Benchmarking state-of-the-art p-value aggregation methods

Alberto Fernández-de-Marcos • Department of Statistics, UC3M
Eduardo García-Portugués • Department of Statistics, UC3M

Combining p-values to test multiple null hypotheses with a single test is of fundamental importance in statistics. A wide variety of methods exist and continue to emerge in the literature. Early methods were designed to aggregate independent p-values, such as the combination method of Fisher (1934). More recent approaches seek validity under dependence structures among the p-values. The Cauchy Combination Test (CCT), proposed by Liu & Xie (2020), was developed to be powerful under sparse signals. However, in the presence of p-values close to 1, the method can become ineffective. To address this limitation, the Positive Cauchy Combination Test (PCCT) was recently proposed by Ouyang et al. (2025). Having a form very similar to the CCT, it overcomes the aforementioned drawback. Meanwhile, the Harmonic Mean p-value (HMP) was proposed in Wilson (2019), and empirically shown to be valid under certain dependence scenarios. Later, Wilson (2020) extended the HMP by using generalized means.

This thesis project aims to analyze these state-of-the-art methods, both theoretically and empirically. This includes: (1) understanding and explaining precisely the mathematics of the methods; (2) performing selected numerical experiments from the references and constructing original ones to showcase the empirical properties of the methods, identifying their strengths and weaknesses; (3) choosing one of these two applications: (i) compare the methods to aggregate existing hypothesis tests for uniformity testing problems with directional data, or (ii) compare the methods when applied to genomic data (see, e.g., Visscher et al. (2012) and Okbay et al. (2016)).

References

Fisher, R. A. (1934). *Statistical methods for research workers* (Vol. 5). Oliver; Boyd.

Liu, Y., & Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.*, *115*(529), 393–402. https://doi.org/10.1080/01621459.2018.1554485

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.*, *48*(6), 624–633. https://doi.org/10.1038/ng.3552

Ouyang, Y., Liu, X., Zhu, L., & Xu, W. (2025). Effective positive Cauchy combination test. *arXiv*. https://doi.org/10.48550/arXiv.2410.10345

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *Am. J. Hum. Genet.*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc. Natl. Acad. Sci. U.S.A.*, *116*(4), 1195–1200. https://doi.org/10.1073/pnas.1814092116

Wilson, D. J. (2020). Generalized mean p-values for combining dependent tests: Comparison of generalized central limit theorem and robust risk analysis. *Wellcome Open Res.*, *5*, 55. https://doi.org/10.12688/wellcomeopenres.15761.1