The Annals of Probability
1981, Vol. 9, No. 4, 684–692

A REMAINDER TERM ESTIMATE FOR THE NORMAL APPROXIMATION IN CLASSICAL OCCUPANCY

By Gunnar Englund

Royal Institute of Technology

Let balls be thrown successively at random into N boxes, such that each ball falls into any box with the same probability $1/N$. Let Z_n be the number of occupied boxes (i.e., boxes containing at least one ball) after n throws. It is well known that Z_n is approximately normally distributed under general conditions. We give a remainder term estimate, which is of the correct order of magnitude. In fact we prove that

$$0.087/\max(3, DZ_n) \leqq \sup_x | P(Z_n < x) - \Phi((x - EZ_n)/DZ_n)| \leqq 10.4/DZ_n.$$

1. Introduction and formulation of main results. Balls are thrown successively at random into N boxes, such that each ball falls into any box with the same probability $1/N$, independently of what happens to the other balls. Set

$$Z_n = \text{the number of occupied boxes (i.e., boxes containing at least one ball) after } n \text{ throws.} \tag{1.1}$$

The problem of finding the distribution of Z_n, called the (classical) occupancy problem, has been treated extensively in the literature, and we refer to Johnson and Kotz (1977) for an account. In particular we have the following formulas (D^2 denoting variance)

$$EZ_n = N\{1 - (1 - 1/N)^n\}, \tag{1.2}$$

$$D^2Z_n = N(1 - 1/N)^n - N(1 - 2/N)^n - N^2(1 - 1/N)^{2n} + N^2(1 - 2/N)^n. \tag{1.3}$$

It is well known that Z_n is approximately normally distributed under general conditions, see Weiss (1958), Rényi (1962) or Johnson and Kotz (1977) Chapter 6. Our main aim in this paper is to derive the following estimates for the accuracy of the normal distribution approximation. Φ denotes as customary the normal distribution function.

THEOREM 1. *For $k = 1, 2, 3, \cdots, N$, $n = 1, 2, \cdots$ we have*

$$\text{(a) } \sup_k | P(Z_n < k) - \Phi((k - EZ_n)/DZ_n)| \leqq 10/DZ_n \tag{1.4}$$

$$\text{(b) } \sup_{-\infty<x<\infty} | P(Z_n < x) - \Phi((x - EZ_n)/DZ_n)| \leqq 10.4/DZ_n \tag{1.5}$$

$$\text{(c) } \sup_x | P(Z_n < x) - \Phi((x - EZ_n)/DZ_n)| \geqq 0.087/\max(3, DZ_n) \tag{1.6}$$

Note that (1.6) shows that the bounds in (1.4) and (1.5) are of the correct order of magnitude in DZ_n. Next we comment on the "strength" of the inequalities (1.5) and (1.6) as a tool for establishing asymptotic normality of Z_n. Consider a sequence of ball throwing situations indexed by i, $i = 1, 2, \cdots$. We use the convention that an index i attached to a quantity means that it relates to situation i. In particular N_i denotes the number of boxes in sitututation i.

COROLLARY TO THEOREM 1. *We have, where $\mathscr{L}$ denotes distribution and $\to_{\mathscr{D}}$ denotes*

Received August 9, 1979; revised March 31, 1980.
AMS 1970 *subject classifications*. 60F05.
Key words and phrases. Classical occupancy, normal approximation, remainder term.

convergence in distribution

$$\mathscr{L}((Z_{n_i} - EZ_{n_i})/DZ_{n_i}) \to_{\mathscr{D}} N(0, 1) \tag{1.7}$$

if and only if

$$DZ_{n_i} \to \infty \quad \text{as} \quad i \to \infty. \tag{1.8}$$

The result in the corollary was proved by Rényi (1962). We would also like to point to the remainder term estimates presented in Kolchin (1966). Throughout this paper we will use the notation

$$p = \frac{n}{N}. \tag{1.9}$$

2. Basic ideas in the proof of Theorem 1. First note that (1.5) follows from (1.4) and the simple inequality

$$\sup_{|x-y|\leqq 1} |\Phi(x/DZ_n) - \Phi(y/DZ_n)| \leqq (2\pi)^{-1/2}/DZ_n. \tag{2.1}$$

We therefore concentrate on (1.4). When k is sufficiently far away from EZ_n the inequality (1.4) is almost trivial. Assume that

$$k \leqq EZ_n - \frac{1}{\sqrt{10}}(DZ_n)^{3/2}. \tag{2.2}$$

Then Chebyshev's inequality easily yields that $P(Z_n < k) \leqq 10/DZ_n$. Furthermore the inequality $\Phi(-a) \leqq a^{-2}$, $a > 0$, yields that $\Phi((k - EZ_n)/DZ_n) \leqq 10/DZ_n$. By combining these estimates we see that (1.4) is true if (2.2) holds. The case

$$k \geqq EZ_n + \frac{1}{\sqrt{10}}(DZ_n)^{3/2} \tag{2.3}$$

can be treated quite analogously.

Hence it suffices to prove (1.4) for k:s such that

$$|(k - EZ_n)/DZ_n| \leqq \sqrt{DZ_n}/\sqrt{10} \tag{2.4}$$

Following an idea from Rényi (1962) we introduce the following random variables. We let

(2.5) U_j be the time, which the increasing sequence $Z_0, Z_1, Z_2, \cdots$ spends in the state "exactly j boxes are occupied", $j = 0, 1, 2, \cdots, N-1$.

Then

(2.6) $U_0, U_1, U_2, \cdots, U_{N-1}$ are independent random variables

$$P(U_j = m) = \left(1 - \frac{j}{N}\right)\left(\frac{j}{N}\right)^{m-1}, \qquad m = 1, 2, \cdots, j = 0, 1, \cdots, N-1, \tag{2.7}$$

i.e., $U_j - 1$ has a geometric distribution with parameter $(1 - j/N)$. Set

$$V_k = \textstyle\sum_{j=0}^{k-1} U_j, \qquad k = 1, 2, \cdots, N. \tag{2.8}$$

The interpretation of V_k is that V_k is the time needed to reach the state "exactly k boxes are occupied". We have, and this is the desired representation,

$$P(Z_n \geqq k) = P(V_k \leqq n). \tag{2.9}$$

From (2.9) we get

$$\begin{aligned}|P(Z_n < k) - \Phi((k - EZ_n)/DZ_n)| &\leqq |P(V_k > n) - \Phi((EV_k - n)/DV_k)| \\ &\quad + |\Phi((EV_k - n)/DV_k) - \Phi((k - EZ_n)/DZ_n)|.\end{aligned} \tag{2.10}$$

Our program is to show that both terms to the right in (2.10) are dominated by quantities of the type C/DZ_n.

We shall therefore need various estimates of the moments of the Z-, V- and U-variables, and these will be derived in the subsequent Sections 3 and 4. The proof of (1.4) will be concluded in Section 5. Section 6 contains the surprisingly easy proof of (1.6).

3. Estimates of EZ_n and DZ_n. We first note that in proving (1.4) we can without loss of generality assume

$$D^2Z_n \geqq 100, \tag{3.1}$$

since if (3.1) does not hold then (1.4) is trivial. From (1.3) we see that $D^2Z_n \leqq N(1-1/N)^n - N^2((1-1/N)^{2n} - (1-2/N)^n) \leqq Ne^{-p}$, i.e., (3.1) implies

$$Ne^{-p} \geqq 100, \tag{3.2}$$

$$N \geqq 100. \tag{3.3}$$

Our main aim in this section is to prove the following two lemmas.

LEMMA 3.1. *Define $r_1(n, N)$ by the relation*

$$EZ_n = N(1-e^{-p}) + r_1(n, N). \tag{3.4}$$

Then if (3.3) *holds we have*

$$0 \leqq r_1(n, N) \leqq 0.511\, pe^{-p}. \tag{3.5}$$

LEMMA 3.2. *Define $r_2(n, N)$ by the relation*

$$D^2Z_n = Ne^{-p}\{1 - e^{-p}(1+p)\}\{1 + r_2(n, N)\}. \tag{3.6}$$

Then if (3.3) *holds we have*

$$|r_2(n, N)| \leqq \frac{6.13\, pe^{-p}}{Ne^{-p}\{1 - e^{-p}(1+p)\}}. \tag{3.7}$$

If (3.1) *holds we have*

$$|r_2(n, N)| \leqq 0.024. \tag{3.8}$$

PROOF OF LEMMA 3.1. By Taylor expansion we have (using $1 - x = \exp(\log(1-x))$) $e^{-nx} \geqq (1-x)^n \geqq \exp(-nx - \tfrac{1}{2} nx^2(1-x)^{-2})$, $0 \leqq x < 1$. This together with the inequality $0 \leqq 1 - e^{-y} \leqq y$, $y \geqq 0$ yields

$$0 \leqq e^{-nx} - (1-x)^n \leqq \frac{nx^2}{2(1-x)^2} e^{-nx}, \qquad 0 \leqq x < 1,\ n = 0, 1, 2, \cdots \tag{3.9}$$

With $x = 1/N$ and observing (1.2) and (3.3) we obtain

$$0 \leqq r_1(n, N) = N\left(e^{-p} - \left(1 - \frac{1}{N}\right)^n\right) \leqq \frac{pe^{-p}}{2(1-0.01)^2} \leqq 0.511\, pe^{-p}. \tag{3.10}$$

PROOF OF LEMMA 3.2. By (1.3) and (3.6) we have

$$\begin{aligned} Ne^{-p}\{1 - e^{-p}(1+p)\}r_2(n, N) &= N\left\{\left(1 - \frac{1}{N}\right)^n - e^{-p}\right\} + N\left\{e^{-2p} - \left(1 - \frac{2}{N}\right)^n\right\} \\ &\quad + N\left(pe^{-2p} - N\left(1 - \frac{1}{N}\right)^{2n} + N\left(1 - \frac{2}{N}\right)^n\right). \end{aligned} \tag{3.11}$$

By Taylor expansion we have (using $1 - x = \exp(\log(1-x))$)

$$\exp\left(-nx-\frac{1}{2}nx^2-\frac{nx^3}{3(1-x)^3}\right) \leqq (1-x)^n \leqq \exp\left(-nx-\frac{1}{2}nx^2\right), \quad 0 \leqq x \leqq 1, \tag{3.12}$$

which together with the inequality $0 \leqq 1 - e^{-y} \leqq y$, $y \geqq 0$ yields

$$\begin{aligned} 0 &\leqq \exp\left(-nx-\frac{1}{2}nx^2\right) - (1-x)^n \\ &\leqq \exp\left(-nx-\frac{1}{2}nx^2\right)\left(1-\exp\left(-\frac{nx^3}{3(1-x)^3}\right)\right) \\ &\leqq \frac{nx^3}{3(1-x)^3}\exp(-nx). \end{aligned} \tag{3.13}$$

By using (3.9) in the first two terms in (3.11) and (3.13) twice on the last one (with $x = 1/N$ and $2/N$) we obtain (note (3.3))

$$\begin{aligned} Ne^{-p}\{1-e^{-p}(1+p)\}|\, r_2(n, N)\,| &\leqq \frac{pe^{-p}}{2(1-0.01)^2} + \frac{2pe^{-2p}}{(1-0.02)^2} \\ &\quad + N^2\left(\frac{8}{3\times 0.98^3}\frac{n}{N^3}e^{-2p} + \frac{2}{3\times 0.99^3}\frac{n}{N^3}e^{-2p}\right) \\ &\quad + N^2\left|\frac{p}{N} + \exp\left(-\frac{2p}{N}\right) - \exp\left(-\frac{p}{N}\right)\right| e^{-2p}. \end{aligned} \tag{3.14}$$

By the simple inequality $|\,u + e^{-2u} - e^{-u}| \leqq 2u^2$, $u \geqq 0$ (which is easily proved by Taylor expansion) where we set $u = p/N$

$$|\, r_2(n, N)\,| \leqq \frac{(0.511 + 5.61\, e^{-p} + 2pe^{-p})pe^{-p}}{Ne^{-p}\{1-e^{-p}(1+p)\}} \leqq \frac{6.13\, pe^{-p}}{Ne^{-p}\{1-e^{-p}(1+p)\}} \tag{3.15}$$

proving (3.7). By inserting this in (3.6) and using $pe^{-p} \leqq e^{-1}$ and (3.1) we get

$$Ne^{-p}\{1-e^{-p}(1+p)\} \geqq D^2Z_n - 6.13pe^{-p} \geqq 100 - \frac{6.13}{e} \geqq 97.7. \tag{3.16}$$

By combining (3.16) and (3.7) we get (again using $pe^{-p} \leqq e^{-1}$)

$$|\, r_2(n, N)| \leqq \frac{6.13}{97.7e} \leqq 0.024 \tag{3.17}$$

□

4. Some estimates of the moments of the U- and V-variables. Throughout this section U and V denote the random variables which were defined in (2.6)–(2.8), (2.11) and (2.12).

LEMMA 4.1. *For $k < N$ define $r_3(k, N)$ and $r_4(k, N)$ by*

$$EV_k = -N\log(1-k/N) - r_3(k, N), \tag{4.1}$$

$$D^2V_k = N\{k/(N-k) + \log(1-k/N)\} - r_4(k, N). \tag{4.2}$$

Then we have

$$0 \leqq r_3(k, N) \leqq ½\, k/(N-k), \tag{4.3}$$

$$0 \leqq r_4(k, N) \leqq ½\, Nk/(N-k)^2. \tag{4.4}$$

LEMMA 4.2. *For $k < N$ we have*

$$\textstyle\sum_{j=0}^{k-1} E\,|\, U_j - EU_j|^3 \leqq 2\dfrac{Nk^2}{(N-k)^2}. \tag{4.5}$$

PROOFS. By (2.7) we have the following well known formulas

$$EU_j = (1 - j/N)^{-1}, \qquad D^2U_j = (j/N)(1 - j/N)^{-2}. \tag{4.6}$$

From (2.8) and (4.6) we get

$$EV_k = \textstyle\sum_{j=0}^{k-1} (1 - j/N)^{-1} = N \sum_{\nu=N-k+1}^{N} 1/\nu, \tag{4.7}$$

while (2.8), (2.6) and (4.6) yield

$$D^2V_k = \textstyle\sum_{j=0}^{k-1} (j/N)(1 - j/N)^{-2} = N \sum_{\nu=N-k+1}^{N} (N - \nu)/\nu^2. \tag{4.8}$$

By applying the following two estimates (which are easily proved by the Euler-McLaurin summation formula) we obtain (4.3) and (4.4).

$$0 \leqq \log\left(\frac{m}{n}\right) - \textstyle\sum_{\nu=n+1}^{m} 1/\nu \leqq \frac{1}{2}\left(\frac{1}{n} - \frac{1}{m}\right), \qquad 1 \leqq n < m. \tag{4.9}$$

$$0 \leqq m\left(\frac{1}{n} - \frac{1}{m}\right) - \log\left(\frac{m}{n}\right) - \textstyle\sum_{\nu=n+1}^{m} (m - \nu)/\nu^2 \leqq \frac{1}{2}(m - n)/n^2, \quad 1 \leqq n < m. \tag{4.10}$$

The estimate (4.5) follows analogously from the following results (4.11) and (4.12)

$$\textstyle\sum_{\nu=n+1}^{m} (m - \nu)/\nu^3 \leqq \frac{1}{2} m\left(\frac{1}{n} - \frac{1}{m}\right)^2, \qquad 1 \leqq n < m. \tag{4.11}$$

If the random variable X has the distribution $P(X = k) = \rho(1 - \rho)^{k-1}$, $k = 1, 2, \cdots 0 < \rho \leqq 1$ then

$$E\,|\,X - EX\,|^3 \leqq 4\,\frac{1 - \rho}{\rho^3}. \tag{4.12}$$

The estimate (4.12) follows since $\left(\text{note } EX = \frac{1}{\rho},\ E(\rho X)^2 = 2 - \rho,\ E(\rho X)^3 = 6 - 6\rho + \rho^2\right)$

$$\begin{aligned} E\,|\,\rho X - 1\,|^3 &\leqq E(\rho X + 1)(\rho X - 1)^2 \\ &= E(\rho X)^3 - E(\rho X)^2 - E(\rho X) + 1 = (1 - \rho)(4 - \rho). \end{aligned} \tag{4.13}$$

(In fact it can be shown by tedious calculations that the constant 4 in (4.12) can be replaced by $(12/e - 2)$, and that this constant is the best possible.) Hence Lemma 4.1 and Lemma 4.2 are proved. □

Thereby we have obtained our basic moment estimates. However, as stated in Section 2, we are interested in the behaviour of EV_k, DV_k and other quantities when k lies in the vicinity of EZ_n, more precisely, cf. (2.4), when k satisfies

$$k = EZ_n + \theta(DZ_n)^{3/2}, \qquad |\,\theta\,| \leqq 1/\sqrt{10} \tag{4.14}$$

Our next task will therefore be to recast the estimates in Lemmas 4.1 and 4.2 for k:s of the type (4.14). We continue to use the notation

$$p = n/N. \tag{4.15}$$

LEMMA 4.3. *Let k satisfy* (4.14) *and define $r_5(n, N, \theta)$ and $r_6(n, N, \theta)$ by*

$$EV_k = n + \theta N(e^p/N)^{1/4}\{1 - e^{-p}(1 + p)\}^{3/4} + r_5(n, N, \theta), \tag{4.16}$$

$$D^2V_k = Ne^p\{1 - e^{-p}(1 + p)\}\{1 + r_6(n, N, \theta)\}. \tag{4.17}$$

Then, if also (3.1) *is satisfied we have*

$$|\,r_5(n, N, \theta)| \leqq 1.63e^p(1 - e^{-p}) + 0.66\theta^2\sqrt{N}\,e^{p/2}\{1 - e^{-p}(1 + p)\}^{3/2}, \tag{4.18}$$

$$|\,r_6(n, N, \theta)| \leqq 2.29\,|\,\theta\,|\{N(1 - e^{-p}(1 + p)\}^{-1/4} + 1.08\{Ne^{-p}(1 - e^{-p}(1 + p))\}^{-1}, \tag{4.19}$$

$$|\,r_6(n, N, \theta)| \leqq 0.242 \tag{4.20}$$

LEMMA 4.4. *Let k satisfy* (4.14) *and assume that* (3.1) *holds. Then*

$$\sum_{j=0}^{k-1} E\,|\,U_j - EU_j\,|^3 \leqq 2.63Ne^{2p}(1-e^{-p})^2. \tag{4.21}$$

To prove these lemmas we shall need the following auxiliary results.

LEMMA 4.5. *Let k be as in* (4.14) *and define $r_7(n, N, \theta)$ and $r_8(n, N, \theta)$ by*

$$1 - k/N = e^{-p}\{1 + r_7(n, N, \theta)\}, \tag{4.22}$$

$$k/N = (1-e^{-p})\{1 + r_8(n, N, \theta)\}. \tag{4.23}$$

Then

$$r_7(n, N, \theta) = -\theta(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{3/4}\{1 + r_2(n, N)\}^{3/4} - e^p r_1(n, N)/N. \tag{4.24}$$

If also (3.1) *is satisfied we have*

$$|\,r_7(n, N, \theta)\,| \leqq 1.018\,|\,\theta\,|(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{3/4} + 0.511\frac{p}{N}, \tag{4.25}$$

$$|\,r_7(n, N, \theta)| \leqq 0.107, \tag{4.26}$$

$$|\,r_8(n, N, \theta)| \leqq |\,r_7(n, N, \theta)|/(e^p - 1), \tag{4.27}$$

$$|\,r_8(n, N, \theta)| \leqq 0.024. \tag{4.28}$$

PROOF OF LEMMA 4.5. By using (4.14), (3.4) and (3.6) we get

$$\begin{aligned} k/N = EZ_n/N + \theta(DZ_n)^{3/2}/N = 1 - e^{-p} + r_1/N \\ + \theta N^{3/4}e^{-3p/4}\{1 - e^{-p}(1+p)\}^{3/4}(1+r_2)^{3/4}/N. \end{aligned} \tag{4.29}$$

Hence

$$1 - k/N = e^{-p}(1 - \theta(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{3/4}(1+r_2)^{3/4} - e^p r_1/N) \tag{4.30}$$

proving (4.24). By using (3.8) and (3.5) we obtain (4.25). To obtain (4.26) dominate p/N by e^p/N and apply (4.14) and (3.2). Turning to (4.27) and (4.28) we note that by (4.22) and (4.23) we have

$$r_8 = -\,r_7/(e^p - 1). \tag{4.31}$$

which proves (4.27). In virtue of (4.27), (4.25) and the fact that $p \leqq e^p(1 - e^{-p})$

$$|\,r_8\,| \leqq 1.018\,|\,\theta\,|(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{-1/4}\frac{1 - e^{-p}(1+p)}{e^p - 1} + \frac{0.511}{N} \tag{4.32}$$

By using the simple estimate $(1 - e^{-p}(1+p))/(e^p - 1) \leqq \frac{1}{2}e^{-1}$, $p > 0$ and the same reasoning as above we easily obtain (4.28). □

PROOF OF LEMMA 4.3. By inserting (4.22) into (4.1) we get

$$EV_k = -N\log(e^{-p}\{1 + r_7\}) - r_3 = n - Nr_7 + N\{r_7 - \log(1 + r_7)\} - r_3. \tag{4.33}$$

Hence by (4.24), the inequality $|\,x - \log(1+x)\,| \leqq \frac{1}{2}x^2/(1 - |\,x\,|)$, $|\,x\,| < 1$ and (4.26)

$$\begin{aligned} |\,r_5\,| \leqq N\,|\,\theta\,|(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{3/4}\,|(1 + r_2)^{3/4} - 1\,| + e^p r_1 \\ + \tfrac{1}{2}N(r_7)^2(1 - 0.107)^{-1} + |\,r_3\,|. \end{aligned} \tag{4.34}$$

By (3.7), (3.8), the inequality $|(1 + x)^{3/4} - 1| \leqq \frac{3}{4}\,|\,x\,|(1 - |\,x\,|)^{-1/4}$, $|\,x\,| < 1$, (3.5), (4.3)

$$\begin{aligned} |\,r_5\,| \leqq N\,|\,\theta\,|(e^p/N)^{1/4}\{1 - e^{-p}(1+p)\}^{3/4}\frac{3}{4}\frac{6.13e^p(1 - e^{-p})}{N\{1 - e^{-p}(1+p)\}}(1 - 0.024)^{-1/4} \\ + 0.511e^p(1 - e^{-p}) + 0.63N(r_7)^2 + \frac{1}{2}(k/N)(1 - k/N)^{-1}. \end{aligned} \tag{4.35}$$

In virtue of (4.25) we get

$$N(r_7)^2 \leqq 1.018^2\theta^2(e^p/N)^{1/2}\{1-e^{-p}(1+p)\}^{3/2}N + 2\times 1.018\times 0.511\times|\theta|(e^p/N)^{1/4}p + 0.511^2\frac{p^2}{N} \tag{4.36}$$

By (4.14), (3.2) and the inequality $p \leqq e^p(1-e^{-p})$, $p>0$ this yields

$$N(r_7)^2 \leqq 1.037\theta^2\sqrt{N}\,e^{p/2}\{1-e^{-p}(1+p)\}^{3/2} + 0.107e^p(1-e^{-p}). \tag{4.37}$$

Hence from (4.35), (4.37) and Lemma 4.5 we see that (using (3.16) and (4.14))

$$|r_5| \leqq 0.47e^p(1-e^{-p}) + 0.511e^p(1-e^{-p}) + 0.66\theta^2\sqrt{N}\,e^{p/2}\{1-e^{-p}(1+p)\}^{3/2} + 0.07e^p(1-e^{-p}) + \tfrac{1}{2}(1-e^{-p})(1+0.024)e^p(1-0.107)^{-1}, \tag{4.38}$$

which yields (4.18). In order to prove (4.19) and (4.20) we use (4.22) and (4.23) in (4.2) and obtain

$$\begin{aligned} D^2V_k &= N(e^p(1-e^{-p})(1+r_8)/(1+r_7) + \log(e^{-p}\{1+r_7\})) - r_4 \\ &= N(e^p-1-p+(e^p-1)\{(1+r_8)/(1+r_7)-1\} + \log(1+r_7) - r_4/N). \end{aligned} \tag{4.39}$$

By (4.17), (4.26) and the inequality $|\log(1+x)| \leqq |x|/(1-|x|)$, $|x|<1$, we have

$$\begin{aligned} |r_6| &\leqq \frac{e^p-1}{e^p-1-p}(|r_7|+|r_8|)(1-0.107)^{-1} \\ &\quad + (e^p-1-p)^{-1}|r_7|(1-0.107)^{-1} \\ &\quad + (e^p-1-p)^{-1}|r_4|/N. \end{aligned} \tag{4.40}$$

By (4.27) and (4.4) we get

$$\begin{aligned} |r_6| &\leqq 1.12\{1-e^{-p}(1+p)\}^{-1}(1+e^{-p})|r_7| \\ &\quad + \tfrac{1}{2}\{1-e^{-p}(1+p)\}^{-1}(k/N)(1-k/N)^{-2}\cdot e^{-p}/N. \end{aligned} \tag{4.41}$$

In virtue of Lemma 4.5 we have

$$(k/N)\cdot(1-k/N)^{-2} \leqq (1+0.024)(1-0.107)^{-2}e^{2p}(1-e^{-p}). \tag{4.42}$$

Inserting (4.42) and (4.25) into (4.41) we obtain

$$\begin{aligned} |r_6| &\leqq (1.12)(2)(1.018)\{Ne^{-p}(1-e^{-p}(1+p))\}^{-1/4}|\theta| \\ &\quad + (1.12)(2)(0.511)\{Ne^{-p}(1-e^{-p}(1+p))\}^{-1}pe^{-p} \\ &\quad + 0.65(1-e^{-p})\{Ne^{-p}(1-e^{-p}(1+p))\}^{-1}, \end{aligned} \tag{4.43}$$

which by elementary inequalities yields (4.19). To get (4.20) from (4.19) we use (4.14) and (3.16). □

Proof of Lemma 4.4. Combining Lemma 4.2 and Lemma 4.5 we obtain

$$\textstyle\sum_{j=0}^{k-1} E\,|U_j - EU_j|^3 \leqq 2N(1-e^{-p})^2e^{2p}(1+0.024)^2(1-0.107)^{-2}, \tag{4.44}$$

which yields (4.21).

5. Conclusion of the proof of (1.4). The estimate (1.4) follows readily from (2.10) and the following two lemmas. Recall the remark in Section 2 saying that it suffices to consider k-values satisfying (4.14) and the remark in Section 3 that it suffices to consider the case (3.1).

LEMMA 5.1. *For k as in* (4.14) *we have when* (3.1) *is satisfied*

$$|P(V_k > n) - \Phi((EV_k - n)/DV_k)| \leqq 6.46/DZ_n. \tag{5.1}$$

LEMMA 5.2. *For k as in* (4.14) *we have when* (3.1) *is satisfied*

$$|\Phi((EV_k - n)/DV_k) - \Phi((k - EZ_n)/DZ_n)| \leqq 3.11/DZ_n. \tag{5.2}$$

PROOF OF LEMMA 5.1. By the Berry-Esseen theorem (see e.g. Loève (1977), page 300 and van Beek (1972)), Lemma 4.4 and Lemma 4.3 we obtain

$$\begin{aligned} &|P(V_k > n) - \Phi((EV_k - n)/DV_k)| \\ &\leqq 0.7975 \times 2.63Ne^{2p}(1 - e^{-p})^2(DV_k)^{-3} \\ &\leqq 2.10Ne^{2p}(1 - e^{-p})^2(Ne^p\{1 - e^{-p}(1 + p)\}(1 - 0.242))^{-3/2} \\ &\leqq 3.19\frac{(1 - e^{-p})^2}{1 - e^{-p}(1 + p)}\{Ne^{-p}(1 - e^{-p}(1 + p))\}^{-1/2}. \end{aligned} \tag{5.3}$$

By using (3.6), (3.8), the fact that $(1 - e^{-p})^2/\{1 - e^{-p}(1 + p)\} \leqq 2, p > 0$ we obtain (5.1) from (5.3). □

In the proof of Lemma 5.2 we shall need the following estimate.

LEMMA 5.3. *Under the assumptions of Lemma* 5.2 *we have*

$$|(k - EZ_n)/DZ_n - (EV_k - n)/DV_k| \leqq 2.50\theta^2 + 2.15/DZ_n. \tag{5.4}$$

PROOF OF LEMMA 5.3. (4.14) together with Lemmas 3.2 and 4.3 yield

$$\begin{aligned} \Delta &= |(k - EZ_n)/DZ_n - (EV_k - n)/DV_k| \\ &\leqq |\theta(Ne^{-p})^{1/4}\{1 - e^{-p}(1 + p)\}^{1/4}(1 + r_2)^{1/4} \\ &\quad - \theta \cdot N(e^p/N)^{1/4}\{1 - e^{-p}(1 + p)\}^{3/4} \\ &\quad \cdot \{Ne^p(1 - e^{-p}(1 + p))\}^{-1/2}(1 + r_6)^{-1/2}| \\ &\quad + |r_5|\{Ne^p(1 - e^{-p}(1 + p))\}^{-1/2}(1 - 0.242)^{-1/2}. \end{aligned} \tag{5.5}$$

In virtue of (4.18) and the simple inequalities $|(1 + x)^{1/4} - 1| \leqq \frac{1}{4}|x|(1 - |x|)^{-3/4}$, $|x| < 1$, and $|(1 + x)^{-1/2} - 1| \leqq \frac{1}{2}|x|(1 - |x|)^{-3/2}$, $|x| < 1$, we obtain

$$\begin{aligned} \Delta \leqq |\theta|\{Ne^{-p}(1 - e^{-p}(1 + p))\}^{1/4}(0.25|r_2|(1 - |r_2|)^{-3/4} + \tfrac{1}{2}|r_6|(1 - |r_6|)^{-3/2}) \\ + 1.88\{Ne^{-p}(1 - e^{-p}(1 + p))\}^{-1/2} + 0.76\theta^2. \end{aligned} \tag{5.6}$$

By using (3.7), (3.8), (3.16), (4.19), (4.20) and (4.14) this yields

$$\Delta \leqq 2.50\theta^2 + 2.12\{Ne^{-p}(1 - e^{-p}(1 + p))\}^{-1/2}. \tag{5.7}$$

In virtue of (3.6) and (3.8) we get

$$\Delta \leqq 2.50\theta^2 + 2.15/DZ_n, \tag{5.8}$$

proving Lemma 5.3. □

PROOF OF LEMMA 5.2. By the mean value theorem we have

$$\Phi(x) - \Phi(y) = (2\pi)^{-1/2}(x - y)\exp(-\tfrac{1}{2}\{x + \delta(y - x)\}^2), \qquad 0 < \delta < 1. \tag{5.9}$$

Hence we obtain with $x = (k - EZ_n)/DZ_n$ and $y = (EV_k - n)/DV_k$

$$| \Phi((k - EZ_n)/DZ_n) - \Phi((EV_k - n)/DV_k)| \tag{5.10}$$
$$\leqq (2\pi)^{-1/2}(2.50\theta^2 + 2.15/DZ_n)\exp(-\tfrac{1}{2}\{\theta\sqrt{DZ_n} + \delta((EV_k - n)/DV_k - \theta\sqrt{DZ_n})\}^2).$$

If $|\theta| \sqrt{DZ_n} \leqq 1.5$ then the right hand side of (5.10) is dominated by

$$(2\pi)^{-1/2} (2.50 \times 1.5^2 + 2.15)/DZ_n \leqq 3.11/DZ_n. \tag{5.11}$$

If, on the other hand, $|\theta| \sqrt{DZ_n} \geqq 1.5$ then by Lemma 5.3, (3.1) and (4.14) we have

$$|(EV_k - n)/DV_k - \theta\sqrt{DZ_n}| \leqq |\theta| \sqrt{DZ_n} \left(\frac{2.50 \, |\theta|}{\sqrt{DZ_n}} + \frac{2.15}{|\theta| \sqrt{DZ_n} \cdot DZ_n}\right) \tag{5.12}$$
$$\leqq 0.394 \cdot |\theta| \sqrt{DZ_n},$$

yielding that the right hand side of (5.10) is dominated by

$$(2\pi)^{-1/2} 2.50(DZ_n)^{-1}\theta^2 DZ_n \exp(-\tfrac{1}{2}\theta^2 DZ_n\{1 - 0.394\}^2) \tag{5.13}$$
$$+ (2\pi)^{-1/2} 2.15 \exp(-\tfrac{1}{2} \times 1.5^2\{1 - 0.394\}^2)/DZ_n,$$

which by the simple inequality $xe^{-\alpha x} \leqq (\alpha e)^{-1}$ yields that (5.10) is dominated by $2.99/DZ_n$ when $|\theta| \sqrt{DZ_n} \geqq 1.5$. This combined with (5.11) yields (5.2). □

6. Proof of (1.6). Chebyshev's inequality yields

$$P(EZ_n - \sqrt{3}\, DZ_n < Z_n < EZ_n + \sqrt{3}\, DZ_n) \geqq \tfrac{2}{3}. \tag{6.1}$$

As the distribution of Z_n is concentrated on the integers $1, 2, \cdots, \min(n, N)$ and as the interval $(EZ_n - \sqrt{3}\, DZ_n, EZ_n + \sqrt{3}\, DZ_n)$ contains at most $[2\sqrt{3}\, DZ_n + 1]$ integers, (6.1) yields that the largest point mass $p(n, N)$ in the distribution of Z_n satisfies

$$p(n, N) \geqq \frac{\tfrac{2}{3}}{2\sqrt{3}\, DZ_n + 1}. \tag{6.2}$$

Furthermore, as $\Phi(x)$ is continuous we have

$$\sup_{-\infty<x<\infty} | P(Z_n < x) - \Phi((x - EZ_n)/DZ_n)| \geqq \tfrac{1}{2} p(n, N). \tag{6.3}$$

(1.6) now follows from (6.3), (6.2) and the elementary inequality

$$(6\sqrt{3}\, x + 3)^{-1} \geqq (6\sqrt{3} + 1)^{-1}/\max(x, 3) \geqq 0.087/\max(x, 3). \tag{6.4}$$ □

Acknowledgments. I wish to express my gratitude to my teacher, Professor Bengt Rosén, and to Peter Högfeldt for valuable comments and stimulating discussions during the course of this work.

REFERENCES

van Beek, P. (1972). An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **23** 187–196.

Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Applications.* Wiley, New York.

Kolchin, V. F. (1966). The speed of convergence to limit distributions in the classical ball problem. *Theo. Probability Appl.* **11** 128–140.

Loève, M. (1977). *Probability Theory I.* 4th Ed. Springer, New York.

Rényi, A. (1962). Three new proofs and a generalization of a theorem of Irving Weiss. *Publ. Math. Inst. Hung. Acad. Sci.* **7** 203–214.

Weiss, I. (1958). Limiting distributions in some occupancy problems. *Ann. Math. Statist.* **29** 874–884.

Department of Mathematics
Royal Institute of Technology
S-100 44 Stockholm, Sweden