Longitudinal First Hitting-Time Models with Extension to Neural Network

Mei-Ling Ting Lee

University of Maryland, College Park

ABSTRACT

Disease progression in a patient can be described mathematically as a stochastic process. The patient experiences a failure event when his/her disease progression first reaches a critical threshold level. This happening defines the failure event itself and the first hitting time (FHT) is the event time. First hitting-time based threshold regression (TR) models incorporate regression functions for parameters of the underlying stochastic process. The TR models are intuitive and do not require the proportional hazards assumption, therefore represent a realistic alternative to the Cox model. Recently the TR model has been extended to the Levy family with stationary independent increments and a cumulant generating function. Extension to neural network applications will also be discussed.

Keywords: Levy processes, non-proportional hazards, Wiener processes

Time-Dependent Pseudo R-Squared for Assessing Predictive Performance in Competing Risks Data

Zian Zhaung, Wen Su, Eric Kawaguchi, and **Gang Li***

Biostatistics and Computational Medicine, University of California at Los Angeles

ABSTRACT

Evaluating and validating the performance of prediction models is a fundamental task in statistics, machine learning, and their diverse applications. However, developing robust performance metrics for competing risks time-to-event data poses unique challenges. We first highlight how certain conventional predictive performance metrics for competing risks time-to-event data, such as the C-index, Brier Score, and time-dependent AUC, can yield unexpected results when comparing predictive performance between different prediction models. To address this research gap, we introduce a novel time-dependent pseudo R-squared measure to evaluate the predictive performance of a predictive cumulative incidence function over a restricted time domain under right-censored competing risks time-to-event data. Specifically, we first propose a population-level time-dependent pseudo R-squared measures for the competing risk event of interest and then define their corresponding sample versions based on right-censored competing risks time-to-event data. We investigate the asymptotic properties of the proposed measure and demonstrate its advantages over conventional metrics through comprehensive simulation studies and three real-data applications.

Keywords: Right censoring; Competing risks; Explained variance; Predictive performance

Assessing Transcriptomic Heterogeneity of Single-Cell RNASeq Data by Bulk-Level Gene Expression Data

Khong-Loon Tiong[1#], Dmytro Luzhbin[1#], **Chen-Hsiang Yeang[1*]**

[1]*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*

[#]*Co-first authors*

[*]*Corresponding author, chyeang@stat.sinica.edu.tw*

Abstract

Motivation

Single-cell RNA sequencing (sc-RNASeq) data illuminate transcriptomic heterogeneity but also possess a high level of noise, abundant missing entries and sometimes inadequate or no cell type annotations at all. Bulk-level gene expression data lack direct information of cell population composition but are more robust and complete and often better annotated. We propose a modeling framework to integrate bulk-level and single-cell RNASeq data to address the deficiencies and leverage the mutual strengths of each type of data and enable a more comprehensive inference of their transcriptomic heterogeneity. Contrary to the standard approaches of factorizing the bulk-level data with one algorithm and (for some methods) treating single-cell RNASeq data as references to decompose bulk-level data, we employed multiple deconvolution algorithms to factorize the bulk-level data, constructed the probabilistic graphical models of cell-level gene expressions from the decomposition outcomes, and compared the log-likelihood scores of these models in single-cell data. We term this framework *backward deconvolution* as inference operates from coarse-grained bulk-level data to fine-grained single-cell data. As the abundant missing entries in sc-RNASeq data have a significant effect on log-likelihood scores, we also developed a criterion for inclusion or exclusion of zero entries in log-likelihood score computation.

Results

We selected six deconvolution algorithms and validated backward deconvolution in four datasets. In the insilico mixtures of mouse sc-RNASeq data, the log-likelihood scores of the deconvolution algorithms were strongly anticorrelated with their errors of mixture coefficients and cell type specific gene expression signatures. In the true bulk-level mouse data, the sample mixture coefficients were unknown but the loglikelihood scores were strongly correlated with accuracy rates of inferred cell types. In datasets of breast cancer and low-grade gliomas (LGG), we compared the log-likelihood scores of three simple hypotheses about the gene expression patterns of the cell types underlying the tumor subtypes. The model that tumors of each subtype were dominated by one cell type persistently outperformed an alternative model that each cell

type had elevated expression in one gene group and tumors were mixtures of those cell types. The results indicate that backward deconvolution serves as a sensible model selection tool for deconvolution algorithms and facilitates discerning hypotheses about cell type compositions underlying heterogeneous specimens such as tumors.

Availability and implementation

We have implemented the backward deconvolution algorithm in R, and deposited the source codes and their description on github.com/chyeang/backward-deconvolution/.

Keywords: Single-cell RNASeq data, Deconvolution, Probabilistic graphical models, Heterogeneity

Back to Sessions List

A Unified Framework for Statistical Inference and Power Analysis of Single and Comparative F_β Scores

Chih-Yuan Hsu, Qi Liu, Yu Shyr

Department of Biostatistics, Vanderbilt University Medical Center, Nashville 37203, TN, USA

ABSTRACT

Machine learning and artificial intelligence are increasingly applied to medical diagnostics and clinical decision-making. To evaluate model performance, the F1 score and its generalized form, the F score, are widely used as they balance precision and sensitivity. However, rigorous statistical inference and power analysis for the F1 and F scores remain limited. In this study, we propose psF1, a unified and comprehensive framework for interval estimation, hypothesis testing, and power and sample size calculation for both single and comparative F1 and F scores. psF1 leverages exact probability distributions as well as normal approximations for large sample sizes to provide valid statistical inference and power analyses. Extensive simulations demonstrate the accuracy and robustness of psF1 across a range of sensitivity, precision, and sample size scenarios. We further showcase its practical utility through real-world biomedical classification tasks. This framework enables principled evaluation and comparison of classifiers using F1 and F scores with reliable uncertainty quantification and informed sample size planning. psF1 is freely available at http://github.com/cyhsuTN/psF1.

Keywords: F1 score; F score; Interval estimation; Hypothesis testing; Power and sample size calculation