Discussion of *Stability Selection* by Meinshausen and Bühlmann

Rajen Shah and Richard Samworth

University of Cambridge

We congratulate the authors for their innovative and thought-provoking paper. Here we propose a minor variant of the subsampling algorithm that is the basis of stability selection. Instead of drawing individual subsamples at random, we advocate drawing disjoint pairs of subsamples at random. This variant appears to have favourable properties.

Below, we use the same notation as the paper. Our method of subsampling involves splitting $\{1, \ldots, n\}$ into two halves at random and picking a subset of size $\lfloor n/2 \rfloor$ in each half. Repeating this M times, we obtain a sequence of subsets $I_1, \ldots, I_{2M}$ with $I_{2i} \cap I_{2i-1} = \emptyset,\ i = 1, \ldots, M$. For $k \in \{1, \ldots, p\}$., define

$$\tilde{\Pi}^{\lambda}_{k,M} = \frac{1}{2M} \sum_{i=1}^{2M} \mathbb{1}\{k \in \hat{S}^{\lambda}(I_i)\}.$$

Similarly to the stability selection algorithm, we select variable k when $\max_{\lambda \in \Lambda} \tilde{\Pi}^{\lambda}_{k,M} \geq \pi_{\text{thr}}$. Then:

(a) Letting V_M be the number of falsely selected variables, $\mathbb{E}(V_M)$ satisfies the same upper bound as in Theorem 1 of the paper. Briefly, defining

$$\hat{\Pi}^{\text{simult},\lambda}_{k,M} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}\{k \in \hat{S}^{\lambda}(I_{2i-1}) \cap \hat{S}^{\lambda}(I_{2i})\},$$

the result corresponding to Lemma 1 of the paper is

$$0 \leq \frac{1}{M} \sum_{i=1}^{M} (1 - \mathbb{1}\{k \in \hat{S}^{\lambda}(I_{2i-1})\})(1 - \mathbb{1}\{k \in \hat{S}^{\lambda}(I_{2i})\}) = 1 - 2\tilde{\Pi}^{\lambda}_{k,M} + \hat{\Pi}^{\text{simult},\lambda}_{k,M}.$$

The arguments of Lemma 2 and Theorem 1 then follow through since $\mathbb{E}(\hat{\Pi}^{\text{simult},\lambda}_{k,M}) = \mathbb{E}(\hat{\Pi}^{\text{simult},\lambda}_{k})$. Thus we have the same error control as in the paper for finite M, as well as the infinite subsampling case.

(b) Simulations suggest that we obtain a slight decrease in the Monte Carlo variance. A heuristic explanation is that, when n is even, each observation is contained in the same number of subsamples. This minimises the sum of the pairwise intersection sizes of our subsamples.

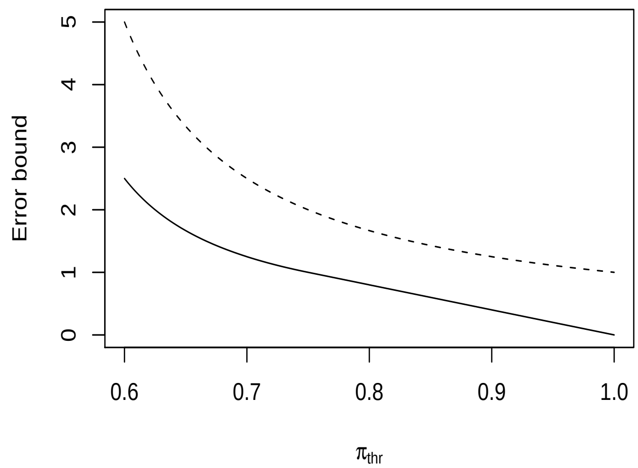

Fig. 1. We plot the factor multiplying q_Λ^2/p against π_{thr} for each of the bounds: the bound of Theorem 1 (dashed) and the new bound with $M = \infty$ (solid).

(c) With essentially no extra computational cost, we obtain estimates of simultaneous selection probabilities, which can also be useful for variable selection; see Fan, Samworth and Wu (2009).

(d) If, in addition to the assumptions of Theorem 1, we also assume that the distribution of $\max_{\lambda \in \Lambda} \hat{\Pi}_{k,M}^{\text{simult},\lambda}$ is unimodal, we obtain improved bounds:

$$\mathbb{E}(V_M) \leq \begin{cases} \dfrac{1}{2(2\pi_{\text{thr}} - 1 - \frac{1}{2M})} \dfrac{q_\Lambda^2}{p} & \text{if } \pi_{\text{thr}} \in \left(\frac{q_\Lambda^2}{p^2} + \frac{1}{2},\ \frac{3}{4}\right] \\ \dfrac{4(1 - \pi_{\text{thr}} + \frac{1}{2M})}{1 + \frac{1}{M}} \dfrac{q_\Lambda^2}{p} & \text{if } \pi_{\text{thr}} \in \left(\frac{3}{4},\ 1\right]. \end{cases}$$

For a visual comparison between this bound and that of Theorem 1, see Figure 1. The improvement suggests that using sample splitting with this bound can lead to more accurate error control than using standard stability selection.

(e) This new bound gives guidance about the choice of M. For instance, when $\pi_{\text{thr}} = 0.6$, choosing $M > 52$ ensures that the bound on $\mathbb{E}(V_M)$ is within 5% of its limit as $M \to \infty$. When $\pi_{\text{thr}} = 0.9$, choosing $M > 78$ has the same effect.

References

Fan, J., Samworth R. and Wu, Y. (2009), Ultrahigh dimensional feature selection: beyond the linear model, *J. Machine Learning Research*, **10**, 2013–2038.