Conference on "Quantitative methods in statistics, biostatistics and actuarial sciences"

Poster - Abstract

Conditional estimation for dependent functional data

AUTHOR

BATTEY Heather (University of Bristol)
Joint work with A. Sancetta

ABSTRACT

Although traditionally statisticians regard data as taking values in a finite dimensional space, there are many situations in which observations are more naturally viewed as discretised realisations of random functions. This view has led to the relatively new field of functional data analysis (FDA). In this poster I will present a new approach to FDA that relaxes the i.i.d. assumption often assumed, in favour of the more general Markov property. I will present a nonparametric procedure that exploits the time dependence of past functional data to predict the future. I will discuss the theoretical properties of this estimator, and will illustrate the finite sample performance by way of a Monte Carlo study.

Relationship between antibody levels and clinical protection from influenza H3N2: A study based on 4 phase III trials

AUTHOR

BENOIT Anne (Université catholique de Louvain)
Joint work with W. Dewé, C. Legrand, A. Sancetta

ABSTRACT

In the context of the clinical development of influenza vaccines, the Haemagglutinin Inhibition (HI) titers indicating the amount of antibodies against a specific influenza strain are used as markers for efficacy in the earlier phases (I and II). However, the association between HI titres and infection risk has yet to be fully characterized.

To do so, we analyzed immunogenicity and efficacy data pooled from 4 phase III trials. More specifically, we studied the link between the HI titers and the efficacy endpoint (influenza infection) for the type of influenza virus having the highest number of cases. We also estimated the impact of various covariates, among which are age and strain circulation levels as a marker for virus exposure.

Logistic regressions with covariates selection were used to characterize the HI titers – infection risk relationship. Further, ROC curves were built in order to estimate cutoff points for the HI titres responses linked to a desired protection level.

Historical efficacy endpoints for the ratio between pre and post-vaccination titers seemed to be confirmed as a 4-fold increase in the specific strain titer was estimated as inducing a 50% decrease in the risk of infection. While age and circulation levels appeared to affect the baseline risk of infection, no effect of those covariates could be shown on the slope of the titers - infection risk curve.

In conclusion, for the type of influenza virus of interest based on the data available, older age and higher circulation of the virus seem to influence the risk of infection in absence of strain specific antibodies. Per contra, the benefit of vaccination, in term of infection risk reduction, doesn't appear to vary in the studied populations. While our estimated relative gain conferred by vaccination is consistent with the sero-conversion definition used in clinical trials, HI titers thresholds should be adapted according to age and expected virus exposure intensity in order to achieve a fixed desired protection level.

Stochastic modeling of gas and electricity derivatives

AUTHOR

BABAJAN George (Université catholique de Louvain)
Joint work with P. Devolder

ABSTRACT

Energy markets are known to be very volatile compared to other markets and the need of financial derivatives to hedge risk exposure becomes more important with time as the markets mature and become more liquid. One of the aims of our work is to describe these markets with a special emphasis on their financial derivatives and the stochastic models that capture market prices and other stylized facts of their price dynamics. Contrary to equity, the gas and power price dynamics have several unique features that have to be matched by using jump processes as noise factors.

Afirst design of a risk-neutral (market based) model is presented. Derivatives are priced according to Monte-Carlo simulations and Fourier-based techniques using the analytical tractability of the model. Finally, a calibration case-study on electricity option data is used to illustrate the applicability of the model.

A second modelling approach focuses on the spot price behaviour of gas and its stylized facts. Mean reverting Ornstein-Uhlenbeck jump processes are used as a basic modelling tool and its implications on the implied volatility curve is illustrated using a case-study on the UK market.

The stochastic processes involved in the preceeding models are of increasing complexity. One can start by using the \simple» Brownian motion as basic noise and add a mean-reverting effect to model commodities instead of equity. A third layer would be changing the way time evolves by allowing a stochastic clock that would change the nature of the resulting mean-reverting processes. And by adding together several of such processes one gets very exible models already documented in literature. Some new ideas will be discussed during the poster session where a final layer will be add to the modeling procedure.

Finally, all these modelling approaches are shown to have the potential to be used for electricity contingent claim pricing. Power derivatives can be priced and hedged using derivatives of gas and possibly other fuels.

Estimation of the memory parameter of non-linear time series

AUTHOR

CLAUSEL Marianne (Université de Grenoble)
Joint work with F. Roueff, C. Tudor, M. S. Taqqu

ABSTRACT

Let $X = (X_k)$ be a real valued process not necessary stationary. The first-order difference is $[\Delta X]_k = X_k - X_{k-1}$ and $\Delta^K X$ is defined recursively. The time series X is said to have memory parameter $d \in \mathbb{R}$ if for any integer $K > d - 1/2$ the process $[\Delta^K X]$ is weakly stationary with spectral density function $f\Delta^K X$

$$f\Delta^K X = |1 - e^{-i\lambda}|^{-2d} \, f^*_{\Delta^K X}(\lambda)$$

where $f^*_{\Delta^K X}$ is a spectral density with null memory, that is smooth and strictly positive at $\lambda = 0$.

Several methods of estimation of the long–memory parameter have been proposed. Wavelet analysis is one of the most popular. It has proved to be efficient in the Gaussian and linear case. A quite natural question is then to wonder at what point this approach can be extended to non–Gaussian and non–linear cases. Here we consider a long memory time series of the form $G(X)$ where $G \in L^2(\mathbb{R}, e^{-x^2/2}\,dx)$ and X is a Gaussian series. Using Wiener chaos expansion of the process, we both study the behavior of the wavelets coefficients at large scales and this of the estimators of their empirical variance. We can then deduce the behavior of the estimators of the long memory parameter in a semiparametric framework. We proved that several phenomena can occur depending both on the Hermite rank of function G and the coefficients of function G in Hermite polynomial expansion. We then focus on two large classes of examples where the asymptotic behavior can be quite similar to the Gaussian case or completely different.

Estimation of a density using real and artificial data

AUTHOR

FELBER Tina (Technical University Darmstadt)
Joint work with L. Devroye, M. Kohler

ABSTRACT

Let $X, X_1, X_2, \ldots$ be independent and identically distributed R^d-valued random variables and let $m : R^d \to R$ be a measurable function such that a density f of $Y = m(X)$ exists. Given a sample of the distribution of (X, Y) and additional independent observations of X we are interested in estimating f. We apply a regression estimate to the sample of (X, Y) and use this estimate to generate additional artificial observations of Y. Using these artificial observations together with the real observations of Y we construct a density estimate of f by using a convex combination of two kernel density estimates. It is shown that if the bandwidths satisfy the usual conditions and if in addition the supremum norm error of the regression estimate converges almost surely faster towards zero than the bandwidth of the kernel density estimate applied to the artificial data, then the convex combination of the two density estimates is L_1-consistent. The performance of the estimate for finite sample size is illustrated by simulated data, and the usefulness of the procedure is demonstrated by applying it to a density estimation problem in a simulation model.

Key words and phrases:

Density estimation, L_1-error, nonparametric regression, consistency.

Estimating inequality-based measures of kurtosis, with actuarial and financial applications in view

AUTHOR

FIORI Anna Maria (Technical University Darmstadt)
Joint work with D. Beltrami

ABSTRACT

Although the standard fourth moment coefficient is routinely computed as "the kurtosis" of a distribution, the measure is not easily interpreted and has been a subject of considerable debate in statistical literature (e.g. Fiori and Zenga [2]). The financial community has recently joined in the debate (Kim and White [4], Bonato [1]) calling for more robust estimators of kurtosis in distributions of stock market returns.

In this work we consider alternative measures of right and left kurtosis which arise from a recent characterization of kurtosis as inequality at either side of the median (Zenga [5], Fiori [3]). Based on Gini indices, the new measures apply to both symmetric and asymmetric distributions, their interpretation is clear and they are consistent with common risk perceptions of investors and risk managers.
For purposes of statistical inference, we construct empirical estimators of those right and left kurtosis measures by ratios of L-statistics and discuss their asymptotic properties. Strong consistency and asymptotic normality are proved under broad based assumptions on the underlying distribution, and an explicit formula for the asymptotic variance is presented. While the sampling variance of the conventional kurtosis coefficient is related to the population moment of order eight, the new measures can be consistently estimated under the milder requirement that second moments are finite. In addition to an explicit formula for the asymptotic variance, we estimate the latter by bootstrap techniques and design extensive Monte Carlo simulations to compare the conventional kurtosis measure with its right and left counterparts in finite samples. These results may be useful in actuarial and financial contexts, in which the existence of higher moments is still an open question. An empirical application shows that the proposed measures are likely to provide both a more reliable and a more sophisticated picture of the kurtosis risk embedded in conditional distributions of stock market returns.

1. M. Bonato. Robust estimation of skewness and kurtosis in distributions with infinite higher moments. Finance Research Letters, 2011 (in press).
2. A.M. Fiori and M. Zenga. Karl Pearson and the origin of kurtosis. International Statistical Review, 77, 2, 40-50, 2009.
3. A.M. Fiori. Measuring kurtosis by right and left inequality orders. Communications in Statistics – Theory and Methods, 37, 17, 2665-2680, 2008.
4. T.H. Kim and H. White. On more robust estimation of skewness and kurtosis. Finance Research Letters, 1, 56-73, 2004.
5. M. Zenga. Kurtosis. In: S. Kotz, C.B. Read, N. Balakrishnan and B. Vidakovic (Eds.), Encyclopedia of Statistical Sciences (Second Edition). New York: John Wiley and Sons, 2006.

Combining thresholding rules: a new way to improve the performance of wavelet estimators

AUTHOR

FREYERMUTH Jean-Marc (Katholieke Universiteit Leuven)
Joint work with F. Autin, R. Von Sachs

ABSTRACT

We address the situation where we cannot differentiate wavelet-based threshold estimators because their sets of *well-estimated* functions (maxisets) are not nested. As a generic solution, we propose in (3) to proceed via a combination of these estimators in order to achieve new estimators which perform better in the sense that the involved maxisets contain the union of the previous ones. As an example, we propose to combine vertical- and horizontal-block thresholding estimators that are already known to perform well (see e.g., (1, 2)). We discuss the limitations of our method, and we check our theoretical results through specific numerical experiments.

Fixed design regression estimation based on experimental and articially generated data

AUTHOR

FURER Dmytro (Technical University Darmstadt)
Joint work with M. Kohler

ABSTRACT

Here we study least squares estimates based on experimental and artificially generated data. The articially generated data comes from already undertaken estimates on the basis of similar experiments. It is investigated unter which condition the rate of convergence of least squares estimates applied to this data is better then the rate of convergence of least squares estimates applied to the experimental data.

Estimation of the optimal design of a nonlinear parametric regression problem via Monte Carlo experiments

AUTHOR

HERTEL Ida (Technical University of Damstadt)
Joint work with M. Kohler

ABSTRACT

A Monte Carlo method for estimation of the optimal design of a nonlinear parametric regression problem is presented. The basic idea is to produce via Monte Carlo values of the error of a parametric regression estimate for randomly chosen designs and randomly chosen parameters and to use nonparametric regression to estimate from this data the design for which the maximal error with respect to all possible parameter values is minimal. A theoretical result concerning consistency of this estimate of the optimal design is presented and the method is used to find an optimal design for an experimental fatigue test.

KEYWORDS:

Optimal design, nonparametric regression, consistency

Bandwidth selection for kernel density estimation with doubly truncated data

AUTHOR

MOREIRA Carla (University of Vigo)
Joint work with I. Van Keilegom

ABSTRACT

In this work we introduce and compare several bandwidth selection procedures for kernel density estimation of a random variable that is sampled under random double truncation. The work is motivated by the fact that this type of incomplete data is often encountered in studies in astronomy and medicine. The bandwidth selection procedures we study are appropriate modifications of the normal reference rule, the least squares cross-validation procedure, two types of plug-in procedures, and a bootstrap based method. The methods are first shown to work from a theoretical point of view. A simulation study is then carried out to assess the finite sample behavior of these five bandwidth selectors. We also illustrate the use of the various practical bandwidth selectors by means of data regarding the luminosity of quasars in astronomy.

KEYWORDS:

Bandwidth selection; bootstrap; cross-validation; double truncation; kernel density estimation; normal reference rule; plug-in.

Dynamic pseudo-observations: a robust approach to dynamic prediction in competing risks

AUTHOR

NICOLAIE Mioara Alina (Leiden University Medical Center)
Joint work with H. C. Van Houwelingen, H. Putter

ABSTRACT

Prediction models are used as a basis for treatment decisions and to communicate prognosis to patients. Traditionally, these were designed to render prediction probabilities of events of interest over time accounting only for baseline information. However, in clinical management it is equally relevant to approach prediction dynamically, that is to update prediction at later times in the follow-up, based on new clinical information of the patient.

In this paper, we propose a new approach to the problem of dynamic prediction of survival data in the presence of competing risks as an extension of the landmark model for ordinary survival data of van Houwelingen (2007). The key feature of our method is the introduction of dynamic pseudo-observations constructed from the prediction probabilities at different landmark prediction times. They specifically address the issue of estimating covariate effects directly on the cumulative incidence scale in competing risks. A flexible generalized linear model based on these dynamic pseudo-observations and a GEE approach to estimate the baseline and covariate effects will result in the desired dynamic predictions and robust standard errors.

Our approach has a number of attractive features. It naturally selects only the event history needed for prediction, comprised in the landmark data set, and, on the other hand, it focuses directly on the prediction probabilities of interest, avoiding in this way complex modeling of cause-specific hazards or subdistribution hazards. As a result, it is robust against departures from these omnibus models. From a computational point of view an advantage of our approach is that it can be fitted with existing statistical software and that a variety of link functions and regression models can be considered, once the dynamic pseudo-observations have been estimated.

We illustrate our approach on a real data set of chronic myeloid leukemia patients after bone marrow transplantation. We build a dynamic prognostic tool for prediction of two competing events, relapse and non-relapse mortality, based on prognostic factors at baseline and an intermediate clinical event, acute GvHD. We specifically show how time-dependent covariates which exhibit time-varying effects on the cumulative incidence scale can flexibly be incorporated in the modeling.

Asymptotics versus bootstrapping in errors-in-variables models with dependent errors

AUTHOR

PEŠTA Michal (Charles University in Prague)

ABSTRACT

An intensive research in actuarial and financial science has sprung up for methods to handle measurement errors or disturbances in input and output data simultaneously. Errors-in-variables (EIV) serve as a regression modeling technique, where both dependent and independent variables are considered to be measured with errors.

An EIV regression model with dependent errors is considered and a total least squares (TLS) estimate is constructed. Its consistency and asymptotic normality for weak dependent observations (α- and φ-mixing) are proved [2]. TLS estimate is highly nonlinear and, moreover, the asymptotic variance depends on unknown quantities, which cannot be estimated. Because of this, many statistical procedures for constructing confidence intervals and testing hypotheses cannot be applied. One possible solution to this dilemma is bootstrapping [3]. Justification for use of the moving block bootstrap (MBB, [1]) technique is given.

The results are illustrated through a simulation study. An application of this approach to real data is presented.

Individual loss reserving with the multivariate skew normal model

AUTHOR

MATHIEU Pigeon (Université catholique de Louvain)
Joint work with K. Antonio, M. Denuit

ABSTRACT

In general insurance, the evaluation of future cash ows and solvency capital has become increasingly important. To assist in this process, the present paper proposes an individual discrete-time loss reserving model describing the occurrence, the reporting delay, the time

to the first payment, and the cash ows associated with the settlement process of each individual claim. The approach uses development factors similar to those of the standard chain-ladder method. These are parametrically modeled by the Multivariate Skew Normal distribution. Empirical analyses using a realistic portfolio and out-of-sample prediction tests demonstrate the relevance of the model proposed.

Parfm: parametric frailty models in R

AUTHOR

ROTOLO Federico (Université catholique de Louvain)
Joint work with M. Munda, F. Rotoloy, C. Legrand

ABSTRACT

Frailty models are getting more and more popular to account for overdispersion and/or clustering in survival data. When the form of the baseline hazard is somehow known in advance, the parametric estimation approach can be used advantageously. Nonetheless, there is no unified widely available software that deals with the parametric frailty model. The new parfm package remedies that lack by providing a wide range of parametric frailty models in R. The Gamma, Inverse Gaussian, and Positive Stable frailty distributions can be specified, together with five different baseline hazards. Parameter estimation is done by maximising the marginal log-likelihood, with right-censored and possibly left-truncated data.

KEYWORDS

Parametric frailty models, survival analysis, gamma, positive stable, inverse gaussian, weibull, exponential, gompertz, loglogistic, lognormal, R, parfm.

EGARCH models with fat tails, skewness and leverage

AUTHOR

SUCARRAT Genaro (Cambridge University)
Joint work with A. Harvey

ABSTRACT

An EGARCH model in which the conditional distribution is heavy tailed and skewed is proposed. The properties of the model, including unconditional moments, autocorrelations and the asymptotic distribution of the maximum likelihood estimator, are obtained. Evidence for skewness in conditional t-distribution is found for a range of returns series and the model is shown to give a better fit than the corresponding skewed-t GARCH model.

KEYWORDS

General error distribution; heteroskedasticity; leverage; score; Student's t, two components

Copula-based dependence structure tests

AUTHOR

SZNAJDER Dominik (Katholieke Universiteit Leuven)
Joint work with I. Gijbels

ABSTRACT

This poster presents recent results on testing for dependence structures by means of copula functions. We consider several general classes of dependence structures, namely quadrant dependence, tail monotonicity and stochastic monotonicity. They are of particular interest in empirical studies in finance, insurance and econometrics.

The discussed testing problems are important since a prior knowledge of a specific dependence structure is a qualitative restriction that should be taken into account in further statistical analysis, e.g., when choosing an appropriate copula function to model the dependence structure.

The considered dependence structures can be seen as the features of the underlying copula of the data generating random vector. Thus, the proposed test statistics are built as functionals of the empirical copula estimator. The statistical inference is based on resampling from a smooth constrained nonparametric copula estimator. Such an estimator is obtained by local polynomial smoothing of the initial constrained estimator and by transforming its partial derivatives by rearrangement techniques.

The proposed methodology is generic and exible and can be applied to other dependence concepts, which can be expressed as shape constraints on the copula function.

We apply the proposed testing procedures to several real data examples available in literature.

Pension valuation and solvency

AUTHOR

TASSA Habiba (Université catholique de Louvain)
Joint work with P. Devolder

ABSTRACT

Risk measurement as applicable for insurers or banks can also be considered for pension fund liabilities.

As the premiums collected from policyholders, are invested in risky or risk-less assets, the idea is to study, in parallel, the evolution of contributions and the level of liabilities by using various stochastic models in continuous time in order to estimate solvency capital for some important risks faced by pension funds: market risk and inflation risk (with correlation together).

In the case of defined benefit pension Scheme, some results will be presented by using IAS norms, based on the Projected Unit Credit cost method including a risk measure approach, to compute contributions and actuarial liabilities.

The models are developed first in a geometric Brownian motion environment and afterwards, using a Lévy process for the asset.