The Annals of Probability
1982, Vol. 10, No. 3, 663–671

A BERRY-ESSEEN BOUND FOR AN OCCUPANCY PROBLEM

BY M. P. QUINE AND J. ROBINSON

University of Sydney

A Berry-Esseen bound is given for the rate of convergence to normality of the number of empty boxes when balls are distributed independently and at random to boxes with possibly unequal probabilities. The method of proof uses the equivalence of this distribution to a certain conditional distribution based on independent Poisson random variables. Then methods based on the characteristic function of this conditional distribution are used to obtain the result.

1. Introduction. Let N balls be distributed independently and at random into n boxes in such a way that each ball has probability p_k of landing in the kth box, $k = 1, \cdots, n$. Let S denote the number of empty boxes. We give Berry-Esseen bounds for the departure from normality of the distribution function of S.

We will use the following notation:

$$\alpha_k = Np_k,\ k = 1, \cdots, n; \quad \alpha = N/n = n^{-1} \textstyle\sum_k \alpha_k;$$
$$D^2 = n^{-1} [\textstyle\sum_k e^{-\alpha_k}(1 - e^{-\alpha_k}) - N^{-1}(\sum_k \alpha_k e^{-\alpha_k})^2].$$

It is easy to show that if $N, n \to \infty$ in such a way that $N \max_k p_k^2 \to 0$ (or if $nD^2 \to \infty$ and $\max_k \alpha_k/\alpha$ is bounded) then $ES \sim \sum_k e^{-\alpha_k}$ and $V(S) \sim nD^2$. In the symmetric case ($\alpha_k = \alpha$, $k = 1, \cdots, n$),

$$D^2 = e^{-\alpha}\{1 - (1 + \alpha)e^{-\alpha}\}.$$

Englund (1981) has shown in the symmetric case that

$$\sup_x |P((S - E(S))/V^{1/2}(S) \le x) - \Phi(x)| \le C/V^{1/2}(S)$$

with $0.087 \le C \le 10.4$ if $V(S) > 9$. However his method does not appear to be capable of extension to the more general case considered here. Corollary 1 below is asymptotically equivalent to Englund's result.

Holst (1972) has proved a limit theorem for sums of scores based on occupancy numbers which implies the following result: if $N, n \to \infty$, $\alpha \to \alpha_0$, $0 < \alpha_0 < \infty$, $0 < \liminf_n D^2 \le \limsup_n D^2 < \infty$, and $np_k \le C$ for some constant C, for all n and k, then S is asymptotically normal. Corollary 2 improves these conditions.

In what follows, $C_1, C_2, C_3, \cdots$ denote absolute positive constants, and Φ denotes the standard normal distribution function.

THEOREM. *Suppose that* $\max_k \alpha_k \le C_1 \alpha$. *Then*

$$\sup_x \left| P\left(\frac{S - \sum_k e^{-\alpha_k}}{n^{1/2}D} \le x\right) - \Phi(x) \right| \le \frac{C_2}{n^{1/2}D}. \tag{1}$$

The following corollaries follow easily.

COROLLARY 1. *In the symmetric case,*

$$\sup_x \left| P\left(\frac{S - ne^{-\alpha}}{n^{1/2}D} \le x\right) - \Phi(x) \right| \le \frac{C_3}{n^{1/2}D}. \tag{2}$$

Received February 1981; revised August 1981.

AMS 1970 *subject classification*. Primary, 60F05.

Key words and phrases. Berry-Esseen bound, rate of convergence, occupancy problems, central limit theorem.

COROLLARY 2. *Suppose $N, n \to \infty$ in such a way that $n^{1/2}D \to \infty$ and $np_k \le C$ for some constant C, for all n and k. Then S is asymptotically normal.*

2. Proofs. As shown in Loève (1955, page 285), the left hand side of (1) is dominated by

$$\frac{2}{\pi}\int_0^{\eta n^{1/2}D} \frac{|\psi_n(u) - e^{-u^2/2}|}{u}\,du + \frac{24}{\eta n^{1/2}D}(2/\pi)^{1/2}, \tag{3}$$

where

$$\psi_n(u) = E\exp\{iu(S - \textstyle\sum_k e^{-\alpha_k})/n^{1/2}D\},$$

and $\eta > 0$ will be specified later.

Let $Y_1', \cdots, Y_n'$ be independent Poisson random variables with parameters $\alpha_1, \cdots, \alpha_n$; let $X_k' = I(Y_k' = 0)$, $k = 1, \cdots, n$. Then the distribution of S is the conditional distribution of $\sum_k X_k'$ given $\sum_k Y_k' = N$. However, to avoid a possibly degenerate covariance structure of (X_k', Y_k'), we deal instead of X_k' with

$$Z_k' = X_k' + \gamma Y_k', \quad k = 1, \cdots, n,$$

where $\gamma = \sum_k \alpha_k e^{-\alpha_k}/N$. Note that

$$EY_k' = \alpha_k, \quad VY_k' = \alpha_k, \quad EZ_k' = e^{-\alpha_k} + \gamma\alpha_k, \quad VZ_k' = e^{-\alpha_k}(1 - e^{-\alpha_k}) + \gamma^2\alpha_k - 2\gamma\alpha_k e^{-\alpha_k},$$

$$\operatorname{Cov}(Z_k', Y_k') = -\alpha_k e^{-\alpha_k} + \gamma\alpha_k.$$

Put $Y_k = (Y_k' - EY_k')/\alpha^{1/2}$, $Z_k = (Z_k' - EZ_k')/D$ and $Y = n^{-1/2}\sum_k Y_k$, $Z = n^{-1/2}\sum_k Z_k$. Then the characteristic function of (Y_k, Z_k) is

$$\begin{aligned} g_k(w, t) &= E\exp(iY_k' w\alpha^{-1/2} + iZ_k' tD^{-1})\exp(-i\alpha_k w/\alpha^{1/2} - i(e^{-\alpha_k} + \gamma\alpha_k)t/D) \\ &= e^{-\alpha_k}[\exp(\alpha_k e^{iw/\alpha^{1/2} + \gamma it/D}) + e^{it/D} - 1] \\ &\quad \times \exp(-i\alpha_k w/\alpha^{1/2} - i(e^{-\alpha_k} + \gamma\alpha_k)t/D), \end{aligned} \tag{4}$$

and that of (Y, Z) is

$$Ee^{ivY + iuZ} = \textstyle\prod_k g_k(n^{-1/2}v, n^{-1/2}u).$$

Note also that $EY = EZ = 0$, $VY = VZ = 1$, $\operatorname{Cov}(Y, Z) = 0$.

According to a theorem of Bartlett (1938),

$$\begin{aligned} E(e^{iuZ} \mid \textstyle\sum_k Y_k' = N) &= [2\pi P(\textstyle\sum_k Y_k' = N)]^{-1}\displaystyle\int_{-\pi}^{\pi} E\exp\{ix(\textstyle\sum_k Y_k' - N) + iuZ\}\,dx \\ &= c_N \int_{-\pi N^{1/2}}^{\pi N^{1/2}} Ee^{ivY + iuZ}\,dv, \end{aligned} \tag{5}$$

where we have put $v = xN^{1/2}$ and where

$$c_N = \frac{N!e^N}{2\pi N^{N+1/2}} = (2\pi)^{-1/2}e^{O(1/N)}. \tag{6}$$

Furthermore,

$$\begin{aligned} \psi_n(u) &= E\{\exp[iu(\textstyle\sum_k X_k' - \sum_k e^{-\alpha_k})/Dn^{1/2}] \mid \sum_k Y_k' = N\} \\ &= E[e^{iuZ} \mid \textstyle\sum_k Y_k' = N]. \end{aligned} \tag{7}$$

We combine results (5) through (7) as a lemma.

LEMMA 1. *The characteristic function* ψ_n *satisfies*

$$\psi_n(u) = c_N \int_{-\pi N^{1/2}}^{\pi N^{1/2}} \prod_k g_k(n^{-1/2}v, n^{-1/2}u)\, dv.$$

From Lemma 1 and the Mean Value Theorem

$$\psi_n(u) - e^{-u^2/2} = e^{-u^2/2}(e^{u^2/2}\psi_n(u) - 1) = e^{-u^2/2}u\left[\frac{d}{ds}e^{s^2/2}\psi_n(s)\right]_{s=\theta u}$$

$$= ue^{-u^2/2}c_N \int_{-\pi N^{1/2}}^{\pi N^{1/2}} \left[\frac{d}{ds}e^{s^2/2}\prod_k g_k(n^{-1/2}v, n^{-1/2}s)\right]_{s=\theta u} dv,$$

where $0 \le \theta \le 1$. Now write

(8) $$u^{-1}(\psi_n(u) - e^{-u^2/2}) = I_1(u) + I_2(u)$$

where

(9) $$I_1(u) = c_N \int_{|v| \le \varepsilon N^{1/2}} e^{-(u^2+v^2)/2} \left[\frac{d}{ds}e^{v^2/2+s^2/2}\prod_k g_k(n^{-1/2}v, n^{-1/2}s)\right]_{s=\theta u} dv$$

and

(10) $$I_2(u) = c_N \int_{\varepsilon N^{1/2} < |v| < \pi N^{1/2}} e^{-u^2/2} \left[\frac{d}{ds}e^{s^2/2}\prod_k g_k(n^{-1/2}v, n^{-1/2}s)\right]_{s=\theta u} dv,$$

where $0 < \varepsilon < \pi$ will be specified later.

To bound I_1, we need a lemma which is very similar to some results of Sections 8 and 9 of Bhattacharya and Rao (1976) except that the region of interest is a rectangle, which can be far from square if α is large or small. Lemma 2 could only be obtained from their results in the case of fixed or bounded α. The proof is closely related to proofs given in [2] so we defer this to the Appendix.

LEMMA 2. *In the region*

$$R = \left\{(v, s) : |v| < \frac{2}{9}\ell_{1n}^{-1},\ |s| < \frac{2}{9}\ell_{2n}^{-1}\right\}$$

we have

(11) $$\left|\frac{d}{ds}e^{s^2/2+v^2/2}\prod_k g_k(n^{-1/2}v, n^{-1/2}s)\right| \le C_4(|s| + |v| + 1)^3(\ell_{1n} + \ell_{2n})e^{\frac{11}{24}(v^2+s^2)},$$

where $\ell_{1n} = n^{-3/2}\sum_k E|Y_k|^3$, $\ell_{2n} = n^{-3/2}\sum_k E|Z_k|^3$, *whenever*

(12) $$\ell_{1n} < 12^{-3/2}, \quad \ell_{2n} < 12^{-3/2}.$$

To use this result we need bounds on the third moments.

LEMMA 3.

(13) $$\ell_{1n} \le 3C_1 \max(\alpha^{1/2}, 1)/(n\alpha)^{1/2},$$

(14) $$\ell_{2n} \le (34C_1)/Dn^{1/2}$$

and

(15) $$\ell_{1n} + \ell_{2n} \le C_5/Dn^{1/2}.$$

PROOF. Let $\sum'$, $\sum''$ denote the sums over those k for which $\alpha_k < 1$ and $\alpha_k \geq 1$, respectively. Then

$$\textstyle\sum' E\,|\,Y'_k - \alpha_k\,|^3 = \sum' \sum_{j=0}^{\infty} |\,j - \alpha_k\,|^3\, \alpha_k^j e^{-\alpha_k}/j! = \sum' (2\alpha_k^3 e^{-\alpha_k} + \alpha_k) \leq 2 \sum' \alpha_k$$

since $2x^2e^{-x} < 1$ for $0 \leq x \leq 1$, and

$$\textstyle\sum'' E\,|\,Y'_k - \alpha_k\,|^3 \leq \sum'' \{E\,|\,Y'_k - \alpha_k\,|^4\}^{3/4} = \sum'' \{\alpha_k(1 + 3\alpha_k)\}^{3/4} \leq 3 \sum'' \alpha_k^{3/2}.$$

So (13) follows from

$$\ell_{1n} \leq 3(n\alpha)^{-3/2}\,[\textstyle\sum' \alpha_k + \sum'' \alpha_k^{3/2}] \leq 3C_1^{1/2} \max(\alpha^{1/2}, 1)/(n\alpha)^{1/2}.$$

To derive (14), note that we can write

$$nD^2 = \textstyle\sum_k e^{-\alpha_k}(1 - (1 + \alpha_k)e^{-\alpha_k}) + \sum_k \alpha_k(\gamma - e^{-\alpha_k})^2, \tag{16}$$

and

$$Z'_k - EZ'_k = X'_k + e^{-\alpha_k}Y'_k - (1 + \alpha_k)e^{-\alpha_k} + (\gamma - e^{-\alpha_k})(Y'_k - \alpha_k),$$

whence

$$E\,|\,Z'_k - EZ'_k\,|^3 \leq 4E\,|\,X'_k + e^{-\alpha_k}Y'_k - (1 + \alpha_k)e^{-\alpha_k}\,|^3 + 4\,|\,\gamma - e^{-\alpha_k}\,|^3 E\,|\,Y'_k - \alpha_k\,|^3. \tag{17}$$

Now the sum over k of the first expectation on the right in (17) is

$$\textstyle\sum_k [(1 - (1 + \alpha_k)e^{-\alpha_k})^3 e^{-\alpha_k} + \alpha_k^4 e^{-4\alpha_k} + \sum_{j=2}^{\infty} |\,j - 1 - \alpha_k\,|^3 e^{-4\alpha_k}\alpha_k^j/j!] = S_1 + S_2 + S_3,$$

say. We have

$$S_1 \leq \textstyle\sum_k [1 - (1 + \alpha_k)e^{-\alpha_k}]e^{-\alpha_k} \leq nD^2 \tag{18}$$

from (16). Also, since $\alpha_k^2 e^{-2\alpha_k} \leq e^{-2} \leq 1/7$,

$$S_2 \leq \frac{1}{7} \textstyle\sum_k \alpha_k^2 e^{-2\alpha_k} \leq \frac{2}{7}\, nD^2 \tag{19}$$

using $\frac{1}{2}x^2e^{-x} \leq 1 - (1 + x)e^{-x}$ and (16). As for S_3, we have

$$\begin{aligned}
&\textstyle\sum' \sum_{j=2}^{\infty} |\,j - 1 - \alpha_k\,|^3 e^{-4\alpha_k}\alpha_k^j/j! \\
&\quad= \textstyle\sum' e^{-3\alpha_k}[-2\alpha_k - 1 + (1 + \alpha_k)^3 e^{-\alpha_k} + \alpha_k^4 e^{-\alpha_k}] \\
&\quad\leq \textstyle\sum' e^{-3\alpha_k}[-2\alpha_k - 1 + (1 + 2\alpha_k + \alpha_k^2)(1 + \alpha_k)e^{-\alpha_k} + \alpha_k^4 e^{-\alpha_k}] \\
&\quad\leq \textstyle\sum' e^{-4\alpha_k}[\alpha_k^2(1 + \alpha_k) + \alpha_k^4]
\end{aligned}$$

since $(1 + x)e^{-x} \leq 1$. This sum is less than $\sum' \alpha_k^2 e^{-2\alpha_k}$, since $e^{-2x}(1 + x + x^2) \leq 1$ for $x > 0$. Also

$$\begin{aligned}
\textstyle\sum'' \sum_{j=2}^{\infty} |\,j - 1 - \alpha_k\,|^3 e^{-4\alpha_k}\alpha_k^j/j! &= \textstyle\sum'' e^{-3\alpha_k}[E\,|\,Y'_k - 1 - \alpha_k\,|^3 - (1 + \alpha_k)^3 e^{-\alpha_k} - \alpha_k^4 e^{-\alpha_k}] \\
&\leq \textstyle\sum'' e^{-3\alpha_k}[E\,|\,Y'_k - 1 - \alpha_k\,|^4]^{3/4} \\
&= \textstyle\sum'' e^{-3\alpha_k}[\alpha_k(1 + 3\alpha_k) - 4\alpha_k + 6\alpha_k + 1]^{3/4} \\
&\leq 2 \textstyle\sum'' \alpha_k^2 e^{-2\alpha_k},
\end{aligned}$$

so that

$$S_3 \leq 2 \textstyle\sum_k \alpha_k^2 e^{-2\alpha_k} \leq 4nD^2. \tag{20}$$

Next, as in the derivation of (13),

$$\textstyle\sum_k |\,\gamma - e^{-\alpha_k}\,|^3\, E\,|\,Y'_k - \alpha_k\,|^3 \leq 3 \sum' |\,\gamma - e^{-\alpha_k}\,|^3 \alpha_k + 3 \sum'' |\,\gamma - e^{-\alpha_k}\,|^3 \alpha_k^{3/2}.$$

For $\alpha_k \leq 1$,

$$|\,\gamma - e^{-\alpha_k}\,| \leq 1.$$

Now suppose $\alpha_k > 1$. If $\alpha > 1$,

$$\gamma \alpha_k^{1/2} \le C_1^{1/2} \alpha^{-1/2} \le C_1^{1/2}$$

since

$$\gamma = \sum_k \alpha_k e^{-\alpha_k}/n\alpha \le \alpha^{-1}.$$

If $\alpha \le 1$,

$$\gamma \alpha_k^{1/2} \le C_1^{1/2} \alpha^{1/2} \le C_1^{1/2},$$

since $\gamma \le 1$. So

$$|\gamma - e^{-\alpha_k}| \alpha_k^{1/2} \le C_1^{1/2}.$$

Thus

$$\sum_k |\gamma - e^{-\alpha_k}|^3 E|Y_k' - \alpha_k|^3 \le 3C_1^{1/2} \sum_k (\gamma - e^{-\alpha_k})^2 \alpha_k \le 3C_1^{1/2} nD^2. \tag{21}$$

Now (14) follows from (17) through (21).

Finally, since $D^2 \le 1$ and $D^2 \le n^{-1} \sum_k e^{-\alpha_k}(1 - e^{-\alpha_k}) \le n^{-1} \sum_k \alpha_k = \alpha$, the bound in (13) is dominated by

$$3 \max\left(\frac{C_1}{n^{1/2}D}, \frac{1}{n^{1/2}D}\right) \le \frac{3C_1}{n^{1/2}D},$$

and (15) follows from this and (14).

Assume henceforth that

$$\varepsilon \le \begin{cases} \varepsilon_0, & \alpha \le 1, \\ \alpha^{-1/2}\varepsilon_0, & \alpha > 1 \end{cases} \tag{22}$$

with

$$\varepsilon_0 = 2/(27C_1) \quad \text{and} \quad \eta \le \tfrac{2}{9}(34C_1)^{-1}. \tag{23}$$

Lemma 4. *If $Dn^{1/2} > 12^{3/2}C_5$ and (22) and (23) hold, then*

$$\int_0^{\eta n^{1/2}D} |I_1(u)|\, du \le \frac{C_6}{n^{1/2}D}.$$

Proof. Since $|v| < \varepsilon N^{1/2}$, $|s| < \eta n^{1/2}D$, (22) and (23) and Lemma 3 ensure that $(v, s) \in R$ as specified in Lemma 2. So from (9),

$$\int_0^{\eta n^{1/2}D} |I_1(u)|\, du \le c_N \int_0^{\eta n^{1/2}D} \left[\int_{|v|<\varepsilon N^{1/2}} C_4(|u| + |v| + 1)^3 \times \frac{C_5}{n^{1/2}D} e^{-(v^2+u^2)/24}\, dv\right] du$$
$$\le \frac{C_6}{n^{1/2}D}.$$

Lemma 5. *If (23) holds and $\eta = \eta_0 = \varepsilon_0^2/90$, then*

$$\int_0^{\eta_0 n^{1/2}D} |I_2(u)|\, du \le \frac{C_7}{n^{1/2}D}.$$

Proof. First consider the case $\alpha \le 1$. From (4),

$$\begin{aligned} |g_k(n^{-1/2}v, n^{-1/2}s)|^2 = e^{-2\alpha_k}\{&e^{2\alpha_k \cos\tau} + 2[(1 - \cos\zeta)(1 - \cos\delta_k) \\ &+ (e^{\alpha_k \cos\tau} - 1)\cos\delta_k(\cos\zeta - 1) \\ &+ e^{\alpha_k \cos\tau}\sin\delta_k \sin\zeta]\} \end{aligned} \tag{24}$$

where $\tau = N^{-1/2}v + \gamma s/(n^{1/2}D)$, $\zeta = s/n^{1/2}D$ and $\delta_k = \alpha_k \sin \tau$. Since $\eta \le 1/50$ from (23), and since $|\tau| \le \pi + \eta$, we have $\cos \tau - 1 \le -\tau^2/5$. Since

$$\eta < \tfrac{1}{2}\varepsilon_0, \tag{25}$$

and $\gamma < 1$, we have $|\tau| > \frac{1}{2}\varepsilon_0$. So from (24),

$$\begin{aligned}|g_k(n^{-1/2}v, n^{-1/2}s)|^2 &\le e^{-\alpha_k\varepsilon_0^2/10} + 2e^{-2\alpha_k}(\tfrac{1}{4}\eta^2\alpha_k^2 + \tfrac{1}{2}\alpha_k e^{\alpha_k}\eta^2 + e^{\alpha_k}\alpha_k\eta)\\ &\le 1 - \frac{\alpha_k\varepsilon_0^2}{10}\left(1 - \frac{\alpha_k\varepsilon_0^2}{20}\right) + 3\alpha_k\eta\\ &< 1 - \alpha_k\varepsilon_0^2/30,\end{aligned}$$

since $\alpha_k\varepsilon_0^2/20 < \frac{1}{3}$, so long as

$$\eta \le \varepsilon_0^2/90. \tag{26}$$

In this case

$$|g_k(n^{-1/2}v, n^{-1/2}s)| \le e^{-\alpha_k\varepsilon_0^2/60}. \tag{27}$$

Now

$$\begin{aligned}\left|\frac{d}{ds}\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| &\le \textstyle\sum_k n^{-1}[EZ_k^2|s| + E|Y_kZ_k||v|]\,\Pi_{k'\ne k}|g_{k'}(n^{-1/2}v, n^{-1/2}s)|\\ &\le \textstyle\sum_k n^{-1}[EZ_k^2|s| + E|Y_kZ_k||v|]\exp\left\{-\frac{1}{60}\varepsilon_0^2(n\alpha - \alpha_k)\right\}\\ &\le (|s| + |v|)\exp\left\{-\frac{1}{60}\varepsilon_0^2(n\alpha - C_1)\right\}\\ &\le C_8(n\alpha)^{1/2}\exp\left\{-\frac{1}{60}\varepsilon_0^2 n\alpha\right\}.\end{aligned} \tag{28}$$

So

$$\begin{aligned}\left|\frac{d}{ds}e^{1/2s^2}\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| &\le |s|e^{s^2/2}e^{-n\alpha\varepsilon_0^2/60} + C_8 e^{s^2/2}(n\alpha)^{1/2}e^{-n\alpha\varepsilon_0^2/60}\\ &\le C_9 e^{s^2/2 - n\alpha\varepsilon_0^2/60}(n\alpha)^{1/2},\end{aligned}$$

and from (10),

$$\begin{aligned}\int_0^{\eta n^{1/2}D}|I_2(u)|\,du &\le c_N\int_{u=0}^{\eta n^{1/2}D}\int_{v=0}^{\pi N^{1/2}} C_9(n\alpha)^{1/2}e^{-n\alpha\varepsilon_0^2/60}\,du\,dv\\ &\le C_{10}/(n\alpha)^{1/2} \le C_{10}/(n^{1/2}D).\end{aligned} \tag{29}$$

Now suppose $\alpha > 1$. From (4)

$$|g_k(vn^{-1/2}, sn^{-1/2})| \le e^{-\alpha_k}(e^{\alpha_k\cos\tau} + \eta). \tag{30}$$

Since $|v| > \varepsilon_0 n^{1/2}$, $\eta < \frac{1}{2}\varepsilon_0$ and $\gamma < \alpha^{-1} < \alpha^{-1/2}$,

$$\left|\frac{\gamma s}{n^{1/2}D}\right| < \gamma\eta < \frac{1}{2}\gamma\varepsilon_0 < \frac{1}{2}|v|(n\alpha)^{-1/2},$$

so $|\tau| > \frac{1}{2}|v|(n\alpha)^{-1/2}$. Thus if $\alpha_k \ge \frac{1}{2}\alpha$,

$$\begin{aligned}|g_k(n^{-1/2}v, n^{-1/2}s)| &\le e^{-\alpha_k v^2/(20N)} + \eta e^{-\alpha_k}\\ &= e^{-\alpha_k v^2/(40N)}\{e^{-\alpha_k v^2/(40N)} + \eta e^{-\alpha_k(1 - v^2/(40N))}\}\\ &\le e^{-\alpha_k v^2/(40N)}\{e^{-v^2/(80n)} + \eta e^{-\alpha_k(1-\pi^2/40)}\}.\end{aligned}$$

In this case, then,

$$|g_k(n^{-1/2}v, n^{-1/2}s)| \le e^{-v^2/(80n)} \tag{31}$$

so long as

$$\eta < (1 - e^{-\varepsilon_0^2/80})e^{1/2(1-\pi^2/40)}. \tag{32}$$

Now set η equal to $\eta_0 = \varepsilon_0^2/90$, so that (32), and *a fortiori* (23), (25) and (26), hold.
Let m be the number of $\alpha_k \ge \frac{1}{2}\alpha$. Then

$$\tfrac{1}{2}(n-m)\alpha + mC_1\alpha \ge \tfrac{1}{2}(n-m)\alpha + \textstyle\sum_{\alpha_k \ge \alpha/2} \alpha_k \ge n\alpha$$

so that $m \ge n/(2C_1 - 1)$. Using this, (28) and (31), we get

$$\begin{aligned}\left|\frac{d}{ds} e^{s^2/2}\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| &\le (2|s| + |v|)e^{s^2/2}e^{-v^2/[80(2C_1-1)]}\\ &\le 2|v|e^{s^2/2}e^{-C_{11}v^2},\end{aligned}$$

say, since $|s| \le \eta_0 n^{1/2}D \le \frac{1}{2}\varepsilon_0 n^{1/2} \le \frac{1}{2}|v|$. So from (10),

$$\int_0^{\eta_0 n^{1/2}D} |I_2(u)|\,du \le c_N \int_{u=0}^{\eta_0 n^{1/2}D} \int_{|v|>\varepsilon_0 n^{1/2}} 2|v|e^{-C_{11}v^2}\,du\,dv \le \frac{C_{12}}{Dn^{1/2}}.$$

The lemma follows on setting $C_7 = \max(C_{10}, C_{12})$.
It follows from (8) and Lemmas 4 and 5 that (3) is dominated by

$$2\pi^{-1}(C_6 + C_7 + 24(2/\pi)^{-1/2}\eta_0^{-1})/(n^{1/2}D) = \frac{C_{13}}{n^{1/2}D},$$

so long as $Dn^{1/2} > 12^{3/2}C_5$. The constant C_{13} can be increased if necessary to a value C_2 so that (1) is true for all values of $n^{1/2}D$.

APPENDIX

PROOF OF LEMMA 2. The lemma will be proved for arbitrary independent pairs of random variables (Y_k, Z_k), with $EY_k = EZ_k = 0$, $E|Y_k|^3 < \infty$, $E|Z_k|^3 < \infty$, $\sum_k EY_k^2 = \sum_k EZ_k^2 = n$ and $\sum_k EY_kZ_k = 0$, with ℓ_{1n}, ℓ_{2n} as defined in the lemma. Let

$$R_1 = \{(v, s) : |v| < \tfrac{1}{3}\ell_{1n}^{-1/3}, |s| < \tfrac{1}{3}\ell_{2n}^{-1/3}\},$$

and let $R_2 = R - R_1$. First consider R_1. Let

$$h(v, s) = \textstyle\sum_k \log g_k(n^{-1/2}v, n^{-1/2}s) + \frac{1}{2}v^2 + \frac{1}{2}s^2.$$

Then

$$\left|\frac{d}{ds} e^{v^2/2+s^2/2}\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| \le \left|\frac{d}{ds} h(v, s)\right| e^{|h(v,s)|}. \tag{A1}$$

Now

$$\begin{aligned}|g_k(n^{-1/2}v, n^{-1/2}s) - 1| &\le \tfrac{1}{2}n^{-1}E(vY_k + sZ_k)^2\\ &\le \tfrac{1}{2}(n^{-3/2}\textstyle\sum_k E|vY_k + sZ_k|^3)^{2/3}\\ &\le 2^{1/3}(|v|^3\ell_{1n} + |s|^3\ell_{2n})^{2/3}.\end{aligned} \tag{A2}$$

So in R_1,

$$|g_k(n^{-1/2}v, n^{-1/2}s) - 1| \le \tfrac{1}{2}. \tag{A3}$$

Since $|\log(1+x) - x| \le x^2$ for $|x| \le \frac{1}{2}$,

$$\begin{aligned}|h(v, s)| \le \textstyle\sum_k |g_k(n^{-1/2}v, n^{-1/2}s) - 1 - \frac{1}{2}n^{-1}EY_k^2v^2 - \frac{1}{2}n^{-1}EZ_k^2s^2|\\ + \textstyle\sum_k |g_k(n^{-1/2}v, n^{-1/2}s) - 1|^2.\end{aligned}$$

From (A2)

$$|g_k(n^{-1/2}v, n^{-1/2}s) - 1|^2 \le \tfrac{1}{4}(4|v|^3\ell_{1n} + 4|s|^3\ell_{2n})^{1/3}(n^{-3/2}E|vY_k + sZ_k|^3)$$
$$\le \tfrac{1}{6}n^{-3/2}E|vY_k + sZ_k|^3.$$

Also

$$\textstyle\sum_k |g_k(n^{-1/2}v, n^{-1/2}s) - 1 - \tfrac{1}{2}n^{-1}EY_k^2v^2 - \tfrac{1}{2}n^{-1}EZ_k^2s^2| \le \tfrac{1}{6}n^{-3/2}\sum_k E|vY_k + sZ_k|^3.$$

So for $(v, s) \in R_1$,

$$|h(v, s)| \le \tfrac{1}{3}n^{-3/2}\textstyle\sum_k E|vY_k + sZ_k|^3 \le \tfrac{8}{27}(v^2 + s^2) \le \tfrac{11}{24}(v^2 + s^2). \tag{A4}$$

Further

$$\left|\frac{d}{ds}h(v, s)\right| \le A_1 + A_2$$

where

$$A_1 = \left|\textstyle\sum_k \dfrac{d}{ds}g_k(n^{-1/2}v, n^{-1/2}s) + s\right|$$
$$= |\textstyle\sum_k EiZ_kn^{-1/2}(e^{ivY_kn^{-1/2}+isZ_kn^{-1/2}} - 1 - ivY_kn^{-1/2} - isZ_kn^{-1/2})|$$
$$\le \frac{1}{2}n^{-3/2}\textstyle\sum_k E|Z_k||vY_k + sZ_k|^2$$
$$\le (v^2 + s^2)(\ell_{1n} + \ell_{2n})$$

and

$$A_2 = \left|\textstyle\sum_k [g_k(n^{-1/2}v, n^{-1/2}s)]^{-1}[1 - g_k(n^{-1/2}v, n^{-1/2}s)]\dfrac{d}{ds}g_k(n^{-1/2}v, n^{-1/2}s)\right|$$
$$\le n^{-3/2}\textstyle\sum_k |EiZ_k(e^{ivY_kn^{-1/2}+isZ_kn^{-1/2}} - 1)|E(vY_k + sZ_k)^2$$

from (A2) and (A3). Thus

$$A_2 \le n^{-2}\textstyle\sum_k (|v|E|Y_kZ_k| + |s|EZ_k^2)E(vY_k + sZ_k)^2$$
$$\le n^{-2}\textstyle\sum_k E^{1/2}Z_k^2(|v|E^{1/2}Y_k^2 + |s|E^{1/2}Z_k^2)^3$$
$$\le 4(n^{-1}\max_k EZ_k^2)^{1/2}n^{-3/2}(|v|^3 + |s|^3)(\textstyle\sum_k E|Y_k|^3 + \sum_k E|Z_k|^3)$$
$$\le 4(|v|^3 + |s|^3)(\ell_{1n} + \ell_{2n}),$$

since $n^{-1}\max_k EZ_k^2 \le n^{-1}\sum_k EZ_k^2 = 1$. So

$$\left|\frac{d}{ds}h(v, s)\right| \le 4(|v| + |s| + 1)^3(\ell_{1n} + \ell_{2n}). \tag{A5}$$

Substituting (A4) and (A5) in (A1) gives (11) in R_1.

Now consider R_2. We have

$$\left|\frac{d}{ds}e^{v^2/2+s^2/2}\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| \le e^{v^2/2+s^2/2}\textstyle\sum_{k'}|\Pi_{k\ne k'}g_k(n^{-1/2}v, n^{-1/2}s)|$$
$$\times\left|\frac{d}{ds}g_{k'}(n^{-1/2}v, n^{-1/2}s)\right| + |s|e^{v^2/2+s^2/2}|\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)|. \tag{A6}$$

Now

$$|g_k(n^{-1/2}v, n^{-1/2}s)| \le 1 - \tfrac{1}{2}n^{-1}E(vY_k + sZ_k)^2 + \tfrac{3}{8}n^{-3/2}E|vY_k + sZ_k|^3, \tag{A7}$$
$$\le \exp[-\tfrac{1}{2}n^{-1}E(vY_k + sZ_k)^2 + \tfrac{3}{8}n^{-3/2}E|vY_k + sZ_k|^3],$$

since for $n^{-1}E(vY_k + sZ_k)^2 < 2$,

$$|g_k(n^{-1/2}v, n^{-1/2}s)| \le 1 - \tfrac{1}{2}n^{-1}E(vY_k + sZ_k)^2 + \tfrac{1}{6}n^{-3/2}E|vY_k + sZ_k|^3$$

and for $n^{-1}E(vY_k + sZ_k)^2 \ge 2$,

$$n^{-3/2}E|vY_k + sZ_k|^3 \ge (n^{-1}E(vY_k + sZ_k)^2)^{3/2} \ge \tfrac{4}{3}n^{-1}E(vY_k + sZ_k)^2,$$

so the right hand side of (A7) is greater than 1. Thus in R,

$$|\Pi_{k\neq k'}g_k(n^{-1/2}v, n^{-1/2}s)|$$
$$\le \exp[-\tfrac{1}{2}v^2 - \tfrac{1}{2}s^2 + \tfrac{3}{2}(|v|^3\ell_{1n} + |s|^3\ell_{2n}) + \tfrac{1}{2}n^{-1}E(vY_{k'} + sZ_{k'})^2]$$

and

$$\tfrac{1}{2}n^{-1}E(vY_{k'} + sZ_{k'})^2 \le v^2\ell_{1n}^{2/3} + s^2\ell_{2n}^{2/3} < \tfrac{1}{12}(v^2 + s^2)$$

from (12). So

$$\text{(A8)} \qquad |\Pi_{k\neq k'}g_k(n^{-1/2}v, n^{-1/2}s)| \le e^{-(v^2+s^2)/12},$$

and so clearly,

$$\text{(A9)} \qquad |\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)| \le e^{-(v^2+s^2)/12}.$$

Now

$$\text{(A10)} \qquad \begin{aligned}\textstyle\sum_k \left|\frac{d}{ds}g_k(n^{-1/2}v, n^{-1/2}s)\right| &= \textstyle\sum_k |EiZ_kn^{-1/2}(e^{ivY_kn^{-1/2}+isZ_kn^{-1/2}} - 1)| \\ &\le \textstyle\sum_k (|s|n^{-1}EZ_k^2 + |v|n^{-1}E|Y_kZ_k|) \le |v| + |s|.\end{aligned}$$

Substituting (A8), (A9), (A10) in (A6) gives

$$\begin{aligned}\left|\frac{d}{ds}e^{v^2/2+s^2/2}\,\Pi_k g_k(n^{-1/2}v, n^{-1/2}s)\right| &\le (|v| + 2|s|)e^{5(v^2+s^2)/12} \\ &\le C_4(|v| + |s| + 1)^3(\ell_{1n} + \ell_{2n})e^{11(v^2+s^2)/24}\end{aligned}$$

for (v, s) in R_2, which gives (11) in R_2 immediately.

REFERENCES

[1] Bartlett, M. S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc.* **13** 62–67.

[2] Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal approximation and asymptotic expansions.* Wiley, New York.

[3] Englund, G. (1981). A remainder term estimate for the normal approximation in classical occupancy. *Ann. Probability* **9** 684–692.

[4] Holst, L. (1971). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* **59** 137–145.

[5] Loève, M. (1955). *Probability theory.* Van Nostrand, New York.

Department of Mathematical Statistics
University of Sydney
N.S.W. 2006
Australia