Z. Wahrscheinlichkeitstheorie verw. Gebiete
60, 381–391 (1982)

Zeitschrift für
**Wahrscheinlichkeitstheorie**
und verwandte Gebiete
© Springer-Verlag 1982

# On the Coupon Collector's Remainder Term

Gunnar Englund

Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm 70, Sweden

**Summary.** Each element in a finite population $\pi$ is assigned a "bonus value", i.e. a real number. Elements are selected from $\pi$ by simple random sampling with replacement and with equal draw probabilities. Each time we receive a "new" element, i.e. an element which has not been previously selected, we receive the corresponding bonus. Let $W_n$ denote the bonus sum after $n$ selections. It is well known that $W_n$ is approximately normally distributed under mild conditions. We give a remainder term estimate of the Berry-Esseen type for this normal distribution approximation.

## 1. Introduction, Formulation of Main Results and Outline of Proofs

Consider a finite population $\pi=(1, 2, \ldots, N)$, for which there is a "bonus value", i.e. a real number, associated with each element. Let the bonus values be $\mathbf{a}=(a_1, a_2, \ldots, a_N)$. In the sequel we will use the following notation.

$$m=\frac{1}{N}\sum_{s=1}^{N} a_s, \tag{1.1}$$

$$\sigma^2=\frac{1}{N-1}\sum_{s=1}^{N}(a_s-m)^2, \tag{1.2}$$

$$\gamma^3=\frac{1}{N-1}\sum_{s=1}^{N}|a_s-m|^3. \tag{1.3}$$

Elements are selected from $\pi$ by simple random sampling with replacement and with equal draw probabilities. Each time we get a "new" element, i.e. an element which has not been previously selected, we receive the corresponding bonus. Let

$$W_n=\text{the bonus sum after } n \text{ selections.} \tag{1.4}$$

It is well known that $W_n$ under mild conditions is approximately normally distributed, see e.g. Rosén (1969). Our main aim in this paper is to establish

0044-3719/82/0060/0381/$02.20

the following remainder term estimate for this normal distribution approximation.

**Theorem 1.** *With assumptions and notation as above and with*

$$p=\frac{n}{N} \tag{1.5}$$

*we have*

$$\sup_{-\infty<x<\infty}\left|P(W_n\leqq x)-\Phi\left(\frac{x-EW_n}{DW_n}\right)\right|\leqq(8+1.76\,C_1)\left(\frac{\gamma}{\sigma}\right)^3\frac{1}{\sqrt{Ne^{-p}(1-e^{-p})}}$$
$$+\min\left(\frac{13.4}{\sqrt{Ne^{-p}\{1-e^{-p}(1+p)\}}},\,1.1\left(\frac{m}{\sigma}\right)^2 p\right), \tag{1.6}$$

*where $C_1$ is the universal constant appearing in Theorem 2 below.*

Next we present some basic steps and ingredients in the proof of Theorem 1. Let

$$Z_n=\text{the number of distinct elements from } \pi \text{ obtained in } n \text{ selections,} \tag{1.7}$$

and let further $J_1, J_2, \ldots$ be the sequence of lables for the successive new elements obtained during the drawing procedure. Set

$$S_k=\sum_{j=1}^{k} a_{J_j}. \tag{1.8}$$

Then we have the following relation. Here and in the sequel $\mathscr{L}(X)$ denotes the distribution of the random variable $X$.

$$\mathscr{L}(W_n)=\mathscr{L}(S_{Z_n}). \tag{1.9}$$

The following proposition is an obvious consequence of the assumption of equal draw probabilities.

**Proposition 1.**

$$\mathscr{L}(S_{Z_n}\,|\,Z_n=k)=\mathscr{L}(S_k). \tag{1.10}$$

Note that the distribution of $S_k$ is that of the sample sum in a size $k$ sample drawn without replacement from $\pi$. We shall derive (1.6) by an appropriate application of (1.9), (1.10), the law of total probability and the following two estimates.

**Theorem 2** (Höglund (1978)). *There is a universal constant $C_1$ such that*

$$\sup_x\left|P(S_k\leqq x)-\Phi\left(\frac{x-ES_k}{DS_k}\right)\right|\leqq C_1\left(\frac{\gamma}{\sigma}\right)^3\frac{1}{\sqrt{k\left(1-\frac{k}{N}\right)}}. \tag{1.11}$$

**Theorem 3** (Englund (1981)).

$$\frac{0.087}{(3\sqrt{3}DZ_n)}\leqq\sup_x\left|P(Z_n\leqq x)-\Phi\left(\frac{x-EZ_n}{DZ_n}\right)\right|\leqq\frac{10.4}{DZ_n}. \tag{1.12}$$

The course of the paper is as follows. In the next section we present some auxiliary results, and in Sect. 3 we complete the proof of Theorem 1.

To the best of our knowledge no numerical value for $C_1$ in Theorem 2 has been established. The method of Höglund seems to yield a very large value, but there is reason to believe that $C_1$ is of moderate size, i.e. about the same value as the constant in the traditional Berry-Esseen theorem for sums of independent random variables (where the current world record seems to be 0.7975, cf. van Beek (1972)).

We want to point out Theorem 4.10 in Lanke (1975), where by similar methods a remainder term estimate for the so-called Basu estimator was derived. In our notation the Basu estimator is $W_n/Z_n$.

We conclude this section with some remarks on the "strength" of the result in Theorem 1.

**1.** Note that if there are constants $K_1$ and $K_2$ such that

$$0<K_1\leqq\frac{n}{N}\leqq K_2<\infty, \tag{1.13}$$

then the remainder term in (1.6) can be written in the form

$$C(K_1, K_2)\left(\frac{\gamma}{\sigma}\right)^3\frac{1}{\sqrt{n}}, \tag{1.14}$$

where $C(K_1, K_2)$ is a quantity only depending on $K_1$ and $K_2$. This is a remainder term similar to that of the classical Berry-Esseen theorem (cf. Feller (1971) p. 542).

**2.** If $m=0$, then the last term in (1.6) vanishes and the bound in (1.6) reduces to

$$C\left(\frac{\gamma}{\sigma}\right)^3\frac{1}{\sqrt{Ne^{-p}(1-e^{-p})}}. \tag{1.15}$$

Hence, if $m=0$ and $n/N\leqq K_2<\infty$ the bound in (1.15) is of the type $C(K_2)(\gamma/\sigma)^3/\sqrt{n}$, where $C(K_2)$ is a quantity only depending on $K_2$. Note that this incorporates the case $n\ll N$.

**3.** We shall show below that the first term to the right in (1.6) is "necessary" by considering the following situation. Let the bonuses be $a_i=1$, $i=1, 2, \ldots, N/2$, and $a_i=-1$, $i=N/2+1, \ldots, N$ (we assume that $N$ is even for simplicity). Then we have $m=0$, $\sigma\cong 1$ and $\gamma\cong 1$ and (1.6) reduces to

$$\frac{8+1.76\,C_1}{\sqrt{Ne^{-p}(1-e^{-p})}}. \tag{1.16}$$

By Čebyšev's inequality we have (note that $EW_n=0$, cf. Lemma 2.2)

$$P(-\sqrt{3}\,DW_n<W_n<\sqrt{3}\,DW_n)\geqq\tfrac{2}{3}. \tag{1.17}$$

Since the bonuses are integer-valued, $W_n$ is also integer-valued. The interval $(-\sqrt{3}DW_n, \sqrt{3}DW_n)$ contains at most $[2\sqrt{3}DW_n+1]$ of the point masses in the distribution of $W_n$. Hence (1.17) yields that the largest point mass $p(n, N)$ satisfies

$$p(n, N) \geqq \frac{\frac{2}{3}}{2\sqrt{3}DW_n+1}. \tag{1.18}$$

From the subsequent Lemmas 2.2, 2.3 and 2.4 it can be deduced that

$$D^2 W_n = \sigma^2 N e^{-p}(1-e^{-p})(1+r(n, N)), \tag{1.19}$$

where $|r(n, N)| \leqq 0.1$. Hence, since $\sigma \cong 1$, we have

$$p(n, N) \geqq \frac{\frac{2}{3}}{2\sqrt{3.3}\sqrt{N e^{-p}(1-e^{-p})}+1}. \tag{1.20}$$

Furthermore since $\Phi(x)$ is continuous we have

$$\sup_x \left| P(W_n \leqq x) - \Phi\left(\frac{x-EW_n}{DW_n}\right)\right| \geqq \tfrac{1}{2} p(n, N). \tag{1.21}$$

By combining (1.21) with (1.20) we see that the first term to the right in (1.6) is indeed "required".

**4.** An interesting particular case of coupon collection is that of classical occupancy, i.e. when $a_i = 1$, $i = 1, 2, \ldots, N$. By letting the bonus values be $a_i = 1 \pm \varepsilon$ where $\varepsilon \to 0$, it can fairly easily be shown that we do not "lose" the particular case of classical occupancy, i.e. Theorem 3 is in a sense "essentially" (apart from the numerical value of the constant) a particular case of Theorem 1.

**5.** As stated in **2** above, the term $1.1\,(m/\sigma)^2 p$ has a reducing effect on the bound in (1.6) when $m=0$, but it can also have a considerable reducing effect when $m \neq 0$. So e.g. if $m=\sigma=1$ and $p \cong 1/\sqrt{n}$ the present bound (1.6) is of the order of magnitude $C/\sqrt{n}$, while omission of the term above would yield a bound of the order $C/n^{1/4}$.

## 2. Some Auxiliary Results

By well known results for sampling without replacement from a finite population we have the following result.

**Lemma 2.1.**

$$ES_k = km, \tag{2.1}$$

$$D^2 S_k = \sigma^2 k\left(1-\frac{k}{N}\right). \tag{2.2}$$

From (1.9), (1.10), (2.1), (2.2) and well-known formulas for the moments of sums with randomly many terms the following formulas are readily derived.

**Lemma 2.2.**

$$EW_n = mEZ_n, \tag{2.3}$$

$$D^2 W_n = \sigma^2 N \frac{EZ_n}{N}\left(1-\frac{EZ_n}{N}\right)+m^2 D^2 Z_n - \sigma^2 \frac{D^2 Z_n}{N}. \tag{2.4}$$

In proving Theorem 1 we can without loss of generality assume that

$$Ne^{-p}(1-e^{-p}) \geqq 32 \tag{2.5}$$

which can be seen by the following argument. By Jensen's inequality we have for $N \geqq 2$ (the case $N=1$ is degenerate)

$$\left(\frac{\gamma}{\sigma}\right)^3 \geqq \sqrt{\frac{N-1}{N}} \geqq \frac{1}{\sqrt{2}}. \tag{2.6}$$

Hence Theorem 1 is trivial, in the sense that the right hand side of (1.6) exceeds 1, if (2.5) is not fulfilled. We will in the sequel use (2.5) without explicitly mentioning it each time. Next we state two results from Englund (1981). Note that (2.5) implies that $N \geqq 128$.

**Lemma 2.3.** *Define* $r_1(n, N)$ *by the relation*

$$EZ_n = N(1-e^{-p}) + r_1(n, N). \tag{2.7}$$

*If* $N \geqq 100$, *we have*

$$0 \leqq r_1(n, N) \leqq 0.511\, pe^{-p}. \tag{2.8}$$

**Lemma 2.4.** *Define* $r_2(n, N)$ *by the relation*

$$D^2 Z_n = Ne^{-p}(1-e^{-p}(1+p))(1+r_2(n, N)). \tag{2.9}$$

*If* $N \geqq 100$, *we have*

$$|r_2(n, N)| \leqq \frac{6.13\, pe^{-p}}{Ne^{-p}(1-e^{-p}(1+p))}. \tag{2.10}$$

*Remark.* By (2.10) and the elementary inequality $1-e^{-p}(1+p) \geqq \frac{1}{2}(1-e^{-p})^2$, $p \geqq 0$, we obtain (using also (2.5) and $p \leqq e^p - 1$, $p \geqq 0$)

$$|r_2(n, N)| \leqq \frac{6.13\, pe^{-p}}{\frac{1}{2}Ne^{-p}(1-e^{-p})^2} \leqq 12.3 \cdot \frac{p}{e^p-1} \cdot \frac{1}{Ne^{-p}(1-e^{-p})} \leqq 0.39. \quad \square \tag{2.11}$$

**Lemma 2.5.** *Let* $X$, $Y$, $Z$ *and* $Z'$ *be random variables such that* $Z$ *is independent of* $X$, $Z'$ *is independent of* $Y$ *and* $\mathcal{L}(Z) = \mathcal{L}(Z')$. *Then we have, for any real number* $a$

$$\sup_x |P(aX+Z \leqq x) - P(aY+Z' \leqq x)| \leqq \sup_x |P(X \leqq x) - P(Y \leqq x)|. \tag{2.12}$$

*Proof.* The case $a=0$ is trivial. For $a>0$ we have

$$|P(aX+Z\leqq x)-P(aY+Z'\leqq x)|=\left|\int_{-\infty}^{\infty}(P(aX+z\leqq x)-P(aY+z\leqq x))\,dF_Z(z)\right|$$
$$\leqq\int_{-\infty}^{\infty}\left|P\left(X\leqq\frac{x-z}{a}\right)-P\left(Y\leqq\frac{x-z}{a}\right)\right|dF_Z(z), \tag{2.13}$$

which readily yields (2.12). The case $a<0$ can be treated quite similarily. □

Next we introduce some additional notation. Let

$$\mathcal{K}(n,N)=\{k:|k-EZ_n|\leqq(Ne^{-p}(1-e^{-p}))^{3/4},\quad 1\leqq k\leqq N\} \tag{2.14}$$
$$q(n,N)=[EZ_n], \tag{2.15}$$

where $[\cdot]$ denotes integer part. For notational convenience we supress $N$ and write $q(n)$ instead of $q(n,N)$.

**Lemma 2.6.** *If* (2.5) *holds, then*

$$q(n)\geqq 0.96\,N(1-e^{-p})\quad\text{and}\quad N-q(n)\geqq 0.99\,Ne^{-p}. \tag{2.16}$$

*Proof.*

$$q(n)=N(1-e^{-p})+(EZ_n-N(1-e^{-p}))+(q(n)-EZ_n). \tag{2.17}$$

By using (2.7), (2.8), (2.15) and (2.5) we get

$$\frac{q(n)}{N(1-e^{-p})}=1+\frac{r_1(n,N)}{N(1-e^{-p})}-\frac{EZ_n-q(n)}{N(1-e^{-p})}\geqq 1-\frac{1}{N(1-e^{-p})}\geqq 1-\frac{1}{32}\geqq 0.96. \tag{2.18}$$

In the same manner we obtain

$$\frac{N-q(n)}{Ne^{-p}}=1-\frac{r_1(n,N)}{Ne^{-p}}+\frac{EZ_n-q(n)}{Ne^{-p}}\geqq 1-\frac{0.511\,pe^{-p}}{Ne^{-p}}\geqq 1-\frac{0.511}{32\,e}\geqq 0.99. \tag{2.19}$$

where we also used the inequality $pe^{-p}\leqq 1/e$, $p\geqq 0$. □

**Lemma 2.7.** *If* (2.5) *holds and if* $k\in\mathcal{K}(n,N)$, *we have*

$$\frac{k}{N}\geqq 0.57\,(1-e^{-p})\quad\text{and}\quad 1-\frac{k}{N}\geqq 0.57\,e^{-p}. \tag{2.20}$$

*Proof.* We easily get from (2.14) that for $k\in\mathcal{K}(n,N)$ we have

$$\frac{k}{N}\geqq 1-e^{-p}+\frac{r_1}{N}-\frac{(Ne^{-p}(1-e^{-p}))^{3/4}}{N}\geqq(1-e^{-p})\left(1-\frac{e^{-p}}{\sqrt[4]{Ne^{-p}(1-e^{-p})}}\right)$$
$$\geqq(1-e^{-p})\left(1-\frac{1}{\sqrt[4]{32}}\right)\geqq 0.57\,(1-e^{-p}). \tag{2.21}$$

Similarily we get

$$1-\frac{k}{N}\geqq e^{-p}\left(1-\frac{0.511\,pe^{-p}}{Ne^{-p}}-\frac{1-e^{-p}}{\sqrt[4]{Ne^{-p}(1-e^{-p})}}\right)$$
$$\geqq e^{-p}\left(1-\frac{0.511}{32\,e}-\frac{1}{\sqrt[4]{32}}\right)\geqq 0.57\,e^{-p}. \quad \square \tag{2.22}$$

**Lemma 2.8.** *If* (2.5) *holds, then*

$$\left|1-\frac{DS_k}{DS_{q(n)}}\right|\leqq 1.06\,\frac{1+|k-EZ_n|}{Ne^{-p}(1-e^{-p})}. \tag{2.23}$$

*Proof.* By using (2.2) we get (also using $|1-\sqrt{x}|\leqq|1-x|$, $x\geqq 0$)

$$\left|1-\frac{DS_k}{DS_{q(n)}}\right|=\left|1-\sqrt{\frac{k(N-k)}{q(n)(N-q(n))}}\right|\leqq\left|1-\frac{\frac{k}{N}\left(1-\frac{k}{N}\right)}{\frac{q(n)}{N}\left(1-\frac{q(n)}{N}\right)}\right|$$
$$=\frac{\left|\frac{q(n)}{N}-\frac{k}{N}\right|\left|1-\frac{q(n)+k}{N}\right|}{\frac{q(n)}{N}\left(1-\frac{q(n)}{N}\right)}\leqq\frac{1+|k-EZ_n|}{Ne^{-p}(1-e^{-p})}\,\frac{N(1-e^{-p})}{q(n)}\,\frac{Ne^{-p}}{N-q(n)}. \tag{2.24}$$

Application of (2.16) yields (2.23). $\square$

**Lemma 2.9.** *If* $x>0$ *and* $y>0$, *then we have for any real number* $z$

$$\left|\Phi\left(\frac{z}{x}\right)-\Phi\left(\frac{z}{y}\right)\right|\leqq 1.25\left|\frac{x}{y}-1\right|. \tag{2.25}$$

The lemma holds with the constant $1+(2\pi e)^{-1/2}<1.25$.

*Proof.* Since the left hand side of (2.25) is dominated by 1, we can without loss of generality assume that

$$\left|\frac{x}{y}-1\right|\leqq\frac{1}{1.25}. \tag{2.26}$$

By Taylor expansion we obtain

$$\Delta=\left|\Phi\left(\frac{z}{x}\right)-\Phi\left(\frac{z}{y}\right)\right|=\left|\frac{z}{x}-\frac{z}{y}\right|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\left(\frac{1}{x}+\theta\left(\frac{1}{y}-\frac{1}{x}\right)\right)^2\right), \tag{2.27}$$

where $0<\theta<1$. By using (2.26) we obtain

$$\Delta\leqq\frac{1}{\sqrt{2\pi}}\left|1-\frac{x}{y}\right|\cdot\frac{|z|}{x}\exp\left(-\frac{1}{2}\left(\frac{z}{x}\right)^2\left(1-\frac{1}{1.25}\right)^2\right), \tag{2.28}$$

which by using the simple inequality $|y|\cdot\exp(-\alpha^2 y^2)\leqq 1/\alpha\sqrt{2e}$, $\alpha>0$, yields (2.25). $\square$

## 3. Proof of Theorem 1

From (1.9) and (1.10) we obtain by conditioning on $Z_n$

$$\begin{aligned}
&\left|P(W_n \leqq x) - \Phi\left(\frac{x-EW_n}{DW_n}\right)\right| \leqq P(Z_n \notin \mathscr{K}(n,N)) \\
&\quad + \left|\sum_{k\in\mathscr{K}(n,N)} P(S_k \leqq x)\, P(Z_n = k) - \Phi\left(\frac{x-EW_n}{DW_n}\right)\right| \leqq P(Z_n \notin \mathscr{K}(n,N)) \\
&\quad + \sup_{k\in\mathscr{K}(n,N)} \left|P(S_k \leqq x) - \Phi\left(\frac{x-km}{DS_k}\right)\right| \\
&\quad + \left|\sum_{k\in\mathscr{K}(n,N)} P(Z_n = k)\cdot\left(\Phi\left(\frac{x-km}{DS_k}\right) - \Phi\left(\frac{x-km}{DS_{q(n)}}\right)\right)\right| \\
&\quad + P(Z_n \notin \mathscr{K}(n,N)) + \left|\sum_k \Phi\left(\frac{x-km}{DS_{q(n)}}\right)\cdot P(Z_n = k) - \Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right)\right| \\
&\quad + \left|\Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right) - \Phi\left(\frac{x-EW_n}{DW_n}\right)\right| \\
&\leqq 2P(Z_n \notin \mathscr{K}(n,N)) + \sup_{k\in\mathscr{K}(n,N)} \left|P(S_k \leqq x) - \Phi\left(\frac{x-km}{DS_k}\right)\right| \\
&\quad + \sum_{k\in\mathscr{K}(n,N)} \left|\Phi\left(\frac{x-km}{DS_k}\right) - \Phi\left(\frac{x-km}{DS_{q(n)}}\right)\right| \cdot P(Z_n = k) \\
&\quad + \left|\sum_k \Phi\left(\frac{x-km}{DS_{q(n)}}\right) P(Z_n = k) - \Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right)\right| \\
&\quad + \left|\Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right) - \Phi\left(\frac{x-EW_n}{DW_n}\right)\right| \\
&\triangleq A_1 + A_2(x) + A_3(x) + A_4(x) + A_5(x). \qquad (3.1)
\end{aligned}$$

We shall show that these five terms can be estimated as follows ($C_1$ denoting the universal constant in Theorem 2).

$$A_1 \leqq \frac{2.78}{\sqrt{Ne^{-p}(1-e^{-p})}}, \qquad (3.2)$$

$$A_2(x) \leqq 1.76\, C_1 \left(\frac{\gamma}{\sigma}\right)^3 \frac{1}{\sqrt{Ne^{-p}(1-e^{-p})}}, \qquad (3.3)$$

$$A_3(x) \leqq \frac{1.8}{\sqrt{Ne^{-p}(1-e^{-p})}}, \qquad (3.4)$$

$$A_4(x) \leqq \frac{13.4}{\sqrt{Ne^{-p}(1-e^{-p}(1+p))}}, \qquad (3.5)$$

$$A_4(x) \leqq 1.1 \left(\frac{m}{\sigma}\right)^2 p, \qquad (3.6)$$

$$A_5(x) \leqq \frac{0.56}{\sqrt{N e^{-p}(1-e^{-p})}}. \tag{3.7}$$

The estimates (3.2)–(3.7) together with (3.1) and (2.6) prove Theorem 1 since $(2.78+1.8+0.56)\sqrt{2} \leqq 8$. We now turn to the task of proving (3.2)–(3.7).

*Proof of* (3.2). By (2.14) and Čebyšev's inequality we obtain

$$P(Z_n \notin \mathscr{K}(n, N)) \leqq \frac{D^2 Z_n}{(N e^{-p}(1-e^{-p}))^{3/2}} \leqq \tag{3.8}$$

by using Lemma 2.4, (2.11) and the inequality $1-e^{-p}(1+p) \leqq 1-e^{-p}$, $p \geqq 0$,

$$\leqq \frac{1.39\, N e^{-p}(1-e^{-p}(1+p))}{(N e^{-p}(1-e^{-p}))^{3/2}} \leqq \frac{1.39}{\sqrt{N e^{-p}(1-e^{-p})}}, \tag{3.9}$$

proving (3.2) □

*Proof of* (3.3). By Theorem 2 we have

$$\sup_{k \in \mathscr{K}(n, N)} \left| P(S_k \leqq x) - \Phi\left(\frac{x-km}{DS_k}\right) \right| \leqq C_1 \left(\frac{\gamma}{\sigma}\right)^3 \sup_{k \in \mathscr{K}(n, N)} \frac{1}{\sqrt{k\left(1-\frac{k}{N}\right)}}. \tag{3.10}$$

By Lemma 2.7 we have

$$\sup_{k \in \mathscr{K}(n, N)} \frac{1}{\sqrt{k\left(1-\frac{k}{N}\right)}} \leqq \frac{1}{\sqrt{0.57 \times 0.57}} \times \frac{1}{\sqrt{N e^{-p}(1-e^{-p})}}, \tag{3.11}$$

which yields (3.3). □

*Proof of* (3.4). Observing Lemma 2.9 and Lemma 2.8 we get for $k \in \mathscr{K}(n, N)$

$$\left| \Phi\left(\frac{x-km}{DS_k}\right) - \Phi\left(\frac{x-km}{DS_{q(n)}}\right) \right| \leqq 1.25 \times 1.06\, \frac{1+|k-EZ_n|}{N e^{-p}(1-e^{-p})}, \tag{3.12}$$

yielding

$$A_3(x) \leqq 1.325\, \frac{1+E|Z_n-EZ_n|}{N e^{-p}(1-e^{-p})} \leqq \frac{1.325}{N e^{-p}(1-e^{-p})} + \frac{1.325\, DZ_n}{N e^{-p}(1-e^{-p})}. \tag{3.13}$$

Simple use of (2.5), Lemma 2.4 and (2.11) as in (3.9) yields (3.4). □

*Proof of* (3.5). By definition we have

$$A_4(x) = \left| \sum_{k=0}^{\infty} \Phi\left(\frac{x-km}{DS_{q(n)}}\right) P(Z_n = k) - \Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right) \right|. \tag{3.14}$$

Let $U$ denote a standard normal random variable which is independent of $Z_n$. Then we have

$$\sum_{k=0}^{\infty} \Phi\left(\frac{x-km}{DS_{q(n)}}\right) P(Z_n = k) = P(mZ_n + UDS_{q(n)} \leqq x). \tag{3.15}$$

If $V$ is a normally distributed r.v. with $EV=EZ_n$ and $DV=DZ_n$, which is independent of $U$, we get

$$\Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)}+m^2 D^2 Z_n}}\right)=P(mV+UDS_{q(n)}\leqq x). \tag{3.16}$$

By inserting (3.15) and (3.16) into (3.14) and using Lemma 2.5 we obtain (3.5) from Theorem 3 and Lemma 2.4 (see also the remark after Lemma 2.4). □

*Proof of* (3.6). We easily get

$$\begin{aligned} A_4(x)\leqq &\left|\sum_k \Phi\left(\frac{x-km}{DS_{q(n)}}\right) P(Z_n=k)-\Phi\left(\frac{x-mEZ_n}{DS_{q(n)}}\right)\right| \\ &+\left|\Phi\left(\frac{x-mEZ_n}{DS_{q(n)}}\right)-\Phi\left(\frac{x-mEZ_n}{\sqrt{D^2 S_{q(n)}+m^2 D^2 Z_n}}\right)\right| \triangleq A_4'(x)+A_4''(x). \end{aligned} \tag{3.17}$$

By Taylor expansion (where $0<\theta_k<1$) we have

$$\begin{aligned} \Phi\left(\frac{x-km}{DS_{q(n)}}\right)=&\Phi\left(\frac{x-mEZ_n}{DS_{q(n)}}\right)+\frac{m}{DS_{q(n)}}(EZ_n-k)\,\Phi'\left(\frac{x-mEZ_n}{DS_{q(n)}}\right) \\ &+\frac{m^2}{2D^2 S_{q(n)}}(EZ_n-k)^2\,\Phi''\left(\frac{x-mEZ_n}{DS_{q(n)}}+\theta_k\left(\frac{x-km}{DS_{q(n)}}-\frac{x-mEZ_n}{DS_{q(n)}}\right)\right), \end{aligned} \tag{3.18}$$

which yields by Lemma 2.4, (2.11) and the inequality $|\Phi''(x)|\leqq 1/\sqrt{2\pi e}$

$$\begin{aligned} A_4'(x)&\leqq\frac{m^2}{2D^2 S_{q(n)}} D^2 Z_n \sup_x|\Phi''(x)| \\ &\leqq\frac{1.39}{2\sqrt{2\pi e}}\left(\frac{m}{\sigma}\right)^2\frac{1-e^{-p}(1+p)}{1-e^{-p}}\,\frac{Ne^{-p}}{N-q(n)}\,\frac{N(1-e^{-p})}{q(n)}. \end{aligned} \tag{3.19}$$

By using the inequality $1-e^{-p}(1+p)\leqq p(1-e^{-p})$ and (2.16) this yields

$$A_4'(x)\leqq 0.18\left(\frac{m}{\sigma}\right)^2 p. \tag{3.20}$$

In order to estimate $A_4''(x)$ we use Lemma 2.9 to get

$$A_4''(x)\leqq 1.25\left|\frac{\sqrt{D^2 S_{q(n)}+m^2 D^2 Z_n}}{DS_{q(n)}}-1\right|\leqq 0.625\,\frac{m^2 D^2 Z_n}{D^2 S_{q(n)}}, \tag{3.21}$$

where we used the inequality $|\sqrt{1+x}-1|\leqq\frac{1}{2}x$, $x\geqq 0$ in the last step. Repeating the argument in (3.19) and (3.20) we see that

$$A_4''(x)\leqq\frac{0.625\times 1.39}{0.99\times 0.96}\left(\frac{m}{\sigma}\right)^2 p\leqq 0.92\left(\frac{m}{\sigma}\right)^2 p. \tag{3.22}$$

By combining (3.22) and (3.20) we get (3.6). □

*Proof of* (3.7). By using (2.4) we get

$$D^2 W_n = D^2 S_{q(n)} + m^2 D^2 Z_n + \sigma^2 \left( EZ_n \left(1 - \frac{EZ_n}{N}\right) - q(n)\left(1 - \frac{q(n)}{N}\right) - \frac{D^2 Z_n}{N}\right). \quad (3.23)$$

Hence we get

$$|D^2 W_n - (D^2 S_{q(n)} + m^2 D^2 Z_n)| \leqq \sigma^2 \left| EZ_n - q(n) + \frac{q^2(n) - E^2 Z_n}{N}\right| + \sigma^2 \frac{D^2 Z_n}{N}$$

$$\leqq \sigma^2 |EZ_n - q(n)| \cdot \left|1 - \frac{q(n) + EZ_n}{N}\right| + 1.39\, \sigma^2 \leqq 2.39\, \sigma^2, \quad (3.24)$$

where we used Lemma 2.4, (2.11) and (2.15). Since $EW_n = mEZ_n$ (cf. (2.3)) we get by using Lemma 2.9

$$A_5(x) = \left| \Phi\left(\frac{x - mEZ_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}}\right) - \Phi\left(\frac{x - mEZ_n}{DW_n}\right)\right|$$

$$\leqq 1.25 \left| \frac{DW_n}{\sqrt{D^2 S_{q(n)} + m^2 D^2 Z_n}} - 1\right|. \quad (3.25)$$

By using the inequality $|x-1| \leqq |x^2 - 1|$, $x \geqq 0$, we get from (3.25), (3.24) and (2.16)

$$A_5(x) \leqq \frac{1.25 \times 2.39\, \sigma^2}{D^2 S_{q(n)}} = \frac{1.25 \times 2.39}{N \frac{q(n)}{N}\left(1 - \frac{q(n)}{N}\right)} \leqq \frac{3.15}{Ne^{-p}(1 - e^{-p})}, \quad (3.26)$$

By using (2.5) we easily obtain (3.7) from (3.26). □

*Acknowledgement.* I wish to express my gratitude to my teacher Bengt Rosén for valuable comments and stimulating discussions.

## References

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

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

Feller, W.: An Introduction to Probability Theory and its Applications, Vol. II, 2nd Ed. New York: Wiley 1971

Höglund, T.: Sampling from a finite population. A remainder term estimate. Scand. J. Statist. **5**, 69–71 (1978)

Kolchin, V., Sevastyanov, B., Chistyakov, V.: Random Allocations. New York: John Wiley 1978

Lanke, J.: Some contributions to the theory of survey sampling. Ph.D. thesis, Dept. of Math. Statist. Univ. of Lund, Sweden 1975

Rosén, B.: Asymptotic normality in a coupon collector's problem. Z. Wahrscheinlichkeitstheorie verw. Gebiete **12**, 256–279 (1969)

Received February 27, 1981; in revised form June 20, 1981

*Note Added in Proof.* Dr. Malcolm Quine of Univ. of Sydney informs me that he has obtained the preliminary estimate 48 of $C_1$ in Theorem 2.