Title: Model selection bias in genome-wide genetic studies.

Presenter: Lei Sun

Associate Professor
Departments of Public Health Sciences and Statistics
University of Toronto

The primary concern motivating the research is the observed difficulty in replicating initial claims of gene discoveries in genetic studies. One main contributing factor is that the traditional estimates of gene-effect size are often grossly upward-biased due to selection of both significant results and competing genetic variants, leading to optimistic power and sample size calculations for the replication studies. I will discuss several model-free resampling-based estimators that were originally proposed by Sun and Bull (2005) and subsequently extended by Wu et al. (2005; 2006). The proposed method can be applied to the original dataset without the necessity of an additional independent sample, and the estimators were shown to substantially reduce the estimation bias. The performance is similar to that of the likelihood-based approach of Zollner and Pritchard (2007) (Faye et al. 2007). However, one caveat is that the variances of the proposed estimators in both cases are considerably higher than the original naive estimator, rendering highly variable estimates of sample size for replication studies, even if the root mean squared errors are lower. The problem seems to be closely related to the work of Leeb and Potscher (2006), which showed that one could not estimate the unconditional distribution of a post-model-selection estimator with reasonable accuracy even asymptotically. We introduce a Bayesian framework incorporating prior information to further reduce the bias and decrease the variance of the estimates. This is joint work with Professors Shelley Bull and Radu Craiu, and graduate students Longyang Wu and Laura Faye.

References

Faye L, Sun L, Bull SB (2007).
Reducing selection bias: efficiency and robustness of parametric and non-parametric effect estimation.
Abstract 2152 presented at the annual meeting of The American Society of Human Genetics, October 25, 2007, San Diego, California.
Available from http://www.ashg.org/genetics/ashg07s/index.shtml

Leeb H, Potscher B (2006). Can one estimate the conditional distribution of post-model-selection estimators? Annals of Statistics

34:2554-2591.

Sun L, Bull SB (2005). Reduction of selection bias in genome-wide genetic studies by resampling. Genetic Epidemiology 28:352-367.

Wu LY, Lee SSF, Shi HS, Sun L, Bull SB (2005). Resampling methods to reduce the selection bias in genetic effect estimation in genome-wide scans. BMC Genetics 6:S24.

Wu LY, Sun L, Bull SB (2006). Locus-specific heritability estimation via the bootstrap in linkage scans for quantitative trait loci. Human Heredity 62:84-96.

Zollner and Pritchard (2007). Overcoming the Winner's curse: estimating penetrance parameters from case-control data. American Journal of Human Genetic 80:605-615.