The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 1057–1068

SEQUENTIAL MAXIMUM LIKELIHOOD ESTIMATION OF THE SIZE OF A POPULATION[1]

By Ester Samuel

The Hebrew University and Purdue University

1. Introduction and summary. Consider the following model, the practical applications of which will be discussed elsewhere. An urn contains an unknown number, N, of white balls, and no others. An estimate of N is desired, based on the following sampling procedure. Balls are drawn at random, one at a time, from the urn. A white ball is colored black before it is returned, a black ball is returned unchanged. The ball is always returned before the next ball is drawn. We are interested in two problems: (i) what stopping rule t to use to terminate sampling, and (ii) how to estimate N after we stop.

The present problem (also in a more general setup) has been considered by several authors, notably L. A. Goodman [5], Chapman [1], Darroch [3] and Darling and Robbins [2]. We shall refer to their results in the sequel.

Let w_i, b_i denote the (random) number of white balls, black balls, respectively, observed in the first i draws ($w_i + b_i = i$). We shall consider mainly the following stopping rules.

Rule A. Let $A > 0$ be a fixed integer. $t_A = A$.

Rule B. Let $B > 0$ be a fixed integer. $t_B = \inf\{i \mid b_i = B\}$.

Rule C. Let $C > 0$ be fixed.

$$t_C = \inf\{i \mid b_i \geqq Cw_i\} = \inf\{i \mid i \geqq (C+1)w_i\}.$$

Rule D. Let $-\infty < D < \infty$ be fixed.

$$t_D = \inf\{i \mid b_i \geqq \max(1, w_i \log w_i + w_i D)\}$$
$$= \inf\{i \mid i \geqq \max(w_i + 1, w_i \log w_i + w_i(D+1))\}.$$

Rule E. Let $\{D_j\}$ be such that $\lim D_j = \infty$.

$$t_E = \inf\{i \mid b_i \geqq \max(1, w_i \log w_i + w_i D_{w_i})\}.$$

Since $w_i \leqq N$ each of these rules is bounded, and thus clearly stops with probability one.

Of these rules, Rule B has been investigated most. See [5], [1] and [3]. Rule D has been considered in a recent paper [2] by Darling and Robbins, who show that for any $0 < \alpha < 1$ and a suitable choice of D one can have $P_N(W_D = N) \geqq 1 - \alpha$ uniformly in N, where W_D is the total of white balls observed before stopping. (See Section 6).

Received 13 April 1967; revised 4 October 1967.

[1] This research was supported in part by the National Science Foundation under Grant GP-07631.

The motivation for consideration of rules A to E stems from a theorem on the limiting distribution, as $N \to \infty$, of w_i, in a sample of *fixed size* i, for various relationships between i and N. We restate the theorem here, since we shall need part of it in the sequel. Different parts of the theorem have been proved by various authors. See for example Rényi [8], where also proper references are given. Let $u_i = N - w_i$ = number of unobserved white balls in the sample of size i (= number of white balls in the urn, after i draws). Since there are linear relationships between w_i, b_i, u_i we shall express the limiting distribution for one of these variables only. Let Φ denote the distribution function of a standard normal variable. We have,

THEOREM 1. *Let* $N \to \infty$.

Case A. If $i = (2N\lambda_N)^{\frac{1}{2}}$ *where* $\lambda_N \to 0$ *then* $P_N(b_i = 0) \to 1$.

Case B. If $i = (2N\lambda_N)^{\frac{1}{2}}$ *where* $\lambda_N \to \lambda$ *and* $0 < \lambda < \infty$ *then*

$$P_N(b_i = k) \to e^{-\lambda}\lambda^k/k! \qquad k = 0, 1, \cdots.$$

Case C. If $i = a_N N$ *where* $\xi_N N^{-\frac{1}{2}} < a_N < \log N - \mu_N$ *and* $\xi_N \to \infty$, $\mu_N \to \infty$ *then*

$$P_N((w_i - Ew_i)/(\text{Var } w_i)^{\frac{1}{2}} \leqq x) \to \Phi(x), \qquad -\infty < x < \infty.$$

Case D. If $i = N \log N + Na_N$ *where* $a_N \to a$ *and* $-\infty < a < \infty$ *then*

$$P_N(u_i = k) \to e^{-\lambda}\lambda^k/k!, \qquad k = 0, 1, \cdots, \text{ where } \lambda = e^{-a}.$$

Case E. If $i = N \log N + Na_N$ *where* $a_N \to \infty$, *then* $P_N(w_i = N) \to 1$.

We consider mainly the Maximum Likelihood Estimate (MLE) of N, denoted by $\hat{N}$. It turns out that if when we stop we have seen w white and b black balls, then $\hat{N} = \hat{N}(w, b)$ and *does not depend on the stopping rule* used, though the distribution of $\hat{N}$ clearly will depend on the stopping rule. The value of $\hat{N}(w, b)$ is discussed in Section 2. In Section 3 we briefly consider Rule A. It satisfies $P_N(\hat{N} = \infty) > 0$ for all $N \geqq A$. Rule B is discussed in Section 4 and it is shown that $2B\hat{N}/N$ has an asymptotic chi square distribution with $2B$ degrees of freedom, (to be denoted χ^2_{2B}), as $N \to \infty$. In Section 5 Rule C is considered, and bounds on the distribution of $(\hat{N} - N)N^{-\frac{1}{2}}$ in terms of the normal distribution are given. Rules D and E are considered in Section 6. Let $[x]^*$ be the largest integer not exceeding x. For Rule D it is shown that $\hat{N} - N + [\lambda]^*$ has an asymptotic Poisson distribution with parameter $\lambda = \exp(-D - 1)$ and for Rule E $P_N(\hat{N} = N) \to 1$. The exact and asymptotic distributions of the corresponding t's is also considered, and is closely related to the distribution of $\hat{N}$.

2. Maximum likelihood estimation of N. For any stopping rule t, the probability of having observed exactly w white and b black balls when we stop clearly depends on t as well as on N. Let $P_N(w, b \mid t)$ denote this probability. It can be shown that

$$P_N(w, b \mid t) = (N)_w h(w, b)/N^{w+b}, \qquad w = 1, 2, \cdots, b = 0, 1, \cdots, \tag{2.1}$$

where $(N)_i = N(N-1) \cdots (N - i + 1)$, and where $h(w, b)$ depends on t but not on N. (We shall see the particular form of (2.1) for Rules A to C later). Thus for any w, b such that $h(w,b) \neq 0, \hat{N}(w,b)$ is the positive integer which maximizes $(N)_w/N^{w+b}$. Maximum likelihood estimation for our problem has been considered by Darroch [3], and for a mathematically equivalent problem by Lewontin and Prout in [6]. (The claim of asymptotic normality of the MLE in [6], p. 221, is clearly generally false, as seen from Section 3 in conjunction with Theorem 1.) Direct inspection yields

$$(2.2) \quad \hat{N}(w, 0) = \infty \quad \text{for} \quad w \geqq 2, \qquad \hat{N}(1, b) = 1 \quad \text{for} \quad b \geqq 1.$$

(The case $w = 1, b = 0$ is of no interest, since the first ball drawn is always white, and thus more than one draw must take place in order to obtain information about N.)

We shall treat $(N)_w/N^{w+b}$ as a function of a positive real variable N, $N > w - 1$. We have

$$(2.3) \quad d\{(N)_w/N^{w+b}\}/dN = \{\textstyle\sum_{j=0}^{w-1} (N - j)^{-1} - (w + b)N^{-1}\}(N)_w N^{-(w+b)}.$$

Equating the right hand side of (2.3) to zero yields that the maximum value $\tilde{N} = \tilde{N}(w, b)$, satisfies

$$(2.4) \qquad (w + b)/\tilde{N}(w, b) = \textstyle\sum_{j=0}^{w-1} (\tilde{N}(w, b) - j)^{-1}.$$

Clearly (2.4) has a unique finite solution, and

$$\hat{N}(w, b) = [\tilde{N}(w, b)]^* \quad \text{or} \quad [\tilde{N}(w, b)] \quad \text{or} \quad \text{both},$$

where $[x]$ is the smallest integer not less than x. The right hand side of (2.4) is less than $\log \{\tilde{N}/(\tilde{N} - w)\}$ and greater than $\log \{(\tilde{N} + 1)/(\tilde{N} - w + 1)\}$. The solution of

$$(2.5) \qquad (w + b)x^{-1} = \log \{x/(x - w)\}$$

is given by $x = (w + b)/m(s)$, where $s = w/(w + b)$ and

$$(2.6) \qquad m(s) \quad \text{is the solution of} \quad s = (1 - e^{-m})/m$$

and can be obtained from existing tables. Thus $\hat{N}(w, b)$ is given approximately by

$$(2.7) \qquad \hat{N}(w, b) \approx (w + b)/m(s) \quad \text{where} \quad s = w/(w + b).$$

The interpretation of (2.7) is that the MLE is approximately proportional to the sample size, with the proportionality factor a function only of the proportion of white balls in the sample drawn.

More accurate information about $\hat{N}$ can be obtained by considering the ratio of $(N)_w/N^{w+b}$ to $(N - 1)_w/(N - 1)^{w+b}$, which we denote by $g_{w,b}(N)$, and when no confusion is likely, by $g(N)$.

$$(2.8) \qquad g_{w,b}(N) = N(N - w)^{-1}(1 - N^{-1})^{w+b}, \qquad N \geqq w.$$

Since $(N)_w/N^{w+b}$ is a continuous function with a unique maximum at $\tilde{N}$, and is strictly increasing for $N < \tilde{N}$, strictly decreasing for $N > \tilde{N}$, there exists a unique real value N^* such that

$$(2.9) \quad g(N^*) = 1, \qquad g(N) > 1 \text{ for } N < N^* \text{ and } g(N) < 1 \text{ for } N > N^*.$$

Thus $\hat{N}(w, b) = [N^*(w, b)]^*$, except when N^* is an integer, in which case $\hat{N}$ is not unique, and can be taken to be N^* or $N^* - 1$.

For any function, $k(w, b)$, one can determine whether $\hat{N}(w, b) \geqq [k(w, b)]^*$ or $\hat{N}(w, b) < k(w, b)$ by computing $g_{w,b}(k(w, b))$ and noting if it is $\geqq 1$, or < 1, respectively. Notice that always $\hat{N}(w, b) \geqq w$.

3. Rule A. For a fixed sample size A, the asymptotic distribution of w_A is given in Theorem 1, for the various relationships between A and N. The exact distribution is given by

$$(3.1) \quad P_N(w_A = k) = \binom{N}{k} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (j/N)^A = (N)_k S_A^{(k)}/N^A, \qquad k = 1, \cdots, N,$$

where $S_A^{(k)}$ is a Stirling number of the second kind, and is defined by the identity (3.1). See for example [4], p. 92. (Compare (3.1) and (2.1).)

Assertion 1. *For every* fixed A *and Rule* A

$$(3.2) \quad P_N(\hat{N} = \infty) > 0 \quad \textit{for all} \quad N \geqq A, \quad \textit{and} \quad \lim_{N\to\infty} P_N(\hat{N} = \infty) = 1.$$

$$(3.3) \qquad \textit{For every} \text{ fixed } N \lim_{A\to\infty} P_N(\hat{N} = N) = 1.$$

Proof. (3.2) is immediate from (2.2) and Case A of Theorem 1. The assertion can actually be strengthened to a corresponding statement for every uniformly (in N) bounded stopping rule, and other cases confirming with Case A of Theorem 1. We shall show that, for every fixed w, $\hat{N}(w, b) = w$ for all b sufficiently large. Then (3.3) follows, since clearly for fixed N, $\lim_{A\to\infty} P_N(w_A = N) = 1$.

Set $b = w^2$, and $N = w + 1$ in (2.8), to get $g_{w,w^2}(w + 1) < (w + 1)e^{-w} \leqq 2e^{-1} < 1$. Thus $\hat{N}(w, w^2) \leqq w$. But for any w, b $\hat{N}(w, b) \geqq w$, and the assertion follows. (In Section 6 it will be clear how the value $b = w^2$ can be improved upon.)

4. Rule B. The distribution of t_B is given by (cf. [3], (16))

$$(4.1) \quad P_N(t_B = k) = (N)_{k-B}(k - B)S_{k-1}^{(k-B)}/N^k, \qquad k = B + 1, \cdots, B + N.$$

This follows directly from (3.1), since $t_B = k$ if and only if $w_{k-1} = k - B$ and the last draw results in a black ball. The asymptotic distribution of t_B is given by

$$(4.2) \qquad \lim_{N\to\infty} P_N(t_B^2/N \leqq x) = F_{2B}(x), \qquad -\infty < x < \infty,$$

where F_{2B} is the distribution function of a χ_{2B}^2 variable. (4.2) is a special case of [5], Theorem 6. (See also [1], Section 3).

Assertion 2. *For every fixed* B *and Rule* B

$$\lim_{N\to\infty} P_N(2B\hat{N}/N \leqq x) = F_{2B}(x), \qquad -\infty < x < \infty.$$

PROOF. We shall show that

$$[(t_B - B)^2/2B]^* \leqq \hat{N} < t_B^2/2B \tag{4.3}$$

and thus the assertion follows from (4.2). (4.3) follows if we show that for every w and $b > 0$

$$[w^2/2b]^* \leqq \hat{N}(w, b) < (w + b)^2/2b. \tag{4.4}$$

To prove (4.4) it suffices to show that

$$g_{w,b}(w^2/2b) > 1 \quad \text{and} \quad g_{w,b}((w + b)^2/2b) < 1, \tag{4.5}$$

where $g_{w,b}$ is defined in (2.8), and where the second inequality must hold for all w and $b > 0$, but where the first inequality must hold only for $w > 2b$, since if $w \leqq 2b$ then $[w^2/2b]^* \leqq w \leqq \hat{N}(w, b)$, and thus clearly the left hand inequality of (4.4) holds for $w \leqq 2b$.

Set $k = w + b$. Using $1 - x < e^{-x}$ we have

$$\begin{aligned} g_{w,b}((w + b)^2/2b) \\ = k^2(k^2 - 2b(k - b))^{-1}(1 - 2bk^{-1})^k < k^2(k^2 - 2b(k - b))^{-1}e^{-2b/k}. \end{aligned} \tag{4.6}$$

Now for every fixed b the right hand side of (4.6) tends to 1 as $k \to \infty$, and its derivative with respect to k is positive. Thus the second part of (4.5) follows. Similarly, (using $1 - x > \exp\{-x/(1 - x)\}$)

$$\begin{aligned} g_{w,b}(w^2/2b) &= w^2(w^2 - 2bw)^{-1}(1 - 2bw^{-2})^{w+b} \\ &> w^2(w^2 - 2bw)^{-1} \exp -2b(w + b)(w^2 - 2b)^{-1}, \end{aligned} \tag{4.7}$$

and the right hand side of (4.7) tends to 1, for every fixed b, as $w \to \infty$, and its derivative with respect to w is negative for all $w > 2b$, and thus the first part of (4.5) follows.

Actually (4.4) can be strengthened to

$$[(w + 2b/3 - 1)^2(2b)^{-1}]^* \leqq N(w,b) \leqq [w^2/2b + w(\tfrac{2}{3} - 1/3b) + b/6 - \tfrac{1}{3}] \tag{4.8}$$

and (4.3) can be replaced accordingly. The proof of (4.8) is similar to that of (4.4), but the algebra becomes tedious.

For Rule B there exists a unique uniformly minimum variance unbiased estimator, (UMVUE), given by Goodman [5], Section 3. Set $W_B = t_B - B$ = number of white balls seen until stopping. Goodman shows that the UMVUE is given by

$$W_B^2/2B + (\tfrac{2}{3} - 1/6B)W_B + P_1/P_2 \tag{4.9}$$

where P_1 and P_2 are polynomials of degree $2B - 1$ in W_B. The MLE and the UMVUE thus have the same asymptotic distribution. (4.8) yields for the MLE

$$W_B^2/2B + (\tfrac{2}{3} - 1/B)W_B + C_1(B) \leqq \hat{N} \leqq W_B^2/2B + (\tfrac{2}{3} - 1/3B)W_B + C_2(B)$$

where $C_1(B)$ and $C_2(B)$ do not depend on W_B. This should be compared with (4.9). Darroch [3], p. 351, shows that the UMVUE equals $S_{t_B}^{(W_B)}/S_{t_B-1}^{(W_B)}$. This follows directly from (4.1), since summation of (4.1) over $k = B + 1, \cdots, B + N$ yields 1 for every N and B. Darroch also considers the MLE, but does not obtain its asymptotic distribution.

5. Rule C. Let W_C denote the number of white balls seen until stopping. t_C can take on only the values $[(C + 1)k]$, $k = 1, \cdots, N$, (the square bracket is superfluous for integer C), and $t_C = [(C + 1)k]$ if and only if $W_C = k$. Set $C + 1 = \gamma$. The exact distribution of t_C is given by

$$(5.1) \quad P_N(W_C = k) = P_N(t_C = [\gamma k]) = (N)_k h_C(k)/N^{[\gamma k]}, \quad k = 1, \cdots, N,$$

where $h_C(k)$ are constants given by

$$(5.2) \quad h_C(1) = 1, \quad h_C(k) = \{1 - \textstyle\sum_{i=1}^{k-1} (k)_i h_C(i) k^{-[\gamma i]}\} k^{[\gamma k]} (k!)^{-1}, \quad k \geqq 2.$$

The proof of (5.1) is as follows. There are $(N)_k$ possibilities of drawing k distinct white balls, and $N^{[\gamma k]}$ ways of drawing any sample of size $[\gamma k]$. We have denoted by $h_C(k)$ the number of ways of ordering k distinct elements, (allowing repetitions), in $[\gamma k]$ places, in such an order that counting from the left, the number of repetitions among the j first elements remains less than C times the number of distinct elements among the j first, for all $j < [\gamma k]$. Since for every $N = 1, 2, \cdots$ (5.1) is a distribution, i.e. the sum over its elements is one, the induction formula (5.2) follows.

It is easily seen that a unique UMVUE exists also for Rule C. It is a function of W_C, which we denote by $a_C(W_C)$. The sequence $a_C(k)$ is the (unique) solution of the equations

$$(5.3) \quad \textstyle\sum_{i=1}^{k} a_C(i)(k)_i h_C(i) k^{-[\gamma i]} = k, \qquad k = 1, 2, \cdots .$$

The solution is given by the inductive formula

$$(5.4) \quad a_C(1) = 1, \quad a_C(k) = \{k - \textstyle\sum_{i=1}^{k-1} a_C(i)(k)_i h_C(i) k^{-[\gamma i]}\} k^{[\gamma k]}/h_C(k)k!,$$

$k \geqq 2$. (The UMVUE is not only integer valued. For example $a_1(3) = 4.8$)

Consider now the MLE. If upon stopping we have w white and b black balls then w, b must satisfy $b \geqq Cw$ and $b - 1 < Cw$, and thus

$$(5.5) \quad (\gamma + 1/w)^{-1} < w/(w + b) \leqq \gamma^{-1}$$

and equality holds on the right hand side of (5.5) whenever C is an integer. Thus for Rule C the approximation (2.7) yields approximately $\hat{N} \approx t_C/m(\gamma^{-1})$, i.e. the MLE is approximately proportional to the stopping time, with the proportionality factor depending on C only. How close this approximation actually is can be seen from

Assertion 3. *Let s be fixed, let $m(s)$ be as in* (2.6) *and let*

$$(5.6) \quad H(s) = (1 - sm(s))/(s - 1 + sm(s)).$$

Then

$$(5.7) \quad [(w + b)/m(s) - H(s)]^* \leqq \hat{N}(w, b) < (w + b)/m(s)$$

for all w, b *with* $w/(w + b) = s$.

PROOF. The right hand side inequality follows since for $w/(w+b) = s$

$$g_{w,b}((w+b)/m(s)) = (1 - sm(s))^{-1}(1 - m(s)(w+b)^{-1})^{w+b} < (1 - sm(s))^{-1}e^{-m(s)} = 1.$$

On the other hand, substituting $(w+b)/m(s) - H$ for N in (2.8), where $H > 0$ is some fixed constant, and writing $i = w + b$, $w = si$ yields for $i > Hm(s)$ (the only values of interest)

$$g_{w,b}(w+b)(m(s))^{-1} - H) = (i - Hm(s))\{i(1 - sm(s)) - Hm(s)\}^{-1} (1 - m(s)(i - Hm(s))^{-1})^{i} \rightarrow_{i\to\infty} 1.$$

Differentiating the above with respect to i and using the inequality $\log(1 - x) < -x$ $(0 < x < 1)$ yields, after some algebra, that the derivative is less than

$$(1 - m(s)(i - Hm(s))^{-1})^{i-1}(m^2(s)(i\{(1 - sm(s))(H+1) - sH\} - (1-s)m(s)H(H+1))(i - Hm(s))^{-1}\{i(1 - sm(s)) - Hm(s)\}^{-2}.$$

For this to be negative for all $i > Hm(s)$ the value in the curly brackets in the numerator must be nonpositive. Equating the curly bracket to zero and solving for H yields (5.6) which is easily seen to be positive. (5.7) follows.

In order to consider the asymptotic distribution of $\hat{N}$, we need to know (whenever C is not an integer) how close $m((\gamma + 1/w)^{-1})$ is to $m(\gamma^{-1})$, for large w. (See (5.5)). This can be obtained by differentiating $m(s)$, as given by (2.6). Some algebra yields

$$(5.8)\quad \lim_{w\to\infty} w\gamma\{m((\gamma + 1/w)^{-1}) - m(\gamma^{-1})\} = m(\gamma^{-1})/(m(\gamma^{-1}) - C).$$

We have not succeeded in obtaining an exact asymptotic distribution for t_C and $\hat{N}$, but we proceed to give bounds on the limiting distribution in terms of the standard normal distribution. Let

$$(5.9)\quad \sigma^2 = m(\gamma^{-1})(\gamma - m(\gamma^{-1}))/(m(\gamma^{-1}) - C), \quad \text{and} \quad \sigma^{*2} = \sigma^2/(m(\gamma^{-1}))^2.$$

It becomes apparent later that there are good reasons to believe in the correctness of the following

Conjecture. For Rule C and every fixed $C > 0$

$$(5.10)\quad \lim_{N\to\infty} P_N((t_C - Nm(\gamma^{-1}))N^{-\frac{1}{2}}\sigma^{-1} \leqq x) = \Phi(x), \quad -\infty < x < \infty,$$

and

$$(5.11)\quad \lim_{N\to\infty} P_N((\hat{N} - N)N^{-\frac{1}{2}}\sigma^{*-1} \leqq x) = \Phi(x), \quad -\infty < x < \infty.$$

If the correctness of (5.10) is established, then (5.11) follows from Assertion 3, (5.5) and (5.8).

We have

ASSERTION 4. *For Rule* C *and every fixed* $C > 0$

$$(5.12)\quad \liminf_{N\to\infty} P_N((t_C - Nm(\gamma^{-1}))N^{-\frac{1}{2}}\sigma^{-1} \leqq x) \geqq \Phi(x), \quad -\infty < x < \infty,$$

and

$$(5.13)\quad \liminf_{N\to\infty} P_N((\hat{N} - N)N^{-\frac{1}{2}}\sigma^{*-1} \leqq x) \geqq \Phi(x), \quad -\infty < x < \infty.$$

PROOF. (5.13) follows from (5.12), Assertion 3, (5.5) and (5.8). To show (5.12), notice that, for every k, $w_{[\gamma k]} \leqq k$ implies $t_C \leqq \gamma k$. Set $k_N = [\alpha N + \beta N^{\frac{1}{2}} x]^*$. We shall find the values of α and β for which Case C of Theorem 1 yields a useful approximation, namely

$$P(w_{[\gamma k_N]} \leqq k_N) \to \Phi(x). \tag{5.14}$$

We have

$$\begin{aligned} Ew_{[\gamma k_N]} &= N(1 - e^{-\alpha\gamma}) + N^{\frac{1}{2}} x\gamma\beta e^{-\alpha\gamma} + o(N^{\frac{1}{2}}), \\ \operatorname{Var} w_{[\gamma k_N]} &= N\{e^{-\gamma\alpha}(1 - (1 + \gamma\alpha)e^{-\gamma\alpha})\} + o(N). \end{aligned} \tag{5.15}$$

Hence for (5.14) to hold we must have $\alpha = (1 - e^{-\alpha\gamma})$, which together with (2.6) yields $\alpha = m(\gamma^{-1})/\gamma$. With this value of α we have from (5.15), (2.6) and the definition (5.9) $\operatorname{Var} w_{[\gamma k_N]} = N\sigma^2(m(\gamma^{-1}) - C)^2/\gamma^2 + o(N)$. Solving for β, substituting the value $m(\gamma^{-1})/\gamma$ for α, now yields $\beta = \sigma/\gamma$. Thus we get, for $k_N = [Nm(\gamma^{-1})/\gamma + \sigma N^{\frac{1}{2}}/\gamma]^*$,

$$P_N(t_C \leqq Nm(\gamma^{-1}) + x\sigma N^{\frac{1}{2}}) \geqq P_N(w_{[\gamma k_N]} \leqq k_N) \to \Phi(x),$$

which yields (5.12).

We proceed to obtain an upper bound on $P_N(t_C \leqq [\gamma k])$. For simplicity we shall assume that C is an integer. Then $t_C = \gamma k$ implies that the last γ balls drawn were black, and the first $\gamma(k - 1)$ draws resulted in exactly k white balls. Thus, by (3.1)

$$\begin{aligned} P_N(W_C = k) = P_N(t_C = \gamma k) &\leqq k^\gamma (N)_k S_{\gamma(k-1)}^{(k)}/N^{\gamma k} \\ &= \binom{N}{k}(k/N)^{\gamma k} k!\, S_{\gamma(k-1)}^{(k)}/k^{\gamma(k-1)} \end{aligned} \tag{5.16}$$

with strict inequality for all $k > 1$. Set

$$v_C(k) = k!\, S_{\gamma(k-1)}^{(k)}/k^{\gamma(k-1)}. \tag{5.17}$$

We shall obtain an approximation of $v_C(k)$, using the result of Moser and Wyman [7], by which

$$\begin{aligned} S_{\gamma(k-1)}^{(k)} \sim (\gamma(k-1))!\,(e^{n(k)} - 1)^k(\{n(k)\}^{\gamma(k-1)}k!)^{-1} \\ \cdot\{(e^{n(k)} - 1)(2\pi\gamma(k-1)(e^{n(k)} - 1 - n(k)))^{-1}\}^{\frac{1}{2}}, \end{aligned} \tag{5.18}$$

where, for abbreviation, we have let $n(k) = m(k/\gamma(k-1))$. (See (2.6).) (The approximation to $S_n^{(k)}$ given in [7] cannot generally be taken as a limit statement. It can, however, when k, n tend to ∞ so that $k/n \to \alpha$ where $0 < \alpha < 1$.) By Stirling's formula, (5.17) and (5.18) we have

$$\begin{aligned} v_C(k) \sim \{(e^{n(k)} - 1)e^{-\gamma}(1 - e^{-n(k)})^{-\gamma}\}^k \\ \cdot e^\gamma(1 - e^{-n(k)})^\gamma\{(e^{n(k)} - 1)(e^{n(k)} - 1 - n(k))^{-1}\}^{\frac{1}{2}}. \end{aligned} \tag{5.19}$$

Clearly $n(k) \to m(\gamma^{-1})$ as $k \to \infty$. The rate of convergence can be obtained through differentiation of $m(s)$. This yields

$$k\{n(k) - m(\gamma^{-1})\} \to -m(\gamma^{-1})/\{m(\gamma^{-1}) - C\}, \tag{5.20}$$

and some algebra shows that (5.20) implies

$$\{(1 - e^{-m(\gamma^{-1})})(1 - e^{-n(k)})^{-1}\}^{\gamma k} \to \exp\{-\gamma(m(\gamma^{-1}) - \gamma)(m(\gamma^{-1}) - C)^{-1}\},$$

$$\{(e^{n(k)} - 1)(e^{m(\gamma^{-1})} - 1)^{-1}\}^{-k} \to \exp\{-\gamma(m(\gamma^{-1}) - C)^{-1}\}. \tag{5.21}$$

Thus (5.19), (5.21) and some algebra yield

$$v_C(k) \sim P_C^{\,k}(1 - e^{-m(\gamma^{-1})})^{\gamma}(m(\gamma^{-1}) - C)^{-\frac{1}{2}}, \tag{5.22}$$

where $\quad P_C = (e^{m(\gamma^{-1})} - 1)\{e(1 - e^{-m(\gamma^{-1})})\}^{-\gamma}.$

Detailed analysis shows that (5.22) can actually be strengthened to

$$v_C(k) = P_C^{\,k}(1 - e^{-m(\gamma^{-1})})^{\gamma}(m(\gamma^{-1}) - C)^{-\frac{1}{2}}(1 + O(i^{-1})). \tag{5.23}$$

This yields

ASSERTION 5. *For Rule* C *and integer* C

$$\limsup_{N\to\infty} P_N((t_C - Nm(\gamma^{-1}))N^{-\frac{1}{2}}\sigma^{-1} \leqq x) \leqq M_C\Phi(x), \quad -\infty < x < \infty, \tag{5.24}$$

and

$$\limsup_{N\to\infty} P_N((\hat{N} - N)N^{-\frac{1}{2}}\sigma^{*-1} \leqq x) \leqq M_C\Phi(x), \quad -\infty < x < \infty, \tag{5.25}$$

where

$$M_C = \{1 - e^{-m(\gamma^{-1})}\}^{\gamma}(m(\gamma^{-1}) - C)^{-1} \downarrow 1 \quad as \quad C \to \infty. \tag{5.26}$$

PROOF. (5.25) follows from (5.24) and Assertion 3. The proof of (5.24) is similar to Feller's proof [4], p. 169–173, of the DeMoivre-Laplace theorem, and we therefore only outline it briefly. For fixed C set $q(k) = \binom{N}{k}(k/N)^{\gamma k}v_C(k)$, and $k = \delta_k + m(\gamma^{-1})N/\gamma$. Stirling's formula and some algebra yield

$$q(k) \sim e^{f(k)}(1 - e^{-m(\gamma^{-1})})^{\gamma}\{N(m(\gamma^{-1}) - C) \cdot 2\pi(m(\gamma^{-1})/\gamma + \delta_k/N)(1 - m(\gamma^{-1})/\gamma - \delta_k/N)\}^{-\frac{1}{2}}, \tag{5.27}$$

where

$$f(k) = -\delta_k^2(2N)^{-1}\gamma^2(m(\gamma^{-1}) - C)\{m(\gamma^{-1})(\gamma - m(\gamma^{-1}))\}^{-1} - \tfrac{1}{6}\delta_k^3N^{-2}\gamma^2(m(\gamma^{-1})^{-2} + (\gamma - m(\gamma^{-1}))^{-2}) + \cdots.$$

Suppose $\delta_k^3/N^2 \to 0$. (This implies $\delta_k/N \to 0$). Let $\varphi(x) = (2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$. Using the notation (5.9) and (5.26) we can rewrite (5.27) as

$$q(k) \sim M_CN^{-\frac{1}{2}}\gamma\sigma^{-1}\varphi(\delta_kN^{-\frac{1}{2}}\gamma\sigma^{-1}). \tag{5.28}$$

Approximating the Riemann sum by the corresponding integral, yields, for any integers α_N, β_N satisfying

$$(5.29)\quad (\alpha_N - Nm(\gamma^{-1})/\gamma)^3/N^2 \to 0, \qquad (\beta_N - Nm(\gamma^{-1})/\gamma)^3/N^2 \to 0,$$

$$(5.30)\quad \textstyle\sum_{k=\alpha_N}^{\beta_N} q(k) - M_C\{\Phi(\{\beta_N - Nm(\gamma^{-1})\gamma^{-1}\}N^{-\frac{1}{2}}(\sigma/\gamma)^{-1}) \\ - \Phi(\{\alpha_N - Nm(\gamma^{-1})\gamma^{-1}\}N^{-\frac{1}{2}}(\sigma/\gamma)^{-1})\}^{-1} \to 0.$$

One can show that α_N can be chosen as above to satisfy $\sum_{k=1}^{\alpha_N-1} P_N(W_C = k) \to 0$. Therefore (5.24) follows from (5.16) and (5.30) upon taking $\beta_N = [Nm(\gamma^{-1})/\gamma + xN^{\frac{1}{2}}\sigma/\gamma]$. It is easy to see that $\lim M_C = 1$. Lengthy algebra also shows that $dM_C/dC < 0$, and hence (5.26) follows.

Some values of M_C are of interest. We have $M_1 = 1.070$, $M_2 = 1.013$, $M_3 = 1.003$, $M_9 = 1 + 2 \times 10^{-7}$.

It is worth while to remark that the true variance, divided by N, need not necessarily tend to the "asymptotic variance", σ^2, given by (5.9). For example, for $C \leqq 1$ (5.1) yields $P_N(t_C = 2) = 1/N$, and thus the first term alone adds to the actual variance approximately $Nm(\gamma^{-1})^2$, whereas it clearly has no influence on the asymptotic distribution.

6. Rules D and E. The definition of t_D, as given, rather than $\inf \{i \mid b_i \geqq w_i \log w_i + w_i D\}$, is necessary in order to prevent us from stopping with one observation only, i.e., with no black balls observed, which we would have to, whenever $D \leqq 0$, according to the latter definition. For $D > 0$ the modification is redundant. A similar remark holds for Rule E. t_D can take on the values $k + 1$ for $k = 1, \cdots, [e^{-D}]^*$, and the values $[k \log k + k(D + 1)]$ for $k = [e^{-D}]^* + 1, \cdots, N$. We shall consider D as fixed, and abbreviate notation by setting

$$(6.1)\quad \begin{aligned} a_k &= k + 1 && \text{for} \quad k \leqq e^{-D} \\ &= k \log k + k(D + 1) && \text{for} \quad k > e^{-D}. \end{aligned}$$

The exact distribution of t_D can be obtained along the same lines as the distribution of t_C was obtained in Section 5, and similarly the UMVUE can be obtained. We shall not consider this in detail, but shall find the asymptotic distribution of t_D and $\hat{N}$. It is immediate from Section 4 that $P_N(t_D > e^{-D}) \to 1$ as $N \to \infty$. Let U_D be the number of unobserved white balls when we stop. Then $t_D = [a_k]$ if and only if $U_D = N - k$. Rule D was first considered by Darling and Robbins, in a recent paper [2]. They prove that

$$(6.2)\quad P_N(U_D = 0) \to e^{-\lambda} \quad \text{where} \quad \lambda = e^{-(D+1)},$$

and thus suggest W_D, the total of white balls observed, as an estimator for N, (for D chosen large enough). A slight modification of their proof yields the strengthening of (6.2) to become

Assertion 6. *For Rule* D *and every fixed* D, $-\infty < D < \infty$,

$$(6.3)\quad P_N(U_D = j) \to e^{-\lambda}\lambda^j/j!, \quad j = 0, 1, \cdots, \quad \textit{where } \lambda = e^{-(D+1)}.$$

Proof. If is easy to show that for every fixed k, $k = 0, 1, \cdots,$

$$(6.4)\quad \limsup_{N\to\infty} P_N(U_D \leqq k) \leqq \textstyle\sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!.$$

Notice that $U_D \leqq k$ implies $u_{[a_{N-k}]} \leqq k$ (where u_i is defined in Section 1), and thus $P_N(U_D \leqq k) \leqq P_N(u_{[a_{N-k}]} \leqq k)$. For fixed k define D_N by

$$a_{N-k} = N \log N + N(D_N + 1). \tag{6.5}$$

Then by (6.1) and simple algebra it follows that $D_N \to D$ as $N \to \infty$. The condition of Case D of Theorem 1 is thus fulfilled, and yields (6.4). It is much more difficult to show that

$$\liminf_{N\to\infty} P(U_D \leqq k) \geqq \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!. \tag{6.6}$$

Since the proof is very similar to the proof of [2], we shall not repeat it here, so as to save space.

As is well known, the maximal term (mode) of a Poisson distribution with parameter λ is the $[\lambda]^*$th term (except when λ is an integer, and both the λth and $(\lambda - 1)$st terms are maximal.) It therefore seems plausible that unless t_D stops very early the value of the MLE will be approximately $W_D + [e^{-(D+1)}]^*$. We have

ASSERTION 7. *For any* $-\infty < D < \infty, w > e^{-D}$,

$$\hat{N}(w, [w \log w + wD]) \geqq w + [e^{-(D+1)}]^* \tag{6.7}$$

and

$$\hat{N}(w, [w \log w + wD]) = w + [e^{-(D+1)}]^* \quad \textit{for all} \quad w > K_D. \tag{6.8}$$

PROOF. We use (2.8) with $N = w + a, a > 0$. This yields

$$\begin{aligned}(w + a)a^{-1}(1 - (w + a)^{-1})^{w\log w + w(D+1)} \\ \geqq g_{w,[w\log w + wD]}(w + a) \\ > (w + a)a^{-1}(1 - (w + a)^{-1})^{w\log w + w(D+1)+1},\end{aligned} \tag{6.9}$$

and as $w \to \infty$ all members of (6.9) tend to $e^{-(D+1)}/a$. The derivative of the term to the left in (6.9) is negative for all $w \geqq e^{-(D+1)}$, and thus setting $a = e^{-(D+1)}$ yields (6.7). Also the derivative of the right hand term in (6.9) is negative. For any $a > e^{-(D+1)}$ the limit in (6.9) is less than 1, and thus (6.8) follows for all w sufficiently large. The constant K_D depends only on D. For all $D > 0$, $w = 1$ and $a = 1$ the value of the left hand side of (6.9) becomes $(\frac{1}{2})^D < 1$, and since the function on the left in (6.9) is for every a and D, a decreasing function, Assertion 6 can be strengthened to yield

$$\hat{N}(w, [w \log w + wD]) = w \quad \text{for} \quad w = 1, 2, \cdots, \quad \text{whenever} \quad D > 0. \tag{6.10}$$

Assertions 6 and 7 yield

ASSERTION 8. *For Rule* D *and any* $-\infty < D < \infty$

$$\lim_{N\to\infty} P_N(\hat{N} = N - j + [\lambda]^*) = e^{-\lambda}\lambda^j/j!, \quad j = 0, 1, \cdots, \quad \textit{where} \quad \lambda = e^{-(D+1)}. \tag{6.11}$$

For Rule E

$$\lim_{N\to\infty} P_N(\hat{N} = N) = 1.$$

Thus the estimator proposed in [2] coincides with the MLE, for $D > 0$, and its distribution is given by (6.11).

Acknowledgment. The author is indebted to Professor Robbins for raising the problem of finding the MLE for Rule B, which initiated the present investigation. She also wishes to express her gratitude to Professors Darling and Robbins for letting her see a copy of [2] prior to its publication.

REFERENCES

[1] CHAPMAN, D. G. (1954). The estimation of biological populations. *Ann. Math. Statist.* **25** 1–15.

[2] DARLING, D. A. and ROBBINS, H. (1967). Finding the size of a finite population. *Ann. Math. Statist.* **38** 1392–1398.

[3] DARROCH, J. N. (1958). The multiple recapture census. I. Estimation of a closed population. *Biometrika* **45** 343–359.

[4] FELLER, W. (1957). *An Introduction to Probability and its Applications*. **1** (2nd Edition). Wiley, New York.

[5] GOODMAN, L. A. (1953). Sequential sampling tagging for population size problems. *Ann. Math. Statist.* **24** 56–69.

[6] LEWONTIN, R. C. and PROUT, T. (1956). Estimation of the number of different classes in a population. *Biometrics* **12** 211–223.

[7] MOSER, L. and WYMAN, M. (1958). Stirling numbers of the second kind. *Duke Math. J.* **25** 29–43.

[8] RÉNYI, A. (1962). Three new proofs and a generalization of a theorem of Irving Weiss. *Public. Math. Inst. Hung. Acad. Sciences Ser. A* **7** 203–214.