A Generalisation of the Geometric Algorithm for Sampling Business Populations

Patricia Gunning, *School of Computing, Dublin City University, Ireland*
Jane M. HORGAN, *School of Computing, Dublin City University, Ireland*, E-mail: jhorgan@computing.dcu.ie
Gary Keogh, *School of Computing, Dublin City University, Dublin, Ireland*

Key words: Horvitz-Thompson Estimator, Optimum Allocation, Pareto Distribution, Stratification

Mathematical Subject Classification: 62D05, 62E17, 60E05

Abstract:

In business surveys, populations are often skewed in that a small number of high-valued units for a large share of the total value, and a large number of low-valued units account for a small share of the total. Such populations are often described by the distribution named after the 19^{th} century Italian economist Vilfredo Pareto, who used it to model the considerable skewness in the distribution of wealth. He hypothesised that the distribution of wealth could obey the "80-20 rule", in which 20% of the population own 80% of the wealth: this was Pareto's empirical observation of Italy at the time. The Pareto distribution is also known as the "power law", and occurs naturally in business populations. Stratification before sampling of skewed distributions such as those often leads to vast improvements in the precision of the estimator [1].

The algorithm, which we propose for stratifying business populations computes the boundaries as a geometric progression. Gunning and Horgan [2] investigated its efficiency for stratifying a step function with uniform distribution within each stratum. In this paper, we extend this analysis for stratifying the Pareto distribution, the typical model for skewed data that arises in business situations. We show that geometric stratification of this distribution gives equal coefficients of variation in the strata. While not satisfying the optimum conditions of Delanius it does however compare favourably with the commonly used cumulative root frequency method of stratum construction in terms of the precision of the Horvitz-Thompson estimator of the total. We illustrate the efficiency using real business populations.

References

[1] Cox, C.G., Binder, B.N, CHinnappa, Christianan, A., Colledge, M.J. and Kotts, P.S. (1995). Business Survey Methods, New York: Wiley.

[2] Gunning, P. and Horgan, J. (2004). A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations, *Survey Methodology*, 30, 2, 159-165.