Perspective
Volume 12:11, 2021

Journal of Biometrics and Biostatistics

ISSN: 2155-6180 Open Access

An Original Factual Technique for Demonstrating Covariate Impacts in Bisulfite Sequencing Determined Proportions of DNA Methylation

Siddiqui Fasil Akbar *

Department of Entomology, Pennsylvania State University, Pennsylvania, USA

Description

Parametric estimation of the Cumulative Incidence Function (CIF) is considered for competing risks data subject to interval censoring. Existing parametric models of the CIF for right censored competing risks data are adapted to the general case of interval censoring. Maximum likelihood estimators for the CIF are considered under the assumed models, extending earlier work on nonparametric estimation. A simple naive likelihood estimator is also considered that utilizes only part of the observed data. The naive estimator enables separate estimation of models for each cause, unlike full maximum likelihood in which all models are fit simultaneously. The naive likelihood is shown to be valid under mixed case interval censoring, but not under an independent inspection process model, in contrast with full maximum likelihood which is valid under both interval censoring models. In simulations, the naive estimator is shown to perform well and yield comparable efficiency to the full likelihood estimator in some settings. The methods are applied to data from a large, recent randomized clinical trial for the prevention of mother-to-child transmission of HIV.

This article targets the estimation of a time-dependent association measure for bivariate failure times, the Conditional Cause-Specific Hazards Ratio (CCSHR), which is a generalization of the Conditional Hazards Ratio (CHR) to accommodate competing risks data. We model the CCSHR as a parametric regression function of time and event causes and leave all other aspects of the joint distribution of the failure times unspecified. We develop a pseudo-likelihood estimation procedure for model fitting and inference and establish the asymptotic properties of the estimators. We assess the finite-sample properties of the proposed estimators against the estimators obtained from a moment-based estimating equation approach [1]. Data from the Cache County study on dementia are used to illustrate the proposed methodology. Wetake a semiparametric approach in fitting a linear transformation model to aright censored data when predictive variables are subject to measurement errors.

We construct consistent estimating equations when repeated measurements of a surrogate of the unobserved true predictor are available. The proposed approach applies under minimal assumptions on the distributions of the true covariate or the measurement errors. We derive the asymptotic properties of the estimator and illustrate the characteristics of the estimator in finite sample performance via simulation studies. We apply the method to analyze an AIDS clinical trial data set that motivated the work [2]. Estimation of the covariance structure for irregular sparse longitudinal data has been studied by many authors in recent years but typically using fully parametric specifications. In addition, when data are collected from several groups over time, it is known that assuming the same or completely different covariance matrices over groups can lead to loss of efficiency and/or bias. Nonparametric approaches have been proposed for estimating the covariance matrix for regular univariate longitudinal data by sharing information across the groups under study. For the irregular case, with longitudinal measurements that are bivariate or multivariate, modeling becomes more difficult. In this article, to model bivariate sparse longitudinal data from several groups, we propose a flexible covariance structure via a novel matrix stick- breaking process for the residual covariance structure and a Dirichlet process mixture of normals for the random effects [3]. Simulation studies are performed to investigate the effectiveness of the proposed approach over more traditional approaches.

Related Sampling Designs

Investigators commonly gather longitudinal data to assess changes in responses over time and to relate these changes to within-subject changes in predictors. With rare or expensive outcomes such as uncommon diseases and costly radiologic measurements, outcome-dependent, and more generally outcome-related, sampling plans can improve estimation efficiency and reduce cost.

***Corresponding author:** Siddiqui FA, Department of Entomology, Pennsylvania State University, 546 Ag Sciences & Industries Bldg, Pennsylvania, USA, E-mail: fas5130@psu.edu*

***Copyright**: © 2021 Siddiqui FA. This is an open-access article distributed under the terms of the creative commons attribution license which permits unrestricteduse, distribution and reproduction in any medium, provided the original author and source are credited.*

Received: Aug 02, 2021; **Accepted**: Aug 16, 2021; **Published**: Aug 23, 2021

Longitudinal follow up of subjects gathered in an initial outcome-related sample can then be used to study the trajectories of responses over time and to assess the association of changes in predictors within subjects with change in response. In this article, we develop two likelihood-based approaches for fitting Generalized Linear Mixed Models (GLMMs) to longitudinal data from a wide variety of outcome-related sampling designs [4]. The second approach is an adaptation of standard conditional likelihood methods and is limited to random intercept models with a canonical link. Data from a study of attention deficit hyperactivity disorder in children motivates the work and illustrates the findings. In order to make a Missing At Random (MAR) or ignorability assumption realistic, auxiliary covariates are often required. However, the auxiliary covariates are not desired in the model for inference.

Typical multiple imputation approaches do not assume that the imputation model marginalizes to the inference model. This has been termed "uncongenial". In order to make the two models congenial (or compatible), we would rather not assume a parametric model for the marginal distribution of the auxiliary covariates, but we typically do not have enough data to estimate the joint distribution well non-parametrically.

In addition, when the imputation model uses a non-linear link function (e.g. the logistic link for a binary response), the marginalization over the auxiliary covariates to derive the inference model typically results in a difficult to interpret form for the effect of covariates [5]. In this article, we propose a fully Bayesian approach to ensure that the models are compatible for incomplete longitudinal data by embedding an interpretable inference model within an imputation model and that also addresses the two complications described above. We evaluate the approach via simulations and implement it on a recent clinical trial.

References

1. Bradford WZ, Cohen AH and Leff JA. "Selection of clinical meaningful primary end-points in phase 3 clinical trials in idiopathic pulmonary fibrosis (letter)". Am J Respir Crit Care Med 187 (2013): 1269-1270.
2. Joseph G. Ibrahim and Ming-Hui Chen. "Power Prior Distributions for Regression Models". Statistical Science 15 (2000): 46-60.
3. Joseph G. Ibrahim, Ming-Hui Chen, Yeongjin Gwon, and Fang Chenc. "The power prior: theory and applications". Statistics in Medicine 34 (2015): 3724-3749.
4. Joseph G. Ibrahim, Ming-Hui Chen and Debajyoti Sinha. "On Optimality Properties of the Power Prior". J Am Stat Assoc 98 (2003): 204-213
5. Fan J, Lv J (2008) Sure independence screening for ultrahigh dimensional feature space. J R Statist Soc B 70: 849-911.

How to cite this article: Siddiqui, Fasil Akbar (2021). An original factual technique for demonstrating covariate impacts in bisulfite sequencing determined proportions of DNA methylation. J Biom Biostat 12:11.