Scientific
Research

Efficiency of Two-Stage Adaptive Cluster Sampling Design in Estimating Fringe-Eared Oryx

Jesse Wachira Mwangi*, Mohamed Esha Salim

Mathematics Department, Faculty of Science, Egerton University, Njoro, Kenya
Email: *jdmwangi@egerton.ac.ke

ABSTRACT

Two-stage adaptive cluster sampling and two-stage conventional sampling designs were used to estimate population total of Fringe-Eared Oryx that are clustered and sparsely distributed. The study region was Amboseli-West Kilimanjaro and Magadi-Natron cross boarder landscape between Kenya and Tanzania. The study region was partitioned into different primary sampling units with different secondary sampling units that were of different sizes. Results show that two-stage adaptive cluster sampling design is efficient compared to simple random sampling and the conventional two-stage sampling design. The design is less variable compared to the conventional two-stage sampling design.

Keywords: Non-Overlapping Scheme; Cluster Sampling; Horvitz-Thompson Estimator

1. Introduction

The Fringe-eared Oryx is one of the subspecies of Oryx gazelle. They range from the grassland of Northern Eastern Tanzania to Southern Kenya [1] and they are conservation dependent species. Currently, the species is not listed as an endangered species but its populations are declining [1]. This is partly because in the past, they were hunted for their tough skin by European settlers [2] and recently they are being gunning down by poachers for their horns. The other causes for their decline are uncontrolled hunting for sport [1] and just like other wildlife, human encroachment to protected areas and land use changes [3] have led to the reduction in their numbers. The Fringe-Eared Oryx lives in clustered herds that vary in sizes from a few to hundreds of individuals [2] and are sparsely distributed [4]. They are mobile and can travel for long distances in search for food [1] since rainfall within Amboseli ecosystem is unevenly distributed [3]. To ensure adequate conservation and management of Fringe-Eared Oryx, knowledge of their numbers is critical.

Adaptive cluster sampling (ACS) was suggested for sampling populations that are clustered and rare [5]. ACS in which the initial sample is selected by simple random sampling with or without replacement was described by [5], where the initial sample is selected by systematic sampling was described by [6]. Ratio estimators in ACS were described by [7]. Smith [8] applied ACS to estimate density of wintering Waterfowl while [9] applied ACS to

estimate abundance within local populations of low abundance plants.

Cochran [10] described two-stage (TS) sampling that is suitable when sampling large areas [11]. This is as a result of lower costs incurred when sampling selected PSUs rather than moving across the study area to perform sampling. PSUs [12] described two-stage adaptive cluster sampling (TSA) which is a combination of ACS and TS sampling designs. Smith [11] applied TSA to estimate density and abundance of fresh water mussels in the Upper Mississippi River using simulated data.

In this study, we evaluate the efficiency of TS and TSA sampling designs in estimating population total of Fringe-Eared Oryx in Amboseli-West Kilimanjaro and Magadi-Natron cross border landscape using data from aerial survey conducted in 2010 by KWS and other partners. Section two briefly describes the theory of TSA design, section three gives the method for the current study, section four gives results, and section five has the discussion and conclusion.

2. Two-Stage Adaptive Cluster Sampling Design

In the TSA design, which was proposed by [12], information from a pilot survey is used to design an efficient sampling procedure by controlling the final sample size and sampling cost. In this design, the study region consisting of N units is partitioned into M_P primary sampling units (PSUs) each with N_i secondary sampling units (SSUs) where $i = 1, 2, \cdots, M_p$. Let y_{ij} be the ob-

---

*Corresponding author.

served value in j^{th} secondary unit of ith primary unit and let $\tau_i = \sum_{j=1}^{N_i} y_{ij}$ be the sum of the observed values in the ith primary unit.

In the first stage, m_p primary units are selected from M_P units by simple random sampling (SRS) without replacement. In the second stage n_i secondary units are selected from primary unit i, where $i = 1, 2, \cdots, m_p$, using simple random sampling without replacement. If any selected secondary unit meets a condition $y_{ij} > c$, neighboring units are added. If any of the added units meets the condition, still more units are added. The procedure continues until there are clusters of units each with units that meet the condition (network) surrounded by units that do not meet the condition (edge units). A neighborhood of a unit consists of itself plus its adjacent units on the North, East, West and South. The procedure results in two situations [12], where clusters intersect more than one PSUs (overlapping scheme) and where clusters are truncated at the boundary of PSUs (non-overlapping scheme).

Salehi and Seber [12], derived two design unbiased estimators, Horvitz-Thompson (HT) and Hansen-Hurwitz (HH) estimators. We consider the non-overlapping scheme since the estimators are more efficient compared to overlapping scheme [12]. We also make use of the HT estimator since it is more efficient compared to HH estimator [5,12].

Let I_{ik} be an indicator function that is one if the initial sample intersect network k and zero otherwise, let K_i be the number of networks in primary unit i, y_{ik}^* the sum of observed values in network k of primary unit i and $m_{p_{ik}}$ be the number of units in network k in primary unit i, then the partial inclusion probability, that is the probability that units in the initial sample of primary unit I intersect network k, is given by [12]

$$\pi_{ik} = 1 - \binom{N_i - m_{p_{ik}}}{n_i} \Big/ \binom{N_i}{n_i}.$$

The HT estimator is thus given by [12]

$$\hat{\mu}_{HT} = M_P N^{-1} \sum_{i=1}^{m_p} \hat{\tau}_i / m_p$$

where $\hat{\tau}_i = \sum_{k=1}^{K_i} y_{ik}^* I_{ik} / \pi_{ik}$. The second order partial inclusion probability is given by [12]

$$\pi_{ikl} = 1$$
$$- \left\{ \binom{N_i - m_{pik}}{n_i} + \binom{N_i - m_{pil}}{n_i} - \binom{N_i - m_{pik} - m_{pil}}{n_i} \right\} \Big/ \binom{N}{n_i}$$

Let the variance for the ith PSUs be

$$s_{M_P}^2 = \sum_{i=1}^{m_p} \left(\tau_i - \hat{\bar{\tau}}_i \right)^2 \Big/ \left(m_p - 1 \right)$$

and the variance for SSUs be $\hat{V}_i / N^2$ where

$$\hat{V}_i = \sum_{k=1}^{k_i} \sum_{l=1}^{k_i} y_{ik}^* y_{il}^* \left(\pi_{ikl} - \pi_{ik} \pi_{il} \right) / \left(\pi_{ikl} \pi_{ik} \pi_{il} \right),$$

with k_i the number of distinct networks intersected by primary unit i, an unbiased estimator of the variance of HT is given by [12]

$$\tilde{v}\left(\hat{\mu}_{HT}\right) = N^{-2} m_p^{-1} \left\{ M_P \left(M_P - m_p \right) s_{M_P}^2 + M_P \sum_{i=1}^{m_p} \hat{V}_i \right\}.$$

3. Method

Amboseli-West Kilimanjar/Magadi–Natron cross-border landscape stretches between Kenya and Tanzania. The landscape was placed on a rectangular grid which is, for the purpose of this study, the study area. The study area was divided into 24×18 squared units each measured 69.44 km^2. Similar to [13], the study area was partitioned into 6, 12 and 18 PSUs based on the survey report [4] on the degree of rarity and clusteredness of Fringe-Eared Oryx in the study area. This enabled the study of large, medium and small sized PSUs where each measured 4999.68 km^2, 2499.84 km^2 and 1666.56 km^2 respectively. For the adaptive design, the condition to adaptively add units was $y_{ij} > 1$.

The sample sizes for large sized PSUs, m_{p1}, were 2 and 3, for medium sized PSUs were 4 and 6 and for small sized PSUs were 6 and 9. Initial sample sizes for secondary units, n_1, were 2, 3 and 4 in all cases for adaptive design. The sample size for SSUs for the non-adaptive design was set up to the effective sample size for the adaptive design.

Two-stage adaptive and non-adaptive designs were considered. In each case, re-sampling was carried out 1000 times. Efficiency was obtained as

$\text{eff}\left(TSA\right) = v\left(\hat{\mu}_{SRS}\right) / \tilde{v}\left(\hat{\mu}_{TSA}\right)$ for adaptive and

$\text{eff}\left(TS\right) = v\left(\hat{\mu}_{SRS}\right) / v\left(\hat{\mu}_{TS}\right)$ for non-adaptive design. Population total and variance were obtained as

$\text{Total}_{TSA} = \hat{\mu}_{HT_{TSA}} \times N$ and $v\left(\text{Total}_{TSA}\right) = v\left(\hat{\mu}_{HT_{TSA}}\right) \times N^2$

for adaptive design and $\text{Total}_{TS} = \hat{\mu}_{HT_{TS}} \times N$ and

$v\left(\text{Total}_{TS}\right) = v\left(\hat{\mu}_{HT_{TS}}\right) \times N^2$ for non-adaptive designs respectively. For adaptive design only, the effective sample size was obtained as $E\left(v\right) = n_f / 1000$, that is, the average number of units in the final sample over the number of re-sampling. Estimates were obtained using Program SAMPLE [11].

4. Results

Results are shown in **Table 1** below. From the table, TSA

Table 1. Efficiency (E), population total (T) and coefficient of variation (CV) of TSA and TS designs for a study region divided into M_p PSUs with a sample size of m_p PSUs, n_1 initial sample size for adaptive and a sample size n units for conventional design.

M_P	m_p	n_1	n	E_{TSA}	E_{TS}	T_{TSA}	T_{TS}	CV_{TSA}	CV_{TS}
6	2	2	3	1.368	0.763	206	225	2.254	2.997
6	2	3	5	1.379	0.860	203	199	1.843	2.438
6	2	4	7	1.424	0.876	202	201	1.628	2.000
6	3	2	3	1.563	0.827	189	209	1.848	2.466
6	3	3	5	1.574	0.938	191	204	1.495	1.849
6	3	4	7	1.608	0.944	198	192	1.294	1.629
12	4	2	3	1.675	0.890	198	199	1.650	2.176
12	4	3	4	1.696	0.899	200	207	1.334	1.795
12	4	4	6	1.729	0.939	191	191	1.174	1.485
12	6	2	3	1.669	0.931	195	202	1.301	1.682
12	6	3	4	1.732	0.987	203	197	1.048	1.462
12	6	4	6	1.825	0.997	192	196	0.920	1.186
18	6	2	3	1.689	0.955	199	200	1.291	1.718
18	6	3	4	1.699	0.986	191	184	1.115	1.550
18	6	4	5	1.712	1.000	185	184	1.016	1.366
18	9	2	3	1.703	1.035	200	190	1.000	1.374
18	9	3	4	1.789	1.044	192	188	0.844	1.201
18	9	4	6	1.898	1.049	201	188	0.734	1.029

design has higher efficiency compared to the non-adaptive (TS) design. The adaptive design is efficient compared to SRS irrespective of the number or size of PSUs. Efficiency for the non-adaptive design is attained only when the study region is divided into 18 PSUs and using a sample size of 9 for PSUs.

For both adaptive and non-adaptive designs, Efficiency increases with increasing in number of units in PSUs in the study region as well as the sample size of PSUs and initial sample size/sample size for SSUs.

While the survey recorded a total of 196 Fringe-Eared Oryx [4], the population was estimated to range from 189 to 206 when adaptive design is used and from 192 to 225 upon use of non-adaptive design when the study region is partitioned into 6 PSUs. When the study region is partitioned into 12 PSUs, the total population for Fringe-Eared Oryx was estimated to range from 191 to 203 and from 191 to 207 when adaptive and non-adaptive designs are used respectively. Finally, Fringe-Eared Oryx are estimated to be between 185 and 201 using adaptive design and between 184 and 200 using non adaptive design for a study region that is partitioned into 18 PSUs.

Variation decreases with increasing initial sample size for SSUs, initial sample size for PSUs and the size of

PSUs for both designs. TSA design is more reliable since it reports lower values of coefficient of variation compared to TS design.

For the adaptive design, the effective sample size is a function of the unit size of PSUs, sample size of PSUs and initial sample size of SSUs.

5. Discussion and Conclusions

Salehi and Seber [12] proposed a TSA due to its suitability in performing a pilot survey so as to design an experiment with a given precision of estimation. TS designs are economical designs when sampling such a large study area since they reduce movement within the large study area [11]. Since populations of Fringe-Eared Oryx are declining [4] and noting the vastness of the study area, TSA could be applied to for conservation purposes. Small initial sample sizes (2, 3 and 4) for secondary units were used for adaptive design; this was to avoid the negative variances on individual PSUs [12].

Results show that the size, M_P, and the sample size, m_p, for PSUs as well as the initial sample size/sample size for SSU have an effect on the efficiency and amount of variance on the two-stage sampling for both adaptive

and non-adaptive designs. The adaptive design resulted in higher efficiency with lower variance compared to the non-adaptive design. These results agree with [5] that adaptive designs result if higher efficiency and lower variance compared to conventional sampling designs.

For this study, highest precision is attained when the study region was partitioned into 18 PSUs each with 24 SSUs and where initial sample size for PSUs, m_1, is 9. The TSA still has a drawback of the final sample size. Further controlling the final sample size with this design deserves further study.

6. Acknowledgement

REFERENCES

[1] Woodland Park Zoo, "Fringe-Eared Oryx," 1893. http://lib.colostate.edu/wildlife/item.php?id=132164

[2] V. L. Terrill, "Group of Fringe-Eared Oryx in Grass," Vol. 15, No. 5, 2012. http://lib.colostate.edu/wildlife/item.php?id=132164

[3] M. M. Okello, "The Contraction of Wildlife Dispersal Areas by Human Structures and Activities in Mbirikani Group Ranch in the Amboseli Ecosystem, Kenya," *International Journal of Biodiversity and Conservation*, Vol. 4, No. 6, 2012, pp. 243-259.

[4] Kenya Wildlife Service, "Aerial Total Count: Amboseli-West Kilimanjaro/Magadi-Natron Cross Border Landscape," 2010. http://www.kws.org/export/sites/kws/info/publications/census_reports/Amboseli_West_kili_Magadi_Natron_2010_census_report.pdf

[5] S. K. Thompson, "Adaptive Cluster Sampling," *Journal of American Statistical Association*, Vol. 85, No. 412, 1990, pp. 1050-1059. doi:10.1080/01621459.1990.10474975

[6] S. K. Thompson, "Adaptive Cluster Sampling: Designs with Primary and Secondary Units," *Biometrics*, Vol. 47, No. 3, 1991, pp. 1103-1115. doi: 10.2307/2532662

[7] A. L. Dryver and C. T. Chao, "Ratio Estimators in Adaptive Cluster Sampling," *Environmetrics*, Vol. 18, No. 6, 2007, pp. 607-620. doi:10.1002/env.838

[8] D. R. Smith, M. J. Conroy and D. H, Brakhage, "Efficiency of Adaptive Cluster Sampling for Estimating Density of Wintering Waterfowl," *Biometrics*, Vol. 51, No. 2, 1995, pp. 777-788. doi: 10.2307/2532964

[9] T. Philippi, "Adaptive Cluster Sampling for Estimation of Abundances within Local Populations of Low Abundance Plants," *Ecology*, Vol. 86, No. 5, 2005, pp. 1091-1100. doi: 10.1890/04-0621

[10] W. G. Cochran, "Sampling Techniques," 3rd Edition, Wiley, New York, 1977.

[11] D. R. Smith, J. T. Rogala, B. R. Gray, S. J. Zigler and T. J. Newton, "Evaluation of Single and Two-Stage Adaptive Sampling Designs for Estimation of Density and Abundance of Freshwater Mussels in a Large River," *River Research and Applications*, Vol. 27, No. 1, 2011, pp. 122-133. doi:/10.1002/rra.1334

[12] M. M. Salehi and G. A. F. Seber, "Two-Stage Adaptive Cluster Sampling," *Biometrics*, Vol. 53, No. 3, 1997, pp. 959-970. doi: 10.2307/2533556

[13] M. M. Salehi and D. R. Smith, "Two-Stage Sequential Sampling: A Neighborhood-Free Adaptive Sampling Procedure," *Journal of Agricultural*, *Biological and Environmental Statistics*, Vol. 10, No. 1, 2005, pp. 84-103. doi: 10.1198/108571105X28183