A discussion on "Sure independence screening for ultrahigh dimensional feature space" by J. Fan and L. Lv, Christian P. Robert, CEREMADE, Université Paris Dauphine and CREST, INSEE

While I appreciate the "tour de force" involved in the paper, including the proof that $\mathbb{P}(\mathcal{M}_\star \subset \mathcal{M}_\gamma)$ converges, I cannot but get an overall feeling of slight disbelief about the statistical consequences of the results contained in the paper: in short, I basically question the pertinence of assuming a "true" model in settings when $p \gg n$...

When constructing a statistical model like the regression model at the core of the paper, it is highly improbable that there exists a *single* model, e.g. a *single* subset of regressors that explains the data. Therefore, to assume, as the authors do, (a) that there exists such a subset and (b) that a statistical procedure will pick the "right" regressors when applied in a context where $p \gg n$ strikes me as implausible or only applicable in formalised settings such as orthogonal regressors. If confronted by the opposite, I question the final relevance of the asymptotic results in terms of statistical meaning. Once again, those mathematical valid asymptotic results seem to be orthogonal to the purposes of statistical modelling. In most practical settings, considering a large number p of potential regressors implies that a wide range of alternative submodels will enjoy the same *predictive* properties, especially if $n \ll p$ because, in this setting, an *explicative* model is in my opinion statistically meaningless. Significant variables may be identified in such cases but not a single monolithic collection of those.

A decisional approach that focuses on the decisional consequences of model selection rather than assuming the existence of a single "true" model would seem to be more appropriate, especially because it naturally accounts for correlation among covariates. In addition, using a loss function allows for a rational definition of "important variables", instead of the $0-1$ dichotomy found in the paper. That traditional model choice procedures suffer from computational difficulties and are in practice producing suboptimal solutions is a recognised problem, even though more efficient explorations techniques are under development (Hans et al., 2007a, 2007b; Berger et al., 2008; Bottolo and Richardson, 2008). In addition, adopting a more sensible predictive perspective means that missing the exploration of the full submodel space is only relevant if better fitting models are omitted. This is more than a philosophical difference of perspectives, since it has direct consequences on the way inference is conducted and since the overall simplicity of the hard threshold is more convincing for practitioners than more elaborate modellings.

References

Liang, F., Paulo, R,. Molina, G,. Clyde, M.A. and Berger, J.O. (2008) Mixtures of g Priors for Bayesian Variable Selection. *J. American Statistical Association*, 481 (1), 410-423.

Bottolo, L. and Richardson, S. (2008) Evolutionary Stochastic Search. Technical report, Imperial College London.

Hans, C., Dobra, A. and West, M. (2007a) Shotgun stochastic search in regression with many predictors. *J. American Statistical Association*, 480 (2), 507-516.

Hans, C., Wang, Q., Dobra, A. and West, M. (2007b) SSS: High-dimensional Bayesian regression model search. *Bulletin of the International Society for Bayesian Analysis* 24, 8-9.