DOWNLOAD PDF NONPARAMETRIC AND SEMIPARAMETRIC MODELS

Chapter 1 : Semiparametric model - Wikipedia

The book does naturally split into two parts: Nonparametric models (histogram, kernel density estimation, nonparametric regression) and semiparametric models (generalized regression, single index models, generalized partial linear models, additive and generalized additive models).

Abstract Interval censoring arises frequently in clinical, epidemiological, financial and sociological studies, where the event or failure of interest is known only to occur within an interval induced by periodic monitoring. We formulate the effects of potentially time-dependent covariates on the interval-censored failure time through a broad class of semiparametric transformation models that encompasses proportional hazards and proportional odds models. We consider nonparametric maximum likelihood estimation for this class of models with an arbitrary number of monitoring times for each subject. We devise an EM-type algorithm that converges stably, even in the presence of time-dependent covariates, and show that the estimators for the regression parameters are consistent, asymptotically normal, and asymptotically efficient with an easily estimated covariance matrix. Current-status data, EM algorithm, Interval censoring, Linear transformation model, Nonparametric likelihood, Proportional hazards, Proportional odds, Semiparametric efficiency, Time-dependent covariate 1. Introduction Interval-censored data arise when the event or failure of interest is known only to occur within a time interval. Such data are commonly encountered in disease research, where the ascertainment of an asymptomatic event is costly or invasive and so can take place only at a small number of monitoring times. Likewise, biopsies are performed on patients at clinic visits to determine the occurrence or recurrence of cancer. There are several types of interval-censored data. The simplest and most studied type is called case-1 or current-status data, which involves only one monitoring time per subject and is routinely found in cross-sectional studies. The fact that the failure time is never observed exactly poses theoretical and computational challenges in semiparametric regression analysis of such data. The estimators are obtained by the iterative convex minorant algorithm, which becomes unstable for large datasets. For the proportional odds model with case-1 and case-2 data, Rabinowitz et al. None of the existing work accommodates time-dependent covariates or can handle case- or mixed-case interval censoring. In this paper we consider interval censoring in the most general form, that is, mixed-case data. We study nonparametric maximum likelihood estimation for a broad class of transformation models that allows time-dependent covariates and includes the proportional hazards and proportional odds models as special cases. We develop an EM-type algorithm, which is demonstrated to perform satisfactorily in a wide variety of settings, even with time-dependent covariates. The theoretical development requires careful treatment of the time trajectories of covariate processes and the joint distribution for an arbitrary sequence of monitoring times. Transformation models and likelihood construction Let denote the failure time, and let denote a -vector of potentially time-dependent covariates. Under the semiparametric transformation model, the cumulative hazard function for conditional on.

Chapter 2 : Nonparametric statistics - Wikipedia

The semiparametric modeling technique compromises the two aims, flexibility and simplicity of statistical procedures, by introducing partial parametric components. These components allow to match structural conditions like e.g. linearity in some variables and may be used to model the influence of discrete variables.

Definitions[edit] The statistician Larry Wasserman has said that "it is difficult to give a precise definition of nonparametric inference". The first meaning of nonparametric covers techniques that do not rely on data belonging to any particular distribution. These include, among others: As such it is the opposite of parametric statistics. It includes nonparametric descriptive statistics , statistical models , inference and statistical tests. Order statistics , which are based on the ranks of observations, is one example of such statistics and these play a central role in many nonparametric approaches. For example, the hypothesis a that a normal distribution has a specified mean and variance is statistical; so is the hypothesis b that it has a given mean but unspecified variance; so is the hypothesis c that a distribution is of normal form with both mean and variance unspecified; finally, so is the hypothesis d that two unspecified continuous distributions are identical. It will have been noticed that in the examples a and b the distribution underlying the observations was taken to be of a certain form the normal and the hypothesis was concerned entirely with the value of one or both of its parameters. Such a hypothesis, for obvious reasons, is called parametric. Hypothesis c was of a different nature, as no parameter values are specified in the statement of the hypothesis; we might reasonably call such a hypothesis non-parametric. Hypothesis d is also non-parametric but, in addition, it does not even specify the underlying form of the distribution and may now be reasonably termed distribution-free. Notwithstanding these distinctions, the statistical literature now commonly applies the label "non-parametric" to test procedures that we have just termed "distribution-free", thereby losing a useful classification. The second meaning of non-parametric covers techniques that do not assume that the structure of a model is fixed. Typically, the model grows in size to accommodate the complexity of the data. In these techniques, individual variables are typically assumed to belong to parametric distributions, and assumptions about the types of connections among variables are also made. These techniques include, among others: Applications and purpose[edit] Non-parametric methods are widely used for studying populations that take on a ranked order such as movie reviews receiving one to four stars. The use of non-parametric methods may be necessary when data have a ranking but no clear numerical interpretation, such as when assessing preferences. In terms of levels of measurement , non-parametric methods result in ordinal data. As non-parametric methods make fewer assumptions, their applicability is much wider than the corresponding parametric methods. In particular, they may be applied in situations where less is known about the application in question. Also, due to the reliance on fewer assumptions, non-parametric methods are more robust. Another justification for the use of non-parametric methods is simplicity. In certain cases, even when the use of parametric methods is justified, non-parametric methods may be easier to use. Due both to this simplicity and to their greater robustness, non-parametric methods are seen by some statisticians as leaving less room for improper use and misunderstanding. The wider applicability and increased robustness of non-parametric tests comes at a cost: In other words, a larger sample size can be required to draw conclusions with the same degree of confidence. Non-parametric models[edit] Non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. The term non-parametric is not meant to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. A histogram is a simple nonparametric estimate of a probability distribution. Kernel density estimation provides better estimates of the density than histograms. Nonparametric regression and semiparametric regression methods have been developed based on kernels , splines , and wavelets. Data envelopment analysis provides efficiency coefficients similar to those obtained by multivariate analysis without any distributional assumption. KNNs classify the unseen instance based on the K points in the training set which are nearest to it. A support vector machine with a Gaussian kernel is a nonparametric large-margin classifier. Methods[edit] Non-parametric or distribution-free inferential statistical methods are mathematical

procedures for statistical hypothesis testing which, unlike parametric statistics , make no assumptions about the probability distributions of the variables being assessed. The most frequently used tests include Andersonâ€"Darling test:

Chapter 3 : Parametric, Semiparametric, and Nonparametric Models â€" thirdorderscientist

Wolfgang Hardle, Marlene MÂ¨ uller, StefanÂ¨ Sperlich, Axel Werwatz Nonparametric and Semiparametric Models An Introduction February 6, Springer.

Chapter 4 : Maximum likelihood estimation for semiparametric transformation models with interval-censore

Semiparametric models lie in the grey area between parametric and non-parametric models. To specify a semiparametric model, you must specify both a finite-dimensional vector of parameters, and an infinite-dimensional function.

Chapter 5 : Marlene MÃ¼ller - R material for "Nonparametric and Semiparametric Models"

Introduction.- Histogram.- Nonparametric Density Estimation.- Nonparametric Regression.- Semiparametric and Generalized Regression Models.- Single Index Models.- Generalized Partial Linear Models.

Chapter 6 : Semiparametric regression - Wikipedia

Note: The package has been renamed to gplm (the former name was KernGPLM).Until now, the gplm package is not required for the R scripts above, however this is to be changed in the future (hopefully:)).