discussion with benjamin

Showing the use of knock-off covariates and comparaison between theoritical and empircal propreties.

covariate: variable related to outcome variable

The difference between:

Classical inference

1. Devise a model
2. Collect data
3. Test hypotheses

Post selection inference

With a selected model (which we choose according to a specific metric, leads to bias, anyway...) we then do hypothesis testing to know confidence or other measure about parameters/model.

What we usually do is sample splitting, use one for selection and the other for inference.

1. Collect data
2. Select a model
3. Test hypotheses

Covariate knock-off

Objectif: Select subset of variables relevant for Y (or select covariates) while bounding the False Discovery rate.

Allows for model selection and hypothesis testing at he same time?

Notes:

U-statistics: un-biased estimator

V-statistics: not unbiased but similarly asymptotically then U-stats

---

Perform study with:

1 - simulated data: covariates and output sharing non-linear representation like: $X_i \sim \mathcal{N}(I_d, \beta.\Sigma)$ where $\beta \in \{0,1\}^d$
2 - $p >> n$
3 - Real data from UCI, low/high dimensional

to show:

1 - that the difference between the expected and the observed is bounded
2 - Accuracy of different methods
3 - Compare with existing methods: PC, HSIC
4 - The False Detection Rate (but this could be the same as 1))

And we can use:

1 - Tobias HSIC code. Ideas of just replacing HSIC by Kendal, spearman or MMD.
2 - Jazza's paper (find full reference, and also maybe not Jazza, it could be Zhang, author of one of the references)
3 - Ref [3] *Model-Free Feature Screening and FDR Control with Knockoff Features* provides code. In particular, python code for gene knock off data simulations. Should be ok to modify existing W_i statistic and TR.
4 - Ref[1] for knock off *Controlling the false discovery rate via knockoffs*, The Annals of Statistics

- https://web.stanford.edu/group/candes/knockoffs/
- https://en.wikipedia.org/wiki/Knockoffs_(statistics)
- https://www.stat.cmu.edu/~ryantibs/journalclub/knockoff.pdf
- https://imaging-in-paris.github.io/semester2019/slides/w3/Thirion.pdf (in practice and applied to GWAS)

5 - Ref[4] for formula origin *Some new measures of dependence for random variables based on Spearman's ρ and Kendall's τ*
6 - Book asymptotic statistics, of Van der Wanz