Statistics and Computing manuscript No.
(will be inserted by the editor)

A new Universal Resample Stable Bootstrap-based Stopping Criterion in PLS Components Construction

Jérémy Magnanensi · Frédéric Bertrand · Myriam Maumy-Bertrand · Nicolas Meyer

Received: date / Accepted: date

PLS-PR: Constraint insertion

Due to the specificities of the Poisson distribution and its inverse link function exp, we added the following constraint:

$$-1 \leqslant \sum_{k=1}^{3} r_k \leqslant 5,$$

where r_k is a realisation of a random variable $R_k \sim \mathcal{N}\left(0, \sigma_k^2\right)$.

Indeed, the usual simulation process algorithmically performs realisations, noted θ_i, $1 \leqslant i \leqslant n$, of a random variable θ so that each response $y_i = \mathcal{P}(exp(\theta_i))$. The distribution of θ is directly link to the sum of r_1, r_2 and r_3, among others, so that, without constraints on these parameters, this process leads to zero-inflated models combined with a heavy-tailed distribution. Let Z be a simplified approximation of θ. Adding this constraint leads to the following theoretical results. Let:

$$Z = 0.5\left(\sum_{k=1}^{3} R_k + R_5\right) \tag{1}$$

$$\Rightarrow \begin{cases} \mathbb{E}\left(Z | -1 \leqslant \sum_{k=1}^{3} R_k \leqslant 5\right) = 0.985 \\ \mathbb{V}\text{ar}\left(Z | -1 \leqslant \sum_{k=1}^{3} R_k \leqslant 5\right) = 0.745 \\ \text{sd}\left(\sum_{k=1}^{3} R_k | -1 \leqslant \sum_{k=1}^{3} R_k \leqslant 5\right) = 1.727 \end{cases} \tag{2}$$

We simulated two similar datasets, containing 1000 observations. These datasets were simulated respectively with and without constraint. As expected, without constraint no less than 490 out of the 1000 response values are equal to 0 and the maximal simulated one is equal to 111074934. By adding this constraint, the obtained response vector only contains 147 responses equal to 0 and 23 as a maximal value. The two types of distributions (with and without constraint) linked to θ and $\mathbf{y}$ are graphically displayed in Fig.1

Frédéric Bertrand
Institut de Recherche Mathématique Avancée, LabEx IRMIA
Université de Strasbourg et CNRS, Strasbourg, France
E-mail: frederic.bertrand@math.unistra.fr

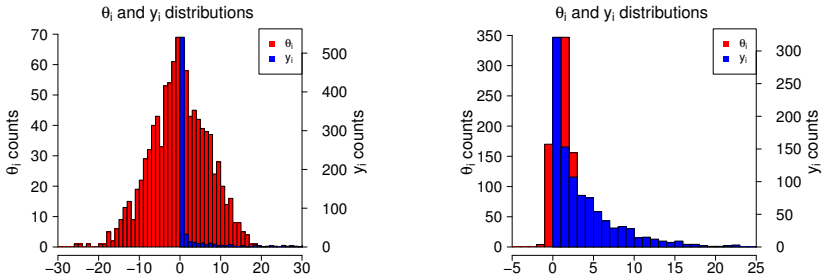

Fig. 1: Left: Distributions without constraint; Right: Distributions with constraint