OPTIMIZING MULTIVARIATE SAMPLE ALLOCATION USING AUXILIARY VARIABLES IN STRATIFIED SAMPLING

D. PUMPUTIS

The study focuses on improving sample allocation methods in stratified surveys involving multiple study variables. When a population is divided into homogeneous strata, an optimal distribution of samples across these strata is crucial for achieving high precision of the estimators used, while minimizing survey cost. Traditional approaches such as Neyman allocation [5] often perform suboptimally when multiple correlated variables are studied simultaneously. To address this, the paper explores several optimization-based methods that account for both survey cost and precision while leveraging auxiliary variables — readily available variables correlated with the study variables — to enhance estimation accuracy.

The study compares seven algorithms for multivariate optimal sample allocation: the integer programming algorithm (IPA) [2], Bethel's algorithm (BA) [1], generalized simulated annealing (GSAA) [8], constrained optimization by linear approximations (COBYLA) [6], particle swarm optimization (PSO) [4], biased random key genetic algorithm (BRKGA) [3], and differential evolution optimization (DEO) [7]. These methods are applied to simulated and real survey populations with varying correlation strengths between auxiliary and study variables.

Simulation results show that all algorithms perform well when the correlations between auxiliary and study variables are high, but required sample sizes increase substantially as correlations weaken. The IPA achieves the exact global minimum cost and serves as a benchmark for comparison. Among the heuristic methods, DEO performs most consistently, achieving results closest to the global minimum in 86% of the tested cases, followed by BRKGA and PSO.

In addition to efficiency, the paper investigates algorithmic convergence behavior, assessing how rapidly and reliably each method reaches its optimal or near-optimal solution. DEO shows stable convergence and low sensitivity to parameter tuning, while GSAA and PSO occasionally exhibit slower convergence or premature stagnation. COBYLA performs well for smooth objective functions but can miss the global minimum in multimodal cases.

Overall, the findings highlight that using correlated auxiliary variables improves estimation precision and reduces survey cost in multivariate stratified sampling. Moreover, metaheuristic optimization methods — especially differential evolution (DEO) — combine high convergence reliability with computational robustness, making them suitable for large, complex survey designs.

References

[1] Bethel J. *An optimum allocation algorithm for multivariate surveys* // Proceedings of the Survey Research Methods Section. — 1985. — P. 209—212. — http://www.asasrms.org/ Proceedings/papers/1985_035.pdf.

[2] Brito J.A., do Nascimento Silva P.L., Semaan G.S., Maculan N. *Integer programming formulations applied to optimal allocation in stratified sampling* // Survey Methodology. — 2015. — V. 41, No 2. — P. 427—442. — https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015002/article/ 14249-eng.pdf?st=P7ZqwcD1.

[3] Gonçalves J.F., Resende M.G.C. *Biased random-key genetic algorithms for combinatorial optimization* // Journal of Heuristics. — 2011. — V. 17, No 5. — P. 487—525. — https://doi.org/10.1007/s10732-010-9143-1.

[4] Kennedy J., Eberhart R. *Particle swarm optimization* // Proceedings of ICNN'95 - International Conference on Neural Networks. — 1995. — P. 1942-–1948. — https://doi.org/10.1109/ICNN.1995.488968

[5] Neyman J. *On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection (with discussion)* // Journal of the Royal Statistical Society. — 1934. — V. 97. — P. 558—625. — https://doi.org/10.2307/2342192.

[6] Powell M.J.D. *A direct search optimization method that models the objective and constraint functions by linear interpolation* // in S. Gomez, J. P. Hennart (Eds.), Advances in Optimization and Numerical Analysis. — 1994. — P. 51—67. — https://doi.org/10.1007/978-94-015-8330-5.

[7] Storn R., Price K. *Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces* // Journal of Global Optimization. — 1997. — V. 11. — https://doi.org/10.1023/A:1008202821328.

[8] Tsallis C., Stariolo D.A. *Generalized simulated annealing* // Physica A: Statistical Mechanics andits Applications. — 1996. — V. 233, No 1. — P. 395—406. — https://doi.org/10.1016/S0378-4371(96)00271-3.

Vilnius Gediminas Technical University (VILNIUS TECH), Vilnius, Lithuania

Email address: dalius.pumputis@vilniustech.lt