sim ck = cdn\theta_n^2$ and so here the effect of an "increasing variance" reappears. Because $2b = d\{c - \log(1+c)\}$, we have $cd > 2b$; moreover, cd is growing when d becomes larger and $cd \downarrow 2b$ if $d \downarrow 0$. □

LEMMA 4.1. *Let Y_n be mult $(n, p_1, \dots, p_k)$ distributed. Assume that*

$$\limsup_{k\to\infty} k \sum_{i=1}^{k} p_i^2 < \infty \quad \text{and} \quad k = k(n) = o(n) \quad \text{as } n \to \infty,$$

then

$$\frac{\sum_{i=1}^{k} Y_{ni}^2 - E_p \sum_{i=1}^{k} Y_{ni}^2}{n^2/k} \to_{P_p} 0.$$

PROOF. Let $\varepsilon > 0$. We have

$$P_p\left(\left|\frac{\sum_{i=1}^{k} Y_{ni}^2 - E_p \sum_{i=1}^{k} Y_{ni}^2}{n^2/k}\right| > \varepsilon\right) \le \frac{1}{\varepsilon} \frac{E_p\left|\sum_i \left(Y_{ni}^2 - E_p Y_{ni}^2\right)\right|}{n^2/k}$$

$$\le \frac{\sum_i E_p |Y_{ni}^2 - E_p Y_{ni}^2|}{\varepsilon n^2/k} \le \frac{\sum_i \left\{\operatorname{Var}_p Y_{ni}^2\right\}^{1/2}}{\varepsilon n^2/k} \le \frac{\sqrt{11} \sum_i \left\{1 + (np_i)^{3/2}\right\}}{\varepsilon n^2/k}.$$

Hölder's inequality yields

$$\sum_i p_i^{3/2} \le \left\{\sum_i \left(p_i^{3/2}\right)^{4/3}\right\}^{3/4} \left(\sum_i 1\right)^{1/4} = \left(\sum_i p_i^2\right)^{3/4} k^{1/4},$$

implying

$$\frac{\sum_i \left\{1 + (np_i)^{3/2}\right\}}{n^2/k} \le \frac{k^2}{n^2} + \frac{k^{3/4}\left(\sum_i p_i^2\right)^{3/4}}{(n/k)^{1/2}} \to 0 \quad \text{as } n \to \infty. \qquad \square$$

Acknowledgment. The author is most grateful to B. F. Schriever for computational assistance and to J. Oosterhoff for stimulating discussions.

REFERENCES

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

EFRON, B. and TRUAX, D. (1968). Large deviations in exponential families. *Ann. Math. Statist.* **39** 1402–1424.

FU, J. C. (1983). On an upper bound of Sanov's inequality for the probability of large deviations. *Statistics and Probability Letters* **1** 107–113.

HOEFFDING, W. (1965a). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–401.

HOEFFDING, W. (1965b). On probabilities of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–219.

KALLENBERG, W. C. M. (1978). *Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families.* Math. Centre Tracts, No. 77. Amsterdam.

KALLENBERG, W. C. M. (1981). Bahadur deficiency of likelihood ratio tests in exponential families. *J. Multivariate Anal.* **11** 506–531.

KALLENBERG, W. C. M. (1983). Asymptotic efficiency and deficiency of tests. In *Proceedings of the 44th Session of the ISI, Madrid 1983*, Vol. L, book 2, 1173–1189.

KALLENBERG, W. C. M., OOSTERHOFF, J. and SCHRIEVER, B. F. (1985). The number of classes in chi-square goodness of fit tests. *Jour. Amer. Statist. Assoc.* **13** 727–742.

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

QUINE, M. P. and ROBINSON, J. (1985). Efficiencies of chi-square and likelihood ratio goodness-of-fit tests, *Ann. Statist.* **13** 727–742.

SHIRAHATA, S. (1976). Tests for independence in infinite contingency tables. *Ann. Statist.* **4** 542–553.

STECK, G. P. (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist.* **2** 237–284.

TUMANYAN, S. H. (1954). On the asymptotic distribution of the chi-square criterion. *Dokl. Akad. Nauk. SSSR* **94** 1011–1012.

DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
FREE UNIVERSITY
POSTBUS 7161
1007 MC AMSTERDAM
THE NETHERLANDS