Program
Multitask ECODEP Conference
IHP- February 12-14, 2024

Monday 12 Morning: Time series for ecology

10:00-10:10 **Gérard Benarous** (NYU, New-York) Opening
10:10-10:50 **Konstantinos Fokianos** (Cyprus University)
10:50-11:30 **Guillaume Franchi** (ENSAI-CREST, Rennes)
11:30-12:10 **Lionel Truquet** (ENSAI-CREST, Rennes)

Monday 12 Afternoon: Processes

14:00-14:40 **Félix Cheysson** (Sorbonne University, Paris)
14:40-15:20 **José R Leon** (UdelaR, Montevideo, Uruguay)
15:20-16:00 **Vytaute Pilipauskaite** (Aalborg University, Denmark)
16:30-17:10 **Nizar Touzi** (NYU, New-York, USA)
17:10-17:50 **Mathieu Rosenbaum** (Polytechnique, Palaiseau, France)
17:50-18:30 **Adam Jakubowski** (Nicolaus Copernicus University, Torun, Poland)

Tuesday 13 Morning: High dimension

10:00-10:40 **Alexandre Tsybakov** (ENSAE-CREST, Palaiseau)
10:40-11:20 **Imma Curato** (TU, Chemnitz, Germany)
11:20-12:00 **Wei Biao Wu** (University of Chicago)
12:00-12:40 **Anne van Delft** (Columbia University, New-York)

Tuesday 13 Afternoon: Time series

14:00-14:40 **Herold Dehling** (Bochum University, Germany)
14:40-15:20 **Michael Neumann** (University Jena, Germany)
15:20-16:00 **Olivier Wintenberger** (Sorbonne University, Paris)
16:30-17:10 **Hamdi Raissi** (PUCV, Valparaiso)
17:10-17:50 **Sana Louhichi** (Université Grenoble Alpes, France)
17:50-18:00 **Matthieu Cornec** (Bordeaux)

Wednesday 14 Morning: Levy processes

10:00-10:40 **Gennady Samorodnitsky** (Cornell, Ithaca, USA)
10:40-11:20 **Donatas Surgailis** (University of Vilnius)
11:20-12:00 **Alexander Lindner** (University of Ulm, Germany)
12:00-12:40 **Thomas Mikosch** (University of Copenhagen, Denmark)

Hypertext links lead to the abstracts

Multitask ECODEP Conference
IHP
February 12-14, 2024

Hypertext links lead to the abstracts

Abstracts

Félix Cheysson (Sorbonne University, Paris)
Inference of Hawkes processes from imperfect data

Imma Curato (TU, Chemnitz, Germany)
Mixed moving average field guided learning

Herold Dehling (Bochum University, Germany)
Test for independence of long-range dependent time series using distance covariance

Konstantinos Fokianos (Cyprus University)
Nonlinear Network autoregression

Guillaume Franchi (ENSAI-CREST, Rennes)
Dynamic modelling of abundance data

Adam Jakubowski (Nicolaus Copernicus University, Torun, Poland)
Phantom distribution for non-stationary time series as an averaging operation

Oleg Klesov (Kyiv Polytechnic Institute, Ukraine)
Rate of convergence in the strong law of large numbers for records

José R Leon (UdelaR, Montevideo, Uruguay)
Some ideas about modeling real situations using hypoelliptic SDE equations with or without boundary conditions

Alexander Lindner (University of Ulm, Germany)
Quasi-infinitely divisible distributions

Sana Louhichi (Université Grenoble Alpes, France)
On bandwidth selection problems in nonparametric trend estimation under martingale difference errors

Thomas Mikosch (University of Copenhagen, Denmark)
Self-normalized sums of heavy-tailed random vectors

Michael Neumann (University Jena, Germany)
A log-linear model for non-stationary time series of counts

Vytaute Pilipauskaite (Aalborg University, Denmark)
Nonparametric estimation of McKean-Vlasov SDEs via deconvolution

Mathieu Rosenbaum (Polytechnique, Palaiseau, France)
The two square root laws of market impact: rough volatility and the role of sophisticated market participants

Gennady Samorodnitsky (Cornell, Ithaca, USA)
Clustering in large deviations

Donatas Surgailis (University of Vilnius)
Fractionally integrated spatial models and statistical applications

Nizar Touzi (NYU, New-York, USA)
HJB equations on the process space, application to mean field control with common noise

Lionel Truquet (ENSAI-CREST, Rennes)
Stability Properties of Certain Markov Chain Models in Random Environments

Alexandre Tsybakov (CREST, ENSAE, Institut Polytechnique de Paris)
Gradient-free optimization from noisy data

Anne van Delft (Columbia University, New-York)
Detecting moments of significance in nonstationary time series of random geometric objects

Olivier Wintenberger (Sorbonne University, Paris)
Moment conditions for random coefficient AR(∞) under non-negativity assumptions

Wei Biao Wu (University of Chicago)
Fast Algorithms for Estimating Covariance Matrices of Stochastic Gradient Descent Solutions

<u>Félix Cheysson</u> (Sorbonne University, Paris)
Inference of Hawkes processes from imperfect data

Hawkes processes are a family of point processes for which the occurrence of any event increases the probability of further events occurring.
Although the linear Hawkes process, for which a representation in the form of a superposition of branching processes exists, is particularly well studied, difficulties remain in estimating the parameters of the process from imperfect data (noisy, missing or aggregated data), since the usual estimation methods based on maximum likelihood or least squares do not necessarily offer theoretical guarantees or are numerically too costly.
In this talk, we present a spectral approach well-adapted to this context, for which we proved consistency and asymptotic normality in a previous work.
We illustrate this approach for noisy data, and show that identifiability of the model becomes an issue in this case, but can be alleviated under additional assumptions.
Joint work with Anna Bonnet, Miguel Martinez and Maxime Sangnier.

Imma Curato (TU, Chemnitz, Germany)
Mixed moving average field guided learning

Influenced mixed moving average fields are a versatile modeling class for spatio-temporal data. However, their predictive distribution is not generally known. Under this modeling assumption, we define a novel spatio-temporal embedding and a theory-guided machine learning approach that employs a generalized Bayesian algorithm to make ensemble forecasts. We employ Lipschitz predictors and determine fixed-time and any-time PAC Bayesian bounds in the batch learning setting. Performing causal forecast is a highlight of our methodology as its potential application to data with spatial and temporal short and long-range dependence.

<u>Herold Dehling</u> (Bochum University, Germany)

Test for independence of long-range dependent time series using distance covariance

We apply the concept of distance covariance for testing independence of two long-range dependent time series. As test statistic we propose a linear combination of empirical distance cross covariances. We derive the asymptotic distribution of our test statistics, and we show consistency against a large class of alternatives. The asymptotic theory developed in this paper is based on a novel non-central limit theorem for stochastic processes with values in a Hilbert space.
Joint work with Annika Betken, TU Twente, The Netherlands.

Konstantinos Fokianos (Cyprus University)
Nonlinear Network autoregression

We study general nonlinear models for time series networks of integer and continuous valued data. The vector of high dimensional responses, measured on the nodes of a known network, is regressed non-linearly on its lagged value and on lagged values of the neighboring nodes by employing a smooth link function. We study stability conditions for such multivariate process and develop quasi maximum likelihood inference when the network dimension is increasing. In addition, we study linearity score tests by treating separately the cases of identifiable and non-identifiable parameters. In the case of identifiability, the test statistic converges to a chi-square distribution. When the parameters are not-identifiable, we develop a supremum-type test whose *p*-values are approximated adequately by employing a feasible bound and bootstrap methodology.

<u>Guillaume Franchi</u> (ENSAI-CREST, Rennes)
Dynamic modelling of abundance data

In order to understand the dynamic of an ecosystem along time, we propose to construct and study two models of time series on specific compact sets.
The first one deals with relative abundance and is defined on the simplex. Our approach is based on a general construction of infinite memory models, called chains with complete connections, and we study a specific case with a Dirichlet transition kernel.
The second model tries to predict the presence or the absence of each species in the ecosystem along time, and is actually a dynamic version of a probit regression.
In both models, simple conditions ensuring the existence of stationary paths are given, and we discuss inference methods based on likelihood maximization, some other convex optimization, and the use of panel data.

Phantom distribution for non-stationary time series as an averaging operation

Adam Jakubowski
Nicolaus Copernicus University, Toruń, Poland
adjakubo@mat.umk.pl

Abstract

The notion of a phantom distribution function was introduced by O'Brien (1987). Doukhan et al. [1] showed that the existence of a continuous phantom distribution function is a quite common phenomenon among stationary sequences.

It should be noted that phantom distribution functions may exist also for non-stationary sequences and that the corresponding theory was developed in [3] and perfected in [4].

In the present lecture we discuss several examples of *non-stationary* sequences admitting a phantom distribution functions (including e.g. [2]). The conclusion is that if a phantom distribution function exists in the non-stationary case, it admits an interpretation as an averaging operator.

We discuss also other asymptotic representations for maxima of non-stationary sequences.

This is *a joint work* with Paul Doukhan.

References

[1] P. Doukhan, A. Jakubowski and G. Lang. Phantom distribution functions for some stationary sequences. *Extremes* **18 (2015)**, 697—725.

[2] J. Hüsler. Extreme Values of Non-Stationary Random Sequences. *Journal of Applied Probability* **23 (1986)**, 937–950.

[3] A. Jakubowski. An asymptotic independent representation in limit theorems for maxima of nonstationary random sequences. *The Annals of Probability* **21 (1993)**, 819—830.

[4] A. Jakubowski and P. Truszczyński. Quenched phantom distribution functions for Markov chains. Statistics and Probability Letters, **137 (2018)**, 79–83.

RATE OF CONVERGENCE IN THE STRONG LAW OF LARGE NUMBERS FOR RECORDS

O. I. KLESOV, O. V. KOLESNIK

Consider a sequence of independent identically distributed random variables $\{X_k\}$ and the corresponding sequence of *record times*:

$$T_1 = 1, \qquad T_k = \min\{j : \ X_j > X_{T_{k-1}}\}, \quad k > 1.$$

The random variables X_{T_k}, $k \geqslant 1$, are called *records.* The

$$\mu_k = \text{number of records up to moment } k$$

is of interest in this talk. The strong law of large numbers for $\{\mu_k\}$ tells us that

$$\lim_{n\to\infty} \frac{\mu_n}{\ln(n)} = 1 \qquad \text{almost surely.}$$

An improvement of this statement could be called the Marcinkiewicz–Zygmund strong law of large numbers:

$$\lim_{n\to\infty} \frac{\mu_n - \ln(n)}{(\ln(n))^{1/r}} = 0 \qquad \text{almost surely}$$

for any $1 \leqslant r < 2$. We study a similar result in a more general setting called F^α-scheme.

Let X be a random variable and F_X be its distribution function being nondegenerate in the sense that

$$P(X = 0) \neq 0.$$

Let $\{\alpha_k\}$ be a sequence of nonnegative numbers. A sequence of independent random variables $\{X_k\}$ is called the F^α-scheme associated with X and $\{\alpha_k\}$ if

$$F_{X_k} = F_X^{\alpha_k}.$$

Define T_k and μ_k, $k \geqslant 1$, in the same way as above. Then one can prove for the F^α-scheme that

$$\lim_{n\to\infty} \left[\frac{\mu_n}{\ln(A_n)} - \frac{1}{A_n} \sum_{k=1}^{n} \frac{\alpha_k}{A_k} \right] = 0 \qquad \text{almost surely,}$$

The first author is supported in part by DFG.

where $A_n = \alpha_1 + \cdots + \alpha_n$ (see [2]). We propose a method for finding the sequences $\{f(n)\}$ and $\{g(n)\}$ such that

$$\lim_{n\to\infty} \left[\frac{\mu_n}{g(n)} - f(n) \right] = 0 \qquad \text{almost surely.}$$

Based on this result, we are able to solve the same problem for record times $\{T_k\}$. In doing so we use the idea of pseudo-regularly functions, since T_n is an generalized renewal process constructed from μ_n (see [1]).

Література

[1] V. V. Buldygin, K.-H. Indlekofer, O. I. Klesov, J. G. Steinebach (2018). *Pseudo-Regularly Varying Functions and Generalized Renewal Processes.* Springer, Cham, 482+XXIIpp.

[2] P. Doukhan, O. I. Klesov, J. G. Steinebach (2015) *Strong laws of large numbers in an F^α-scheme* in book *Festschrift dedicated to 65th birthday of Prof. Paul Deheuvels*, Springer, Cham, 287–303

National Technical University of Ukraine "Igor Sikorsky Kyiv Politechnic Institute, Kyiv, Ukraine

Email address: voselk@gmail.com

RATE OF CONVERGENCE IN THE STRONG LAW OF LARGE NUMBERS FOR RECORDS

O. I. KLESOV, O. V. KOLESNIK

Consider a sequence of independent identically distributed random variables $\{X_k\}$ and the corresponding sequence of *record times*:

$$T_1 = 1, \qquad T_k = \min\{j : \ X_j > X_{T_{k-1}}\}, \quad k > 1.$$

The random variables X_{T_k}, $k \geqslant 1$, are called *records*. The

$$\mu_k = \text{number of records up to moment } k$$

is of interest in this talk. The strong law of large numbers for $\{\mu_k\}$ tells us that

$$\lim_{n\to\infty} \frac{\mu_n}{\ln(n)} = 1 \qquad \text{almost surely.}$$

An improvement of this statement could be called the Marcinkiewicz–Zygmund strong law of large numbers:

$$\lim_{n\to\infty} \frac{\mu_n - \ln(n)}{(\ln(n))^{1/r}} = 0 \qquad \text{almost surely}$$

for any $1 \leqslant r < 2$. We study a similar result in a more general setting called F^α-scheme.

Let X be a random variable and F_X be its distribution function being nondegenerate in the sense that

$$P(X = 0) \neq 0.$$

Let $\{\alpha_k\}$ be a sequence of nonnegative numbers. A sequence of independent random variables $\{X_k\}$ is called the F^α-scheme associated with X and $\{\alpha_k\}$ if

$$F_{X_k} = F_X^{\alpha_k}.$$

Define T_k and μ_k, $k \geqslant 1$, in the same way as above. Then one can prove for the F^α-scheme that

$$\lim_{n\to\infty} \left[\frac{\mu_n}{\ln(A_n)} - \frac{1}{A_n} \sum_{k=1}^{n} \frac{\alpha_k}{A_k} \right] = 0 \qquad \text{almost surely,}$$

The first author is supported in part by DFG.

where $A_n = \alpha_1 + \cdots + \alpha_n$ (see [2]). We propose a method for finding the sequences $\{f(n)\}$ and $\{g(n)\}$ such that

$$\lim_{n\to\infty} \left[\frac{\mu_n}{g(n)} - f(n) \right] = 0 \qquad \text{almost surely.}$$

Based on this result, we are able to solve the same problem for record times $\{T_k\}$. In doing so we use the idea of pseudo-regularly functions, since T_n is an generalized renewal process constructed from μ_n (see [1]).

Література

[1] V. V. Buldygin, K.-H. Indlekofer, O. I. Klesov, J. G. Steinebach (2018). *Pseudo-Regularly Varying Functions and Generalized Renewal Processes.* Springer, Cham, 482+XXIIpp.

[2] P. Doukhan, O. I. Klesov, J. G. Steinebach (2015) *Strong laws of large numbers in an F^α-scheme* in book *Festschrift dedicated to 65th birthday of Prof. Paul Deheuvels*, Springer, Cham, 287–303

National Technical University of Ukraine “Igor Sikorsky Kyiv Politechnic Institute, Kyiv, Ukraine

E-mail address: voselk@gmail.com

Some ideas about modeling real situations using hypoelliptic SDE equations with or without boundary conditions

Jose R. León*

December 15, 2023

Abstract

Newton's second law writes very simply: $ma = F$, where m is the mass of the mobile, a is its acceleration and F is the force acting on the mobile. If we denote by $x(t) \in \mathbb{R}^d$ the trajectory that the mobile follows and if we assume that the force is written $-\gamma(x, x')x' - \nabla V(x) + dW$, being γ the friction force coefficient, $V : \mathbb{R}^d \to \mathbb{R}$ the potential whose gradient is the force exerted on the mobile and W a standard Brownian motion (BM) in $\mathbb{R}^d$, Newton's second law gives us the equation for the nonlinear harmonic oscillator forcing with a white noise

$$mx''(t) + \gamma(x(t), x'(t))x'(t) + \nabla V(x(t)) = dW(t).$$

We interpret this equation as a second-order stochastic differential equation (SDE). We can also describe the motion in phase space by putting $X(t) = x(t)$ e $Y(t) = x'(t)$ we have the following SDE system

$$\begin{aligned} dX(t) &= Y(t)dt \\ mdY(t) &= -\gamma(X(t), Y(t))Y(t) - \nabla V(X(t)) + dW(t). \end{aligned}$$

In this conference, inspired by the seminal works of Brillinger et al. we consider the solution of the system, the existence of an invariant measure, and the exponential mixing of the solution. However, the mobile will move restricted to bounded domains but only in the space variable. We then dwell on applications of such models: description of animal movement in bounded domains, and diffusion of pollutants in closed water bodies etc. Our emphasis will be on dealing in a mobile restricted to a bounded region $\Omega \subset \mathbb{R}^d$.

*Universidad de La República. IMERL. Motevideo, Uruguay, Escuela de Matemática UCV. Venezuela
rlramos@fing.edu.uy

Alexander Lindner (University of Ulm, Germany)
Quasi-infinitely divisible distributions

A quasi-infinitely divisible distribution is a probability distribution whose characteristic function can be written as the quotient of the characteristic functions of two infinitely divisible distributions. Equivalently, a probability distribution is quasi-infinitely divisible if and only if its characteristic function admits a Lévy-Khintchine representation with a "signed Lévy measure". In this talk we give some examples of quasi-infinitely divisible distributions and study some of their properties.
The talk is based on joint works with Berger, Kutlu, Pan and Sato.

Sana Louhichi (Université Grenoble Alpes, France)
On bandwidth selection problems in nonparametric trend estimation under martingale difference errors

In this talk, we are interested in the problem of smoothing parameter selection in nonparametric curve estimation under dependent errors. We focus on kernel estimation and the case when the errors form a general stationary sequence of martingale difference random variables where neither linearity assumption nor “all moments are finite" are required. We compare the behaviors of the smoothing bandwidths obtained by minimizing either the unknown average squared error, the theoretical mean average squared error, a Mallows-type criterion adapted to the dependent case and the family of criteria known as generalized cross validation (GCV) extensions of the Mallows' criterion. We prove that these three minimizers and those based on the GCV family are first-order equivalent in probability. We give also a normal asymptotic behavior of the gap between the minimizer of the average square error and that of the Mallows-type criterion. This is extended to the GCV family. Finally, we apply our theoretical results to a specific case of martingale difference sequence, namely the Auto-Regressive Conditional Heteroscedastic (ARCH(1)) process. A Monte-carlo simulation study, for this regression model with ARCH(1) process, is conducted.
This talk is based on a common work with Didier A. Girard (CNRS, LJK) and Karim Benhenni (LJK, UGA)

Self-normalized sums of heavy-tailed random vectors

Thomas Mikosch (Copenhagen)

Joint work with Muneya Matsui (Nagoya) and Olivier Wintenberger (Paris, Vienna)

Self-normalized sum processes have attracted attention through the last decades. Most work has been devoted to sums of iid variables, and as normalization one often chooses the sample maximum or standard deviation.

We consider the case of a regularly varying stationary process with infinite second moment. First we derive the joint convergence of sums and self-normalizing quantities, then we consider the convergence of the ratios of the sums and self-normalizating sequences. We also study aspects of the distributions of the limiting ratios such as their moments.

Michael Neumann (University Jena, Germany)
A log-linear model for non-stationary time series of counts

We propose a new model for nonstationary integer-valued time series which is particularly suitable for data with a strong trend. In contrast to popular Poisson-INGARCH models, but in line with classical GARCH models, we propose to pick the conditional distributions from nearly scale invariant families where the mean absolute value and the standard deviation are of the same order of magnitude. As an important prerequisite for applications in statistics, we prove absolute regularity of the count process with exponentially decaying coefficients.
This talk is based on joint work with Anne Leucht (Universität Bamberg).

Vytaute Pilipauskaite (Aalborg University, Denmark)
Nonparametric estimation of McKean-Vlasov SDEs via deconvolution

In this talk we discuss the estimation of the interaction function W' in the drift for a class of McKean-Vlasov stochastic differential equations (SDEs). We introduce a deconvolution estimator that uses the observation at a single time T of the associated system of N interacting particles. As both N and T tend to infinity, we derive logarithmic convergence rates for the MISE of our estimator of W' within a class of functions with polynomially decaying tails. We prove that these rates can be improved to polynomial under a more restrictive assumption that W' has Gaussian tails. This is joint work with C. Amorino, D. Belomestny, M.Podolskij and S.-Y. Zhou.

Mathieu Rosenbaum (Polytechnique, Palaiseau, France)
The two square root laws of market impact: rough volatility and the role of sophisticated market participants

The goal of this work is to disentangle the roles of volume and participation rate in the price response of the market to a sequence of orders. To do so, we use an approach where price dynamics are derived from the order flow via no arbitrage constraints. We also introduce in the model sophisticated market participants having superior abilities to analyse market dynamics. Our results lead to two square root laws of market impact, with respect to executed volume and with respect to participation rate. We also make connection with the rough behaviour of the volatility.
This is joint work with Bruno Durin and Grégoire Szymanski.

Gennady Samorodnitsky (Cornell, Ithaca, USA)
Clustering in large deviations

We describe the cluster of large deviations events that arise when one such large deviations event occurs. We work in the framework of an infinite moving average process with a noise that has finite exponential moments. The difference between short memory and long memory cases is very significant.

Nizar Touzi *(NYU, New-York, USA)*
HJB equations on the process space, application to mean field control with common noise

Motivated by the problem of mean field control with common noise, we consider a general class of partial differential equations on the space of square integrable processes, which appear naturally as the dynamic programming equation for a corresponding class of control problems. By introducing a suitable notion of viscosity solutions, we provide a comparison result between Lipschitz sub and supersolutions. The existence is obtained by a representation result which is standard for such dynamic programming equations.

Lionel Truquet (ENSAI-CREST, Rennes)
Stability Properties of Certain Markov Chain Models in Random Environments

In time series analysis, considerable efforts have been invested in formulating nonlinear models that align with the concept of stationarity. These models should also allow the application of various limit theorems essential for statistical inference.
Markov chain techniques have been extensively employed to investigate the stability properties of numerous autoregressive processes.
In this presentation, we aim to explore the feasibility of integrating exogenous covariates into such dynamical systems. Despite their practical significance, the incorporation of exogenous covariates into autoregressive processes remains inadequately comprehended.
We will introduce stability criteria applicable to select Markov chain models operating within random environments, offering potential solutions to address these complexities.
Our discussion will encompass two primary approaches. The first pertains to extending a classical drift/small set criterion. Notably, in certain observation-driven models detailed in the literature, the small sets theory proves inapplicable. Subsequently, we will introduce a second criterion designed to circumvent this limitation.
Both cases rely on appropriate coupling methodologies to manage the loss of memory within chain iterations, measured in terms of either total variation or Wasserstein-type distances.
The presentation will feature numerous illustrative examples where this theory can be effectively applied.

Alexandre Tsybakov (CREST, ENSAE, Institut Polytechnique de Paris)

Gradient-free optimization from noisy data

We study the problem of estimating the minimizer or the minimal value of a smooth function by exploration of its values under possibly adversarial noise. We consider active (sequential) and passive settings of the problem and various approximations of the gradient descent algorithm, where the gradient is estimated by procedures involving function evaluations in randomized points and a smoothing kernel based on the ideas from nonparametric regression. The objective function is assumed to be either Hölder smooth or Hölder smooth and satisfying additional assumptions such as strong convexity or Polyak--Łojasiewicz condition. In all scenarios, we suggest polynomial time algorithms achieving non-asymptotic minimax optimal or near minimax optimal rates of convergence. The talk is based on joint work with Arya Akhavan, Evgenii Chzhen, Davit Gogolashvili and Massimiliano Pontil.

Olivier Wintenberger (Sorbonne University, Paris)
Moment conditions for random coefficient AR(∞) under non-negativity assumptions

We consider random coefficient autoregressive models of infinite order (AR(∞)) under the assumption of non-negativity of the coefficients.
We develop novel methods yielding sufficient or necessary conditions for finiteness of moments, based on combinatorial expressions of first and second moments. The methods based on first moments recover previous sufficient conditions by Doukhan and Wintenberger (2008) in our setting. The second moment method provides, in particular, a necessary and sufficient condition for finiteness of second moments which is different, but shown to be equivalent to the classical criterion of Nicholls and Quinn (1982) in the case of AR(p) models with finite order p <1. We further illustrate our results through examples.
Joint work with P. Maillard, IMT, Université Toulouse III, France

Wei Biao Wu (University of Chicago)
Fast Algorithms for Estimating Covariance Matrices of Stochastic Gradient Descent Solutions

Stochastic gradient descent (SGD), an important optimization method in machine learning, is widely used for parameter estimation especially in online setting where data comes in stream. While this recursive algorithm is popular for the computation and memory efficiency, it suffers from randomness of the solutions. In this talk we shall estimate the asymptotic covariance matrices of the averaged SGD iterates (ASGD) in a fully online fashion. Based on the recursive estimator and classic asymptotic normality results of ASGD, we can conduct online statistical inference of SGD estimators and construct asymptotically valid confidence intervals for model parameters. The algorithm for the recursive estimator is efficient and only uses SGD iterates: upon receiving new observations, we update the confidence intervals at the same time as updating the ASGD solutions without extra computational or memory cost. This approach fits in online setting even if the total number of data is unknown and takes the full advantage of SGD: computation and memory efficiency. This work is joint with Wanrong Zhu and Xi Chen.