Research results in 2022

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK (team STATIFY).

655, avenue de l'Europe, Montbonnot. 38334 Saint-Ismier Cedex, France

Stephane.Girard@inria.fr

Abstract

This note summarizes my research results in 2022, dealing with neural networks, risk assessment, extreme-value analysis, dimension reduction and Bayesian inference.

1 Joint supervised classification and reconstruction of irregularly sampled satellite image times series

Recent satellite missions have led to a huge amount of earth observation data, most of them being freely available. In such a context, satellite image time series have been used to study land use and land cover information. However, optical time series, like Sentinel-2 or Landsat ones, are provided with an irregular time sampling for different spatial locations, and images may contain clouds and shadows. Thus, pre-processing techniques are usually required to properly classify such data. The proposed approach is able to deal with irregular temporal sampling and missing data directly in the classification process. It is based on Gaussian processes and allows to perform jointly the classification of the pixel labels as well as the reconstruction of the pixel time series. The method complexity scales linearly with the number of pixels, making it amenable in large scale scenarios. Experimental classification and reconstruction results show that the method does not compete yet with state of the art classifiers but yields reconstructions that are robust with respect to the presence of undetected clouds or shadows and does not require any temporal preprocessing [1].

2 Extreme events and neural networks

Feedforward neural networks based on Rectified linear units (ReLU) cannot efficiently approximate quantile functions which are not bounded, especially in the case of heavy-tailed distributions. We thus propose a new parametrization for the generator of a Generative adversarial network (GAN) adapted to this framework, basing on extreme-value theory. We provide an analysis of the uniform error between the extreme quantile and its GAN approximation. It appears that the rate of convergence of the error is mainly driven by the second-order parameter of the data distribution. The above results are illustrated on simulated data and real financial data [2].

A similar investigation has been conducted to simulate fractional Brownian motion with ReLU neural networks [3].

In [4], we propose new parametrizations for neural networks in order to estimate extreme quantiles in both non-conditional and conditional heavy-tailed settings. All proposed neural network estimators feature a bias correction based on an extension of the usual second-order condition to an arbitrary order. The convergence rate of the uniform error between extreme log-quantiles and their neural network approximation is established. The finite sample performances of the non-conditional neural network estimator are compared to other bias-reduced extreme-value competitors on simulated data. It is shown that our method outperforms them in difficult heavy-tailed situations where other estimators almost all fail. Finally, the conditional neural network estimators are implemented to investigate the behaviour of extreme rainfalls as functions of their geographical location in the southern part of France. The results are submitted for publication.

3 Estimation of extreme risk measures

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0, 1)$ corresponds to the upper α-quantile of the loss distribution. Weissman extrapolation device for estimating extreme quantiles (when $\alpha \to 0$) from heavy-tailed distributions is based on two estimators: an order statistic to estimate an intermediate quantile and an estimator of the tail-index. The common practice is to select the same intermediate sequence for both estimators. In [5], we show how an adapted choice of two different intermediate sequences leads to a reduction of the asymptotic bias associated with the resulting refined Weissman estimator. This new bias reduction method is fully automatic and does not involve the selection of extra parameters. Our approach is compared to other bias reduced estimators of extreme quantiles both on simulated and real data.

The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: VaR(α) does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α-quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution.

Risk measures of a financial position are, from an empirical point of view, mainly based on quantiles. Replacing quantiles with their least squares analogues, called expectiles, has recently received increasing attention [6]. The novel expectile-based risk measures satisfy all coherence requirements. Currently available estimators of extreme expectiles are typically biased and hence may show poor finite-sample performance even in fairly large samples. In [7], we focus on the construction of bias-reduced extreme expectile estimators for heavy-tailed distributions. The rationale for our construction hinges on a careful investigation of the asymptotic proportionality relationship between extreme expectiles and their quantile counterparts, as well as of the extrapolation formula motivated by the heavy-tailed context. We accurately quantify and estimate the bias incurred by the use of these relationships when constructing extreme expectile estimators. This motivates the introduction of a class of bias-reduced estimators whose asymptotic properties are rigorously shown, and whose finite-sample properties are assessed on a simulation study and three samples of real data from economics, insurance and finance.

4 Conditional extremal events

Expectiles have recently started to be considered as serious candidates to become standard tools in actuarial and financial risk management. However, expectiles and their sample versions do not benefit from a simple explicit form, making their analysis significantly harder than that of quantiles and order statistics. This difficulty is compounded when one wishes to integrate auxiliary information about the phenomenon of interest through a finite-dimensional covariate, in which case the problem becomes the estimation of conditional expectiles.

We exploit the fact that the expectiles of a distribution F are in fact the quantiles of another distribution E explicitly linked to F, in order to construct nonparametric kernel estimators of extreme conditional expectiles. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions. Applications to simulated data and real insurance data are provided [8]. The extension to functional covariates is investigated in [9].

5 Estimation of multivariate risk measures

Expectiles form a family of risk measures that have recently gained interest over the more common value-at-risk or return levels, primarily due to their capability to be determined by the probabilities of tail values and magnitudes of realisations at once. However, a prevalent and ongoing challenge of expectile inference is the problem of uncertainty quantification, which is especially critical in sensitive applications, such as in medical, environmental or engineering tasks. In [10], we address this issue by developing a novel distribution, termed the multivariate expectilebased distribution (MED), that possesses an expectile as a closed-form parameter. Desirable properties of the distribution, such as log-concavity, make it an excellent fitting distribution in multivariate applications. Maximum likelihood estimation and Bayesian inference algorithms are described. Simulated examples and applications to expectile and mode estimation illustrate the usefulness of the MED for uncertainty quantification.

6 Dimension reduction for extremes

In the context of the PhD thesis of Meryem Bousebata, we propose a new approach, called Extreme-PLS, for dimension reduction in regression and adapted to distribution tails. The objective is to find linear combinations of predictors that best explain the extreme values of the response variable in a non-linear inverse regression model. The asymptotic normality of the Extreme-PLS estimator is established in the single-index framework and under mild assumptions. The performance of the method is assessed on simulated data. A statistical analysis of French farm income data, considering extreme cereal yields, is provided as an illustration [11].

7 Bayesian inference for extreme values

Combining extreme value theory with Bayesian methods offers several advantages, such as a quantification of uncertainty on parameter estimation or the ability to study irregular models that cannot be handled by frequentist statistics. However, it comes with many options that are left to the user concerning model building, computational algorithms, and even inference itself. Among them, the parameterization of the model induces a geometry that can alter the efficiency of computational algorithms, in addition to making calculations involved. In [12], we focus on the Poisson process characterization of extremes and outline two key benefits of an orthogonal parameterization addressing both issues. First, several diagnostics show that Markov chain Monte Carlo convergence is improved compared with the original parameterization. Second, orthogonalization also helps deriving Jeffreys and penalized complexity priors, and establishing posterior propriety. The analysis is supported by simulations, and our framework is then applied to extreme level estimation on river flow data. The results are submitted for publication.

8 Diagnosing convergence of Markov chain Monte Carlo

Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis. Among the most popular methods, the potential scale reduction factor (commonly named $\hat{R}$) is an indicator that monitors the convergence of output chains to a stationary distribution, based on a comparison of the between- and within-variance of the chains. Several improvements have been suggested since its introduction in the 90'ss. We analyse some properties of the theoretical value

R associated to $\hat{R}$ in the case of a localized version that focuses on quantiles of the distribution. This leads to proposing a new indicator, which is shown to allow both for localizing the MCMC convergence in different quantiles of the distribution, and at the same time for handling some convergence issues not detected by other $\hat{R}$ versions. This work is submitted for publication [13].

9 Dimension reduction with Sliced Inverse Regression

Since its introduction in the early 90's, the Sliced Inverse Regression (SIR) methodology has evolved adapting to increasingly complex data sets in contexts combining linear dimension reduction with non linear regression. The assumption of dependence of the response variable with respect to only a few linear combinations of the covariates makes it appealing for many computational and real data application aspects. In [14], we propose an overview of the most active research directions in SIR modeling from multivariate regression models to regularization and variable selection.

References

[1] A. Constantin, M. Fauvel, and S. Girard. Mixture of multivariate Gaussian processes for classification of irregularly sampled satellite image time-series. *Statistics and Computing*, 32:79, 2022.

[2] M. Allouche, S. Girard, and E. Gobet. EV-GAN: Simulation of extreme events with ReLU neural networks. *Journal of Machine Learning Research*, 23(150):1–39, 2022.

[3] M. Allouche, S. Girard, and E. Gobet. Generative model for fBm with deep ReLU neural networks. *Journal of Complexity*, 73:101667, 2022.

[4] M. Allouche, S. Girard, and E. Gobet. Estimation of extreme quantiles from heavy-tailed distributions with neural networks. `https://hal.archives-ouvertes.fr/hal-03751980`, 2022.

[5] M. Allouche, J. El-methni, and S. Girard. A refined Weissman estimator for extreme quantiles. *Extremes*, 2023. to appear.

[6] A. Daouia, S. Girard, and G. Stupfler. Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society series B*, 80:263–292, 2018.

[7] S. Girard, G. Stupfler, and A. Usseglio-Carleve. On automatic bias reduction for extreme expectile estimation. *Statistics and Computing*, 32:64, 2022.

[8] S. Girard, G. Stupfler, and A. Usseglio-Carleve. Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics*, 49(1):78–115, 2022.

[9] S. Girard, G. Stupfler, and A. Usseglio-Carleve. Functional estimation of extreme conditional expectiles. *Econometrics and Statistics*, 21:131–158, 2022.

[10] J. Arbel, S. Girard, H. Nguyen, and A. Usseglio-Carleve. Multivariate expectile-based distribution: properties, Bayesian inference, and applications. *Journal of Statistical Planning and Inference*, 225:146–170, 2023.

[11] M. Bousebata, G. Enjolras, and S. Girard. Extreme Partial Least-Squares. *Journal of Multivariate Analysis*, 194:105101, 2023.

[12] T. Moins, J. Arbel, S. Girard, and A. Dutfoy. Reparameterization of extreme value framework for improved Bayesian workflow. `https://hal.archives-ouvertes.fr/hal-03806159`, 2022.

[13] T. Moins, J. Arbel, A. Dutfoy, and S. Girard. On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic. `https://hal.inria.fr/hal-03600407`, 2022.

[14] S. Girard, H. Lorenzo, and J. Saracco. Advanced topics in sliced inverse regression. *Journal of Multivariate Analysis*, 188:104852, 2022.