STATISTICAL ANALYSIS OF SERIAL NUMBERS

José Galvão LEITE[1]
Eliane Maria Brandemarte MOREIRA[1]
Luis Aparecido MILAN[1]

■ ABSTRACT: Consider a random sample of random size of a sequence of items numbered from 1 to N. We propose conditionally UMVUE estimators for N, the variance of the estimator of N and a confidence interval for N.

■ KEYWORDS: Serial numbers; conditionally UMVUE estimator; conditional confidence intervals.

1 Introduction

Consider a population of elements numbered from 1 to N, where N, the size of the population, is unknown. A random sample of size n is obtained, where n is random and $n \leq N$. Each element of the population has probability p, p unknown, of being selected for the sample. We are interested in estimating N, and p is a nuisance parameter.

As an illustrative example, consider the situation where someone is waiting for a taxi and taking notes of the numbers of those which do not stop. Provided that the taxis are numbered sequentially from 1 to N, and each one has probability p of passing by the point where the observer is, the problem is to estimate the number N of taxis available.

This subject was originally presented by (Ruggles and Brodie, 1947) during World War II for estimating equipment available to the enemy. Goodman (1952) treated this problem for a fixed sample size and Roberts (1967) provided an exact Bayesian approach considering informative stopping rules.

Here, we develop a UMVUE point estimator for N and obtain its conditionally UMVUE estimator for the variance and the interval estimate. We also develop

[1]Departamento de Estatística, Universidade Federal de São Carlos - UFSCar, CEP 13565-905, São Carlos, SP, Brasil. E-mail: *leite@power.ufscar.br*

simulation studies in order to verify the performance of the method for estimating the population size for an example.

In Section 2, we define the UMVUE point estimator for N and obtain its conditionally UMVUE estimator for the variance. The interval estimate is derived in Section 3 and, in Section 4, a simulated example is presented. Some related results are proven in Appendix A.

2 Point estimation

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the order statistics for a sample of size n. The sampling distribution is

$$P(n, X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, \cdots, X_{(n)} = x_{(n)} \mid N, p) = p^n (1-p)^{N-n}, \qquad (1)$$

where $1 \leq n \leq x_{(n)}$ and $1 \leq x_{(1)} < x_{(2)} < \cdots < x_{(n)} \leq N$.

The likelihood function is

$$L(N, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)})) = p^n (1-p)^{N-n}, \qquad (2)$$

for $N \geq x_{(n)}$ and $0 < p < 1$.

Following are four examples of unbiased estimators for the population size:

$$\widehat{N_1} = X_{(n)} + X_{(1)} - 1;$$

$$\widehat{N_2} = X_{(n)} + \frac{1}{n-1} \sum_{i=2}^{n} (X_{(i)} - X_{(i-1)} - 1) = \left(\frac{n}{n-1} \right) X_{(n)} - \frac{1}{n-1} X_{(1)} - 1;$$

$$\widehat{N_3} = X_{(n)} + \frac{1}{n}[(X_{(1)} - 1) + (X_{(2)} - X_{(1)} - 1) + \cdots + (X_{(n)} - X_{(n-1)} - 1)] =$$

$$= X_{(n)} + \frac{1}{n}(X_{(n)} - n) = \left(\frac{n+1}{n} \right) X_{(n)} - 1$$

and

$$\widehat{N_4} = M + (M-1) = 2M - 1,$$

where M is the sampling median.

Following, we show that $\widehat{N_3}$ is, conditionally in n, a Uniformly Minimum Variance Unbiased Estimator, (UMVUE), for N.

Theorem 2.1 *$\widehat{N_3}$, given $n \geq 1$, is UMVUE for N and*

$$Var(\widehat{N_3} \mid n, N, p) = \frac{(N+1)(N-n)}{n(n+2)}. \blacksquare$$

Proof: (i) Given $n \geq 1$, $X_{(n)}$ is a sufficient statistic for N. The conditional likelihood function given is

$$L\left(N, p \mid (x_{(1)}, x_{(2)}, \cdots, x_{(n)})\right) = P(X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, \cdots, X_{(n)} = x_{(n)} \mid n, N, p).$$

¿From (1) and considering that the conditional distribution of n, given N and p, is Binomial(N, p),

$$L\left(N, p \mid (x_{(1)}, x_{(2)}, \cdots, x_{(n)})\right) = \begin{pmatrix} N \\ n \end{pmatrix}^{-1},$$

for $N \geq x_{(n)}$, *i.e.*, the conditional likelihood function depends on the sample $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ only through $x_{(n)}$. By the factorization criteria, $X_{(n)}$ is conditionally sufficient for N.

(ii) Given $n \geq 1$, $X_{(n)}$ is complete for N. Suppose g is a real function such that $E(g(X_{(n)}) \mid n, N, p) = 0$, for any $N \geq n$. Then, for $N = n$, we have

$$E(g(X_{(n)}) \mid n, N, p) = g(n)P(X_{(n)} = n \mid n, N, p) = g(n) = 0.$$

For $N > n$, suppose $g(n) = g(n+1) = \cdots = g(N) = 0$. Then

$$E(g(X_{(n)}) \mid n, N+1, p) = g(N+1)P(X_{(n)} = N+1 \mid n, N+1, p)$$

and since

$$E(g(X_{(n)}) \mid n, N+1, p) = 0,$$

we have

$$g(N+1) = 0.$$

Therefore

$$g(n) = g(n+1) = \cdots = g(N+1) = 0$$

and, given n, $X_{(n)}$ is complete for N.

(iii) Given $n \geq 1$, $\widehat{N_3}$ is an unbiased estimate of N. From (14)

$$E(\widehat{N_3} \mid n, N, p) = E\left(\left(\frac{n+1}{n}\right) X_{(n)} - 1 \mid n, N, p\right) = \left(\frac{n+1}{n}\right) E(X_{(n)} \mid n, N, p) - 1 = N.$$

Thus, from (i), (ii), (iii) and from Lehmann-Scheffé theorem we conclude that, given n, $\widehat{N_3}$ is an UMVUE for N.

¿From (16), the conditional variance of $\widehat{N_3}$, given n, is

$$Var(\widehat{N_3} \mid n, N, p) = Var\left(\left(\frac{n+1}{n}\right) X_{(n)} - 1 \mid n, N, p\right) = \frac{(N+1)(N-n)}{n(n+2)}. \blacksquare$$

Theorem 2.2 $(n, X_{(n)})$ *is sufficient for* (N, p).

Proof: The likelihood function, equation (2), depends on sampled data through $(n, x_{(n)})$. By the factorization criteria, the theorem is proven. ■

Theorem 2.3 (i) *The maximum likelihood estimator, MLE, of* (N, p) *is* $(\widehat{N}, \widehat{p}) = (X_{(n)}, n/X_{(n)})$.

(ii) *The conditional MLE of* N, *given* n, is $\widehat{N_c} = X_{(n)}$.

Proof: (i) From (2) we have

$$\ln L(N, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)})) = n \ln p + (N - n) \ln(1 - p),$$

for $N \geq x_{(n)}$ and $0 < p < 1$. Then

$$\frac{\partial}{\partial p} \ln L(N, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)})) = \frac{n}{p} - \frac{N - n}{1 - p} = 0 \Longleftrightarrow p = \frac{n}{N}.$$

Also, for $N \geq x_{(n)}$ and p fixed,

$$\frac{L(N + 1, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)}))}{L(N, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)}))} = 1 - p < 1,$$

i.e., $L(N, p \mid n, (x_{(1)}, x_{(2)}, \cdots, x_{(n)}))$ is a decreasing function in N. Therefore, $(x_{(n)}, n/x_{(n)})$ is the maximun of the likelihood function.

(ii) From equation (7), the conditional likelihood function, given n, is

$$L(N, p \mid (x_{(1)}, x_{(2)}, \cdots, x_{(n)})) = \begin{pmatrix} N \\ n \end{pmatrix}^{-1},$$

for $N \geq x_{(n)}$. Therefore the maximun of $L(N, p \mid (x_{(1)}, x_{(2)}, \cdots, x_{(n)}))$ is $x_{(n)}$. ∎

Also, from (14) and (16)

$$E(\widehat{N_c} \mid n, N, p) = E(X_{(n)} \mid n, N, p) = \frac{n(N + 1)}{n + 1}$$

and

$$Var(\widehat{N_c} \mid n, N, p) = Var(X_{(n)} \mid n, N, p) = \frac{n(N + 1)(N - n)}{(n + 1)^2(n + 2)}.$$

2.1 Estimation of the variance of the estimator

Following, we present conditionally UMVUE, given n, for the variances of $\widehat{N_3}$ and $\widehat{N_c}$.

Lemma 2.4

$$(i)\, E(X_{(n)}(X_{(n)} - n) \mid n, N, p) = \frac{n(N + 1)(N - n)}{n + 2}, \text{ and}$$

$$(ii)\, Var(X_{(n)}(X_{(n)} - n) \mid n, N, p) = \frac{n(N + 1)(N - n)}{(n + 2)^2\,(n + 4)}[4(N + 1)(N - n) + n(n + 2)].$$

Proof: (i) From (14) and (15) we have

$$E(X_{(n)}(X_{(n)}-n) \mid n,N,p) = \frac{n(N+1)\,(N-n)}{n+2}. \qquad (3)$$

(ii) From (10) and (Feller, 1968, chapter II, p.55), we have

$$E\left((X_{(n)}(X_{(n)}-n))^2 \mid n,N,p\right) = \binom{N}{n}^{-1} \sum_{j=n+1}^{N} j^2(j-n)^2 \binom{j-1}{n-1} =$$

$$= n(N+1)(N-n)\left[\frac{n+1}{n+2} + \frac{(N+2)(N-n-1)}{n+4}\right]. \qquad (4)$$

Using (3) and (4) we obtain

$$Var(X_{(n)}(X_{(n)}-n) \mid n,N,p) = \frac{n(N+1)(N-n)}{(n+2)^2\,(n+4)}[4(N+1)(N-n)+n(n+2)].\blacksquare$$

Theorem 2.5

$$T = \frac{1}{n^2}X_{(n)}(X_{(n)}-n)$$

is conditionally UMVUE, given $n \geq 1$, *for* $Var(\widehat{N_3} \mid n,N,p)$ *and*

$$Var(T \mid n,N,p) = \frac{Var(\widehat{N_3} \mid n,N,p)}{n\,(n+4)}\left[4Var(\widehat{N_3} \mid n,N,p)+1\right].$$

Proof: From lemma (2.4) we have

$$E(T \mid n,N,p) = \frac{(N+1)\,(N-n)}{n(n+2)} = Var(\widehat{N_3} \mid n,N,p).$$

Since, given n, $X_{(n)}$ is sufficient and complete for N (Theorem 2.1) , by the Lehmann-Scheffé Theorem, T is a conditionally UMVUE, given $n \geq 1$, for $Var(\widehat{N_3} \mid n,N,p)$.

Also, from lemma (2.4)

$$Var(T \mid n,N,p) = \frac{Var(\widehat{N_3} \mid n,N,p)}{n\,(n+4)}\left[4Var(\widehat{N_3} \mid n,N,p)+1\right].\blacksquare$$

Theorem 2.6

$$T_c = \frac{1}{(n+1)^2}X_{(n)}(X_{(n)}-n)$$

is conditionally UMVUE, given $n \geq 1$, *for* $Var(\widehat{N_c} \mid n,N,p)$ *and*

$$Var(T_c \mid n,N,p) = \frac{Var(\widehat{N_c} \mid n,N,p)}{n\,(n+4)}\left[4Var(\widehat{N_c} \mid n,N,p)+\left(\frac{n}{n+1}\right)^2\right].$$

Proof: From lemma (2.4) we have

$$E(T_c \mid n, N, p) = \frac{n(N+1)\,(N-n)}{(n+1)^2(n+2)} = Var(\widehat{N_c} \mid n, N, p).$$

Since, given n, $X_{(n)}$ is sufficient and complete for N (Theorem 2.1) , by the Lehmann-Scheffé Theorem, T_c is conditionally UMVUE, given $n \geq 1$, for $Var(\widehat{N_c} \mid n, N, p)$.

Also, from lemma (2.4)

$$Var(T_c \mid n, N, p) = \frac{Var(\widehat{N_c} \mid n, N, p)}{n\,(n+4)} \left[4Var(\widehat{N_c} \mid n, N, p) + \left(\frac{n}{n+1}\right)^2\right]. \blacksquare$$

3 Conditional confidence interval

Given n, N and γ, $(0 < \gamma < 1)$, let m the "least positive integer" such that

$$m \geq N\left(1 - (1-\gamma)^{\frac{1}{n}}\right) \text{ or } \left(1 - \frac{m}{N}\right)^n \leq 1 - \gamma.$$

¿From (10) and (Feller, 1968, chapter II, p.55), we have

$$P(N \geq X_{(n)} + m \mid n, N, p) = P(X_{(n)} \leq N - m \mid n, N, p) =$$

$$= \binom{N}{n}^{-1} \sum_{i=n}^{N-m} \binom{i-1}{n-1} = \binom{N}{n}^{-1} \binom{N-m}{n} =$$

$$= \left(1 - \frac{m}{N}\right)\left(1 - \frac{m}{N-1}\right)\left(1 - \frac{m}{N-2}\right)\cdots\left(1 - \frac{m}{N-n+1}\right) \leq \left(1 - \frac{m}{N}\right)^n \leq 1-\gamma.$$

Since

$$N \geq X_{(n)} + m \Longleftrightarrow N \geq X_{(n)} + N\left(1 - (1-\gamma)^{\frac{1}{n}}\right),$$

it follows that

$$P\left(X_{(n)} \leq N(1-\gamma)^{\frac{1}{n}} \mid n, N, p\right) = P\left(N \geq X_{(n)} + N\left(1 - (1-\gamma)^{\frac{1}{n}}\right) \mid n, N, p\right) =$$

$$= P\left(N \geq X_{(n)} + m \mid n, N, p\right) \leq 1 - \gamma.$$

Thus,

$$P\left(N \leq \frac{X_{(n)} - 1}{(1-\gamma)^{\frac{1}{n}}} \mid n, N, p\right) = P\left(X_{(n)} \geq N(1-\gamma)^{\frac{1}{n}} + 1 \mid n, N, p\right) =$$

$$= 1 - P\left(X_{(n)} < N(1-\gamma)^{\frac{1}{n}} + 1 \mid n, N, p\right) =$$

$$= 1 - P\left(X_{(n)} \leq N(1-\gamma)^{\frac{1}{n}} \mid n, N, p\right) \geq 1 - (1-\gamma) = \gamma,$$

implying that, given n,

$$\left[X_{(n)}, \frac{X_{(n)} - 1}{(1-\gamma)^{\frac{1}{n}}}\right]$$

is a confidence interval for N, with confidence $\geq \gamma$.

4 Example

We illustrate the aplication of the estimators described by simulating nine situations where $N = 1000$ and each item is observed with probability $p = 0.1, 0.2, ..., 0.9$. The observed data is summarized by sample size, n and maximum, $x_{(n)}$, at Table 1.

Table 1 - Simulated data

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n	110	212	299	402	448	600	711	806	897
$x_{(n)}$	1000	999	996	998	1000	1000	1000	999	1000

The estimates of N are presented in Table 2.

Table 2 - Point Estimates and 95% confidence intervals for N, given n

p		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\widehat{N_3}$		1008	1003	998	999	1001	1001	1000	999	1000
$\widehat{N_c}$		1000	999	996	998	1000	1000	1000	999	1000
Conf.Int.	*Lower bound*	1000	999	996	998	1000	1000	1000	999	1000
Conf.Int.	*Upper bound*	1027	1012	1005	1004	1006	1004	1003	1002	1002

The estimated variance of the estimators are given in Table 3.

Table 3 - Estimated variances for estimators of N, given n

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T	73.55	17.49	7.77	3.68	2.75	1.11	0.50	0.30	0.13
T_c	72.23	17.33	7.71	3.66	2.74	1.11	0.57	0.30	0.13

As expected the best estimates are provided by the greatest sample sizes, as we can see in Tables 1, 2 and 3, the least variances are associated with the greatest values of p and consequently the largest sample sizes. When comparing $\widehat{N_c}$ and $\widehat{N_3}$, we notice in Table 3 that they have a similar behaviour.

Acknowledgement

This study was partly supported by CAPES and CNPq, proj. 302813/2004-7.

LEITE, J.G.; MOREIRA, E.B.M. Análise estatística de séries numéricas. *Rev. Mat. Estat.*, São Paulo, v.23, n.1, p.81-91, 2005

- *RESUMO: Considere uma seqüência de itens numerados de 1 a N. Nós propomos estimadores UMVUE para o número de elementos, N, e para a variância do estimador, baseados em uma amostra selecionada ao acaso, de tamanho aleatório, destes itens. Um intervalo de confiança para N também é fornecido.*

- *PALAVRAS-CHAVE: Séries de números; estimador UMVUE condicional; intervalo de confiança condicional.*

Referências

BOSE, R. C. Error detecting and error correcting indexing systems for large serial numbers. *Rev. Inst. Int. Stat.*, Edenburgh, v.34, p. 334-340, 1966.

FELLER W. *An introduction to the theory of probability and its applications.* 3th ed. New York: John Wiley and Sons, 1968. v.1

GOODMAN, L. A. Serial number analysis. *J. Am. Stat. Assoc.*, New York, v.47, p.622-34, 1952.

GUMBEL, E. J. On serial numbers. *Ann. Math. Stat.*, An Arbor, v.14, p.163-178, 1943.

ROBERTS, H. V. Informative stopping rules and inferences about population size. *J. Am. Stat. Assoc.*, New York, v.62, p.763-75, 1967.

RUGGLES, R.; BRODIE, H. An empirical approach to economic intelligence in world war II. *J. Am. Stat. Assoc.*, New York, v.42, p.72-91, 1947.

Appendix

A.1 Conditional sampling distributions

The probability of observing n elements, minimum and maximun is

$$P(n, X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)} \mid N, p) = \binom{x_{(n)} - x_{(1)} - 1}{n-2} p^n (1-p)^{N-n}, \quad (5)$$

for $1 \leq x_{(1)} < x_{(n)} \leq N$ and $2 \leq n \leq x_{(n)} - x_{(1)} + 1$.

The probability of n elements and the minimum for $1 \leq x_{(1)} \leq N$ and $1 \leq n \leq N - x_{(1)} + 1$ is

$$P(n, X_{(1)} = x_{(1)} \mid N, p) = \binom{N - x_{(1)}}{n-1} p^n (1-p)^{N-n}. \qquad (6)$$

The probability of n elements and the maximun for $1 \leq n \leq x_{(n)} \leq N$ is

$$P(n, X_{(n)} = x_{(n)} \mid N, p) = \binom{x_{(n)} - 1}{n-1} p^n (1-p)^{N-n}. \qquad (7)$$

The distribution of $n \mid N$, p is $Binomial(N, p)$.

The conditional probability of the minimum and maximum, given $n \geq 2$, is

$$P(X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)} \mid n, N, p) = \binom{N}{n}^{-1} \binom{x_{(n)} - x_{(1)} - 1}{n-2} \qquad (8)$$

for $1 \leq x_{(1)} \leq N - n + 1$ and $n + x_{(1)} - 1 \leq x_{(n)} \leq N$.

The conditional probability of the minimum, given $n \geq 1$, is

$$P(X_{(1)} = x_{(1)} \mid n, N, p) = \frac{P(n, X_{(1)} = x_{(1)} \mid N, p)}{P(n \mid N, p)} = \binom{N}{n}^{-1} \binom{N - x_{(1)}}{n-1} \qquad (9)$$

for $1 \leq x_{(1)} \leq N - n + 1$.

The conditional probability of the maximun, given $n \geq 1$ is

$$P(X_{(n)} = x_{(n)} \mid n, N, p) = \frac{P(n, X_{(n)} = x_{(n)} \mid N, p)}{P(n \mid N, p)} = \binom{N}{n}^{-1} \binom{x_{(n)} - 1}{n-1} \qquad (10)$$

and $n \leq x_{(n)} \leq N$,

A.2 Properties of sampling distributions

Theorem 4.1 *The conditional mean and variance of* $X_{(1)}$*, given* $n \geq 1$*, are*

$$(i)\, E(X_{(1)} \mid n, N, p) = \frac{N+1}{n+1} \text{ and } (ii)\, Var(X_{(1)} \mid n, N, p) = \frac{n(N+1)(N-n)}{(n+1)^2(n+2)}.$$

Proof: (i) From (9) and (Feller, 1968, chapter II, p.55),

$$E(X_{(1)} \mid n, N, p) = \binom{N}{n}^{-1} \sum_{k=1}^{N-n+1} k \binom{N-k}{n-1} = (N+1) - n \binom{N}{n}^{-1} \sum_{j=n-1}^{N-1} \binom{j+1}{n} =$$

$$= N + 1 - n \binom{N}{n}^{-1} \binom{N+1}{n+1} = \frac{N+1}{n+1}. \qquad (11)$$

(ii) From (9) and (Feller, 1968, chapter II, p.55),

$$E(X_{(1)}^2 \mid n, N, p) = \binom{N}{n}^{-1} \sum_{k=1}^{N-n+1} k^2 \binom{N-k}{n-1}$$

$$= \binom{N}{n}^{-1} \left\{ \sum_{j=n-1}^{N-1} (N^2 - 2Nj) \binom{j}{n-1} + \sum_{j=n-1}^{N-1} [(j+1)(j+2) - 2 - 3j] \binom{j}{n-1} \right\}$$

$$= (N+1)^2 - \frac{n(N+1)}{(n+1)(n+2)} [N(n+3) + (n+4)]. \tag{12}$$

From (11) and (12)

$$Var(X_{(1)} \mid n, N, p) = E(X_{(1)}^2 \mid n, N, p) - E^2(X_{(1)} \mid n, N, p) =$$

$$= (N+1)^2 - \frac{n(N+1)}{(n+1)(n+2)} [N(n+3) + (n+4)] - \frac{(N+1)^2}{(n+1)^2} = \frac{n(N+1)\,(N-n)}{(n+1)^2(n+2)}. \blacksquare \tag{13}$$

Theorem 4.2 *The conditional mean and variance of* $X_{(n)}$*, given* $n \geq 1$*, are*

$$(i)\, E(X_{(n)} \mid n, N, p) = \frac{n(N+1)}{n+1} \text{ and } (ii)\, Var(X_{(n)} \mid n, N, p) = \frac{n(N+1)(N-n)}{(n+1)^2\,(n+2)}.$$

Proof: (i) From (10) and (Feller, 1968, chapter II, p.55),

$$E(X_{(n)} \mid n, N, p) = \binom{N}{n}^{-1} \sum_{k=n}^{N} k \binom{k-1}{n-1} = \frac{n(N+1)}{n+1}. \tag{14}$$

(ii) From (10) and (Feller, 1968, chapter II, p.55),

$$E(X_{(n)}^2 \mid n, N, p) = \binom{N}{n}^{-1} \sum_{k=n}^{N} k^2 \binom{k-1}{n-1} =$$

$$= \binom{N}{n}^{-1} \sum_{k=n}^{N} (k(k+1) - k) \binom{k-1}{n-1} = n\,(N+1) \left(\frac{N+2}{n+2} - \frac{1}{n+1} \right) \tag{15}$$

and from (14) and (15)

$$Var(X_{(n)} \mid n, N, p) = E(X_{(n)}^2 \mid n, N, p) - E^2(X_{(n)} \mid n, N, p)$$

$$= n(N+1) \left(\frac{N+2}{n+2} - \frac{1}{n+1} \right) - \left(\frac{n(N+1)}{n+1} \right)^2 = \frac{n(N+1)(N-n)}{(n+1)^2\,(n+2)}. \blacksquare \tag{16}$$

Theorem 4.3 *The conditional covariance between* $X_{(1)}$ *and* $X_{(n)}$*, given* $n \geq 2$*, is*

$$Cov(X_{(1)}, X_{(n)} \mid n, N, p) = \frac{(N+1)(N-n)}{(n+1)^2(n+2)}.$$

Proof: From (8), (12) and (Feller, 1968, chapter II, p.55),

$$E(X_{(1)} X_{(n)} \mid n, N, p) = \binom{N}{n}^{-1} \sum_{k_1=1}^{N-n+1} \sum_{k_2=n+k_1-1}^{N} k_1 k_2 \binom{k_2 - k_1 - 1}{n-2} =$$

$$= \begin{pmatrix} N \\ n \end{pmatrix}^{-1} \left[(n-1) \sum_{j=n}^{N} (N+2-(j+1)) \begin{pmatrix} j \\ n \end{pmatrix} \right] + E(X_{(1)}^2 \mid n, N, p) =$$

$$= (n-1)(N+2)(N+1) \frac{1}{(n+1)(n+2)} + E(X_{(1)}^2 \mid n, N, p) =$$

$$= \frac{(N+1)}{(n+1)(n+2)} \left[(N+1)(n+1) - 1 \right]. \qquad (17)$$

¿From (11), (14) and (17)

$$Cov(X_{(1)}, X_{(n)} \mid n, N, p) = E(X_{(1)} X_{(n)} \mid n, N, p) - E(X_{(1)} \mid n, N, p) E(X_{(n)} \mid n, N, p) =$$

$$= \frac{(N+1)}{(n+1)(n+2)} \left[(N+1)(n+1) - 1 \right] - \left(\frac{N+1}{n+1} \right) \frac{n(N+1)}{n+1} = \frac{(N+1)(N-n)}{(n+1)^2 (n+2)}. \blacksquare \qquad (18)$$

Received 25.04.2004.

Received in revised form 07.04.2005.