
**Statistics
Statistique**
(Org: **Yannick Baraud** (Nice) and/et **Boris Levit** (Queen's))

SYLVAIN ARLOT, Université Paris-Sud 11, Orsay F-91405, France

V-fold cross-validation improved: V-fold penalization

We investigate the efficiency of V -fold cross-validation (VFCV) for model selection from the non-asymptotic viewpoint, and suggest an improvement on it, which we call “ V -fold penalization”.

First, considering a particular (though simple) regression problem, we will show that VFCV with a bounded V is suboptimal for model selection. The main reason for this is that VFCV “overpenalizes” all the more that V is large. Hence, asymptotic optimality requires V to go to infinity. However, when the signal-to-noise ratio is low, it appears that overpenalizing is necessary, so that the optimal V is not always the larger one, despite of the variability issue. This is confirmed by some simulated data.

In order to improve on the prediction performance of VFCV, we propose a new model selection procedure, called “ V -fold penalization” (penVF). It is a V -fold subsampling version of Efron’s bootstrap penalties, so that it has the same computational cost as VFCV, while being more flexible. In a heteroscedastic regression framework, assuming the models to have a particular structure, penVF is proven to satisfy a non-asymptotic oracle inequality with a leading constant almost one. In particular, this implies adaptivity to the smoothness of the regression function, even with a highly heteroscedastic noise. Moreover, it is easy to overpenalize with penVF, independently from the V parameter. As shown by a simulation study, this results in a significant improvement on VFCV in several non-asymptotic situations.

PROF. MIKLÓS CSÖRGŐ, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6

Long-range dependent vs. i.i.d.r.v.’s based Bahadur–Kiefer, Vervaat and Vervaat error processes

This talk will review recent developments on long-range dependent sequences based processes as in M. Csörgő, B. Szyszkowicz and L. Wang, *Strong invariance principles for sequential Bahadur–Kiefer and Vervaat error processes for long-range dependent sequences*, Ann. Statist. **34**(2006), 1013–1044; *ibid.*, (6) **35**(2007); M. Csörgő and R. Kulik, *Reduction principles for quantile and Bahadur–Kiefer processes of long-range dependent linear sequences*, Probab. Theory Rel., to appear; M. Csörgő and R. Kulik, *Weak convergence of Vervaat and Vervaat error processes of long-range dependent sequences*, J. Theoret. Probab., to appear; and their references. The paper by E. Csáki, M. Csörgő, A. Földes, Z. Shi and R. Zitikis, *Pointwise and uniform asymptotics of the Vervaat error process*, J. Theoret. Probab. **15**(2002), 845–875, constitutes an i.i.d. backdrop for the recent developments as in the just mentioned papers on long-range dependent sequences based processes. For example, it is well known that the Bahadur–Kiefer and Vervaat error processes cannot converge weakly in the i.i.d. case. In contrast to this, when these processes are based on certain long-range dependent sequences, then we, for example, conclude that they do converge weakly to appropriate Dehling–Taqqu type limit processes that are based on H. Dehling, M.S. Taqqu, *The empirical process of some long range dependent sequences with an application to U-statistics*, Ann. Statist. **17**(1989), 1767–1783.

ARNAK DALALYAN, Université Pierre et Marie Curie, Paris 6

Sparsity and sharp PAC-Bayesian bounds for the EW-aggregate

We study the problem of aggregation under the squared loss in the model of regression with deterministic design. We obtain sharp PAC-Bayesian risk bounds for aggregates defined via exponential weights (EW-aggregate), under general assumptions on the distribution of errors and on the functions to aggregate. We then apply these results to derive sparsity oracle inequalities.

This is a joint work with Alexandre Tsybakov.

ELISABETH GASSIAT, Université Paris-Sud 11, Batiment 425, 91405 Orsay Cedex, France

A non parametric Bernstein–Von Mises theorem

Given a probability mass function $\theta = (\theta(i))_{i \in \mathbb{N}^*}$ on the set of positive integers, let the observations be independent and identically distributed according to θ . We consider the asymptotic behaviour of posterior distributions of $(\theta(i))_{1 \leq i \leq k_n}$ where k_n is an increasing sequence tending to infinity, n being the number of observations. Under suitable assumptions on the prior (smoothness and prior mass put on Fisher balls) and on the true probability θ_0 , we prove a non parametric Bernstein–Von Mises theorem: the posterior distribution concentrates around the true θ_0 at speed \sqrt{n} and its variation distance to the Gaussian distribution with variance the inverse of the Fisher information tends to zero in probability.

We use this result for the construction of confidence intervals for functionals of θ such as the Shannon entropy or Renyi entropies. Indeed, Bernstein–Von Mises theorems hold for the posterior distribution of the plug-in estimator.

CHRISTOPHE GIRAUD, Université de Nice, laboratoire J.A. Dieudonné, Parc Valrose, 06108 Nice, France; et INRA, laboratoire MIA, Parc Vilvert, 78352 Jouy-en-Josas, France

Estimation of Gaussian Graphs by Model Selection

Biological systems involve complex networks of interactions between entities such as genes or proteins. One of the challenge of the post-genomic is to infer these networks from high-throughput data produced by recent biotechnological tools. The task is challenging for the statistician due to the very high-dimensional nature of the data. For example, microarrays measure the expression level of a few thousand genes (typically 4000) whereas the sample size n is no more than a few tens.

Valuable tools for analyzing these network of interactions are the Gaussian Graphical Models. The vector $X = (X_1, \dots, X_p)$ of the expression levels of the p genes is modeled by a Gaussian variable in \mathbf{R}^p . Then, the Gaussian Graph has an edge between the genes i and j if and only if X_i is not independent of X_j conditionally on the other variables. The goal of the statistician is to infer these edges from a n -sample of the variables X .

We propose a statistical procedure to estimate the graph of conditional dependences from X . We first introduce a collection of candidate graphs and then select one of them by minimizing a penalized empirical risk. The performance of the procedure is assessed in a non-asymptotic setting without any hypotheses on the covariance matrix. These good theoretical properties of the procedure are confirmed by numerical results. We pay a special attention to the maximal degree D of the graphs that we can handle, which turns to be roughly $n/(2 \log(p/D))$.

HANNA JANKOWSKI, York University, Toronto

Nonparametric Estimation of a Bathtub-Shaped Hazard

In the analysis of lifetime data, a key object of interest is the hazard function, or instantaneous failure rate. One natural assumption is that the hazard is bathtub, or U-shaped (i.e., first decreasing, then increasing). In particular, this is often the case in reliability engineering or human mortality.

In this talk I will discuss the nonparametric MLE and LSE of the hazard function under the additional assumption that it is also convex. These estimators are consistent and converge locally at a rate of $n^{2/5}$. Both iid sampling and iid sampling under right censoring will be considered.

This is joint work with Jon Wellner.

JOSEPH JOHNSON, Queen's

PETER KIM, University of Guelph, Guelph, Ontario N1G 2W1
Multivariate Topological Data Analysis

Assume that a finite set of points is randomly sampled from a subspace of a metric space. Recent advances in computational topology have provided several approaches to recovering the geometric and topological properties of the underlying space. In this talk we take a nonparametric statistical approach to this problem. We assume that the data is randomly sampled from an unknown probability distribution. We define two filtered complexes with which we can calculate the persistent homology of a probability distribution. Using nonparametric statistical density estimators, we show that we can recover the persistent homology of the underlying distribution.

KEITH KNIGHT, University of Toronto
Some asymptotics for elemental regression estimators

Consider a linear regression model $Y_i = x_i^T \beta + \varepsilon_i$ ($i = 1, \dots, n$). Elemental regression estimators are defined to be estimators based on exactly p cases where p is the dimension of the predictors $\{x_i\}$. Some estimators (for example, L_1 estimators) are exactly elemental estimators while in other cases, estimators can be well-approximated by elemental estimators. In this paper, we will consider the asymptotic distribution for the approximation error for a certain class of estimators as well as the asymptotic distribution of the predictors in the best elemental set.

OLEG LEPSKI, Université de Provence, CMI 53, rue F. Joliot Curie, 13453 Marseille, France
Universal estimation routine

We develop a general estimation procedure that is based on selection of estimators from a large collection. An upper bound on the risk described by an arbitrary semi-norm is established, and it is shown that the proposed selection procedure specialized for different collections of estimators leads to minimax and adaptive minimax estimators in various settings.

BORIS LEVIT, Queen's University, Department of Mathematics and Statistics, Kingston, ON K7L 3N6
A fresh look at the minimax

Some new non-asymptotic upper and lower bounds for the minimax quadratic risk will be presented both in the problem of estimating restricted normal means and for the White Gaussian Noise model on the real line.

In the latter case, some general ellipsoidal functional classes will be introduced, including classes of entire functions of exponential type, Paley-Wiener classes of analytic functions, and Sobolev classes. A comparison of the minimax risks for these classes will be discussed, based on the proposed risk bounds, as well as some numerical results.

These results demonstrate that the commonly perceived notion of a relation between the smoothness of unknown function and the accuracy of estimation can be misguided. In particular, the notion of optimal rates of convergence, dominating asymptotic statistics for the last three decades, can be misleading and indeed counterproductive.

JIM RAMSEY, McGill University, 1205 Dr. Penfield Ave., Montreal
Parameter Estimation for Differential Equations: A Generalized Smoothing Approach

We propose a new method for estimating parameters in models defined by a system of non-linear differential equations from noisy measurements on a subset of variables. The approach is based on a modification of data smoothing methods along with a generalization of profiled estimation, and we refer to it as parameter cascading. We derive parameter estimates and confidence intervals for these estimates, and show that these have low bias and good coverage properties, respectively, for data simulated from models in chemical engineering and neurobiology. The performance of the method is demonstrated using real-world data from chemistry and from the progress of the auto-immune disease lupus.

PATRICIA REYNAUD-BOURET, CNRS Ecole Normale Supérieure (DMA)

Threshold methods for Poisson processes with unknown or infinite support

The first problem is to estimate the intensity of a Poisson process with unknown or infinite support. The intensity is assumed to be L^1 and L^2 . We propose a complete data-driven threshold that we calibrate in an almost optimal way and this from a theoretical and practical point of view. The maxiset of the procedure is the intersection of a classical Besov space and a weak Besov space. The weak Besov space corresponds to sparse functions: a very small number of wavelet coefficients are significant but their location is not fixed. Our procedure is minimax on these sets with the correct power of the logarithmic term.

The second problem is to test whether a Poisson process is homogeneous or not on a finite interval. Let us suppose that the alternative belongs to a weak Besov space. This corresponds, roughly speaking, to the case where there are a small number of locations where the alternative is different from the uniform, but the position of these locations are unknown. Then, surprisingly, the separation rate is the same as the minimax estimation rate, meaning that it is as difficult to test as to estimate. Moreover a threshold procedure achieves this rate.

The first problem is a joint work with V. Rivoirard. The second problem is a joint work with M. Fromont and B. Laurent.

MUNI SHANKER SRIVASTAVA, University of Toronto, Department of Statistics, 100 St. George Street, Toronto, Ontario

Analyzing High Dimensional Data with Fewer Observations

In this talk, I consider the problem of analyzing datasets obtained on several characteristics m of an individual, such as in DNA microarrays in which such characteristics are typically in several thousands. The number, however, of subjects N in which these observations can be taken is typically small, often less than 50. Standard multivariate theory requires that m be substantially smaller than N , and very little theory is available to analyze the case when N is less than m . In the last decade or so, results have been obtained for this case also. I will present some of these results.

Using these results, I will analyze two datasets from microarrays. Both datasets require the comparison of two groups, the so-called two-sample test. Often, it is required to find the components (or genes) that may have caused the rejection of the equality of two mean vectors. An alternative to the currently used False Discovery rate (FDR) will be presented.

GILLES STOLTZ, CNRS; Ecole Normale Supérieure, Paris; HEC Paris

Strategies for prediction under imperfect monitoring

We propose simple randomized strategies for sequential prediction under imperfect monitoring, that is, when the forecaster does not have access to the past outcomes but rather to a feedback signal. The proposed strategies are consistent in the sense that they achieve, asymptotically, the best possible average reward. It was Rustichini (1999) who first proved the existence of such consistent predictors. The forecasters presented here offer the first constructive proof of consistency. Moreover, the proposed algorithms are computationally efficient. We also establish upper bounds for the rates of convergence. In the case of deterministic feedback, these rates are optimal up to logarithmic terms.

Joint work with Shie Mannor, McGill University, Montreal, and Gabor Lugosi, ICREA and Universitat Pompeu Fabra, Barcelona.

JEAN-PHILIPPE VERT, Mines ParisTech and Institut Curie, Paris, France

Collaborative filtering in Hilbert spaces with spectral regularization

Collaborative Filtering (CF) refers to the task of learning preferences of customers for products, such as books or movies, from a set of known preferences. More formally, this can be seen as the task of filling missing entries in a matrix where some entries are known. A standard approach to CF is to find a low rank approximation to the matrix. This problem is computationally difficult and some authors have proposed recently to search instead for a low trace norm matrix, which results in a convex optimization problem. We generalize this approach to the estimation of a compact operator, of which matrix estimation is a special case.

We develop a notion of spectral regularization which captures both rank constraint and trace norm regularization, as well as many others. The major advantage of this approach is that it provides a natural method of utilizing side-information, such as age and gender, about the customers (or objects) in question—a formerly challenging limitation of the low-rank approach. We provide a number of algorithms, and test our results on a standard CF dataset with promising results.

This is a joint work with Jacob Abernethy (UC Berkeley), Francis Bach (INRIA), and Theodoros Evgeniou (INSEAD).