ecpc
A flexible approach to co-data learning for various co-data

Mirrelijn van Nee

Epidemiology & Data Science, Amsterdam University Medical Centers, Netherlands

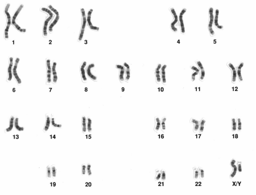
Short course, Channel Network Conference, 2021

Overview

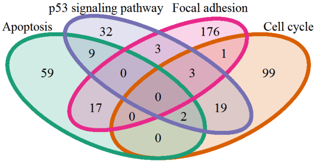
		gren	ecpc
Type of covariate model:	Dense	v	v
	Group-sparse	-	v
	Sparse	v	v/-†
Type of co-data:	Non-overlapping groups	v	v
	Overlapping groups	-	v
	Hierarchical groups	-	v
	Multiple co-data sources	v	v
Hyperparameter shrinkage (many groups)		-	v
Type of response model:	Linear	-	v
	Binary	v	v
	Binomial	v	-
	Survival	-	v

Table: Overview of properties. †Using posterior selection.

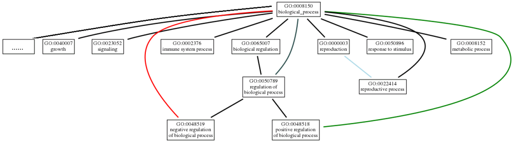
Hypershrinkage is used to accommodate various co-data



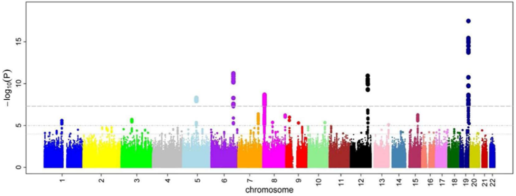
(a) Chromosomes: non-overlapping groups of genes on the same chromosome



(b) Pathways: overlapping groups of interacting genes or molecules



(c) Gene Ontology: structured groups representing relationships in for example biological function



(d) Continuous p-values from an external study

Model

- Generalised linear model for response $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim \pi(\boldsymbol{Y}|X, \boldsymbol{\beta}),\ E(\boldsymbol{Y}) = g^{-1}(X\boldsymbol{\beta})$$

- Normal prior (i.e. ridge penalty) with weighted prior variance:

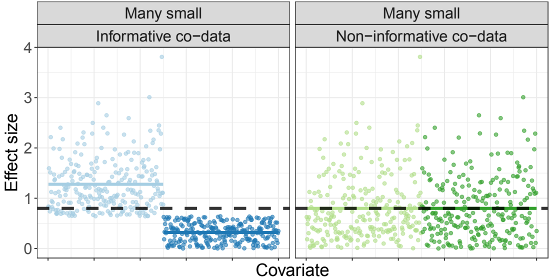
$$\beta_k \overset{ind.}{\sim} N\left(0, \tau^2_{global} \sum_{d=1}^{D} w^{(d)} \left[\boldsymbol{Z}^{(d)}\boldsymbol{\gamma}^{(d)}\right]_k\right),\ k = 1, .., p,$$

  - $Z^{(d)}$: codes group information for each covariate
  - $w^{(d)}$: co-data specific weight
  - $\boldsymbol{\gamma}^{(d)}$: group specific weights for co-data set d

Interpretation empirical Bayes estimates

Empirical Bayes estimates quantify a priori expected effect size globally, in groups and in group sets:

1. $\tau_{global} = \sqrt{\frac{\pi}{2}} E_{\beta_k|\tau^2_{global}}(|\beta_k|) \approx \sqrt{\frac{\pi}{2}} \frac{1}{p} \sum_{k=1}^{p} |\beta_k|$

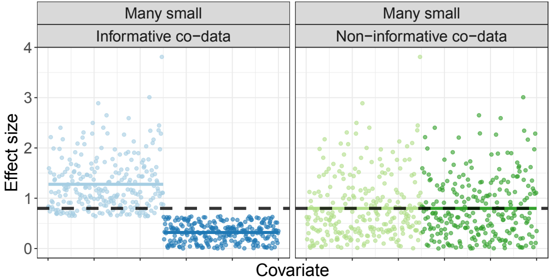


Interpretation empirical Bayes estimates

Empirical Bayes estimates quantify a priori expected effect size globally, in groups and in group sets:

1. $\tau_{global} = \sqrt{\frac{\pi}{2}} E_{\beta_k|\tau^2_{global}}(|\beta_k|) \approx \sqrt{\frac{\pi}{2}} \frac{1}{p} \sum_{k=1}^{p} |\beta_k|$

2. $\sqrt{\gamma_g^{(d)}} = \frac{\sqrt{\frac{\pi}{2}} E_{\beta_k|\tau^2_{global}, \gamma_g^{(d)}}(|\beta_k|)}{\tau_{global}} \approx \frac{\frac{1}{G} \sum_{k \in \mathcal{G}_g^{(d)}} |\beta_k|}{\frac{1}{p} \sum_{k=1}^{p} |\beta_k|}$



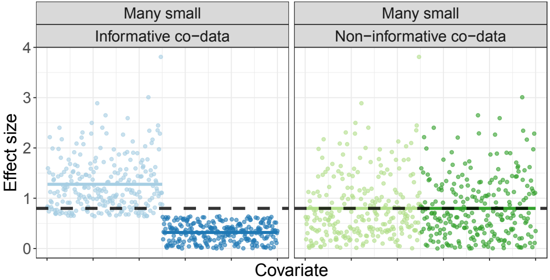
Interpretation empirical Bayes estimates

Empirical Bayes estimates quantify a priori expected effect size globally, in groups and in group sets:

1. $\tau_{global} = \sqrt{\frac{\pi}{2}} E_{\beta_k|\tau^2_{global}}(|\beta_k|) \approx \sqrt{\frac{\pi}{2}} \frac{1}{p} \sum_{k=1}^{p} |\beta_k|$

2. $\sqrt{\gamma_g^{(d)}} = \frac{\sqrt{\frac{\pi}{2}} E_{\beta_k|\tau^2_{global},\gamma_g^{(d)}}(|\beta_k|)}{\tau_{global}} \approx \frac{\frac{1}{G}\sum_{k\in\mathcal{G}_g^{(d)}} |\beta_k|}{\frac{1}{p}\sum_{k=1}^{p} |\beta_k|}$

3. $w^{(1)}, w^{(2)}$



Hypershrinkage: penalised hyperparameter estimates

- Empirical Bayes method of moments boils down to solving a linear system for each co-data set d:

$$A^{(d)}\boldsymbol{\gamma}^{(d)} = \boldsymbol{b}^{(d)}$$

- What if..
  - ..the co-data are not informative?
  - ..we have many groups?
  - ..we have overlapping groups?
  - ..we expect only few groups of covariates to be important?
  - ..the groups are hierarchically ordered?
- Include a suitable penalty on the group level and solve penalised system:

$$\hat{\boldsymbol{\gamma}}^{(d)} = (\tilde{\boldsymbol{\gamma}}^{(d)})_+,\ \tilde{\boldsymbol{\gamma}}^{(d)} = \underset{\boldsymbol{\gamma}^{(d)}}{\text{argmin}}\ ||A^{(d)}\boldsymbol{\gamma}^{(d)} - \boldsymbol{b}^{(d)}||_2^2 + f_{pen}^{(d)}\left(\boldsymbol{\gamma}^{(d)}; \hat{\lambda}^{(d)}\right)$$

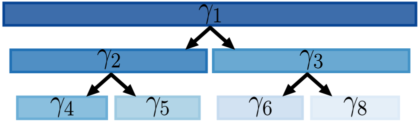
Use random splits of groups to estimate the hypershrinkage penalty

Instead of inefficient *sample* splitting:

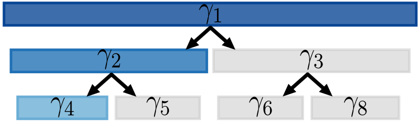
- Split each *group* randomly in two: $\mathcal{G}_{in}$ and $\mathcal{G}_{out}$
- Set up moment-based estimation equations twice: for *in*-part and *out*-part
- Given a set S of random splits, use:

$$\hat{\lambda}^{(d)} = \underset{\lambda^{(d)}}{\operatorname{argmin}} \frac{1}{|\mathcal{S}|} \sum_{\mathcal{S}} ||A_{out}^{(d)} \tilde{\boldsymbol{\gamma}}_{in}^{(d)}(\lambda^{(d)}) - \boldsymbol{b}_{out}^{(d)}||_2^2$$

Example - data-driven discretisation of continuous co-data



(a) Discretise continuous scale in groups in increasingly smaller group sizes



(b) Grey groups are not selected. Use hierarchical lasso to select group weights only if all its parents are selected. (Yan and Bien 2017; Jacob, Obozinski, and Vert 2009; Yang and Zou 2015)

Software

- R-package ecpc
- Download on github using R-package devtools:
  ```
  install.packages("devtools")
  library(devtools)
  install_github("Mirrelijn/ecpc/Rpackage")
  ```
- Function ecpc:
  - estimate global prior variance τ^2_{global}, group weights $\boldsymbol{\gamma}^{(d)}$, group set weights $\boldsymbol{w}$ and regression coefficients $\boldsymbol{\beta}$
  - Input variable hypershrinkage, e.g. "none", "ridge", "lasso" or "hierLasso,ridge"

Data examples in the practical

- Application to classification of lymph node metastasis (LNM) (Te Beest et al. 2017)
- Available co-data:
  1. Signature of genes
  2. P-values from previous study
  3. Correlation with copy number alteration data from the same samples
- Perform a similar analysis twice:
  - First without using any hypershrinkage
  - Then with hypershrinkage to flexibly learn from the co-data

Course information

- More details in (van Nee, Wessels, and van de Wiel 2020)
- Course material available on: `https://github.com/Mirrelijn/Short-course-CNC2021`

Jacob, Laurent, Guillaume Obozinski, and Jean-Philippe Vert (2009). "Group lasso with overlap and graph lasso". In: *Proceedings of the 26th annual international conference on machine learning*. ACM, pp. 433–440.

Te Beest, Dennis E et al. (2017). "Improved high-dimensional prediction with Random Forests by the use of co-data". In: *BMC bioinformatics* 18.1, pp. 1–11.

van Nee, Mirrelijn M, Lodewyk FA Wessels, and Mark A van de Wiel (2020). "Flexible co-data learning for high-dimensional prediction". In: *arXiv preprint arXiv:2005.04010*.

Yan, Xiaohan, Jacob Bien, et al. (2017). "Hierarchical sparse modeling: A choice of two group lasso formulations". In: *Statistical Science* 32.4, pp. 531–560.

Yang, Yi and Hui Zou (2015). "A fast unified algorithm for solving group-lasso penalize learning problems". In: *Statistics and Computing* 25.6, pp. 1129–1141.