Sequential Search Algorithm for Estimation of the Number of Classes in a Given Population

Klass, Michael Jay; Nowicki, Krzysztof

2016

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Sequential Search Algorithm for Estimation of the Number of Classes in a Given Population

MICHAEL J. KLASS, UNIVERSITY OF CALIFORNIA, BERKELEY
KRZYSZTOF NOWICKI, LUND UNIVERSITY
Sequential Search Algorithm
for Estimation of the Number of Classes in a Given Population

Michael J. Klass
University of California, Berkeley
and
Krzysztof Nowicki
Lund University, Sweden

Abstract
Let \( N \) be the number of classes in a population to be estimated. Fix any preassigned error probability \( 0 < \epsilon < \exp(-2) \) (roughly). We present a sequential search algorithm to estimate the exact value of \( N \), with an error probability of at most \( \epsilon \), regardless of the value of \( N \).

Key words and phrases: Unobserved species, estimation of population size, sequential estimation procedure, error probability

1 Introduction
Historically many people consider the classic "species problem", that of estimating the number \( N \) of categories in a given population based on a random sample, first introduced by Fisher et al (1943) and widely studied in ecology and later extended to many other applications: see, for instance, Thisted and Efron (1987) and Mao and Lindsay (2007). Bunge and Fitzpatrick (1993) provide a review of various statistical methods to estimate the number of unseen species.
The estimation of the number of unseen species is a question closely related to the problem of estimating the expected number of new species that will be seen if we take an additional sample of any given size, see Good and Toulmin (1956), Efron and Thisted (1976), Boneh et al (1998).

Several authors proposed estimation methods of the number of classes utilizing sequential random sampling with various stopping rules, see Goodman (1953), Samuel (1968, 1969), Holst (1971) and Nayak and Kundu (2007). In the algorithm we propose the stopping rule is connected to a preassigned error probability \( P(N \neq \hat{N}) \) where \( \hat{N} \) is our estimate of \( N \).

The problem addressed in this paper can be formalized as follows. Let \( N \) be a fixed but unknown positive integer. Let \( X_1, X_2, \ldots \) be i.i.d. random variables which take values in \( \{1, 2, \ldots, N\} \), each of the \( N \) outcomes being equally likely. Fix any preassigned error probability \( 0 < \epsilon < \exp(-2) \) (roughly). We want to estimate the exact value of \( N \), with an error probability of at most \( \epsilon \), regardless of the value of \( N \).

2 The search algorithm

To teach ourselves how to proceed, suppose we begin by asking: When might one decide it would no longer be advantageous to continue searching/sampling (gathering observations) for items yet unseen?

Given that \( j \) different objects have already been recorded, let \( W_{j+1} \) denote the random waiting time describing the number of additional observations that happen to be needed until the \( (j + 1) \)\textsuperscript{th} item surfaced, with \( W_{j+1} = \infty \) if there are only \( j \) items. Formally, define \( T_1 = W_1 = 1 \) and, having defined \( W_j \) for \( 1 \leq j \leq k \), let \( T_j = W_1 + \cdots + W_j \) and define

\[
W_{j+1} = \begin{cases} 
1^{st} i \geq 1 : X_{i+T_j} \notin \{X_1, \ldots, X_{T_j}\} \\
\infty \text{ if no such } i \text{ exists.}
\end{cases}
\]  

(1)

and \( T_{j+1} = W_1 + \cdots + W_{j+1} \).

Suppose \( N = n \) for some integer \( n > 0 \). For each \( n \) we use notation \( W_{n,j+1} \) and \( T_{n,j+1} \). Imagine that we have found \( j \) different objects already and have conducted another \( s \) observations without finding anything new. What information does this provide?

\[
P(W_{n,j+1} > s) = \left( \frac{j}{n} \right)^s, \text{ for } 1 \leq j \leq n - 1.
\]

(2)
Suppose our search has currently found \( j \) objects and then \( W_{j+1} \) exceeds \( \lceil A(j + k) \rceil \) where \( k \) is to be determined. Let

\[
t = 1^{st} \quad j \geq 1 : W_{j+1} > \lceil A(j + k) \rceil
\]

We stop at such time \( t \) and declare that \( \hat{N} \), our best guess as to the value of \( N \), is \( t \). At that point we have conducted

\[
Q \equiv 1 + W_2 + \cdots + W_t + \lceil A(t + k) \rceil
\]

searches. What is the maximum error probability incurred by this rule over all possible values of \( N \geq 1 \)?

Let \( \hat{N} \) be the integer which our procedure guesses. If \( n = 1 \), \( \hat{N} \) is always 1. For \( N = n \geq 2 \)

\[
P(\hat{N} \neq n) = P(\bigcup_{j=1}^{n-1} \{W_{n,j+1} > \lceil A(j + k) \rceil\}). \tag{5}
\]

To upper-bound this expression we introduce the following lemma.

**Lemma 2.1** Let \( E_{j+1}, \ 1 \leq j \leq n - 1 \) be a set of independent events and let \( E_{j+1}^*, \ 1 \leq j \leq n - 1 \) be a set of independent events such that \( P(E_{j+1}) \leq P(E_{j+1}^*) \) for \( 1 \leq j \leq n - 1 \). Then

\[
P(\bigcup_{j=1}^{n-1} E_{j+1}) \leq \sum_{j=1}^{n-1} P(E_{j+1}^*) - P(E_n^*)P(E_{n-1}^*). \tag{6}
\]

**Proof:** First observe that

\[
P(\bigcup_{j=1}^{n-1} E_{j+1}) = 1 - P(\bigcap_{j=1}^{n-1} E_{j+1}^c)
\]

\[
= 1 - \bigcap_{j=1}^{n-1} P(E_{j+1}^c) \quad \text{(by independence)}
\]

\[
\leq 1 - \bigcap_{j=1}^{n-1} P((E_{j+1}^*)^c) \quad \text{(by the assumption)}
\]

\[
= P(\bigcup_{j=1}^{n-1} E_{j+1}^*).
\]

\[
P(\hat{N} \neq n) = P(\bigcup_{j=1}^{n-1} \{W_{n,j+1} > \lceil A(j + k) \rceil\}).
\]

\[
P(\hat{N} \neq n) = P(\bigcup_{j=1}^{n-1} \{W_{n,j+1} > \lceil A(j + k) \rceil\}).
\]

\[
P(\hat{N} \neq n) = P(\bigcup_{j=1}^{n-1} \{W_{n,j+1} > \lceil A(j + k) \rceil\}).
\]

\[
P(\hat{N} \neq n) = P(\bigcup_{j=1}^{n-1} \{W_{n,j+1} > \lceil A(j + k) \rceil\}).
\]
Moreover, by Boole’s inequality
\[ P(\bigcup_{j=1}^{n-1} E_{j+1}^*) \leq P(E_n^* \cup E_{n-1}^*) + \sum_{j=1}^{n-2} P(E_{j+1}^*) \]  
(8)
and by independence \( P(E_n^* \cup E_{n-1}^*) = P(E_n^*) + P(E_{n-1}^*) - P(E_n^*)P(E_{n-1}^*) \) from which (6) follows. \[ \square \]

For \( 1 \leq j \leq n - 1 \), \( A > 0 \) and fixed \( k \geq 1 \) let
\[ E_{n,j+1} = \{ W_{n,j+1} > \lceil A(j+k) \rceil \} \]  
(9)
and
\[ P(E_{n,j+1}^*) = \left( \frac{j}{n} \right)^{A(j+k)}. \]  
(10)
Clearly, \( P(E_{n,j+1}) \leq P(E_{n,j}^*) \). Using Lemma 2.1, (5) can be upper-bounded by
\[ P(\bigcup_{j=1}^{n-1} E_{n,j+1}^*) \leq P(\bigcup_{j=1}^{n-1} E_{n,j+1}^*) \leq \sum_{j=1}^{n-1} P(E_{n,j+1}^*) - P(E_{n,n}^*)P(E_{n,n-1}^*). \]  
(11)
Next we introduce upper-bounds for terms in (11).

**Lemma 2.2** For \( 1 \leq j \leq 2k \) and \( n \geq j + 1 \)
\[ P(E_{n,n-j+1}^*) \leq \exp(-Aj). \]  
(12)

**Proof:** For \( k \geq 1 \), \( 1 \leq j \leq n - 1 \), \( A > 0 \),
\[ P(W_{n,j+1} > \lceil A(j+k) \rceil) = \left( \frac{j}{n} \right)^{\lceil A(j+k) \rceil} \text{ (by definition of } W_{n,j+1} \text{)} \]
\[ \leq \left( \frac{j}{n} \right)^{A(j+k)} \]
\[ = \exp(-A(j+k) \ln(\frac{n}{j})) \equiv P(E_{n,j+1}^*). \]  
(13)
Replacing \( j \) by \( n - j \) and \( n \) by \( x \), for \( x \geq j + 1 \) let
\[ f(x) = -(x - j + k) \ln(\frac{x}{x - j}). \]  
(14)
Notice that $P(E_{n,n-j+1}^*) = \exp(Af(n))$. Then
\[
f'(x) = -\ln\left(\frac{x}{x-j}\right) - (x-j+k)\left(\frac{1}{x} - \frac{1}{x-j}\right)
= -\ln\left(\frac{x}{x-j}\right) - \frac{k-j}{x} + \frac{k}{x-j}
\]
and
\[
f''(x) = -\frac{1}{x} + \frac{1}{x-j} + \frac{k-j}{x^2} - \frac{k}{(x-j)^2}
= \frac{j}{x(x-j)} + \frac{(k-j)(x-j)^2-kx^2}{x^2(x-j)^2}
= \frac{jx^2-j^2x-jx^2-2j(k-j)x+j^2(k-j)}{x^2(x-j)^2}
= \frac{j(j-2k)x+j^2(k-j)}{x^2(x-j)^2}
\leq 0 \quad \text{(for} \quad k \leq j \leq 2k \text{ or if} \quad 1 \leq j \leq k-1 \text{ since} \quad (2k-j)x \geq (k-j)j)\).
\]
Therefore $f(x)$ is concave. Notice that $\lim_{x \to \infty} f'(x) \to 0$. Therefore $f'(x) > 0$ for all $x > j + 1$. Consequently
\[
\sup_{x \geq j+1} f(x) = \lim_{x \to \infty} f(x).
= \lim_{x \to \infty} x \ln(1 - \frac{j}{x})
= -j.
\]
Hence
\[
\sup_{n \geq j+1} P(E_{n,n-j+1}^*) = \lim_{n \to \infty} P(E_{n,n-j+1}^*) = \lim_{n \to \infty} \exp(Af(n)) = \exp(-Aj).
\]
(18)

\textbf{Theorem 2.1} For all $n \geq 2$, $A \geq 2$, $k = 4$
\[
P\left(\bigcup_{j=1}^{n-1} E_{n,j+1}^*\right) < \exp(-A) + \exp(-2A) + \exp(-3A).
\]
(19)
Proof: We treat various $n$ separately.

For $n = 1$ the error probability is zero.

**Case 1:** $2 \leq n \leq 4$ Inequality (12) combined with Boole’s inequality yields

$$P\left( \bigcup_{j=1}^{n-1} E_{n,j+1}^* \right) \leq \exp(-A) + \exp(-2A) + \exp(-3A) \quad (20)$$

**Cases $n \geq 5$**

Invoking (11) and (12), for $n \geq 5$

$$P\left( \bigcup_{j=1}^{n-1} E_{n,j+1}^* \right) < \exp(-A) + \exp(-2A) + \exp(-3A) - P(E_{n,n}^*) P(E_{n,n-1}^*)$$

$$+ \sum_{j=1}^{n-4} P(E_{n,j+1}^*). \quad (21)$$

**Case $n = 5$**

Since

$$P(E_{5,2}^*) = \exp(-5A \ln(5)), \quad (22)$$

$$P(E_{5,5}^*) P(E_{5,4}^*) = \exp(-(\Gamma_{5,5} + \Gamma_{5,4})A)$$

$$> \exp(-5.4A) \text{ (by (36) and (37) below)}$$

$$> \exp(-5.4A \ln(5))$$

$$= P(E_{5,2}^*) \quad (23)$$

so by (21) Theorem 1.1 holds for $n = 5$.

**Case $n = 6$**

Applying (21) for $n = 6$ we need to verify the inequality

$$P(E_{6,2}^*) + P(E_{6,3}) \leq P(E_{6,5}^*) P(E_{6,6}^*). \quad (24)$$
Since $\Gamma_{n,n}$ and $\Gamma_{n,n-1}$ decrease in $n$,

\[
P(E_{6,6}^*)P(E_{6,5}^*) = \exp(-A(\Gamma_{6,6} + \Gamma_{6,5}))
\]

\[
> \exp(-A(\Gamma_{5,5} + \Gamma_{5,4}))
\]

\[
> \exp(-5.4A)
\]

\[
> \exp(-5A\ln(6)) + \exp(-6A\ln(3))
\]

\[
= P(E_{6,2}^*)P(E_{6,3}^*)
\]

which confirms Theorem 1.1 for $n=6$.

**Case $n=7$**

Since $\Gamma_{n,n}$ and $\Gamma_{n,n-1}$ decrease in $n$,

\[
P(E_{7,7}^*)P(E_{7,6}^*) > P(E_{5,5}^*)P(E_{5,4}^*) > \exp(-5.4A)
\]

\[
> \exp(-5A\ln(7)) + \exp(-6A\ln(\frac{7}{2})) + \exp(-7A\ln(\frac{7}{3}))
\]

\[
= P(E_{7,2}^*) + P(E_{7,3}^*) + P(E_{7,4}^*) \quad \text{(for $A \geq 2$),}
\]

whence Theorem 1.1 holds for $n=7$.

**Case $n \geq 8$**

We begin by considering $\sum_{j=1}^{n-4} P(E_{n,j+1}^*)$. Splitting this sum into three parts,

\[
\sum_{j=1}^{n-4} P(E_{n,j+1}^*) = P(E_{n,2}^*) + \sum_{2 \leq j < \frac{n}{2}} P(E_{n,j+1}^*) + \sum_{\frac{n}{2} < j \leq n-4} P(E_{n,j+1}^*). \quad (27)
\]

We will treat each of the sums separately. First,

\[
\sum_{2 \leq j < \frac{n}{2}} P(E_{n,j+1}^*) = \sum_{j=2}^{\left\lfloor \frac{n}{2} \right\rfloor} \exp(-A(j + 4)\ln(\frac{n}{j}))
\]

\[
\leq \frac{\exp(-6A)}{1 - \exp(-A)} \quad \text{(since $\frac{n}{j} \geq 1$).} \quad (28)
\]
Second,
\[
\sum_{\frac{n}{2} < j \leq n-4} P(E_{n,j+1}^*) = \sum_{\frac{n}{2} < j \leq n-4} \exp(-A(j+4) \ln\left(\frac{n}{j}\right))
\]
\[
= \sum_{\frac{n}{2} < j \leq n-4} \exp(-A(n-j+4) \ln\left(\frac{n}{n-j}\right))
\]
\[
\leq \sum_{1 \leq i < n-\left\lceil \frac{n}{2} \right\rceil -3} \frac{\exp(2 \ln(i) - A(n+1-i) \ln\left(\frac{n}{n-i-3}\right))}{i^2} \tag{29}
\]
\[
\leq \max_{1 \leq i < n-\left\lceil \frac{n}{2} \right\rceil -3} \exp(2 \ln(i) - A(n+1-i) \ln\left(\frac{n}{n-i-3}\right)) \sum_{j=1}^{\infty} \frac{1}{j^2}
\]
\[
\leq \frac{\pi^2}{6} \exp\left(-A n \ln\left(\frac{n}{n-4}\right)\right) \text{ (by Lemma 3.2 if } k = 4).\]

Hence,
\[
\sum_{j=1}^{n-4} P(E_{n,j+1}^*) \leq \exp(-5A \ln 8) + \frac{\exp(-6A)}{1 - \exp(-A)} + \frac{\pi^2}{6} \exp\left(-A n \ln\left(\frac{n}{n-4}\right)\right)
\]
\[
= \exp(-5A \ln 8) + \frac{\exp(-6A)}{1 - \exp(-A)} + \frac{\pi^2}{6} P(E_{n,n-3}^*). \tag{30}
\]

Next we show
\[
\exp(-5A \ln(8)) + \frac{\exp(-6A)}{1 - \exp(-A)} \leq P(E_{n,n}^*) P(E_{n,n-1}^*) - \frac{\pi^2}{6} P(E_{n,n-3}^*). \tag{31}
\]

For \( n \geq 8, A \geq 2 \)
\[
P(E_{n,n}^*) P(E_{n,n-1}^*) - \frac{\pi^2}{6} P(E_{n,n-3}^*) \geq \exp(-4.5A) - \frac{\pi^2}{6} \exp(-5A)
\]
\[
= \exp(-4.5A)(1 - \frac{\pi^2}{6} \exp(-0.5A)) \geq \exp(-4.5A)(1 - \frac{\pi^2}{6} \exp(-1)) \tag{32}
\]
\[
> \exp(-5A \ln(8)) + \frac{\exp(-6A)}{1 - \exp(-A)}
\]

whence (31) holds for \( n \geq 8 \) and \( A \geq 2 \).
Remark 2.1 Notice that for all \( j \geq 1 \)
\[
\lim_{n \to \infty} P(E_{n,n-j+1}) = \lim_{n \to \infty} \exp(-[A(n - j + 4)] \ln\left(\frac{n}{n - j}\right)) = \exp(-Aj). \quad (33)
\]
Hence as the number \( N = n \) of objects to be found tends to infinity the probability that we fail to guess their exact number tends to
\[
\lim_{n \to \infty} P\left(\bigcup_{j=1}^{n-1} E_{n,j+1}\right) = \lim_{n \to \infty} \left(1 - P\left(\bigcap_{j=1}^{n-1} E_{n,j+1}^c\right)\right)
= \lim_{n \to \infty} \left(1 - \prod_{j=1}^{n-1} (1 - P(E_{n,n-j+1}))\right)
= 1 - \prod_{j=1}^{\infty} (1 - \exp(-Aj)). \quad (34)
\]
An alternative expression for this limit may be obtained by writing
\[
\prod_{j=1}^{\infty} (1 - \exp(-Aj)) = \exp\left(\sum_{j=1}^{\infty} \ln(1 - \exp(-Aj))\right)
= \exp(-\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\exp(-jA)}{k})
= \exp(-\sum_{k=1}^{\infty} \frac{\exp(-kA)}{k(1 - \exp(-kA))}). \quad (35)
\]

3 Appendix

We lower- and upper-bound \( P(E_{n,k}^*) \) for \( k = n, n-1, \) and \( n-3 \).

Lemma 3.1 Let \( \Gamma_{n,k} = -\frac{1}{A} \ln(P(E_{n,k}^*)) \). For \( A \geq 2 \) and \( n \geq 2 \),
\[
\exp(-A(1 + \frac{7}{2n} + \frac{11}{6n^2} + \frac{5}{4n^2(n-1)})) \leq P(E_{n,n}^*) \leq \exp(-A\left(1 + \frac{7}{2n}\right)) \quad (36)
\]
\[
P(-A(2 + \frac{6}{n} + \frac{20}{3n^2} + \frac{28}{3n^2(n-2)})) \leq P(E_{n,n-1}^*) \leq \exp(-A(2 + \frac{6}{n})) \quad (37)
\]
and
\[
\exp(-A(4 + \frac{8}{n} + \frac{64}{3n^2} + \frac{64}{n^2(n-4)})) \leq P(E_{n,n-3}^*) \leq \exp(-A(4 + \frac{8}{n})). \quad (38)
\]
Proof: We derive upper- and lower-bounds of $P(E^*_{n,k})$ by bounding $\Gamma_{n,k}$ for $k = n$, $k = n - 1$ and $k = n - 3$.

First,

$$\Gamma_{n,n} = (n + 3) \ln\left(\frac{n}{n - 1}\right) = -(n + 3) \ln\left(1 - \frac{1}{n}\right)$$

$$= \sum_{j=1}^{\infty} \frac{1}{j n^{j-1}} + \sum_{j=1}^{\infty} \frac{3}{j n^{j}}$$

$$= 1 + \sum_{j=1}^{\infty} \left(\frac{1}{j + 1} + \frac{3}{j}\right) \frac{1}{n^{j}} \tag{39}$$

Clearly,

$$\Gamma_{n,n} \geq 1 + \frac{7}{2n} \tag{40}$$

and

$$\Gamma_{n,n} = 1 + \frac{7}{2n} + \frac{11}{6n^2} + \sum_{j=3}^{\infty} \frac{4j + 3}{(j + 1) j n^{j}}$$

$$\leq 1 + \frac{7}{2n} + \frac{11}{6n^2} + \frac{5}{4} \sum_{j=3}^{\infty} \frac{1}{n^{j}} \tag{41}$$

$$= 1 + \frac{7}{2n} + \frac{11}{6n^2} + \frac{5}{4n^2(n - 1)}$$

which gives (36). Second,

$$\Gamma_{n,n-1} = (n + 2) \ln\left(\frac{n}{n - 2}\right) = -(n + 2) \ln\left(1 - \frac{2}{n}\right)$$

$$= 2 + \sum_{j=1}^{\infty} \frac{2j}{n^{j}} \left(\frac{2}{j + 1} + \frac{2}{j}\right) \tag{42}$$
which gives (37) by calculations similar to those used in (40) and (41). Third, we lower-bound $P(E_{n,n-3}^*) = \exp(-A\Gamma_{n,n-3})$ by upper-bounding

$$\Gamma_{n,n-3} = n \ln \left( \frac{n}{n-4} \right) = -n \ln \left( 1 - \frac{4}{n} \right)$$

$$= 4 + \sum_{j=1}^{\infty} \frac{A^{j+1}}{(j+1)n^j}$$

$$= 4 + \frac{8}{n} + \frac{64}{3n^2} + \sum_{j=3}^{\infty} \frac{4}{j+1} \left( \frac{1}{n} \right)^j$$

$$\leq 4 + \frac{8}{n} + \frac{64}{3n^2} + \sum_{j=3}^{\infty} \left( \frac{1}{n} \right)^j$$

$$= 4 + \frac{8}{n} + \frac{64}{3n^2} + \frac{64}{n^2(n-4)}$$

which gives (38).

\[\square\]

**Lemma 3.2** Let $A \geq 2$, $k \geq 1$, $B = n - k + 1$, $n > k$ and

$$g(x) = 2 \ln(B - x) - A(x + k) \ln \left( \frac{n}{x} \right)$$

Then

$$\sup_{\frac{n}{B-1} \leq x \leq B} g(x) = g(B - 1) = -An \ln \left( \frac{n}{n-k} \right).$$

**Proof:** Toward this end,

$$g'(x) = -\frac{2}{B - x} - A \ln \left( \frac{n}{x} \right) + A + \frac{Ak}{x}$$

$$= -\frac{2}{B - x} + A \ln \left( \frac{ex}{n} \right) + \frac{Ak}{x}$$

$$\geq -\frac{2}{B - x} + 2 \ln \left( \frac{ex}{n} \right) + \frac{2k}{x}$$

$$\equiv 2h(x).$$

For $x = \frac{n}{c}$ we have $\ln \left( \frac{ex}{n} \right) = 0$. Hence

$$g'(x) \geq \frac{2}{x(B - x)}(Bk - (k + 1)x) \geq 0$$

(47)
for $\frac{n}{c} \leq x \leq \frac{Bk}{k+1}$. We need to prove that $g'(x) \geq 0$ for $\frac{Bk}{k+1} \leq x \leq B-1$.

Consider

$$h'(x) = -\frac{1}{(B-x)^2} + \frac{1}{x} - \frac{k}{x^2}$$

(48)

For $\frac{Bk}{k+1} \leq x < B$

$$h''(x) = -\frac{2}{(B-x)^3} - \frac{1}{x^2} + \frac{2k}{x^3}$$

$$< -\frac{2(k+1)^3}{B^3} + \frac{2(k+1)^3}{B^3k^2}$$

(49)

$$\leq 0 \text{ for } k \geq 1.$$

Hence $h(x)$ is concave on $\frac{Bk}{k+1} \leq x \leq B-1$. Therefore

$$\inf_{\frac{Bk}{k+1} \leq x \leq B-1} h(x) = \min \{h(\frac{Bk}{k+1}), h(B-1)\} \geq 0 \text{ iff } h(B-1) \geq 0$$

(50)

$$h(B-1) = -1 + \ln\left(\frac{c(n-k)}{n}\right) + \frac{k}{n-k} \equiv q(n)$$

(51)

$$q'(n) = \frac{1}{n-k} - \frac{1}{n} - \frac{k}{(n-k)^2}$$

$$= \frac{k}{n(n-k)} - \frac{k}{(n-k)^2}$$

$$= \frac{k(n-k) - nk}{n(n-k)^2} < 0.$$ 

(52)

Thus $q(n) \downarrow \lim_{n \to \infty} q(n) = 0$ whence $h(B-1) > 0$ and so $g(B-1) = \sup_{\frac{1}{2} \leq x \leq B-1} g(x)$.

□

References


[5] Good, I.J. and Toulmin, G.H. The number of new species and the increase in population coverage, when a sample is increased, Biometrika, 43 (1956), 45–63.


