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

Proof of Proposition 2.2.

Proof Let $n = pq + r$ be the euclidean division of n by q. Then, a partition of the dataset for CV is composed of r $(p+1)$-element subsets and $(q-r)$ p-element subsets, noted A_k, $1 \leqslant k \leqslant q$. We set $q_0 = 0$ and let $E = \{q_1, \ldots, q_r\}$, $1 \leqslant q_j \leqslant q$, $\forall j \in [\![1, r]\!]$ be a subset of $\{1, \ldots, q\}$ containing the ordered set of indices of the $(p+1)$-element subsets so that:

$$Card\,(A_k) = \begin{cases} p+1, & \forall k \in E \\ p, & \forall k \in E^c \end{cases}$$

Let us first determine the number of distinct partitions of $\{1, \ldots, n\}$ which can be written as:

$$\{A_1, \ldots, A_{q_1-1}, A_{q_1}, A_{q_1+1}, \ldots, A_{q_r-1}, A_{q_r}, A_{q_r+1}, \ldots, A_q\} \tag{1}$$

To lighten the formulas, let us set $\omega = \{q_j \in E \,|\, q_j - q_{j-1} \neq 1\}$ and $m_j = q_j - q_{j-1} - 1$. Then, by knowing that, for any set containing n elements, the number of distinct p-element subsets of it that can be formed is

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

given by $\binom{n}{p}$, we obtain that the number of distinct partitions of the form (1) is equal to:

$$\begin{aligned} f(n,q) = \prod_{j=1}^{r} \Bigg[\Bigg[& \mathbb{1}_{\{q_j \in \omega\}} \prod_{i=1}^{m_j} \binom{n-(q_{j-1}-(j-1)+(i-1))\,p-(j-1)\,(p+1)}{p} \\ & + \mathbb{1}_{\{q_j \notin \omega\}} \Bigg] \times \binom{n-((q_j-1)-(j-1))\,p-(j-1)\,(p+1)}{p+1} \Bigg] \\ & \times \left[\mathbb{1}_{\{q_r \neq q\}} \prod_{i=1}^{q-q_r} \binom{n-(q_r-r+(i-1))\,p-r\,(p+1)}{p} + \mathbb{1}_{\{q_r = q\}} \right] \end{aligned}$$

$$\begin{aligned} = \prod_{j=1}^{r} \Bigg[\Bigg[& \mathbb{1}_{\{q_j \in \omega\}} \prod_{i=1}^{m_j} \binom{n-(q_{j-1}+i-1)\,p-(j-1)}{p} + \mathbb{1}_{\{q_j \notin \omega\}} \Bigg] \\ & \times \binom{n-(q_j-1)\,p-(j-1)}{p+1} \Bigg] \\ & \times \left[\mathbb{1}_{\{q_r \neq q\}} \prod_{i=1}^{q-q_r} \binom{n-(q_r+(i-1))\,p-r}{p} + \mathbb{1}_{\{q_r = q\}} \right] \end{aligned}$$

$$\begin{aligned} = \prod_{j=1}^{r} \Bigg[\Bigg[& \mathbb{1}_{\{q_j \in \omega\}} \prod_{i=1}^{m_j} \frac{(n-(q_{j-1}+i-1)\,p-(j-1))!}{p!\,(n-(q_{j-1}+(i+1)-1)\,p-(j-1))!} + \mathbb{1}_{\{q_j \notin \omega\}} \Bigg] \\ & \times \frac{(n-(q_j-1)\,p-(j-1))!}{(p+1)!\,(n-((q_j+1)-1)\,p-((j+1)-1))!} \Bigg] \\ & \times \left[\mathbb{1}_{\{q_r \neq q\}} \prod_{i=1}^{q-q_r} \frac{(n-(q_r+(i-1))\,p-r)!}{p!\,(n-(q_r+(i+1)-1)\,p-r)!} + \mathbb{1}_{\{q_r = q\}} \right] \end{aligned}$$

Each second factorial term in the denominator being equal to the numerator of the following product term, we obtain that:

$$\begin{aligned} f(n,q) = \prod_{j=1}^{r} \Bigg[\Bigg[& \mathbb{1}_{\{q_j \in \omega\}} \frac{(n-q_{j-1}p-(j-1))!}{(p!)^{m_j}\,(n-(q_j-1)\,p-(j-1))!} + \mathbb{1}_{\{q_j \notin \omega\}} \Bigg] \\ & \times \frac{(n-(q_j-1)\,p-(j-1))!}{(p+1)!\,(n-((q_j+1)-1)\,p-((j+1)-1))!} \Bigg] \\ & \times \left[\mathbb{1}_{\{q_r \neq q\}} \frac{(n-q_rp-r)!}{(p!)^{q-q_r}\,(n-qp-r)!} + \mathbb{1}_{\{q_r = q\}} \right] \\ = \prod_{j=1}^{r} \Bigg[& \mathbb{1}_{\{q_j \in \omega\}} \frac{(n-q_{j-1}p-(j-1))!}{(p!)^{m_j}\,(p+1)!\,(n-q_jp-j)!} \\ & + \mathbb{1}_{\{q_j \notin \omega\}} \frac{(n-(q_j-1)\,p-(j-1))!}{(p+1)!\,(n-q_jp-j)!} \Bigg] \\ & \times \left[\mathbb{1}_{\{q_r \neq q\}} \frac{(n-q_rp-r)!}{(p!)^{q-q_r}} + \mathbb{1}_{\{q_r = q\}} \right] \end{aligned}$$

However, by definition, if $q_j \notin \omega$ then $q_j - 1 = q_{j-1}$ and $m_j = 0$. Furthermore, if $q_r = q$ then $\frac{(n-q_rp-r)!}{(p!)^{q-q_r}} = 1$, so:

$$
\begin{aligned}
f(n,q) = & \prod_{j=1}^{r} \left[\mathbb{1}_{\{q_j \in \omega\}} \frac{(n - q_{j-1}p - (j-1))!}{(p!)^{m_j}\,(p+1)!\,(n - q_j p - j)!} \right. \\
& \left. + \mathbb{1}_{\{q_j \notin \omega\}} \frac{(n - q_{j-1}p - (j-1))!}{(p!)^{m_j}\,(p+1)!\,(n - q_j p - j)!} \right] \\
& \times \left[\mathbb{1}_{\{q_r \neq q\}} \frac{(n - q_r p - r)!}{(p!)^{q-q_r}} + \mathbb{1}_{\{q_r = q\}} \frac{(n - q_r p - r)!}{(p!)^{q-q_r}} \right] \\
= & \prod_{j=1}^{r} \left[\frac{(n - q_{j-1}p - (j-1))!}{(p!)^{m_j}\,(p+1)!\,(n - q_j p - j)!} \right] \times \frac{(n - q_r p - r)!}{(p!)^{q-q_r}} \\
= & \left(\frac{1}{p!}\right)^{\sum_{j=1}^{r} m_j + q - q_r} \times \left(\frac{1}{(p+1)!}\right)^{r} \times n! \\
= & \left(\frac{1}{p!}\right)^{qr} \times \left(\frac{1}{(p+1)!}\right)^{r} \times n!
\end{aligned}
$$

Finally, since the order of parts creation within each of the two classes of subsets ((p+1) and p-element subsets) is not useful in distinguishing one partition from another in a CV framework, we had to divide our result by the number of distinct permutations it exists in each class, so:

$$
f(n,q) = \frac{n!}{r!\,(q-r)!} \times \left(\frac{1}{(p+1)!}\right)^{r} \times \left(\frac{1}{p!}\right)^{q-r} \tag{2}
$$

This result is therefore not dependant on E.