ALLOCATION OF ADDITIONAL GRID POINTS TO EXISTING OSPATS STRATA

J.J. DE GRUIJTER

1. Introduction

We assume that a grid of N points has been divided into H strata by Ospats. Using the same symbols as in de Gruijter e.a. (JSSM, 2015), this resulted in the following contributions from the strata $\mathcal{S}_1 \cdots \mathcal{S}_H$ to the objective function O:

$$O(\mathcal{S}_h) = \left\{ \sum_{i=1}^{N_{\mathcal{S}_h}-1} \sum_{j=i+1}^{N_{\mathcal{S}_h}} D'^2_{ij} \right\}^{1/2} \tag{1}$$

(cf. Eq. 7 in de Gruijter e.a.) with

$$D'^2_{ij} = \frac{(\tilde{z}_i - \tilde{z}_j)^2}{R^2} + V(e_i) + V(e_j) - 2\,\mathrm{Cov}(e_i, e_j) \tag{2}$$

(Eq. 11 in de Gruijter e.a.).

We assume further that additional grid points (not included in the stratification process) are to be allocated to the existing Ospats strata. The same data are assumed to be available for the additional grid points as for the original points, i.e. predictions and prediction error variances.

The problem at hand is how to optimally allocate a set of N_a additional grid points $\mathrm{x}_1 \cdots \mathrm{x}_{N_a}$ to the Ospats strata $\mathcal{S}_1 \cdots \mathcal{S}_H$.

2. Proposed method of allocating additional grid points

The principle of the proposed allocation method is to assign each additional grid point in turn to one of the strata while minimizing the same objective function O as was used to create the stratification. This means that, for the allocation of point x_t, we have to find the stratum $\mathcal{S}_h$ of which the increase of its contribution to O by adding x_t to $\mathcal{S}_h$ is minimal. The contributions from the strata to O prior to allocation are given by Eq. (1) above. The contributions after allocation of point x_t can be calculated by:

Date: 2016-08-03.

$$O(\mathcal{S}_h \cup \mathrm{x}_t) = \left\{ \sum_{i=1}^{N_{\mathcal{S}_h}-1} \sum_{j=i+1}^{N_{\mathcal{S}_h}} D_{ij}^{'2} + \sum_{i=1}^{N_{\mathcal{S}_h}} D_{it}^{'2} \right\}^{1/2} \tag{3}$$

Equations (1) and (3) render the increase of the contribution $\Delta(O(\mathcal{S}_h \cup \mathrm{x}_t))$:

$$\Delta(O(\mathcal{S}_h \cup \mathrm{x}_t)) = \left\{ \sum_{i=1}^{N_{\mathcal{S}_h}-1} \sum_{j=i+1}^{N_{\mathcal{S}_h}} D_{ij}^{'2} + \sum_{i=1}^{N_{\mathcal{S}_h}} D_{it}^{'2} \right\}^{1/2} - \left\{ \sum_{i=1}^{N_{\mathcal{S}_h}-1} \sum_{j=i+1}^{N_{\mathcal{S}_h}} D_{ij}^{'2} \right\}^{1/2} \tag{4}$$

From Eq. (4) it follows that, to optimally allocate an additional grid point x_t by minimizing $\Delta(\cdot)$ over h, only the squared generalized distances from x_t to the existing grid points ($D_{it}^{'2}$) need to be calculated, because the sums $\sum_{i=1}^{N_{\mathcal{S}_h}-1} \sum_{j=i+1}^{N_{\mathcal{S}_h}} D_{ij}^{'2}$ are given and fixed.

To enable calculation of sample size, Neyman allocation and sampling variance after the allocation of additional grid points, both the sizes and the contributions of the strata to O need to be updated during the allocation process.

3. Prediction of sample size, Neyman allocation, and sampling variance

After allocation of the N_a additional grid points, the total size of the extended grid equals $N_t = N + N_a$. Denote the size of the extended strata by $N_1^{'} \cdots N_H^{'}$, with $\sum_{h=1}^{H} N_h^{'} = N_t$, and their contributions to O by $O_1^{'} \cdots O_H^{'}$. The updated value of O is obtained by summing the contributions: $O^{'} = \sum_{h=1}^{H} O_h^{'}$.

Following Section 3.5 of De Gruijter e.a., the sample size minimally needed to attain a sampling variance not exceeding a given threshold $V_{\max}$, and assuming Neyman allocation, can be predicted by:

$$\tilde{n} = \{\overline{O^{'}}\}^2 / V_{\max} \tag{5}$$

where $\overline{O^{'}} = O^{'}/N^{'}$.

The predicted sample allocation, $\tilde{n}_1 \cdots \tilde{n}_H$, for the same conditions follows then from:

$$\tilde{n}_h = \tilde{n} \cdot \frac{a_h O_h^{'}}{\sum_{h=1}^{H} a_h O_h^{'}} \tag{6}$$

where a_h is the relative size of extended stratum h: $a_h = N_h^{'}/N^{'}$.

Conversely, if a maximum sample size $n_{\max}$ is given, then the sampling variance under Neyman allocation can be predicted by:

$$\tilde{V}(\hat{\bar{z}}) = \{\overline{O^{'}}\}^2 / n_{\max} \tag{7}$$

For any given allocation, $n_1 \cdots n_H$, the sampling variance can be predicted by:

$$\tilde{V}(\hat{\bar{z}}) = \frac{1}{N_t^2} \sum_{h=1}^{H} (O_h^{'})^2 / n_h \tag{8}$$