Discussion on "Multiscale Change Point Inference"

Jiashun Jin and Tracy Ke

Carnegie Mellon University and Princeton University

October 30, 2013

We congratulate the authors on a very interesting paper. The paper sheds lights on a problem of great interest, and the theory and methods developed are potentially useful in many applications.

In [5], we have investigated the problem from the variable selection perspective. Consider a linear model

$$Y = X\beta + z, \qquad X = X_{n,n}, \qquad z \sim N(0, I_n), \tag{1}$$

where $X(i,j) = 1\{j \geq i\}$, $1 \leq i, j \leq n$, and β is a sparse vector, containing a small fraction of nonzeros (i.e., "jumps"). In effect, $X\beta$ is a block-wise constant vector, so (1) is a change-point model. We are interested in identifying all jumps.

Fixing $\vartheta \in (0,1)$ and $r > 0$ and letting $\epsilon_n = n^{-\vartheta}$, $\tau_n = \sqrt{2r\log(n)}$, we consider a "Rare and Weak" setting where

$$\beta_j \overset{iid}{\sim} (1-\epsilon_n)\nu_0 + \epsilon_n \nu_{\tau_n}, \qquad (\nu_a\text{: point mass at } a). \tag{2}$$

See [5] for more general case and see [1, 3, 5] for the subtlety of model (2). The *minimax Hamming selection error* is then

$$\text{Hamm}_n^*(\vartheta, r) = \inf_{\hat\beta} \left[\sum_{j=1}^{n} P\left\{ \text{sgn}(\hat\beta_j) \neq \text{sgn}(\beta_j) \right\} \right].$$

We showed in [5] that

$$\text{Hamm}_n^*(\vartheta, r) = \begin{cases} L_n \cdot n^{(1-\vartheta-r/4)} + o(1), & r/\vartheta \leq 6 + 2\sqrt{10}, \\ L_n \cdot n^{(1-3\vartheta-(r/2-\vartheta)^2/(2r))} + o(1), & r/\vartheta > 6 + 2\sqrt{10}, \end{cases}$$

where L_n is a generic *multi*-log(n) term. This implies a *watershed* phenomenon (also found in [1, 3, 2, 4], but in different settings), which can be captured by so-called *Phase Diagram*; see Figure 1.

Moreover, we developed in [5] a procedure called *CASE* which achieves the minimax rate above. CASE is a two-stage Screen and Clean method, where at the heart is *multivariate χ^2-screening* guided by a sparse graph, constructed based on the matrix X.

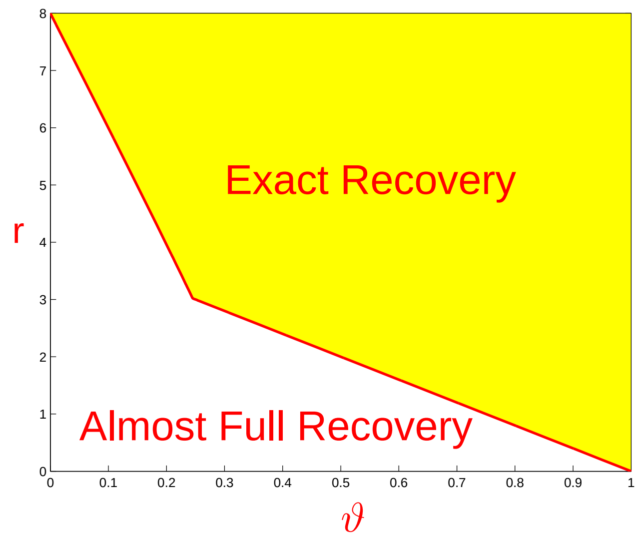

Figure 1: Phase Diagram for change-point model. In the two-dimensional phase space $\{(\vartheta, r) : 0 < \vartheta < 1, r > 0\}$, the curve $r = \rho_0(\vartheta)$ separates the phase space into two sub-regions, where $\rho_0(\vartheta) = \max\{4(1-\vartheta), (4-10\vartheta) + 2\sqrt{[(2-5\vartheta)^2 - \vartheta^2]_+}\}$. For (ϑ, r) in the interior of the region above the curve, it is possible to identify all jumps (say, by using CASE) with high probability. For (ϑ, r) in the interior of the region below the curve, $1 \ll \mathrm{Hamm}_n(\vartheta, r)^* \ll n\epsilon_n$, and it is impossible to identify all jumps, but it is possible to identify most of them (say, by using CASE).

References

[1] Donoho D, Jin, J (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962–994.

[2] Donoho D, Tsaig Y (2008) Fast solutions of ℓ^1-norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theor.* **54**(11), 4789–4812.

[3] Jin J (2009) Impossibility of successful classification when useful features are rare and weak. *Proc. Nat. Acad. Sci.* **106**(22), 8859–8864.

[4] Ingster YI, Pouet C, Tsybakov AB (2009) Sparse classification boundaries. *Phil. Trans. R. Soc. A* **367**, 4427–4448.

[5] Ke T, Jin J, Fan J (2013) Covariance Assisted Screening and Estimation. *arXiv:1205.4645.*