Departamento Académico de Estadística
Río Hondo # 1, Col. Progreso Tizapán,
Ciudad de México, C.P. 01080, Méixco

Seminario Aleatorio

Sesión 341

Design and Analysis of Two-phase Post-Genome-Wide-Association Studies

Dr. Osvaldo Espin García, Senior Biostatistician
Department of Biostatistics
Princess Margaret Cancer Centre
University Health Network

Abstract

Two-phase sampling design and analysis in post-genome-wide association studies (GWASs) can be a cost-effective strategy to infer association between sequencing data and a (quantitative) trait of interest. In phase 1, given a GWAS-identified genetic region, a subsample is selected using the outcome, Y, and an auxiliary covariate, Z, to inform the sampling. Phase 2 consists of collecting expensive targeted-sequence data, G, on the subsample alone, making G missing-by-design and reducing the sequencing costs. Lastly, inference on G is performed via semiparametric maximum likelihood by efficiently utilizing available data from phases 1 and 2. Two main approaches to select the phase 2 subsample are studied: heuristic and optimal designs. Heuristic designs are practical and easy-to-implement sampling strategies that rely on alternative phase 2 sample allocations based on the phase 1 strata distribution. Four heuristic designs are examined: probability proportional to size, balanced, extreme, and a combination of extreme-balanced. Optimal designs, on the other hand, aim to select the phase 2 subsample under a budget constraint. Specifically, these strategies minimize features of the variance-covariance matrix based on predefined optimality criteria found in previous literature and common to experimental design. The proposed strategies implement such minimization via (1) Lagrange multipliers and (2) a genetic algorithm. The proposed methods are evaluated through comprehensive simulation studies. In addition, an illustration of these approaches using the Northern Finland Birth Cohort 1966 is presented. We argue that two-phase studies can drastically reduce the costs of gathering sequencing data without substantial loss of power to detect genetic associations

Viernes 25 de octubre de 2019, 13:00 hrs.

Aula PF101, **Plantel Río Hondo**

El Seminario Aleatorio está destinado tanto a profesores como a estudiantes, por lo que el Departamento de Estadística agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.

En la red*: http://estadistica.itam.mx/es/seminario-aleatorio-de-estad%C3%ADstica*