



A pseudo-likelihood estimator of the Ornstein-Uhlenbeck parameters from suprema observations

Christophette Blanchet-Scalliet, Diana Dorobantu, Benoît Nieto

► To cite this version:

Christophette Blanchet-Scalliet, Diana Dorobantu, Benoît Nieto. A pseudo-likelihood estimator of the Ornstein-Uhlenbeck parameters from suprema observations. *Statistical Inference for Stochastic Processes*, 2024, 10.1007/s11203-024-09307-4 . hal-03847613v5

HAL Id: hal-03847613

<https://hal.science/hal-03847613v5>

Submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A pseudo-likelihood estimator of the Ornstein-Uhlenbeck parameters from suprema observations

Christophette Blanchet-Scalliet* Diana Dorobantu[†]

Benoit Nieto[‡]

January 10, 2024

Abstract. In this paper, we propose an estimator for the Ornstein-Uhlenbeck parameters based on observations of its supremum. We derive an analytic expression for the supremum density. Making use of the pseudo-likelihood method based on the supremum density, our estimator is constructed as the maximal argument of this function. Using weak-dependency results, we prove some statistical properties on the estimator such as consistency and asymptotic normality. Finally, we apply our estimator to simulated and real data.

Keywords. Ornstein-Uhlenbeck process, supremum law, parameters estimation, asymptotic normality.

1 Introduction

Ornstein-Uhlenbeck (OU) processes are often used in various fields such as finance [22, 27], temperature modelling [2, 10, 12, 13], medicine [25], physics [28], neuro-

*Ecole Centrale de Lyon, ICJ UMR5208, CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, Université Jean Monnet, 69130 Ecully, France, christophette.blanchet@ec-lyon.fr

[†]University Lyon 1, ISFA, LSAF (EA 2429) France, diana.dorobantu@univ-lyon1.fr

[‡]Ecole Centrale de Lyon, ICJ UMR5208, CNRS, INSA Lyon, Université Claude Bernard Lyon 1, Université Jean Monnet, 69130 Ecully, France, benoit.nieto@ec-lyon.fr

science [19]. Estimation comes together with simulation and modelling, as models need to be calibrated. Estimators of discretely observed diffusions are presented in [1]. In [19], the authors propose an estimation based on observations of first hitting times.

We consider a stationary OU process $(X_t)_{t \geq 0}$ solution of

$$\begin{cases} dX_t = (\mu - \lambda X_t)dt + \sqrt{\beta}dB_t, \\ X_0 \sim \mathcal{N}\left(\frac{\mu}{\lambda}, \frac{\beta}{2\lambda}\right), \end{cases} \quad (1.1)$$

with parameter $\theta = (\mu, \lambda, \beta) \in \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$, and $(B_t)_{t \geq 0}$ a Brownian motion independent of X_0 . In this paper, we focus on the parameter estimation of $(X_t)_{t \geq 0}$ when we observe (S^1, \dots, S^N) , $N \in \mathbb{N}^*$ a set of suprema observations taken over a single trajectory.

This approach has already been proposed in [7], using an estimator constructed by the least squares method. Few statistical results have been proved on this estimator and it is computationally expensive to deal with. Our goal is to provide a new estimator with good statistical properties and less computational cost.

Even if the observations (S^1, \dots, S^N) are dependent, the sequence has some mixing properties that imply asymptotic independence. So we estimate the parameter θ using a pseudo-likelihood method introduced in the early 70's in [4]. Our estimator is then the maximal argument of the pseudo-likelihood function:

$$\mathcal{L}(\theta) = \prod_{i=1}^N f_{S^i}(S^i, \theta),$$

where f_{S^i} is the probability density function associated to the random variable S^i .

The estimator requires an expression of the probability density f_{S^i} . One contribution of this work is to propose an explicit expression of this density based on the Parabolic cylinder function which is numerically inexpensive. Moreover with this expression we obtain the consistency and asymptotic normality properties of the estimator.

Outline: In Section 2, we present some properties of the OU supremum sequence, then we derive the probability density function of the supremum. The estimation method and the estimator statistical properties are presented in Section 3. In Section 4, some numerical experiments on simulated data are performed. We also present an application of our procedure to a dataset of daily temperature extreme values from Paris [17]. Proofs are collected in Section 5 and some auxiliary results in Appendix.

Notation: We denote \mathbb{N}^* the set of all strictly positive integers. Furthermore, \mathbb{R}^* is the set of all non-zero real numbers and \mathbb{R}_+^* , the set of all strictly positive reals.

2 Some results related to the supremum of an OU process

Let $(X_t)_{t \geq 0}$ an OU process defined by (1.1), $(t_n)_{n \geq 0}$ a sequence of time such that $t_0 = 0$ and for $i \geq 1$, $t_i - t_{i-1} = \Delta$, where $\Delta > 0$ is fixed. We denote $(S^{i,0})_{i \geq 1}$ the following sequence of suprema observations on time windows of size Δ :

$$S^{i,0} = \sup_{s \in [t_{i-1}, t_i]} X_s.$$

2.1 Properties of the suprema sequence $(S^{i,0})_{i \geq 1}$

In order to present the estimation method, we give some properties of the sequence $(S^{i,0})_{i \geq 1}$. As the sequence of observations are dependent, some weak dependence notions as mixing properties (see [9, 14] e.g.) are required to get statistical results on the estimator.

The following result is induced by the properties of the stationary OU process.

Proposition 2.1. *The sequence $(S^{i,0})_{i \geq 1}$ is stationary, ergodic and exponentially ρ -mixing.*

Proof. Since the process $(X_t)_{t \geq 0}$ is stationary and ergodic, then by Theorem 3.5.8 in [24], the sequence $(S^{i,0})_{i \geq 1}$ is also stationary and ergodic.

Using of the definition in [9], for $p \in \mathbb{N}^*$, the sequence $(S^{i,0})_{i \geq 1}$ is ρ -mixing if it verifies:

$$\rho(p) = \sup_{n \geq 1} \sup_{f \in \mathcal{L}^2(\mathcal{F}_0^n), g \in \mathcal{L}^2(\mathcal{F}_{n+p}^{+\infty})} |\text{Corr}(f, g)| \xrightarrow{p \rightarrow +\infty} 0, \quad (2.1)$$

with $\mathcal{F}_0^n = \sigma(S^{i,0}, 1 \leq i \leq n)$ and $\mathcal{F}_{n+p}^{+\infty} = \sigma(S^{i,0}, i \geq n+p)$.

From Theorem 2.1 in [15], for all $0 \leq s \leq t$, for all functions $\tilde{f} : \mathcal{C}^0([0, s], \mathbb{R}) \rightarrow \mathbb{R}$, $\tilde{g} : \mathcal{C}^0([t, +\infty], \mathbb{R}) \rightarrow \mathbb{R}$ such that \tilde{f} and \tilde{g} are square integrable with respect to the law of X_0 , we have:

$$\left| \text{Cov} \left[\tilde{f}((X_u)_{u \leq s}), \tilde{g}((X_v)_{v \geq t}) \right] \right| \leq e^{-\lambda(t-s)} \sqrt{\text{Var} \left[\tilde{f}((X_u)_{u \leq s}) \right] \text{Var} \left[\tilde{g}((X_v)_{v \geq t}) \right]}. \quad (2.2)$$

Therefore, the process $(X_t)_{t \geq 0}$ is exponentially ρ -mixing.

However, the following inclusions are verified:

$$\sigma(S^{i,0}, 1 \leq i \leq n) \subset \sigma(X_t, 0 \leq t \leq t_n) \quad \text{and} \quad \sigma(S^{i,0}, i \geq n+p) \subset \sigma(X_t, t \geq t_{n+p-1}).$$

Then, the exponentially ρ -mixing property of $(S^{i,0})_{i \geq 1}$ is induced by (2.2). \square

2.2 Supremum law

In this section, we give some results on the law of $S^{1,0}$. Since $S^{1,0} = \sup_{s \in [0, \Delta]} X_s$, we can use the existing results on supremum of an OU process. The cumulative distribution function has already been introduced in [7]. The authors used the Bessel formulation (see [3] e.g.) which is numerically expensive. In [6], the authors give the supremum cumulative distribution of the non-stationary OU process with parameter $\theta = (0, \lambda, 1)$ in term of Parabolic cylinder function. Using this result, we easily obtain the cdf and the density of the supremum of a stationary OU process with parameter $\theta = (\mu, \lambda, \beta)$.

Proposition 2.2 (Probability density of the supremum). *Let $(X_t)_{t \geq 0}$ be an OU process solution of (1.1), with parameter $\theta = (\mu, \lambda, \beta) \in \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. For $\Delta > 0$ and $m \in \mathbb{R}$, the probability density of the supremum $S^{1,0}$ is given by*

$$\begin{aligned} f_\Delta(m, \theta) = & -\sqrt{\frac{\lambda}{\beta\pi}} e^{-\frac{\lambda}{\beta}(m-\frac{\mu}{\lambda})^2} \sum_{n \geq 1} e^{-\lambda\nu_{n,m,\theta}\Delta} \left[-\Delta\lambda \frac{D_{\nu_{n,m,\theta}-1}^2 \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)}{\partial_\nu D_{\nu_{n,m,\theta}} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)^2} \right. \\ & + 2 \frac{D_{\nu_{n,m,\theta}-1} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right) \partial_\nu D_{\nu_{n,m,\theta}-1} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)}{\partial_\nu D_{\nu_{n,m,\theta}} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)^2} \\ & \left. - \frac{D_{\nu_{n,m,\theta}-1}^2 \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right) \partial_\nu^2 D_{\nu_{n,m,\theta}} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)}{\partial_\nu D_{\nu_{n,m,\theta}} \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)^3} \right], \end{aligned} \quad (2.3)$$

with $\nu_{n,m,\theta}$ the positive (ordered) zeros of the function $\nu \mapsto D_\nu \left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}} \right)$ and $D_\nu(\cdot)$ the Parabolic cylinder function (see Appendix).

Proof. For $x_0, m \in \mathbb{R}$ such that $m > x_0$ and $\Delta > 0$, by replacing m , x_0 , Δ and λ respectively by $m - \frac{\mu}{\lambda}$, $x_0 - \frac{\mu}{\lambda}$, $\beta\Delta$ and $\frac{\lambda}{\beta}$ in Proposition 3 in [6], we get the cumulative distribution of the supremum for the non-stationary OU process with parameter $\theta = (\mu, \lambda, \beta)$.

Integration with respect to the invariant measure $X_0 \sim \mathcal{N}\left(\frac{\mu}{\lambda}, \frac{\beta}{2\lambda}\right)$ and Formula (A.1) give the cumulative distribution function of $S^{1,0}$ for the stationary OU process:

$$\mathbb{P}(S^{1,0} < m) = -\frac{e^{-(m-\frac{\mu}{\lambda})^2 \frac{\lambda}{\beta}}}{\sqrt{2\pi}} \sum_{n \geq 1} e^{-\lambda \nu_{n,m,\theta} \Delta} \frac{D_{\nu_{n,m,\theta}-1}\left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}}\right)}{\nu_{n,m,\theta} \partial_\nu D_{\nu_{n,m,\theta}}\left(-(m-\frac{\mu}{\lambda})\sqrt{\frac{2\lambda}{\beta}}\right)}. \quad (2.4)$$

Then, making use of Proposition A.1, the series in Equation (2.4) is differentiable. We easily get the probability density of the random variable $S^{1,0}$. \square

The new cdf expression (2.4) is less expensive than the one in [7].

Remark 2.3. For $\Delta > 0$, the support of the random variable $S^{1,0}$ is \mathbb{R} .

3 Estimation problem

In this section, we introduce the estimation method of a stationary OU process parameters. As the sample of observations are weak dependent, the basic idea is to use a pseudo-likelihood approach. The computation of the likelihood is simplified by approximating the joint probabilities of all data by the product of marginal probabilities.

The data $(S^{i,0})_{i \geq 1}$ are collected on disjoint but consecutive time windows of constant size Δ . Proposition 2.1 suggests to sample the initial set of observations by keeping one observation over k to obtain a sequence of data with less dependence between each over. Let $k \in \mathbb{N}^*$ and $r = (k-1)\Delta$, then the set of observations used in the estimation procedure is given by

$$S^{i,r} = \sup_{s \in [t_{i-1} + (i-1)r, t_i + (i-1)r]} X_s.$$

The choice $k = 1$ is equivalent to deal with the initial data $(S^{i,0})_{i \geq 1}$.

Remark 3.1. For all $r = (k-1)\Delta$ with $k \in \mathbb{N}^*$, the sequence $(S^{i,r})_{i \geq 1}$ has the same properties as $(S^{i,0})_{i \geq 1}$.

We consider a stationary OU process with parameter $\theta_0 = (\mu_0, \lambda_0, \beta_0)$ and $\theta_0 \in \Theta$, a compact subset of $\mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. We denote \mathbb{P}_{θ_0} the probability measure associated to marginals from the sequence $(S^{i,r})_{i \geq 1}$.

Let $N \in \mathbb{N}^*$ and $(S^{1,r}, \dots, S^{N,r})$ a sample from the sequence $(S^{i,r})_{i \geq 1}$. The pseudo-likelihood \mathcal{L}_N^r associated to the sub-sequence is given by:

$$\mathcal{L}_N^r(\theta) = \prod_{i=1}^N f_{\Delta}(S^{i,r}, \theta). \quad (3.1)$$

The OU process parameter $\theta_0 = (\mu_0, \lambda_0, \beta_0)$ is estimated by:

$$\hat{\theta}_N = (\hat{\mu}_N, \hat{\lambda}_N, \hat{\beta}_N) = \underset{\theta \in \Theta}{\text{Argmax}} \mathcal{L}_N^r(\theta). \quad (3.2)$$

The results on the sequence $(S^{i,r})_{i \geq 1}$ allow to get statistical properties on the estimator. A basic but necessary result is the identifiability of the statistical model.

Proposition 3.2. *The statistical model $\mathcal{P} = \{\mathbb{P}_{\theta}, \theta \in \Theta\}$ is identifiable.*

The following results state the statistical properties of the estimator.

Theorem 3.3. *Consider an OU process solution of (1.1), with parameter $\theta_0 = (\mu_0, \lambda_0, \beta_0)$. Assume that θ_0 belongs to Θ a compact subset of $\mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. For any $N \in \mathbb{N}^*$, the estimator $\hat{\theta}_N$ defined by (3.2) is consistent:*

$$\hat{\theta}_N \xrightarrow[N \rightarrow +\infty]{\mathbb{P}_{\theta_0}} \theta_0.$$

Using Central Limit Theorem on ρ -mixing sequence of random variable, the asymptotic normality of the estimator follows.

Theorem 3.4. *For any $N \in \mathbb{N}^*$ and $r > \frac{\ln(\frac{5}{3})}{\lambda_0}$, the following convergence is satisfied:*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow[N \rightarrow +\infty]{law} \mathcal{N}_3(0, I_{\theta_0}^{-1}), \quad (3.3)$$

where I_{θ_0} is the Fisher information matrix given by

$$I_{\theta_0} = \mathbb{E}_{\theta_0} \left[\left(\nabla_{\theta} \log f_{\Delta}(\cdot, \theta) \Big|_{\theta=\theta_0} \right) \left(\nabla_{\theta} \log f_{\Delta}(\cdot, \theta) \Big|_{\theta=\theta_0} \right)^T \right].$$

Proofs of Proposition 3.2, Theorems 3.3 and 3.4 are postponed in Section 5.

4 Numerical experiment

In this section, we discuss on the existence of an optimal gap r used in the estimation procedure. Then, we apply our estimation method to a simulated dataset and a real dataset. A part of this section is dedicated to the comparison between our method and the one proposed in [7].

4.1 Trade off between the number of observations and the gap r

For $N \in \mathbb{N}^*$ and $r = (k - 1)\Delta$ with $k \in \mathbb{N}^*$, an important aspect of this study is the choice of the sample as we take the set of observations from the sequence $(S^{i,r})_{1 \leq i \leq \lceil N/k \rceil}$ rather than $(S^{i,0})_{1 \leq i \leq N}$, where $\lceil \cdot \rceil$ is the ceiling function. The right choice of the gap r between supremum observations is given by controlling the upper bound of an appropriate inequality. Indeed for a fixed dataset, creating a gap between the observations of the supremum removes a quantity of observations in the sample used. For $r = (k - 1)\Delta$ and $k \in \mathbb{N}^*$, we have:

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{1}{\lceil N/k \rceil} \log \mathcal{L}_{\lceil N/k \rceil}^r - \frac{1}{\lceil N/k \rceil} \mathbb{E} [\log \mathcal{L}_{\lceil N/k \rceil}^r] \right)^2 \right] \leq \\ & \text{Var}(\log(f_{\Delta}(S^{1,r}, \theta))) \left(\frac{1}{\lceil N/k \rceil} + \frac{2}{\lceil N/k \rceil^2} \sum_{1 \leq i < j \leq \lceil N/k \rceil} \text{Cov}(\log(f_{\Delta}(S^{i,r}, \theta)), \log(f_{\Delta}(S^{j,r}, \theta))) \right). \end{aligned} \quad (4.1)$$

The inequality (4.1) bounds the variance of the pseudo-likelihood. As $(S^{i,r})_{i \geq 1}$ is identically distributed, the quantity $\frac{1}{\lceil N/k \rceil} \mathbb{E} [\log \mathcal{L}_{\lceil N/k \rceil}^r]$ does not depend on k and N . For $N \in \mathbb{N}^*$ observations and a time window of size Δ , there exists an r^* which minimizes the upper bound of this inequality.

Proposition 4.1. *Consider an OU process solution of (1.1) with parameter $\theta_0 = (\mu_0, \lambda_0, \beta_0)$. The optimal upper bound of (4.1) is reached for $r^* = (k^* - 1)\Delta$ with*

$$k^* = \underset{1 \leq k \leq N}{\text{Argmin}} g(\Delta, N, k, \theta_0),$$

and

$$g(\Delta, N, k, \theta_0) = C \left(\frac{1}{\lceil N/k \rceil} + \frac{2}{\lceil N/k \rceil^2} \frac{e^{\lambda_0 \Delta}}{e^{\lambda_0 k \Delta} - 1} \left[\lceil N/k \rceil + e^{\lambda_0 k \Delta} \left(\frac{e^{-\lambda_0 k \lceil N/k \rceil \Delta} - 1}{e^{\lambda_0 \Delta k} - 1} \right) \right] \right)$$

and C a strictly positive constant.

Proof. Using inequality (4.1) and the ρ -mixing property on the sequence $(\log(f_\Delta(S^{i,r}, \theta)))_{1 \leq i \leq \lceil N/k \rceil}$, we obtain the following inequality:

$$\mathbb{E} \left[\left(\frac{1}{\lceil N/k \rceil} \log \mathcal{L}_{\lceil N/k \rceil}^r - \frac{1}{\lceil N/k \rceil} \mathbb{E} [\log \mathcal{L}_{\lceil N/k \rceil}^r] \right)^2 \right] \leq g(\Delta, N, k, \theta_0).$$

To obtain the optimal $r = (k - 1)\Delta$, it is enough to minimize the function g :

$$k^* = \underset{1 \leq k \leq N}{\text{Argmin}} \ g(\Delta, N, k, \theta_0).$$

□

The appropriate gap for the estimation is $r^* = (k^* - 1)\Delta$.

Remark 4.2. *In the case where the minimal argument is not unique, we choose the largest one. As we will have less observations when k^* is larger, the minimisation algorithm will be numerically less expensive. Moreover, the larger k^* , the greater the observations are decorrelated. Thus, the pseudo-likelihood will be more suitable for our model.*

Remark 4.3. *When we deal with a real dataset we do not know the value of the parameter θ_0 . To find r^* , we first need to compute some bounds, see an example in [7].*

4.2 Numerical simulation

Numerical issues emerge from the ν -zeros of $D_\nu(m)$ involved in the probability density. For large values of $|m|$, the ν -zeros are no longer computable through dichotomy. We use the asymptotic expansions (A.5) and (A.6) to evaluate the density for large values of $|m|$. In the pseudo-likelihood maximization, a multi-start method with 10 repeats is also used to reduce the instability.

- Simulated data:

We simulate a stationary OU process with parameter $\theta_{0,1} = (\mu_{0,1}, \lambda_{0,1}, \beta_{0,1}) = (0, 1, 1)$ using an Euler scheme with $T = 10^3$ and $dt = 10^{-3}$. We denote (N, r) , the set of numerical parameters, with N the number of suprema observations and r the gap between these observations. We apply our estimation method for $\Delta = 1$. Repeating this process 100 times, we obtain a sample of our estimator. We also performed the estimation method for the parameter $\theta_{0,2} = (\mu_{0,2}, \lambda_{0,2}, \beta_{0,2}) = (20.9, 0.95, 47.5)$. From Proposition 4.1, for $\theta_{0,1}$ and $\theta_{0,2}$ with $N = 1000$, we have $r^* = 1$. Simulations will be carried out for three different set of numerical parameters, $(1000, 0)$, $(500, 1)$ and $(250, 3)$. Each estimation will be compared using the relative root-mean-square error (RMSE) and the mean-error (ME).

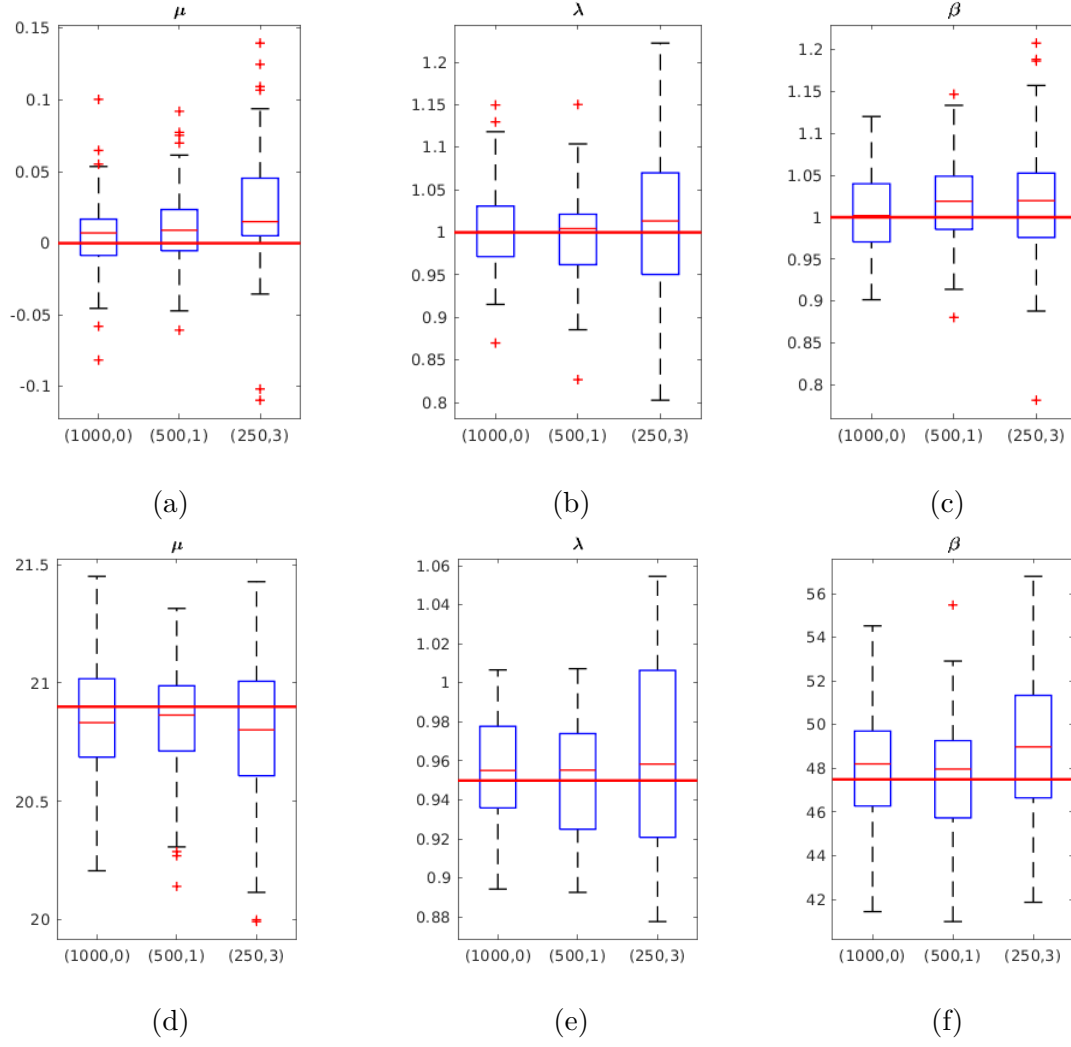


Figure 1: Boxplots of the estimated parameters: (a), (b), (c) $\theta_{0,1} = (\mu_{0,1}, \lambda_{0,1}, \beta_{0,1}) = (0, 1, 1)$; (d), (e), (f) $\theta_{0,2} = (\mu_{0,2}, \lambda_{0,2}, \beta_{0,2}) = (20.9, 0.95, 47.5)$. The red line corresponds to the theoretical value of the parameters.

Numerical parameters	Relative RMSE	ME
(250,3)	(0.0436, 0.0855, 0.0731)	(0.0240,0.0114,0.0201)
(500,1)	(0.0265, 0.0472, 0.0572)	(0.0081,-0.0048,0.0198)
(1000,0)	(0.0279, 0.0507, 0.0473)	(0.0033,0.0044,0.0032)

Table 1: Table of the relative RMSE and ME for the estimator of $\theta_{0,1} = (\mu_{0,1}, \lambda_{0,1}, \beta_{0,1}) = (0, 1, 1)$ with different numerical parameters.

Numerical parameters	Relative RMSE	ME
(250,3)	(0.0154,0.0615 , 0.0701)	(-0.1072,-0.0063,-0.6946)
(500,1)	(0.0109,0.0351 , 0.0557)	(-0.0538,-0.0096,-1.0821)
(1000,0)	(0.0113,0.0348 ,0.0578)	(-0.0482,-0.0074,-1.0693)

Table 2: Table of the relative RMSE and ME for the estimator of $\theta_{0,2} = (\mu_{0,2}, \lambda_{0,2}, \beta_{0,2}) = (20.9, 0.95, 47.5)$ with different numerical parameters.

Relative RMSE are small enough and validate the results on the trade off between the number of observations and the time gap r .

The overestimation/bias on the β estimator comes from the decrease of $\beta \mapsto \mathcal{L}_N^r(\mu, \lambda, \beta)$. As β becomes greater than λ , the pseudo-likelihood function becomes flat in β . Consequently, the β estimator will have a big variation which slowly decreases as the number of observations increases. Better results can be obtain by fixing β and performing the estimation method on parameters μ and λ (2D-estimation).

Using Theorem 3.3 in [23], we look at the probability that $\theta_{0,1}$ and $\theta_{0,2}$, fall in the 95% confidence ellipsoid for the set of numerical parameters (500, 1).

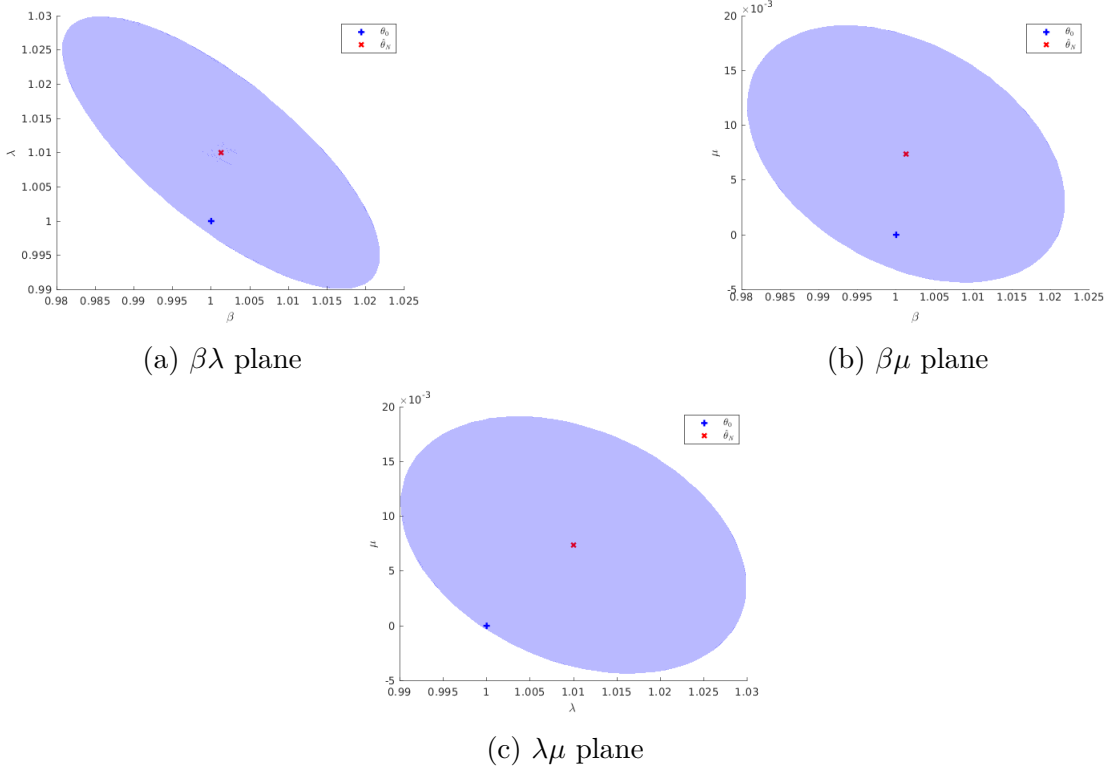


Figure 2: Cut planes of the 95% confidence ellipsoid associate to the estimator of $\theta_{0,1} = (0, 1, 1)$ and the set of numerical parameters $(N, r) = (500, 1)$.

As a consequence of a small variance and a relatively high bias on the β estimator, the parameters $\theta_{0,1}$ and $\theta_{0,2}$ fall infrequently in the 95% confidence ellipsoid. For the 2D-estimation problem (β fixed), both parameters fall in the 95% confidence ellipse with a $\approx 93\%$ probability for the same set of numerical parameters.

- Weather data:

In [16], we can find the daily temperatures of Paris. This dataset is one of the longest we can find, as it started in 1900. In this dataset, we find the daily maximum and minimum temperature measurements as well as the daily average temperature. We choose to focus this study on the maximum temperatures from 15th of June to the 14th of August from 1950 to 1984 (2135 days), using a gap of one day between each daily extreme values.

A multi-start method is also used for the estimation. We obtain the following estimates: $(\mu, \lambda, \beta) = (18.0866, 0.9510, 36.0201)$.

We use the same approaches as [7], to compare the two different estimation methods. The first validating method is the prediction. We take the mean temperature of 14/06/1985 as the starting point for the simulation of our 10-day process and we make confidence intervals on 1000 simulations of the maximum temperatures for each of these days and compare them with the true temperature values (between 15/06/1985 and 24/06/1985). The second method compare the theoretical quantiles with the empirical quantiles.

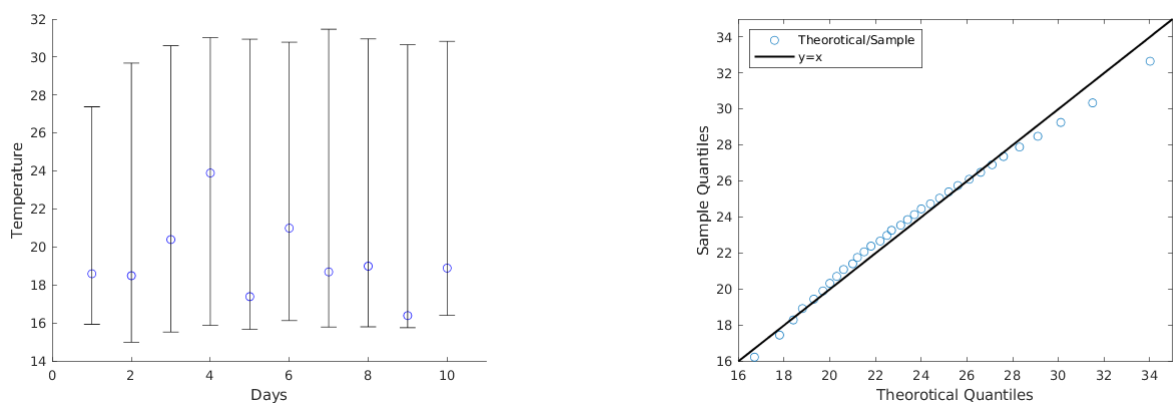


Figure 3: 95% confidence interval for daily extreme temperatures between 15/06/1985 and 24/06/1985 and Quantile-Quantile Graph.

The results are slightly better as [7] for the QQ-plot and the prediction method.

Our method is more efficient in computation time and in accuracy. In [7], the evaluation of the cumulative distribution function required a very expensive Monte Carlo method. Their estimation procedure took a week for the computation of the estimator, our method takes only a few minutes on the same machine.

5 Proofs

This section is dedicated to the proofs of Proposition 3.2, Theorem 3.3 and Theorem 3.4.

5.1 Proof of Proposition 3.2

Identifiability of the statistical model \mathcal{P} depends on the injectivity of the supremum law. For the sake of proof, we denote $(X_t^\theta)_{t \geq 0}$ a stationary OU process with parameter $\theta = (\mu, \lambda, \beta)$ and $S_t^\theta = \sup_{s \leq t} X_s^\theta$. Using (A.2), the cdf of the supremum (2.4) can be rewrite:

$$\mathbb{P}(S_t^\theta < m) = \frac{1}{2\sqrt{\pi}} e^{-(m - \frac{\mu}{\lambda})^2 \frac{\lambda}{\beta}} \sum_{n \geq 1} e^{-\lambda \nu_{n,m,\theta} t} \frac{D_{\nu_{n,m,\theta}-1}^2 \left(-(m - \frac{\mu}{\lambda}) \sqrt{\frac{2\lambda}{\beta}} \right)}{\int_{-\infty}^{(m - \frac{\mu}{\lambda}) \sqrt{\frac{\lambda}{\beta}}} D_{\nu_{n,m,\theta}}^2 (-x\sqrt{2}) dx}. \quad (5.1)$$

We prove the injectivity of the measure \mathbb{P}_θ associated with the random variable S_t^θ for t fixed. We suppose by absurd that there exist $\theta_1 = (\mu_1, \lambda_1, \beta_1)$ and $\theta_2 = (\mu_2, \lambda_2, \beta_2)$ such that $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ *i.e.*:

$$\forall m \in \mathbb{R}, \quad \mathbb{P}(S_t^{\theta_1} < m) = \mathbb{P}(S_t^{\theta_2} < m). \quad (5.2)$$

In particular the equality is satisfied for $m \rightarrow \infty$. Using the asymptotic expansion (A.10) with the three parameters variable change, we get:

$$1 - \frac{e^{-(m - \frac{\mu_1}{\lambda_1})^2 \frac{\lambda_1}{\beta_1}}}{2\sqrt{\pi} \left(m - \frac{\mu_1}{\lambda_1} \right) \sqrt{\frac{\lambda_1}{\beta_1}}} (1 + o(m^{-2+\delta})) = 1 - \frac{e^{-(m - \frac{\mu_2}{\lambda_2})^2 \frac{\lambda_2}{\beta_2}}}{2\sqrt{\pi} \left(m - \frac{\mu_2}{\lambda_2} \right) \sqrt{\frac{\lambda_2}{\beta_2}}} (1 + o(m^{-2+\delta})),$$

with $0 < \delta < 2$. Therefore

$$\begin{aligned} & \frac{\left(m - \frac{\mu_2}{\lambda_2} \right) \sqrt{\frac{\lambda_2}{\beta_2}}}{\left(m - \frac{\mu_1}{\lambda_1} \right) \sqrt{\frac{\lambda_1}{\beta_1}}} e^{-(m - \frac{\mu_1}{\lambda_1})^2 \frac{\lambda_1}{\beta_1} + (m - \frac{\mu_2}{\lambda_2})^2 \frac{\lambda_2}{\beta_2}} = 1 + o(m^{-2+\delta}) \\ \iff & \lim_{m \rightarrow +\infty} m^{2-\delta} \left[\frac{\left(m - \frac{\mu_2}{\lambda_2} \right) \sqrt{\frac{\lambda_2}{\beta_2}}}{\left(m - \frac{\mu_1}{\lambda_1} \right) \sqrt{\frac{\lambda_1}{\beta_1}}} e^{-m^2 \left(\frac{\lambda_1}{\beta_1} - \frac{\lambda_2}{\beta_2} \right) + 2m \left(\frac{\mu_1}{\beta_1} - \frac{\mu_2}{\beta_2} \right) + \frac{\mu_1^2}{\lambda_1 \beta_1} - \frac{\mu_2^2}{\lambda_2 \beta_2}} - 1 \right] = 0. \end{aligned}$$

We deduce that $\frac{\lambda_1}{\beta_1} = \frac{\lambda_2}{\beta_2} = \frac{\lambda}{\beta}$ and $\frac{\mu_1}{\beta_1} = \frac{\mu_2}{\beta_2} = \frac{\mu}{\beta}$ and thus $\nu_{n,m,\theta_1} = \nu_{n,m,\theta_2} = \nu_{n,m}$. Making use of (5.2) and (5.1), we have:

$$\sum_{n \geq 1} \left(e^{-\lambda_1 \nu_{n,m} t} - e^{-\lambda_2 \nu_{n,m} t} \right) \frac{D_{\nu_{n,m}-1}^2 \left(-(m - \frac{\mu}{\lambda}) \sqrt{\frac{2\lambda}{\beta}} \right)}{\int_{-\infty}^{(m - \frac{\mu}{\lambda}) \sqrt{\frac{\lambda}{\beta}}} D_{\nu_{n,m}}^2 (-x\sqrt{2}) dx} = 0.$$

Since $\frac{D_{\nu_{n,m-1}}^2 \left(-(m-\frac{\mu}{\lambda}) \sqrt{\frac{2\lambda}{\beta}} \right)}{\int_{-\infty}^{(m-\frac{\mu}{\lambda}) \sqrt{\frac{2\lambda}{\beta}}} D_{\nu_{n,m}}^2(-x\sqrt{2}) dx} > 0$, then the previous sum is strictly positive if $\lambda_1 < \lambda_2$ (resp. strictly negative if $\lambda_1 > \lambda_2$) because all the terms are strictly positive (resp. strictly negative). We deduce that $\lambda_1 = \lambda_2$, so $\beta_1 = \beta_2$ and $\mu_1 = \mu_2$.

Therefore the measure \mathbb{P}_θ associated with the variable S_t^θ is identifiable.

5.2 Proof of Theorem 3.3

Thereafter, without loss of generality, we suppose $\theta = (0, \lambda, 1)$. In the case of three parameters, the arguments are the same. In the following we write λ instead of θ and Θ is then a compact subset of \mathbb{R}_+^* . To prove the consistency of the estimator $\hat{\theta}_N$, we adapt the proof of Corollary 3.2.9 in [11]. We first prove some regularity properties on the density f_Δ with respect to the parameter λ .

Lemma 5.1. *Let $(X_t)_{t \geq 0}$ be an OU process solution of (1.1), with parameter $\lambda_0 \in \Theta$, the following conditions are satisfied:*

1. *For all $m \in \mathbb{R}$, $\lambda \mapsto \log(f_\Delta(m, \lambda))$ is continuous on Θ .*
2. *For all $\lambda \in \Theta$, there exists a neighborhood V of λ and $G \in L^1(\mathbb{P}_{\lambda_0})$ such that:*

$$\sup_{\eta \in V} |\log(f_\Delta(\cdot, \eta))| \leq G, \quad (5.3)$$

with f_Δ the probability density function of the supremum.

Proof. 1. Continuity of $\lambda \mapsto \log(f_\Delta(m, \lambda))$ for all $m \in \mathbb{R}$ is proved using Proposition A.1 in Appendix.

2. Using Corollary A.4.1 and Corollary A.7.1, there exist $M > 0$, $G_1 \in L^1(\mathbb{P}_{\lambda_0})$ and $G_2 \in L^1(\mathbb{P}_{\lambda_0})$ such that for all η in a neighborhood V of $\lambda \in \Theta$:

$$\forall m \leq -M, \quad |\log(f_\Delta(m, \eta))| \leq G_1(m),$$

$$\forall m \geq M, \quad |\log(f_\Delta(m, \eta))| \leq G_2(m).$$

As $(m, \lambda) \mapsto \log(f_\Delta(m, \lambda))$ is continuous on $\mathbb{R} \times \mathbb{R}_+^*$, we have:

$$\forall (m, \eta) \in [-M, M] \times V, \quad |\log(f_\Delta(m, \eta))| \leq K,$$

with $K > 0$. Then, for all $\eta \in V$ and $m \in \mathbb{R}$:

$$\sup_{\eta \in V} |\log(f_\Delta(m, \eta))| \leq \underbrace{G_1(m)\mathbb{1}_{]-\infty, -M[}(m) + K\mathbb{1}_{[-M, M]}(m) + G_2(m)\mathbb{1}_{]M, +\infty[}(m)}_{G(m)}.$$

We easily prove that $G \in L^1(\mathbb{P}_{\lambda_0})$ and the conclusion holds. □

Combining the results of the previous lemma with Proposition 3.2 and the Ergodic Theorem (see [5] e.g.), we can prove Theorem 3.3. More precisely for a sample $(S^{1,r}, \dots, S^{N,r})$ with $N \in \mathbb{N}^*$, the maximum pseudo-likelihood estimator is the maximal argument of the function:

$$M_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \log \left(\frac{f_\Delta(S^{i,r}, \lambda)}{f_\Delta(S^{i,r}, \lambda_0)} \right).$$

For all $\lambda \in \Theta$, the strict concavity of the logarithm function and Jensen's inequality, give:

$$\mathbb{E}_{\lambda_0} \left[\log \left(\frac{f_\Delta(S^{i,r}, \lambda)}{f_\Delta(S^{i,r}, \lambda_0)} \right) \right] < 0. \quad (5.4)$$

According to the Ergodic Theorem, the following convergence holds:

$$M_N(\lambda) \xrightarrow[N \rightarrow +\infty]{\mathbb{P}_{\lambda_0}\text{-p.s.}} M(\lambda) = \mathbb{E}_{\lambda_0} \left[\log \left(\frac{f_\Delta(\cdot, \lambda)}{f_\Delta(\cdot, \lambda_0)} \right) \right]. \quad (5.5)$$

Using Lemma 5.1, we prove that the convergence in Equation (5.5) is uniform.

To conclude, $M(\lambda)$ is the Kullback-Leibler divergence and according to Proposition 3.2, it reaches its maximum for $\lambda = \lambda_0$, proving that the estimator is consistent.

5.3 Proof of Theorem 3.4

Asymptotic normality is a basic property used in statistics, details on the classical proof can be found in [26]. We mainly use the ρ -mixing property to obtain the asymptotic normality of the estimator. As in the previous section, we consider only the case $\theta = (0, \lambda, 1)$, the three parameters case is obtained by a similar reasoning. We introduce the following notations:

$$l_\lambda(m) = \partial_\lambda \log(f_\Delta(m, \lambda)) \quad , \quad \dot{l}_\lambda(m) = \partial_\lambda^2 \log(f_\Delta(m, \lambda)) \quad \text{and} \quad \ddot{l}_\lambda(m) = \partial_\lambda^3 \log(f_\Delta(m, \lambda)),$$

$$\psi_N(\lambda) = \frac{1}{N} \sum_{i=1}^N l_\lambda(S^{i,r}) \quad , \quad \dot{\psi}_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \dot{l}_\lambda(S^{i,r}) \quad \text{and} \quad \ddot{\psi}_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \ddot{l}_\lambda(S^{i,r}).$$

To obtain the asymptotic normality the first step is to perform a Taylor expansion of the function $\psi(\hat{\lambda}_N)$ around λ_0 :

$$\sqrt{N}(\hat{\lambda}_N - \lambda_0) = \frac{-\sqrt{N}\psi_N(\lambda_0)}{\dot{\psi}_N(\lambda_0) + \frac{1}{2}(\hat{\lambda}_N - \lambda_0)\ddot{\psi}_N(\bar{\lambda}_N)}, \quad (5.6)$$

with $\bar{\lambda}_N$ a point located between $\hat{\lambda}_N$ and λ_0 .

Lemma 5.2. *For $r > \frac{\ln(\frac{5}{3})}{\lambda_0}$, the following convergence is satisfied:*

$$\sqrt{N}\psi_N(\lambda_0) \xrightarrow[N \rightarrow +\infty]{law} \mathcal{N}\left(0, \mathbb{E}_{\lambda_0}\left(l_{\lambda_0}(\cdot)^2\right)\right).$$

Proof. Convergence is obtained using a Central Limit Theorem for ρ -mixing sequence [14, Theorem 2]. Three conditions need to be checked to apply this theorem. The first condition is a direct consequence of the ρ -mixing property of $(S^{i,r})_{i \geq 1}$. The second one is satisfied from the asymptotic expansions and the smoothness of the cumulative distribution given in Appendix A.2.

Concerning the last condition, it is necessary to check that $\lim_{N \rightarrow +\infty} \mathbb{V}\left(\sum_{i=1}^N l_{\lambda_0}(S^{i,r})\right) = +\infty$. We notice that, by inclusion of the sigma-fields, the sequence $(l_{\lambda_0}(S^{i,r}))_{i \geq 1}$ is ρ -mixing. Moreover, according to Lemma 5.4 below, for any $i \geq 1$ the random variable $l_{\lambda_0}(S^{i,r})$ is centered. Using Proposition 1.5.1 in [14], we have:

$$\mathbb{V}\left(\sum_{i=1}^N l_{\lambda_0}(S^{i,r})\right) = N\tilde{\sigma}^2 + O(1), \quad (5.7)$$

with $\tilde{\sigma}^2 = \mathbb{E}[l_{\lambda_0}(S^{1,r})^2] + 2 \sum_{i=1}^{+\infty} \mathbb{E}[l_{\lambda_0}(S^{1,r})l_{\lambda_0}(S^{i,r})]$.

Using the ρ -mixing property, $\tilde{\sigma}^2 \geq \mathbb{E}[l_{\lambda_0}(S^{i,r})^2] \frac{3-5e^{-\lambda_0 r}}{1-e^{-\lambda_0 r}}$. By assumption $r > \frac{\ln(\frac{5}{3})}{\lambda_0}$, then $\tilde{\sigma}^2 > 0$. So, according to (5.7):

$$\lim_{N \rightarrow +\infty} \mathbb{V}\left(\sum_{i=1}^N l_{\lambda_0}(S^{i,r})\right) = +\infty.$$

Since all the conditions are satisfied, we can apply Theorem 2 in [14] and we obtain:

$$\sqrt{N}\psi_N(\lambda_0) \xrightarrow[N \rightarrow +\infty]{law} \mathcal{N}\left(0, \mathbb{E}_{\lambda_0}\left(l_{\lambda_0}(\cdot)^2\right)\right).$$

□

Remark 5.3. *The condition $r > \frac{\ln(\frac{5}{3})}{\lambda_0}$ remains the same in the three parameters case.*

The required conditions on the other quantities involved in the Taylor expansion are listed in the following lemma:

Lemma 5.4. *For all $\lambda \in \Theta$, the function $\lambda \mapsto l_\lambda(m)$ is twice continuously derivable for all $m \in \mathbb{R}$. Moreover, $\mathbb{E}_{\mathbb{P}_{\lambda_0}} |l_{\lambda_0}(\cdot)| < \infty$ and $\mathbb{E}_{\mathbb{P}_{\lambda_0}} |\dot{l}_{\lambda_0}(\cdot)|$ exists and is non-zero. The functions l_{λ_0} and \ddot{l}_{λ_0} are respectively dominated by $G(\cdot)$ and $F(\cdot)$ in $L^1(\mathbb{P}_{\lambda_0})$ for all λ in a neighborhood of λ_0 .*

Proof. Integrability and differentiability conditions are satisfied using the asymptotic expansions and the smoothness of the cumulative distribution in Appendix. Domination conditions are verified using the same reasoning as the one in the proof of Lemma 5.1. \square

Combining Lemmas 5.2, 5.4 and Theorem 3.3, we have the following convergence:

$$\sqrt{N}(\hat{\lambda}_N - \lambda_0) \xrightarrow[N \rightarrow +\infty]{law} \mathcal{N}(0, I_{\lambda_0}^{-1}), \quad (5.8)$$

where I_{λ_0} is the Fisher information.

A Appendix

A.1 Parabolic cylinder function

We recall the definition and some auxiliary results about the Parabolic cylinder function. Some of these results can be found in [6, 8, 18].

For all $x, \nu \in \mathbb{R}$, the Parabolic cylinder function $D_\nu(x)$ is a solution of the differential equation:

$$y''(x) + \left(\nu + \frac{1}{2} - \frac{1}{4}x^2 \right) y(x) = 0.$$

Moreover for $\nu \in \mathbb{C}$ with $\text{Re}(\nu) > 0$ and $z \in \mathbb{C}$, the function $(z, \nu) \mapsto D_\nu(z)$ is a holomorphic function ([18], Chapter 10).

The Parabolic cylinder function satisfies the following relations:

$$\partial_x D_\nu(x) = \nu D_{\nu-1}(x) - \frac{x}{2} D_\nu(x),$$

and for all $\nu, a \in \mathbb{R}$,

$$\int_{-\infty}^a e^{-\frac{x^2}{2}} D_\nu(-x\sqrt{2}) dx = \frac{e^{-\frac{a^2}{2}}}{\sqrt{2}} D_{\nu-1}(-a\sqrt{2}). \quad (\text{A.1})$$

We remind now some properties of the Parabolic cylinder function ν -zeros. Let $n \in \mathbb{N}^*$, we denote $\nu_{n,m}$, the positive (ordered) zeros of the function $\nu \mapsto D_\nu(-m\sqrt{2})$. Then

$$\int_{-\infty}^m D_{\nu_{n,m}}^2(-x\sqrt{2}) dx = -\frac{\nu_{n,m}}{\sqrt{2}} D_{\nu_{n,m}-1}(-m\sqrt{2}) \partial_\nu D_{\nu_{n,m}}(-m\sqrt{2}). \quad (\text{A.2})$$

Furthermore, according to [8]:

$$\partial_m \nu_{n,m} = \sqrt{2} \frac{\partial_x D_{\nu_{n,m}}(-m\sqrt{2})}{\partial_\nu D_{\nu_{n,m}}(-m\sqrt{2})} \quad (\text{A.3})$$

$$= -\frac{2}{\sqrt{\pi} \int_0^\infty e^{-(2\nu_{n,m}+1)u+m^2 \tanh(u)} \operatorname{erfc}\left(-m\sqrt{\tanh(u)}\right) \frac{du}{\sqrt{\sinh(u) \cosh(u)}}}. \quad (\text{A.4})$$

The following asymptotic expansion is verified:

$$\nu_{n,m} \underset{m \rightarrow -\infty}{=} \frac{m^2}{2} - \frac{1}{2} - |m|^{\frac{2}{3}} 2^{-\frac{1}{3}} a_n + O\left(|m|^{-\frac{2}{3}}\right), \quad (\text{A.5})$$

where a_n , $n \in \mathbb{N}^*$ are the zeros of the first kind Airy function. Furthermore, the following convergence is verified:

$$\nu_{n,m} \underset{m \rightarrow +\infty}{\longrightarrow} n - 1. \quad (\text{A.6})$$

Recall that the zeros a_n of the first kind Airy function are all real, negative and satisfy the following inequality ([21]), for $n \geq 1$:

$$-\left(\frac{3\pi}{8}(4n-1)\right)^{\frac{2}{3}} \left(1 + \frac{5}{48\left(\frac{3\pi}{8}(4n-1)\right)^2}\right) < a_n \leq -\left(\frac{3\pi}{8}(4n-1)\right)^{\frac{2}{3}}. \quad (\text{A.7})$$

A.2 Smoothness of the cumulative distribution function

Some smoothness properties are needed to prove the set of derivations and continuities of the functions presented in the proof of consistency and asymptotic normality. These properties will be proved using results on holomorphic functions. Making use of (A.3), we can rewrite the cumulative distribution function as:

$$\mathbb{P}(S^{1,0} < m) = -\frac{e^{-\lambda m^2}}{2\sqrt{\lambda\pi}} \sum_{n \geq 1} e^{-\lambda \nu_{n,m,\lambda} \Delta} \frac{\partial_m \nu_{n,m,\lambda}}{\nu_{n,m,\lambda}^2}. \quad (\text{A.8})$$

Proposition A.1. *The cumulative distribution function of the supremum verifies the following properties:*

1. $(m, \lambda) \mapsto \mathbb{P}(S^{1,0} < m)$ is a smooth function on $\mathbb{R} \times \mathbb{R}_+^*$.
2. For any $k, j \in \mathbb{N}^*$:

$$\partial_\lambda^k \partial_m^j \mathbb{P}(S^{1,0} < m) = \sum_{n \geq 1} \partial_\lambda^k \partial_m^j f_n(\Delta, m, \lambda),$$

with:

$$f_n(\Delta, m, \lambda) = -\frac{e^{-m^2 \lambda - \lambda \nu_{n,m,\lambda} \Delta}}{\sqrt{2\pi}} \frac{D_{\nu_{n,m,\lambda}-1}(-m\sqrt{2\lambda})}{\nu_{n,m,\lambda} \partial_\nu D_{\nu_{n,m,\lambda}}(-m\sqrt{2\lambda})}. \quad (\text{A.9})$$

Proof. We introduce the following notation, $\mathbb{C}_{M_1 \leq |\cdot| \leq M_2}^{\theta_1 \leq \arg \leq \theta_2} = \{z \in \mathbb{C} \text{ s.t. } \theta_1 \leq \arg(z) \leq \theta_2, M_1 \leq |z| \leq M_2\}$. We denote $(z, \lambda) \mapsto \tilde{F}_\Delta(z, \lambda)$ the continuation of the cumulative distribution function on $\mathbb{C} \times \mathbb{C}_{|\cdot| \neq 0}^{-\frac{\pi}{2} < \arg < \frac{\pi}{2}}$. According to the Implicit Function Theorem, the function $(z, \lambda) \mapsto \nu_{n,z,\lambda}$ is holomorphic. Then by composition of holomorphic functions, we deduce that (A.9) is holomorphic.

We can write \tilde{F}_Δ as follows:

$$\begin{aligned} \tilde{F}_\Delta(z, \lambda) &= \tilde{G}_1(z, \lambda) \mathbb{1}_{\mathbb{C}_{|\cdot| > M}^{\frac{\pi}{2} < \arg < \frac{3\pi}{2}} \times \mathbb{C}_{|\cdot| > M}^{-\frac{\pi}{2} \leq \arg < \frac{\pi}{2}}(z, \lambda) + \tilde{G}_2(z, \lambda) \mathbb{1}_{\mathbb{C}_{|\cdot| < M} \times \mathbb{C}_{0 < |\cdot| < M}^{-\frac{\pi}{2} \leq \arg < \frac{\pi}{2}}(z, \lambda) \\ &\quad + \tilde{G}_3(z, \lambda) \mathbb{1}_{\mathbb{C}_{|\cdot| > M}^{-\frac{\pi}{2} < \arg < \frac{\pi}{2}} \times \mathbb{C}_{|\cdot| > M}^{-\frac{\pi}{2} \leq \arg < \frac{\pi}{2}}(z, \lambda), \end{aligned}$$

with M large enough.

For \tilde{G}_1 and \tilde{G}_3 , using the asymptotic expansion (A.5) and the limit (A.6), we easily prove the normal convergence of the associate series.

For \tilde{G}_2 , the normal convergence can be obtained using the following equivalence from [3]:

$$\nu_{n,z,\lambda} \underset{n \rightarrow +\infty}{\sim} 2n - 1 + \frac{4\lambda z^2}{\pi^2} - 2\frac{\sqrt{\lambda z^2}}{\pi} \sqrt{4n - 1 + 4\frac{\lambda z^2}{\pi^2}}$$

and

$$D_\nu(z) \underset{\nu \rightarrow +\infty}{=} \sqrt{2} \left(\nu + \frac{1}{2} \right)^{\frac{\nu}{2}} e^{-(\frac{\nu}{2} + \frac{1}{4})} \cos \left(z \sqrt{\nu + \frac{1}{2}} - \frac{\pi \nu}{2} \right) (1 + O(\nu^{-\frac{1}{2}})).$$

From Theorem 3.2 in [20], the conclusion holds. \square

A.3 Asymptotic expansions

For the integrability conditions required in the Ergodic Theorem, some asymptotic expansions on the cumulative distribution and the probability density of the supremum $S^{1,0}$ are provided. In the following proofs, without loss of generality we assume $\theta = (0, \lambda, 1)$. To return to the three parameters case, we replace m , Δ , λ respectively by $m - \frac{\mu}{\lambda}$, $\beta\Delta$ and $\frac{\lambda}{\beta}$.

A.3.1 For large positive m

Since the zeros $\nu_{n,m,\lambda}$ tend to positive integers when m goes to infinity, then we are able to give an asymptotic expansion for (5.1).

Proposition A.2. *For large positive m , the cumulative distribution function of $S^{1,0}$ has the following asymptotic expansion:*

$$\mathbb{P}(S^{1,0} < m) \underset{m \rightarrow +\infty}{=} 1 - \frac{e^{-\lambda m^2}}{2\sqrt{\pi\lambda m}} (1 + o(m^{-2+\delta})), \quad (\text{A.10})$$

with $0 < \delta < 2$.

Proof. Recall that the cumulative distribution function is given by (5.1). Using Formula (A.6), we obtain:

$$\mathbb{P}(S^{1,0} < m) \underset{m \rightarrow +\infty}{=} \frac{1}{2\sqrt{\pi}} e^{-\lambda m^2} \left[\frac{D_{-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_0^2(-x\sqrt{2})dx} + \sum_{n \geq 1} e^{-\lambda n \Delta} \frac{D_{n-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_n^2(-x\sqrt{2})dx} \right]. \quad (\text{A.11})$$

According to Formula (10.5.4) in [18]:

$$\frac{e^{-\lambda m^2}}{2\sqrt{\pi}} \frac{D_{-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_0^2(-x\sqrt{2})dx} = \frac{1}{2} \left(1 + \operatorname{erf}(m\sqrt{\lambda})\right),$$

where $\operatorname{erf}(m) = \frac{2}{\sqrt{\pi}} \int_0^m e^{-t^2} dt$ is the Error function. Then applying Formulas (4.9.6) and (4.13.4) in [18], we get:

$$\sum_{n=1}^{\infty} e^{-\lambda n \Delta} \frac{D_{n-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_n^2(-x\sqrt{2\lambda})dx} \underset{m \rightarrow +\infty}{=} e^{-\lambda m^2 \frac{1-e^{-\lambda\Delta}}{1+e^{-\lambda\Delta}}} O(1).$$

As we combine these two results, we get:

$$\mathbb{P}(S^{1,0} < m) \underset{m \rightarrow +\infty}{=} \frac{1}{2} \left(1 + \operatorname{erf}(m\sqrt{\lambda})\right) + e^{-\lambda m^2 \frac{1-e^{-\lambda\Delta}}{1+e^{-\lambda\Delta}}} O(1).$$

Since $1 - \operatorname{erf}(m) \underset{m \rightarrow +\infty}{=} \frac{e^{-m^2}}{\sqrt{\pi}m} (1 + O(m^{-2}))$, the conclusion holds. \square

Proposition A.3. *For large positive m , the asymptotic expansion of the probability density of $S^{1,0}$ is given by:*

$$f_{\Delta}(m, \lambda) \underset{m \rightarrow +\infty}{=} \sqrt{\frac{\lambda}{\pi}} e^{-\lambda m^2} \left(1 + m e^{-\lambda m^2 \frac{1-e^{-\lambda\Delta}}{1+e^{-\lambda\Delta}}} O(1)\right), \quad (\text{A.12})$$

with $|O(1)| \leq \frac{4\sqrt{\lambda}e^{-\lambda\Delta}}{\sqrt{\pi(1-e^{-2\lambda\Delta})}}$.

Proof. When m goes to infinity, using (A.11) the derivative of the cumulative distribution function satisfies:

$$f_{\Delta}(m, \lambda) \underset{m \rightarrow +\infty}{=} \sqrt{\frac{\lambda}{\pi}} e^{-\lambda m^2} + \frac{1}{2\sqrt{\pi}} \partial_m \left[e^{-\lambda m^2} \sum_{n \geq 1} e^{-\lambda n \Delta} \frac{D_{n-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_n^2(-x\sqrt{2})dx} \right].$$

Using Formulas (4.9.5) and (4.13.4) in [18], one can prove that:

$$\partial_m \left[e^{-\lambda m^2} \sum_{n \geq 1} e^{-\lambda n \Delta} \frac{D_{n-1}^2(-m\sqrt{2\lambda})}{\int_{-\infty}^{m\sqrt{\lambda}} D_n^2(-x\sqrt{2})dx} \right] \underset{m \rightarrow +\infty}{=} m e^{-\lambda m^2 \frac{2\lambda m^2}{1+e^{-\lambda\Delta}}} O(1),$$

with $|O(1)| \leq \frac{8\lambda e^{-\lambda\Delta}}{\sqrt{\pi(1-e^{-2\lambda\Delta})}}$. \square

Remark A.4. *We have:*

$$f_{\Delta}(m, \lambda) \underset{m \rightarrow +\infty}{=} e^{-\lambda m^2} \sqrt{\frac{\lambda}{\pi}} \left(1 + o(m^{-\alpha})\right),$$

with $\alpha > 0$.

Corollary A.4.1. *For large positive m ,*

$$\log(f_{\Delta}(m, \lambda)) \underset{m \rightarrow +\infty}{=} -\lambda m^2 + \frac{1}{2} \log\left(\frac{\lambda}{\pi}\right) + m e^{-m^2 \lambda \frac{1-e^{-\lambda \Delta}}{1+e^{-\lambda \Delta}}} O(1), \quad (\text{A.13})$$

$$\text{with } |O(1)| \leq \frac{16\sqrt{\lambda}e^{-\lambda \Delta}}{3\sqrt{\pi(1-e^{-2\lambda \Delta})}}.$$

A similar reasoning as the one in the proof of Proposition A.2 and Proposition A.3 may be applied to prove the following results:

Proposition A.5. *For large positive m , the following asymptotic expansions are satisfied:*

1. $\partial_{\lambda} \log(f_{\Delta}(m, \lambda)) \underset{m \rightarrow +\infty}{=} -m^2 + \frac{1}{2\lambda} + m^3 e^{-m^2 \lambda \frac{1-e^{-\lambda \Delta}}{1+e^{-\lambda \Delta}}} O(1).$
2. $\partial_{\lambda}^2 \log(f_{\Delta}(m, \lambda)) \underset{m \rightarrow +\infty}{=} -\frac{1}{2\lambda^2} + m^5 e^{-m^2 \lambda \frac{1-e^{-\lambda \Delta}}{1+e^{-\lambda \Delta}}} O(1).$
3. $\partial_{\lambda}^3 \log(f_{\Delta}(m, \lambda)) \underset{m \rightarrow +\infty}{=} \frac{1}{\lambda^3} + m^7 e^{-m^2 \lambda \frac{1-e^{-\lambda \Delta}}{1+e^{-\lambda \Delta}}} O(1).$

A.3.2 For large negative m

Using (A.4) and (A.5), we can give an asymptotic expansion for the cumulative distribution function of $S^{1,0}$ for large negative m .

Proposition A.6. *For large negative m , the cumulative distribution function of $S^{1,0}$ has the following asymptotic expansion:*

$$\mathbb{P}(S^{1,0} < m) = 2|m\sqrt{\lambda}|^{-3} e^{\frac{-\lambda m^2 - \left(\frac{m^2 \lambda}{2} - \frac{1}{2}\right) \lambda \Delta + |m|^{\frac{2}{3}} \lambda^{\frac{4}{3}} 2^{-\frac{1}{3}} a_1 \Delta}}{\sqrt{\pi}} (1 + o(1)), \quad (\text{A.14})$$

where a_1 is the first zero of the Airy function of the first kind.

Sketch of the Proof. Formula (A.8) gives:

$$\mathbb{P}(S^{1,0} < m) = -\frac{e^{-\lambda m^2 - \lambda \Delta \nu_{1,m,\lambda}}}{2\sqrt{\pi}} \frac{\partial_m \nu_{1,m,\lambda}}{\sqrt{\lambda} \nu_{1,m,\lambda}^2} \left(1 + \sum_{n \geq 2} e^{-\lambda \Delta (\nu_{n,m,\lambda} - \nu_{1,m,\lambda})} \frac{\nu_{1,m,\lambda}^2 \partial_m \nu_{n,m,\lambda}}{\nu_{n,m,\lambda}^2 \partial_m \nu_{1,m,\lambda}} \right).$$

From the asymptotic expansion of ν -zeros for large negative m (A.5) and (A.4), it follows that:

$$\begin{aligned} \frac{\partial_m \nu_{1,m,\lambda}}{\sqrt{\lambda} \nu_{1,m,\lambda}^2} &\underset{m \rightarrow -\infty}{=} -4|m\sqrt{\lambda}|^{-3}(1 + o(1)), \\ \frac{\nu_{1,m,\lambda}^2 \partial_m \nu_{n,m,\lambda}}{\nu_{n,m,\lambda}^2 \partial_m \nu_{1,m,\lambda}} &\underset{m \rightarrow -\infty}{\longrightarrow} 1 \quad \text{if } n < N(m, \lambda) = \lfloor \frac{2\lambda m^2}{3\pi} + \frac{1}{4} \rfloor + 1, \\ \frac{\nu_{1,m,\lambda}^2 \partial_m \nu_{n,m,\lambda}}{\nu_{n,m,\lambda}^2 \partial_m \nu_{1,m,\lambda}} &\underset{m \rightarrow -\infty, n \rightarrow \infty}{\longrightarrow} 0 \quad \text{if } n > N(m, \lambda). \end{aligned}$$

Therefore for all $c > 1$, there exists $M > 0$ such that for all $m < -M$, we get:

$$\sum_{n \geq 2} e^{-\lambda \Delta (\nu_{n,m,\lambda} - \nu_{1,m,\lambda})} \frac{\nu_{1,m,\lambda}^2 \partial_m \nu_{n,m,\lambda}}{\nu_{n,m,\lambda}^2 \partial_m \nu_{1,m,\lambda}} < c \sum_{n \geq 2} e^{-\lambda \Delta (\nu_{n,m,\lambda} - \nu_{1,m,\lambda})}.$$

Using (A.7), we easily prove that for large negative m ,

$$\sum_{n \geq 2} e^{-\lambda \Delta (\nu_{n,m,\lambda} - \nu_{1,m,\lambda})} = e^{-\lambda \Delta (\nu_{2,m,\lambda} - \nu_{1,m,\lambda})} O(1) = o(1).$$

We then conclude the proof with (A.5). □

Using similar arguments, we can prove the following result:

Proposition A.7. *For large negative m , the asymptotic expansion of the probability density of $S^{1,0}$ is given by:*

$$f_\Delta(m, \lambda) = 2|m|^{-2} \frac{e^{-\lambda m^2 - \left(\frac{m^2 \lambda}{2} - \frac{1}{2}\right) \lambda \Delta + |m|^{\frac{2}{3}} \lambda^{\frac{4}{3}} 2^{-\frac{1}{3}} a_1 \Delta}}{\sqrt{\pi}} (2 + \Delta \lambda) (1 + O(|m|^{-\frac{2}{3}})). \quad (\text{A.15})$$

Corollary A.7.1. *For large negative m ,*

$$\log f_\Delta(m, \lambda) = -\lambda m^2 - \left(\frac{m^2 \lambda}{2} - \frac{1}{2}\right) \lambda \Delta + |m|^{\frac{2}{3}} \lambda^{\frac{4}{3}} 2^{-\frac{1}{3}} a_1 \Delta + \log \frac{2(2 + \Delta \lambda)}{\sqrt{\pi}} - 2 \log(|m|) + O(|m|^{-\frac{2}{3}}).$$

A similar reasoning as the one in the proof of Proposition A.6 may be applied to prove the following:

Proposition A.8. *For large negative m , the following asymptotic expansions are satisfied:*

1. $\partial_\lambda \log(f_\Delta(m, \lambda)) = -m^2 - \Delta \left(m^2 \lambda - \frac{1}{2} - |m|^{\frac{2}{3}} \frac{2^{\frac{2}{3}} \lambda^{-\frac{2}{3}}}{3} a_1 \right) + \frac{\Delta}{2+\Delta\lambda} + |m|^{-\frac{2}{3}} O(1).$
2. $\partial_\lambda^2 \log(f_\Delta(m, \lambda)) = -\Delta \left(m^2 + |m|^{\frac{2}{3}} \frac{2^{\frac{5}{3}} \lambda^{-\frac{5}{3}}}{9} a_1 \right) - \frac{\Delta^2}{(2+\Delta\lambda)^2} + |m|^{-\frac{2}{3}} O(1).$
3. $\partial_\lambda^3 \log(f_\Delta(m, \lambda)) = \Delta |m|^{\frac{2}{3}} \frac{5}{27} 2^{\frac{5}{3}} \lambda^{-\frac{8}{3}} a_1 + \frac{2\Delta^3}{(2+\Delta\lambda)^3} + |m|^{-\frac{2}{3}} O(1).$

B Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgements

We are grateful to two referees for their helpful remarks and suggestions.

References

- [1] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.
- [2] P. Alaton, B. Djehiche, and D. Stillberger. On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9(1):1–20, 2002.
- [3] L. Alili, P. Patie, and J. Pedersen. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models*, 21(4):967–980, 2005.
- [4] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

- [5] P. Billingsley. *Ergodic Theory and Information*, volume 1. Wiley New York, 1965.
- [6] C. Blanchet-Scalliet, D. Dorobantu, and L. Gay. Joint law of an Ornstein-Uhlenbeck process and its supremum. *Journal of Applied Probability*, 57(2):541–558, 2020.
- [7] C. Blanchet-Scalliet, D. Dorobantu, L. Gay, V. Maume-Deschamps, and P. Ribereau. Risk assessment using suprema data. *Stochastic Environmental Research and Risk Assessment*, 32(10):2839–2848, 2018.
- [8] C. Blanchet-Scalliet, D. Dorobantu, and B. Nieto. Some properties for ν -zeros of Parabolic Cylinder functions. *Le Matematiche*. to appear.
- [9] R. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.
- [10] D. Brody, J. Syroka, and M. Zervos. Dynamical pricing of weather derivatives. *Quantitative Finance*, 2(3):189–198, 2002.
- [11] D. Dacunha-Castelle and M. Duflo. *Probability and Statistics: Volume II*, volume 2. Springer Science & Business Media, 2012.
- [12] B. Dischel. At last: A model for weather risk. *Energy and Power Risk Management*, 11(3):20–21, 1998.
- [13] F. Dornier and M. Queruel. Caution to the wind. *Energy & Power Risk Management*, 13(8):30–32, 2000.
- [14] P. Doukhan. *Mixing: properties and examples*, volume 85. Springer Science & Business Media, 2012.
- [15] E. Gobet and G. Matulewicz. Parameter estimation of Ornstein-Uhlenbeck process generating a stochastic graph. *Statistical Inference for Stochastic Processes*, 20(2):211–235, 2017.
- [16] A.M.G. Klein Tank, J.B. Wijngaard, G.P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, et al. Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22(12):1441–1453, 2002.

- [17] K. Laaidi, M. Pascal, M. Ledrans, A. Le Tertre, S. Medina, and C. Schonemann. Le système français d’alerte canicule et santé 2004 (Sacs 2004). un dispositif intégré au plan national canicule. *Bulletin épidémiologique hebdomadaire*, (30-31):134–136, 2004.
- [18] N. Lebedev. Special functions and their applications. *Physics Today*, 18(12):70, 1965.
- [19] P. Mullowney and S. Iyengar. Parameter estimation for a leaky integrate-and-fire neuronal model from ISI data. *Journal of Computational Neuroscience*, 24(2):179–194, 2008.
- [20] B.P. Palka. *An Introduction to Complex Function Theory*. Springer Science & Business Media, 1991.
- [21] G. Pittaluga and L. Sacripante. Inequalities for the zeros of the Airy functions. *SIAM Journal on Mathematical Analysis*, 22(1):260–267, 1991.
- [22] J. Robert, V. John, and P. William. Pairs trading. *Quantitative Finance*, 5(3):271–276, 2005.
- [23] G. Saporta. *Probabilités, analyse des données et statistique*. Editions technip, 2006.
- [24] W.F. Stout. *Almost Sure Convergence*. Academic Press Inc, 1974.
- [25] D.C. Trost, E. A. Overman, J.H. Ostroff, W. Xiong, and P. March. A Model for Liver Homeostasis Using Modified Mean-Reverting Ornstein–Uhlenbeck Process. *Computational and Mathematical Methods in Medicine*, 11(3):21, 2010.
- [26] A.W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.
- [27] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.
- [28] M. Wilkinson and A. Pumir. Spherical Ornstein-Uhlenbeck processes. *Journal of Statistical Physics*, 145(1):113–142, 2011.