rstats-gsoc / **gsoc2015**

Watch ▾ 28   ★ Unstar 29   ⑂ Fork 9

Multivariate Survival analysis

gorfinem edited this page 4 days ago · 1 revision

Edit   New Page

Survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach

Summary: Our goal is building an R package for implementing Gorfine et al.'s (2006) approach for estimating the parameters of frailty models using clustered-survival data for various frailty distributions. Frailty models are highly popular for analyzing clustered time-to-event data. The estimation technique of Gorfine et al. (2006) is applicable to any frailty distribution with finite moments and has a number of desirable features. This includes, easy computation and implementation, a non-iterative procedure for estimating the nonparametric cumulative hazard function, estimators that are consistent and asymptotically normal, and direct consistent covariance estimation. The primary aim of this package is to bridge the knowledge gap between statistical theory and practitioners such as epidemiologists.

Description: As an example, consider a disease outcome of a certain chronic disease and the outcome is age at onset. The clusters typically are families, clinical centers, or schools. In family studies, for example, genetic and environmental background that the family members share often leads to correlation among the family members. Therefore, the standard survival tools that assume independence among observations would provide misleading inference. The aim of the package is to provide a friendly tool for practitioners, for correctly analyzing clustered-survival data.

The frailty distributions that we would like to include in this package are: gamma, positive stable, log-normal, inverse Gaussian, and power variance function (PVN).

The package should provide the estimators of: the regression coefficient vector, the baseline hazard function, and the frailty parameter, along with their standard errors.

The output of the package should be similar to the common style of other R packages for survival analysis (e.g. the package `survival`).

Related work: The R packages `survival` and `frailtypack` fits several classical frailty models using a penalized partial likelihood approach. These packages are only limited to gamma, log-normal and t frailty distributions. Moreover, the asymptotic properties of their estimators are not yet fully established. In contrast, our estimation approach is applicable for any frailty distribution with finite moment that would be more general than either existing package provides. Additionally, we showed that our estimators are consistent and asymptotically normal, and have a direct consistent covariance estimator.

The successful project would include:

- Development of an R package that contains computationally efficient algorithm for estimating the parameters of the frailty models and their standard errors, under the frailty distributions listed above.
- Creation of a Journal of Statistical Software article, which enables practitioners to

Pages 27

Find a Page…

Home

Admin_Application

Animint

Convex optimization modeling language

COPY THIS Proposal Template

Covariance Matrix Estimators

Cpp Version of ARPACK and Integration with (Rcpp)Armadillo

factorAnalytics

GEO analysis with Shiny

ggenealogy

ggplotly

InFER

markovchain

matrixStats

Multivariate Survival analysis

Show 12 more pages…

Clone this wiki locally

https://github.com/rstats-g

understand and use the procedure for analyzing clustered-survival data.

Skills required:

- Experience with R package development.
- Experience with numerical integrations.
- Experience with developing computationally efficient algorithms and working with R's profiling tools.
- A background in doctoral-level statistics.

Test:

The successful applicant should be able to demonstrate a basic understanding in multivariate survival analysis by doing the following:

1. Download and install the package `survival` from CRAN.
2. Create simulated datasets of clustered data, as described in Gorfine et al. (2006, Section 4), and analyze each dataset using `survival` and two estimation methods: (a) ignore the intra-cluster dependence (use only `coxph` function); (b) do not ignore the intra-cluster dependence and estimate the parameters using the function `frailty` (in the `survival` package).
3. Summarize the results in a clear table and state your conclusions.
4. Repeat on Steps 2-3 but with 2 covariates in the model (Z1, Z2) instead of only one (Z).
5. Repeat on Steps 2-3 but with the log-normal frailty distribution instead of gamma.
6. Develop a suggested plan for the project, including a timeline for the development of the code, its documentation, and the testing stage.

Mentors:

Malka Gorfine, (@). Li Hsu, (@).

References:

Gorfine, M., Zucker, D. M. and Hsu, L. Prospective survival analysis with a general semiparametric shared frailty model - a pseudo full likelihood approach. Biometrika, 93: 735--741, 2006. journal

Zucker, D. M., Gorfine, M. and Hsu, L. Pseudo full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: asymptotic theory. Journal of Statistical Planning and Inference, 138: 1998--2016, 2008. journal