

COVID-19 Testing: Annotated Bibliography

CDT Madison Teague

October 19, 2020

1 Thompson, Adaptive Cluster Sampling (1990)

In Thompson's paper, "Adaptive Cluster Sampling," he discusses real world applications of adaptive sampling techniques and how to use the data collected with this sampling to create unbiased estimators for the number of a certain species in a population. Some of the advantages from "sequential statistical methods," include, "increased power, lower expected sample size or more controllable precision" [6]. Adaptive cluster sampling is a design in which an initial sample is taken randomly from a population, and the neighbors of a certain unit which satisfies a specific criterion are added to the sample [6]. With adaptive sampling, typical estimators result in bias, so this article uses design unbiased estimators, improved with the Rao-Blackwell method, which are based on how the samples are selected [6]. The article explains why Hansen-Hurwitz estimator has been used thus far as an unbiased estimator for adaptive cluster samples and introduces Horvitz-Thompson estimator as an improvement. Thompson then uses the improved estimators in a short example. His final point is on expected sample size and cost.

In relation to COVID-19, West Point currently uses a form of adaptive cluster sampling when looking at the cadet population. The article mentions two types of "snowball sampling." When used by Kalton and Anderson in their study (1986), a few members of a rare population were asked to identify other members and those members were asked to identify more. When used by Goodman in 1961, the individuals were asked to identify a fixed number of individuals, who were asked the same, and this continued for a fixed number of stages in order to determine "mutual relationships or "social circles" [6]. This is useful because it is how West Point attempts to do some tracing for positive cadet COVID-19 tests.

2 Salehi, Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling (1999)

In this article, Salehi leverages the Rao-Blackwell method theorem to improve the Horvitz-Thompson and Hansen-Hurwitz estimators introduced by Thompson in his 1990 publication. The article walks through the derivation of unbiased estimators based off the HT and HH estimators and then demonstrates their use in several examples. [4]

In relation to COVID-19, this article helps establish a clear understanding of the estimators I could use to help predict the how pervasive the disease is within the corps of cadets.

3 Salehi, Comparison between Hansen-Hurwitz and Horvitz-Thompson estimators for adaptive cluster sampling (2002)

In this article, Salehi compares the properties of HH and HT estimators. He does so with the help of ordered statistics and focuses heavily on network size in his analysis. Salehi concludes that the computation for both HH and HT can be reduced with the adjustments he makes in his paper. Through the simulations run for this paper, Salehi confirmed that the HT estimator is more efficient than the HH estimator. [5]

The adjustments made to the HH and HT estimators will be useful in the COVID-19 experiment as they will be the estimators used and analyzed for the perversion of COVID-19 in the corps of the cadets.

4 Lipton, Practical Selectivity Estimation through Adaptive Sampling (1990)

This article attempts to describe efficient sampling algorithms for size estimation using adaptive sampling. This paper explores provides an example of the use of adaptive sampling for something outside of population estimation. The algorithm used in this paper checks certain data to see if it would be selected in a sample and if other data points surrounding it would fall to the same fate. It does so on selection data and join data to demonstrate the diversity of the algorithm. [3]

5 Kabaghe, Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural Malawi (2017)

This article provides a practical and current example of the use of adaptive sampling to identify disease hotspots. [2] It relates to the current COVID-19 situation because it uses the adaptive sampling to overcome the lack of information due to technological deficiencies, which mimic the asymptomatic cases that are likely to dominate the Corps of Cadets' cases. The method used in this study differs in the sampling used in West Points testing strategy because it uses historical data to target certain areas for sampling, however the general algorithm is similar to that used by West Point. The information and methods in this document provide some of the most useful information for my study, note there is a lot of Bayesian estimation used in this study.

6 Dryver, Partial Systematic Adaptive Cluster Sampling (2012)

This article suggests a new, simplified, format of adaptive cluster sampling. It suggests this new method on the basis of an available unbiased estimator of variance and a desire for simplicity. It is an adaptation of systematic adaptive cluster sampling, which is essential what West Point is currently using to detect COVID-19 within the Corps of Cadets. PSACS suggests the use of one large primary sampling unit and one large secondary sampling unit rather than multiple smaller samples of both types. [1] Overall, this sampling type does not appear to be conducive to the COVID-19 situation at West Point. However, it does provide a useful and simple explanation of the variables used in estimators like the HH and HT estimators for adaptive cluster sampling.

7 Thompson, Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters (2013)

In this book, I have specifically studied chapter 5. This chapter summarizes the use of adaptive cluster sampling to predict the population numbers of rare species in a pre-defined area and the applicability of those estimations to the general area in which that species is predicted to inhabit.[7] This chapter provides several examples of the adaptive cluster sampling technique, and introduces the HH and HT estimators, which will be the main focus of the COVID-19 testing study. It was also used to identify several other sources which are helpful in providing understanding for the theory and methods backing up my thesis.

References

- [1] Arthur L Dryver, Urairat Netharn, and David R Smith. Partial systematic adaptive cluster sampling. *Environmetrics*, 23(4):306–316, 2012.
- [2] Alinune N Kabaghe, Michael G Chipeta, Robert S McCann, Kamija S Phiri, Michèle Van Vugt, Willem Takken, Peter Diggle, and Anja D Terlouw. Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural malawi. *PLoS One*, 12(2):e0172266, 2017.
- [3] Richard J Lipton, Jeffrey F Naughton, and Donovan A Schneider. Practical selectivity estimation through adaptive sampling. In *Proceedings of the 1990 ACM SIGMOD international conference on Management of data*, pages 1–11, 1990.
- [4] Mohammad M Salehi. Rao-blackwell versions of the horvitz-thompson and hansen-hurwitz in adaptive cluster sampling. *Environmental and Ecological Statistics*, 6(2):183–195, 1999.
- [5] Mohammad M Salehi. Comparison between hansen-hurwitz and horvitz-thompson estimators for adaptive cluster sampling. *Environmental and ecological statistics*, 10(1):115–127, 2002.
- [6] Steven K Thompson. Adaptive cluster sampling. *Journal of the American Statistical Association*, 85(412):1050–1059, 1990.
- [7] William Thompson. *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island Press, 2013.