Polygenic score prediction intervals with cross-validation+

Frederick J. Boehm & Xiang Zhou

2022-07-03

The abundance of summary statistics from genome-wide association studies (GWASs) has enabled new methods for genetic predictions (polygenic scores, PGS) of quantitative, binary, and other complex traits in humans. PGS are calculated by summing the genomewide products of estimated effects and SNP genotypes. As such, a PGS is a statistic, a function of the observed data. Recent research has demonstrated large uncertainties associated with PGS (Ding et al. 2022). Yet, most current PGS methods provide only point estimates. In this work, we develop a general strategy for construction of well calibrated prediction intervals for PGS. Our approach is both computationally scaleable and efficient. We adapt a new method for prediction interval calculation called cross-validation+ (Barber et al. 2021). We simulated quantitative and binary traits to study prediction interval calibration under a variety of heritability and polygenicity settings. We then applied our method to 25 traits (16 continuous and nine binary) from the UK Biobank study.

References

Barber, Rina Foygel, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. 2021. "Predictive Inference with the Jackknife+." *The Annals of Statistics* 49 (1): 486–507.

Ding, Yi, Kangcheng Hou, Kathryn S Burch, Sandra Lapinska, Florian Privé, Bjarni Vilhjálmsson, Sriram Sankararaman, and Bogdan Pasaniuc. 2022. "Large Uncertainty in Individual Polygenic Risk Score Estimation Impacts PRS-Based Risk Stratification." *Nature Genetics* 54 (1): 30–39.