Optimal Multiple Testing with Prior Information

Edgar Dobriban[1] Kristen Fortney[2] Stuart K. Kim[2] Art B. Owen[1]

[1]Statistics, Stanford

[2]Developmental Biology & Genetics, Stanford

August 30, 2015

Overview

Motivating example: Genomics of human longevity

New Method

Motivating example: Human longevity

- 25-30 % heritable
- Genome-Wide Association Studies (GWAS)
  - centenarians vs controls
  - test 500K SNPs (single-nucleotide polymorphisms)
  - usually t-test/z-test
- Only one DNA polymorphysm (APOE/TOMM40) associated
- Challenges:
  - Few centenarians
  - Small effects

Our approach

- Leverage information from age-related diseases
  - Focus on smaller set of variants that protect from disease
- Many large prior GWAS:
  - coronary artery disease (CAD) $n >$20K.
  - type II diabetes (T2D)
  - ...
- How?

P-value weighting

- null hypothesis H_i: i-th SNP is not associated to longevity
- Use weighted Bonferroni method:
  - reject H_i if $P_i \leq qw_i$
  - weights $w_i \geq 0$, $\sum_{i=1}^{J} w_i = J$
  - controls FWER at $\alpha := Jq$
- How to choose weights?
- Optimize expected number of discoveries: Spjotvoll (1972); Benjamini and Hochberg (1997); Rubin et al. (2006); Roeder and Wasserman (2009); Eskin (2008); Peña et al. (2011) ...

Gaussian p-value weighting

- for each SNP have a test statistic $T_i \sim \mathcal{N}(\mu_i, 1)$ in current study
- test $H_i : \mu_i = 0$ against $\mu_i < 0$
- information about μ_i from prior studies: $\mu_i \sim \mathcal{N}(\eta_i, \sigma_i^2)$.
- *Bayes weights* maximize expected number of discoveries $R(w)$:

$$\max_{w \in \mathbb{R}^J} \ \mathbb{E}_\mu \mathbb{E}_T R(w)$$

$$\text{s.t. } w_i \geq 0, \ \sum_{i=1}^{J} w_i = J.$$

- non-convex optimization
- formulated by Westfall et al. (1998), algorithm for small J

Main results

- We give efficient methods to find the weights.
- Let $\gamma = (\sigma^2 + 1)^{1/2}$, and

$$c(\eta, \gamma; \lambda) = -\frac{\eta + \gamma\{\eta^2 + 2\left(\gamma^2 - 1\right)\log(\gamma\lambda)\}^{1/2}}{\gamma^2 - 1}.$$

- **Theorem 1**: If the significance level $q \in [0, 1]$ is small enough, then the optimal Bayes weights are

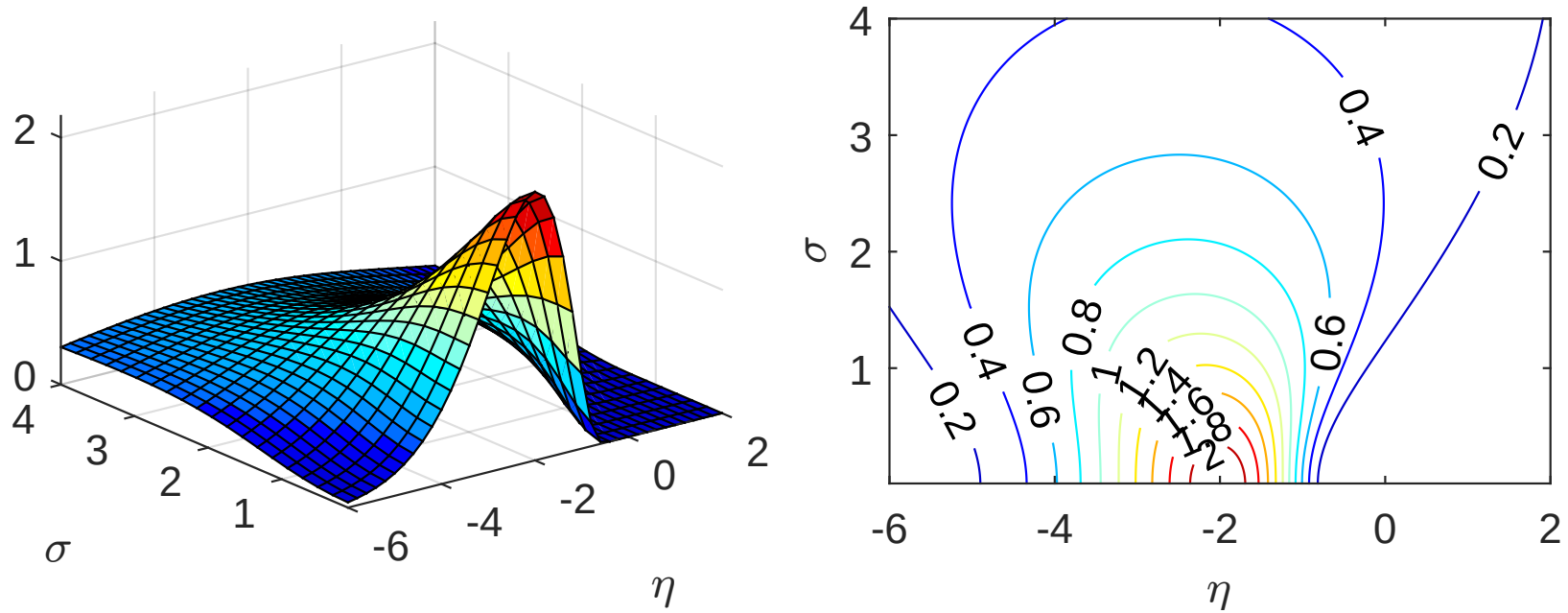
$$w_i = w(\eta_i, \gamma_i; \lambda) = \Phi\{c(\eta_i, \gamma_i; \lambda)\}/q$$

- $\lambda \geq 1$ is the unique constant s.t. $\sum_{i=1}^{J} w(\eta_i, \gamma_i; \lambda) = J$.
  - weights can be found by 1D line search

Main results

- **Theorem 2**: For any significance level q, the weights can be found nearly exactly in $O(J \log J)$ steps.
  - we find it remarkable, non-convex optimization hard in general
  - use duality, reduce to 1D line search

Bayes weights



(a) Surface (b) Contour

Figure : An instance of Bayes weights $w(\eta, \sigma)$, a function of prior mean η and standard error σ. The weight of the i-th p-value is $w(\eta_i, \sigma_i)$.

Further theoretical results

- as $\sigma \to 0$, the Bayes weights tend to the Spjotvoll weights of Spjotvoll (1972); Rubin et al. (2006); Roeder and Wasserman (2009)
- as $\sigma \to \infty$, a symmetry-breaking occurs
  - the weighting scheme $(1/q, \ldots, 1/q, \varepsilon, \ldots, \varepsilon)$ dominates uniform weighting
  - due to the extreme non-convexity of objective
  - reinforces that previous results are remarkable

Empirical results

- apply to GWAS: *iGWAS* or informed GWAS
- longevity: leads to new loci that replicate. Fortney et al., submitted
- simulations: compares favorably to alternatives (more power)
- data analysis: slight advantage over existing methods
  - use GWAS both as prior and current traits
  - several pairs: e.g., eGFR creatinine $\rightarrow$ coronary artery disease (CAD)
  - only our method could improve power, when averaged over all pairs studied
- all computational results reproducible

Simulation

- In Gaussian model, generate $J = 1000$ random means and variances iid $\eta_i \sim \mathcal{N}(0, 1)$, $\sigma_i \sim |\mathcal{N}(0, 1)|$
- Set $q = 10^{-2}$
- Compare weighting methods
  - unweighted testing: $w_i = 1$
  - Bayes weighting: prior parameters $\eta_i, \phi\sigma_i^2$
  - exponential weighting: $w_i \propto \exp(-\beta\eta_i)$, (Roeder et al., 2006)
  - filtering: test top effects $|\eta_i| \geq |M|$, with equal weights
- Compute the power as the objective function divided by J

Simulation

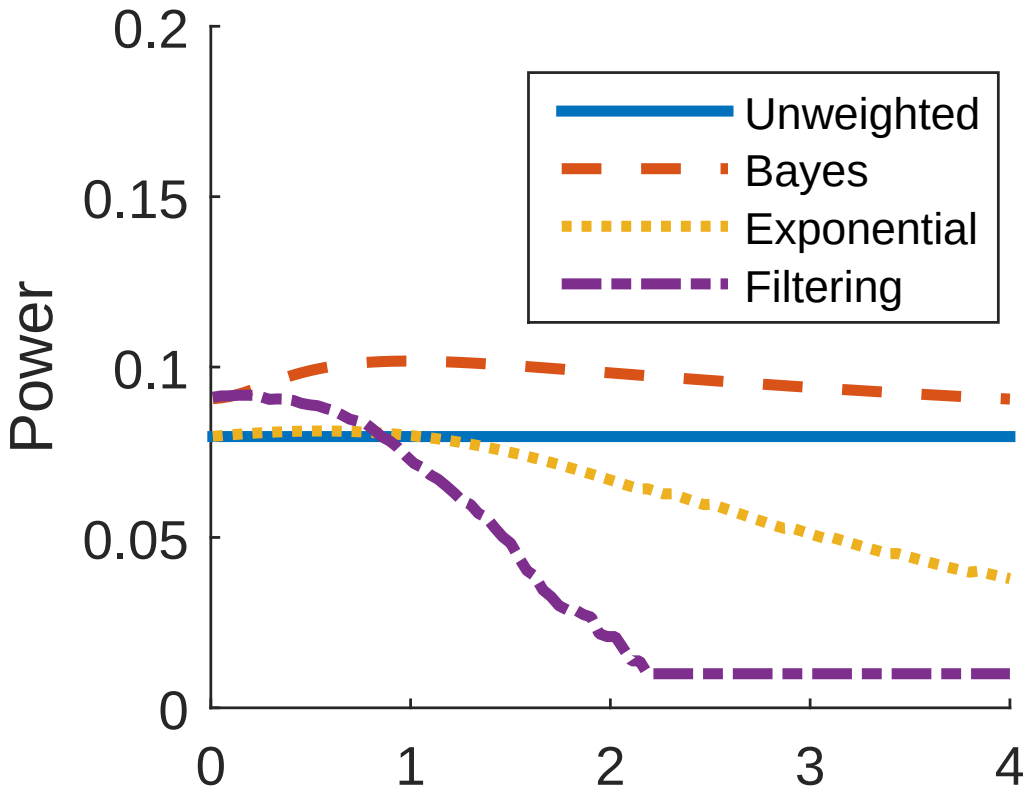


Figure : Power of four p-value weighting methods as a function of their parameter. Bayes as a function of the dispersion ϕ; exponential as a function of β; and filtering as a function of $|M|$.

Conclusion

- new p-value weighting method - Bayes weights/iGWAS
  - control type I error (optimal for FWER control)
  - use prior information
  - Box (1980)'s vision: be Bayesian when predicting and frequentist when testing
- general: Gaussian tests
- easy to use: only summary statistics from prior studies
- computationally efficient

Additional information

- Software
  - R package pweight for p-value weighting on CRAN.
  - MATLAB package from `github.com/dobriban`
- Reference: Dobriban et al. (2015): Optimal multiple testing under a gaussian prior on the effect sizes, *Biometrika*, to appear. *arXiv:1504.02935*.
- Acknowledgements
  - Supported by grants from the AFAR/EMF and the NIH/NIA (AG025941), and NSF grants DMS-1407397, DMS-1418362.
  - Helpful feedback: E. Eskin, J. Habiger, W. Su, J. Wang
- Contact
  - `dobriban@stanford.edu`
  - `web.stanford.edu/~dobriban`
- Thank you!

Benjamini, Y. and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24(3):407–418.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with Discussion). *Journal of the Royal Statistical Society. Series A (General)*, 143(4):383–430.

Dobriban, E., Fortney, K., Kim, S. K., and Owen, A. B. (2015). Optimal multiple testing under a gaussian prior on the effect sizes. *arXiv preprint arXiv:1504.02935*. Biometrika, to appear.

Eskin, E. (2008). Increasing power in association studies by using linkage disequilibrium structure and molecular function as prior information. *Genome research*, 18(4):653–660.

Peña, E. A., Habiger, J. D., and Wu, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *The Annals of Statistics*, 39(1):556–583.

Roeder, K., Bacanu, S.-A., Wasserman, L., and Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *The American Journal of Human Genetics*, 78(2):243–252.

Roeder, K. and Wasserman, L. (2009). Genome-wide significance levels and weighted hypothesis testing. *Statistical Science*, 24(4):398.

Rubin, D., Dudoit, S., and Van der Laan, M. (2006). A method to increase the power of multiple testing procedures through sample splitting. *Statistical Applications in Genetics and Molecular Biology*, 5(1).

Spjotvoll, E. (1972). On the optimality of some multiple comparison procedures. *The Annals of Mathematical Statistics*, 43(2):398–411.

Westfall, P. H., Krishen, A., and Young, S. S. (1998). Using prior information to allocate significance levels for multiple endpoints. *Statistics in Medicine*, 17(18):2107–2119.