ORIGINAL ARTICLE

Journal Section

Discussing CARS

Etienne Roquain[1]

Introduction

I would like to congratulate the authors for this excellent contribution to the multiple testing methodology. This work shows, in a two-sided setting, that reducing a data set to the column of test statistics is suboptimal and that the multiple testing decision can gain a lot from incorporating side information. In a nutshell, the authors proposed to 1) find the procedure, called Oracle CARS, that solves the problem of maximizing the power while controlling the (marginal) FDR; 2) approximate this optimal procedure by a data-driven version, called CARS, by the way of kernel estimators; 3) prove theoretical consistency of CARS when all the parameters are kept fixed and the number m of nulls grows to infinity. Many extensions are discussed and a package is implementing CARS, which makes this method available for practitioners.

The purpose of this discussion is to underline the Bayesian flavor of CARS and discuss a Cauchy-Slab version of it. In particular, since the theoretical framework of the paper seems to exclude the case where the model parameters depend on m, so exclude sparse signal, we consider the problem of obtaining a uniform FDR control under sparsity, which can be formulated as

$$\sup_{\Delta:\|\Delta\|_0\leq s_m} \{\mathrm{FDR}(\Delta)\} \leq \alpha, \quad s_m/m \to 0, \tag{1}$$

where $\Delta \in \mathbb{R}^m$ is the true mean difference and $\|z\|_0$ denotes the number of non-zeros in z. Since the original BH procedure (1) provides such a guarantee, one may ask whether procedures improving upon BH also provide (1).

Stylized CARS Setting

For short, let us consider the following simplified version of the CARS setting. Let $X \sim \mathcal{N}(\mu_x, I_m)$ and $Y \sim \mathcal{N}(\mu_y, I_m)$ be two independent random vectors, for two $\mathbb{R}^m$ mean vectors μ_x and μ_y. We consider the problem of testing simultaneously the nulls $H_{0,i}$: "$\mu_{x,i} = \mu_{y,i}$" against $H_{1,i}$: "$\mu_{x,i} \neq \mu_{y,i}$". We let

$$\theta_i = 1\{\mu_{x,i} \neq \mu_{y,i}\}, \quad 1 \leq i \leq m,$$

the true/false status of the hypotheses. Testing can be done by using the standard test statistics

$$T_1 = (X - Y)/\sqrt{2} \sim \mathcal{N}(\Delta, I_m), \quad \Delta = (\mu_x - \mu_y)/\sqrt{2}.$$

The idea of CARS is to keep the information of the (independent) covariate $T_2 = (X + Y)/\sqrt{2} \sim \mathcal{N}((\mu_x + \mu_y)/\sqrt{2}, I_m)$ in the analysis to help for making the decision.

Their model additionally uses some random effects for the mean couple $(\mu_x, \mu_y) \in \mathbb{R}^{2m}$, which can be interpreted as choosing a particular prior distribution Π on the true parameters. Let $\phi_i = 1\{\mu_{x,i} \neq 0 \text{ or } \mu_{y,i} \neq 0\}$, and generate (θ, ϕ) as follows: θ_i i.i.d. $\sim \mathcal{B}(1-\pi_0)$; $\phi_i \mid \theta_i = 0$ i.i.d. $\sim \mathcal{B}(1-\pi_{0|0})$; $\phi_i = 1$ if $\theta_i = 1$. Also let $\pi_{00} = \pi_0 \pi_{0|0}$ the probability that both means are equal to zero. To complete the description of the prior Π, the components of (μ_x, μ_y) are generated independently as

$$(\mu_{x,i} - \mu_{y,i})/\sqrt{2} \sim \begin{cases} \delta_0 & \text{if } \theta_i = 0 \\ \gamma & \text{if } \theta_i = 1 \end{cases} ; (\mu_{x,i} + \mu_{y,i})/\sqrt{2} \sim \begin{cases} \delta_0 & \text{if } \phi_i = 0 \\ \gamma & \text{if } \phi_i = 1 \end{cases} .$$

At this point, the strategy of CARS is to make the decision by estimating the posterior probability $\Pi(\theta_i = 0 \mid T_{1,i}, T_{2,i})$, which can be seen as implicitly using a non parametric estimator of the slab γ.

Cauchy-slab CARS version

We discuss now the possibility to avoid the non parametric estimation, by fixing γ equal to $\gamma(x) = (2\pi)^{-1/2}(1 - |x|\overline{\Phi}(x)/\phi(x))$, which is the so-called quasi-Cauchy distribution (4). Recent studies suggest that such a prior is particularly suitable to get posterior distributions with good frequentist properties, see e.g. (2) and references therein. An intuitive explanation is that this density has heavy tails, which thus puts mass 'everywhere' and can thus account for any true alternative distribution of the test statistics.

We introduce the Cauchy-slab version of CARS procedure, that rejects $H_{0,i}$ as soon as $q(T_{1,i}, T_{2,i}) \leq \alpha$, for which

$$\begin{aligned} q(t_1, t_2) &= \Pi(\theta_i = 0 \mid |T_{1,i}| \geq t_1, T_{2,i} = t_2) \\ &= \frac{\pi_{00}\overline{\Phi}(t_1)\phi(t_2) + (\pi_0 - \pi_{00})\overline{\Phi}(t_1)g(t_2)}{\pi_{00}\overline{\Phi}(t_1)\phi(t_2) + (\pi_0 - \pi_{00})\overline{\Phi}(t_1)g(t_2) + (1-\pi_0)\overline{G}(t_1)g(t_2)}, \end{aligned}$$

where $g(x) = (\phi * \gamma)(x) = (2\pi)^{-1/2}x^{-2}(1 - e^{-x^2/2})$ and $\overline{G}(s) = \int_s^{+\infty} g(x)dx$. In the last display, the only unknown quantities are the hyper-parameters π_0 and π_{00}. The parameter π_0 (resp. π_{00}) can be easily estimated by marginal maximum likelihood from the sample T_1 (resp. T_2), see (4) and the devoted package (5), because the $T_{1,i}$'s are i.i.d. $\sim \pi_0\phi + (1-\pi_0)g$ (resp. the $T_{2,i}$'s are i.i.d. $\sim \pi_{00}\phi + (1-\pi_{00})g$).

Now, two questions are: 1) does this Cauchy-slab version enjoy the uniform FDR control (1)? 2) does it still improve upon the BH procedure in terms of power? I tend to believe that both answers are positive. First, it seems reasonable to think that the uniform FDR control can be proved by extending the methodology of (3) to the bivariate case. Second, even with the uninformative Cauchy slab, the covariate T_2 can still help T_1 to make the correct decision, as illustrated on Figure 1. These two facts have been confirmed by unreported numerical computations.

Conclusion

Overall, this discussion puts forward the issue of approaching the oracle version of CARS procedure under sparsity. There may be a tradeoff in the choice of the complexity of the slab estimator: while using a non parametric kernel estimator is ambitious from a power point of view, it might be too unstable for a reliable control of the FDR. Instead, using a simpler Cauchy slab introduces a bias that will stabilize FDR, but does reduce power. Finding a principled tradeoff between these two extremes is certainly an interesting avenue for future research.

To conclude, it is a pleasure for me to propose the vote of thanks.

I also warmly acknowledge Ismael Castillo, Sebastian Dölher and Mark van de Wiel for discussions that helped while preparing this note. This work has been supported by ANR-16-CE40-0019 (SansSouci) and ANR-17-CE40-0001 (BASICS).

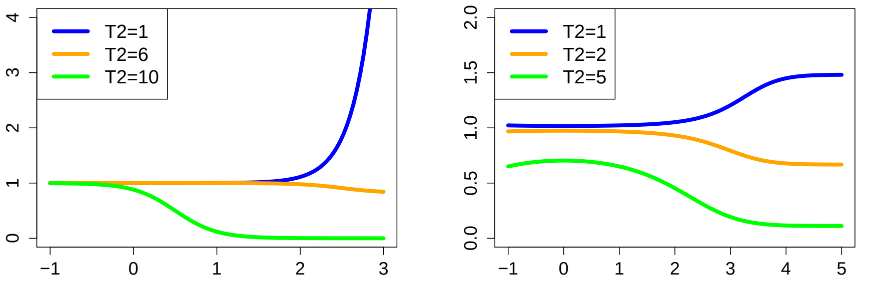

FIGURE 1 Ratio of the posteriors $\frac{\Pi(\theta_i=0 \mid T_{1,i},T_{2,i})}{\Pi(\theta_i=0 \mid T_{1,i})}$ in function of $T_{1,i}$, for different values of $T_{2,i}$. Prior computed either from the true mixture density (left) or from the Cauchy slab (right). In both cases, small (resp. large) values of $T_{2,i}$ helps $T_{1,i}$ to accept (resp. reject) the null.

REFERENCES

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.

[2] I. Castillo and R. Mismer. Empirical bayes analysis of spike and slab posterior distributions. *Electron. J. Statist.*, 12(2):3953–4001, 2018.

[3] I. Castillo and E. Roquain. On spike and slab empirical Bayes multiple testing. *ArXiv e-prints*, page arXiv:1808.09748, Aug. 2018.

[4] I. M. Johnstone and B. W. Silverman. Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, 32(4):1594–1649, 2004.

[5] I. M. Johnstone and B. W. Silverman. EbayesThresh: R Programs for Empirical Bayes Thresholding. *Journal of Statistical Software*, 12(8), 2005.