HAPHAZARD INTENTIONAL SAMPLING TECHNIQUES IN NETWORK DESIGN OF MONITORING STATIONS

Julio Michael Stern, USP - University of Sao Paulo, *jstern@ime.usp.br*
Marcelo Lauretto, USP - University of Sao Paulo, *marcelo.lauretto@gmail.com*
Oswaldo Luiz do Valle Costa, USP - University of Sao Paulo, *oswaldo@lac.usp.br*
Celma de Oliveira Ribeiro, USP - University of Sao Paulo, *celma@usp.br*

Overview

In contemporary empirical science, sampling randomization is the the golden standard to ensure unbiased, impartial, or fair results, see Pearl (2000) and Stern (2008).

Randomization works as a firewall, a technological barrier designed to prevent spurious communication of vested interests or illegitimate interference between parties in the application of interest, that may be a scientific experiment, a legal case, an auditing process, or many other practical applications.

In randomized experiments, a common issue is avoid random allocations yielding groups that differ meaningfully with respect to relevant covariates. This is a critical issue, as the chance of at least one covariate showing a "significant difference" between two treatment groups increases exponentially with the number of covariates.

Methods

Haphazard Intentional Sampling is a statistical technique developed with the specific purpose of yielding sampling techniques that, on one hand, have all the benefits of standard randomization and, on the other hand, avoid exponentially large (and costly) sample sizes.

This approach, proposed at Lauretto et al (2012) and Fossaluza et al (2015), combines intentional sampling using goal optimization techniques with random perturbations that induce good decoupling properties. On one hand, this method has a computational cost that is cubic with the number of covariates. On the other hand this method yields experimental designs that avoid exponentially large sample sizes, allowing great economical benefits that, nevertheless, do not compromise the statistical integrity of the experiment or auditing process.

Results

In this article, we apply the aforementioned Haphazard Intentional Sampling is a statistical techniques to study how to rationally re-engineer networks of measurement stations for atmospheric pollution and/or gas emissions. We show how such re-engineering or re-design can substantially decrease the operation cost of monitoring networks while providing, at the same time, support for arriving at conclusions or taking decisions with the same statistical power as in conventional setups.

Conclusions

This case study illustrates the benefits of Haphazard Intentional Sampling in the planning of statistical experiments and sampling designs, by using the aforementioned techniques in an application concerning the monitoring of atmospheric pollutants in the city of Sao Paulo, Brazil.

References

V.Fossaluza, M.S.Lauretto, C.A.B.Pereira, J.M.Stern. Combining Optimization and Randomization Approaches for the Design of Clinical Trials. In: A.Polpo et al (Org.). Interdisciplinary Bayesian Statistics. 1ed. Heidelberg: Springer, 2015, v. 1, p. 173-184.

M.S.Lauretto, F.Nakano, C.A.B.Pereira, J.M.Stern. Intentional Sampling by Goal Optimization with Decoupling by Stochastic Perturbation. AIP Conference Proceedings, v. 1490, p. 189-201, 2012.

J. Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

J.M.Stern. Decoupling, Sparsity, Randomization and Objective Bayesian Inference. Cybernetics & Human Knowing, v. 15, p. 49-68, 2008.