Biometrika (Year), **Volume**, Number, *p*. 1

Printed in Great Britain

Advance Access publication on Day Month Year

Supplemental Materials: Analytic P-value calculation for the higher criticism test in finite d problems

BY IAN J. BARNETT AND XIHONG LIN

Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.

ibarnett@hsph.harvard.edu xlin@hsph.harvard.edu

1. Q-Q PLOT OF THE HIGHER CRITICISM P-VALUES IN THE LUNG CANCER STUDY

A total of 14396 genes were tested genome-wide with the restriction that each gene must contain more than 1 genetic variants ($d \geq 2$ for each gene). Genes were also pruned of one of each pair of genetic variants that were very strongly correlated and essentially redundant (correlation greater than 0.95). The higher criticism for signal detection is traditionally used as a test of the global null hypothesis: none of the genes are associated with disease. For an example of an attempt to test the global null hypothesis in a genetics context, see Zuber et al. (2012). The higher criticism also can be used for signal identification as done in (**?**)

The top hits as found by the higher criticism in this lung cancer genome-wide association study are supported by the literature. The Q-Q plot shows that these top hits are more significant than we would expect under the null. Using the procedure proposed by Donoho & Jin (2008), the 15496 p-values are corrected for multiple comparisons by applying the higher criticism heirarchically to the the set of all p-values and using the supremum as a threshold for significance. This threshold controls for the false non-discovery rate. By this procedure the top 4 genes are selected as significant.The Q-Q plot appears in Fig. 1.

2. INACCURACY OF THE ASYMPTOTIC DISTRIBUTION OF THE HIGHER CRITICISM IN FINITE d SETTINGS

Asymptotically, one needs only take the supremum over a small subset of $t > 0$ (Donoho & Jin, 2004). However, in finite d settings it is necessary to take the supremum over the full $t > 0$ region. This corresponds to $\epsilon = 0$ and $\delta = 1$, and the asymptotic convergence to the gumbel distribution is even slower in this case than what is observed in Fig. 2. At the $\alpha = 0{\cdot}05$ significance level, the Type I error is $0{\cdot}03$ for $d = 10^4$ if calculated using the asymptotic distribution with $\epsilon = 0{\cdot}01$ and $\delta = 0{\cdot}40$. This inaccurate Type I error rate for even very large d demonstrates how this asymptotic result is not useful for even moderately large d settings.

REFERENCES

DONOHO, D. & JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **34**, 962–994.

DONOHO, D., & JIN, J. (2008). Higher criticism thresholding: optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences*, **105(39)**, 14790-14795.

ZUBER, V., SILVA, A., & STRIMMER, K. (2012). A novel algorithm for simultaneous SNP selection in high-dimensional genome-wide association studies. *BMC bioinformatics* **13**, 284.

[*Received Month* $Year$. *Revised Month* $Year$]

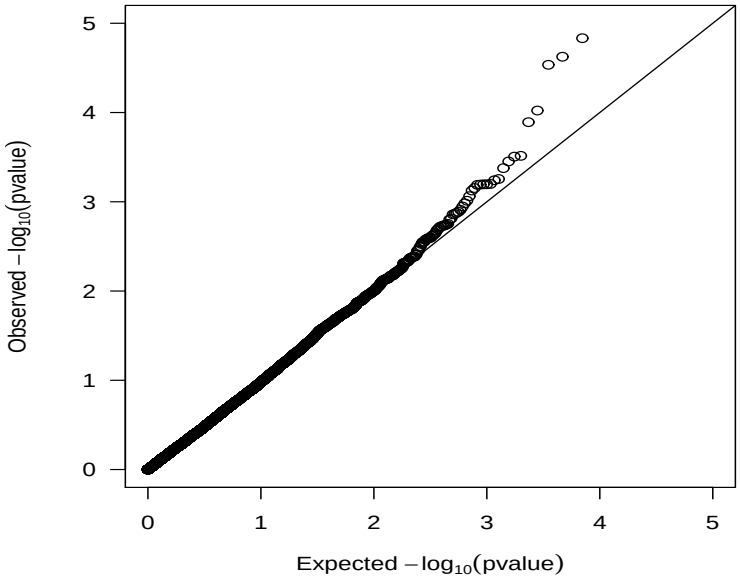

Fig. 1. The Q-Q plot for the gene-level p-values using the higher criticism for the lung cancer case-control genome-wide association study.

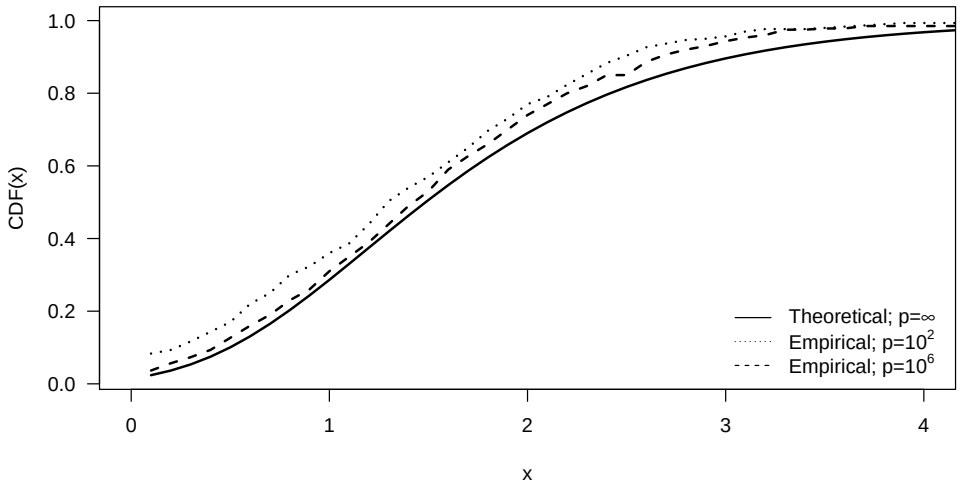

Fig. 2. Comparison of the asymptotic distribution and the empirical distribution of the higher criticism as a function of d. The supremum of the higher criticism test statistic is taken over the range $\Phi^{-1}(1-\delta/2) < t < \Phi^{-1}(1-\epsilon/2)$ where $\epsilon = 0{\cdot}01$ and $\delta = 0{\cdot}40$. Each empirical distribution is constructed using 500 samples of d independent standard normal random variables.