

The harmonic mean p -value in multiple testing problems

Eduardo García-Portugués • Department of Statistics, UC3M

The Harmonic Mean P-value (HMP) is a powerful method to aggregate the p -values associated to testing several null hypotheses in a single, meaningful, p -value. Unlike Fisher (1934)'s aggregation of p -values, it does not assume that the tests are independent, and hence the HMP is more generally applicable in contexts where a large amount of dependent tests are applied, such as in Genome-Wide Association Studies (GWAS). It was recently proposed by Wilson (2019c) (see also Wilson (2019a)) and it offers a conceptually-appealing alternative to other correction procedures that control for multiple testing. Indeed, it is more powerful than common corrections including that by Bonferroni and by Simes (1986) when combining large proportions of all the p -values. It is also more stringent than controlling the false discovery rate (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001), and possesses theoretical robustness to positive correlations between tests and unequal weights. The HMP generates a multi-level test in the sense that a superset of one or more significant tests is certain to be significant and conversely when the superset is non-significant, the constituent tests are certain to be non-significant. It is based on model averaging by mean maximum likelihood and is theoretically well-grounded in a generalized central limit theorem. The R package `harmonicmeannp` (Wilson, 2019b) implements the HMP and gives detailed examples.

This thesis proposal is focused on analyzing the HMP in detail, both theoretically and empirically. This includes: (1) understanding and explaining precisely the mathematics of HMP and its improvements with respect to other correction procedures, fully understanding Wilson (2019c) and Wilson (2019a); (2) replicate some of the numerical experiments in Wilson (2019c) and build original numerical examples showcasing the empirical properties of HMP and its competitors, determining the weaknesses and strengths; (3) use the HMP to aggregate existing hypothesis tests for several testing problems (e.g., one- or two-sample problems) to construct new test by aggregating already existing ones; (4) examine the power of the new tests by simulations; (5) use the HMP-based test in a dataset that was not previously analyzed.

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188. <https://doi.org/10.1214/aos/1013699998>
- Fisher, R. A. (1934). *Statistical methods for research workers* (Fifth edition). Oliver; Boyd.

- Simes, R. J. (1986). An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751–754. <https://doi.org/10.2307/2336545>
- Wilson, D. J. (2019a). Correction for wilson, the harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences of the United States of America*, 116(43), 21948. <https://doi.org/10.1073/pnas.1914128116>
- Wilson, D. J. (2019b). *Harmonicmeanp: Harmonic mean p-values and model averaging by mean maximum likelihood. R package version 3.0.* <https://CRAN.R-project.org/package=harmonicmeanp>
- Wilson, D. J. (2019c). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences of the United States of America*, 116(4), 1195–1200. <https://doi.org/10.1073/pnas.1814092116>