ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 11, No.2
June 2013

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 454 084
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *190 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT included]

- Single issue ………………………………………………………….. **€ 11**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)………. **€ 26**
- Annual subscription (No. 2, No. 3) ……………………………….. **€ 18**

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 121–134

THE EXACT JOINT DISTRIBUTION OF CONCOMITANTS OF ORDER STATISTICS AND THEIR ORDER STATISTICS UNDER NORMALITY

Authors: Ayyub Sheikhi
– Department of Statistics, Shahid Bahonar University of Kerman, Iran
sheikhy.a@uk.ac.ir

Mahbanoo Tata
– Department of Statistics, Shahid Bahonar University of Kerman, Iran
tata@uk.ac.ir

Received: December 2011 Revised: May 2012 Accepted: August 2012

Abstract:

- In this work we derive the exact joint distribution of a linear combination of concomitants of order statistics and linear combinations of their order statistics in a multivariate normal distribution. We also investigate a special case of related joint distributions discussed by He and Nagaraja (2009).

Key-Words:

- *unified skew-normal; L-statistic; multivariate normal distribution; order statistic; concomitant; orthant probability.*

AMS Subject Classification:

- 62G30, 62H10.

1. INTRODUCTION

Suppose that the joint distribution of two n-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ follows a $2n$ dimensional multivariate normal vector with positive definite covariance matrix, i.e.

$$\text{(1.1)} \qquad \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n}\left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \ \sum = \begin{pmatrix} \sum_{\mathbf{xx}} & \sum_{\mathbf{xy}} \\ \sum_{\mathbf{xy}}^T & \sum_{\mathbf{yy}} \end{pmatrix}\right)$$

where $\boldsymbol{\mu}_{\mathbf{x}}$, $\boldsymbol{\mu}_{\mathbf{y}}$ are respectively the mean vectors and $\sum_{\mathbf{xx}}$, $\sum_{\mathbf{yy}}$ are the positive definite variance matrices of $\mathbf{X}$ and $\mathbf{Y}$, while $\sum_{\mathbf{xy}}$ is their covariance matrix. Let $\mathbf{X}_{(n)} = (X_{1:n},\ X_{2:n}, ...,\ X_{n:n})^T$ be the vector of order statistics obtained from $\mathbf{X}$ and $\mathbf{Y}_{(n)} = (Y_{1:n},\ Y_{2:n}, ...,\ Y_{n:n})^T$ be the vector of order statistics obtained from $\mathbf{Y}$. Further, let $\mathbf{Y}_{[n]} = (Y_{[1:n]},\ Y_{[2:n]}, ...,\ Y_{[n:n]})^T$ be the vector of Y-variates paired with the order statistics of $\mathbf{X}$. The elements of $\mathbf{Y}_{[n]}$ are called the concomitants of the order statistics of $\mathbf{X}$.

Nagaraja (1982) has obtained the distribution of a linear combination of order statistics from a bivariate normal random vector where the variables are exchangeable. Loperfido (2008a) has extended the results of Nagaraja (1982) to elliptical distributions. Arellano-Valle and Genton (2007) have expressed the exact distribution of linear combinations of order statistics from dependent random variables. Sheikhi and Jamalizadeh (2011) have showed that for arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$, the distribution of $(X,\ \mathbf{a}^T\mathbf{Y}_{(2)},\ \mathbf{b}^T\mathbf{Y}_{(2)})^T$ is a singular skew-normal and carried out a regression analysis. Yang (1981) has studied the linear combination of concomitants of order statistics. Tsukibayashi (1998) has obtained the joint distribution of $(Y_{i:n}, Y_{[i:n]})$, while He and Nagaraja (2009) have obtained the joint distribution of $(Y_{i:n}, Y_{[j:n]})$ for all $i,\ j = 1,\ 2, ...,\ n$. Goel and Hall (1994) have discussed the difference between concomitants and order statistics using the sum $\sum_{i=1}^n h(Y_{i:n} - Y_{[i:n]})$ for some smooth function h. Recently much attention has been focused on the connection between order statistics and skew-normal distributions (see e.g. Loperfido 2008a and 2008b and Sheikhi and Jamalizadeh 2011). In this article we shall obtain the joint distribution of $\mathbf{a}^T\mathbf{Y}_{(n)}$ and $\mathbf{b}^T\mathbf{Y}_{[n]}$, where $\mathbf{a} = (a_1,\ a_2, ...,\ a_n)^T$ and $\mathbf{b} = (b_1,\ b_2, ...,\ b_n)^T$ are arbitrary vectors in $\mathbb{R}^n$. Since we do not assume independence, our results are more general than those of He and Nagaraja (2009). On the other hand, He and Nagaraja (2009) have not assumed normality.

The concept of the skew-normal distribution was proposed independently by Roberts (1966), Ainger *et al.* (1977), Andel *et al.* (1984) and Azzalini (1985). The univariate random variable Y has a skew-normal distribution if its distribution can be written as

$$\text{(1.2)} \qquad f_Y(y) = 2\varphi\left(y;\ \mu,\ \sigma^2\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right) \qquad y \in \mathbb{R}$$

where $\varphi\left(.;\ \mu,\ \sigma^2\right)$ is the normal density with mean μ and variance σ^2 and $\Phi(.)$ denotes the standard normal distribution function.

Following Arellano-Valle and Azzalini (2006), a d-dimensional random vector $\mathbf{Y}$ is said to have a unified multivariate skew-normal distribution ($\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi},\ \boldsymbol{\delta},\ \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$), if it has a density function of the form

(1.3)
$$f_{\mathbf{Y}}(\mathbf{y}) = \varphi_d(\mathbf{y};\ \boldsymbol{\xi},\ \boldsymbol{\Omega}) \frac{\Phi_m\left(\boldsymbol{\delta} + \boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi});\ \ \boldsymbol{\Gamma} - \boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\right)}{\Phi_m(\boldsymbol{\delta};\ \boldsymbol{\Gamma})} \qquad \mathbf{y} \in \mathbb{R}^d$$

where $\varphi_d(.,\ \boldsymbol{\xi},\ \boldsymbol{\Omega})$ is the density function of a multivariate normal and $\Phi_m(.;\ \Sigma)$ is the multivariate normal cumulative function with the covariance matrix Σ.

If $\Sigma^* = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}$ is a singular matrix we say that the distribution of $\mathbf{X}$ is singular unified skew-normal and write $SSUN_{d,m}(\boldsymbol{\xi},\ \boldsymbol{\delta}, \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$. For more details see Arellano-Valle and Azzalini (2006) and Sheikhi and Jamalizadeh (2011).

In Section 2, we show that for two vectors $\mathbf{a}$ and $\mathbf{b}$, the joint distribution of $\mathbf{a}^T\mathbf{Y}_{(n)}$ and $\mathbf{b}^T\mathbf{Y}_{[n]}$ belongs to the unified multivariate skew-normal family. We also discuss special cases of these distributions under the setting of independent normal random variables. Finally, in section 3 we present a numerical application of our results.

2. MAIN RESULTS

Define $S(\mathbf{X})$ as the class of all permutation of components of the random vector $\mathbf{X}$, i.e. $S(\mathbf{X}) = \{\mathbf{X}^{(i)} = \mathbf{P}_i\mathbf{X};\ i = 1,\ 2, ...,\ n!\}$, where $\mathbf{P}_i$ is an $n \times n$ permutation matrix. Also, suppose $\boldsymbol{\Delta}$ is the difference matrix of dimension $(n-1) \times n$ such that the ith row of $\boldsymbol{\Delta}$ is $\mathbf{e}^T_{n,\,i+1} - \ \mathbf{e}^T_{n,\,i}$, $i = 1, ..., n-1$, where $\mathbf{e}_1,\ \mathbf{e}_2, ...,\ \mathbf{e}_n$ are n-dimensional unit basis vectors. Then $\boldsymbol{\Delta}\mathbf{X} = (X_2 - X_1,\ X_3 - X_2, ...,\ X_n - X_{n-1})^T$. (See e.g. Crocetta and Loperfido 2005).

Further, let $\mathbf{X}^{(i)}$ and $\mathbf{Y}^{(i)}$ be the ith permutation of the random vectors $\mathbf{X}$ and $\mathbf{Y}$ respectively. We write $G_{ij}(\mathbf{t},\ \boldsymbol{\xi},\ \sum) = P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0},\ \ \boldsymbol{\Delta}\mathbf{Y}^{(j)} \geq \mathbf{0}\right)$.

Theorem 2.1. *Suppose the matrix* $\begin{pmatrix} \boldsymbol{\Delta} \\ \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix}$ *is of full rank. Then under the assumption of model (1.1) the cdf of the random vector* $\left(\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}\right)^T$ *is the mixture*

$$F_{\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}}(y_1, y_2) = \sum_{i=1}^{n!}\sum_{j=1}^{n!} F_{SUN}(y_1, y_2;\ \boldsymbol{\xi}_{ij}, \boldsymbol{\delta}_{ij}, \boldsymbol{\Gamma}_{ij}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\Lambda}_{ij})\, G_{ij}(t, \boldsymbol{\xi}, \sum)$$

where $F_{SUN}(.\,;\boldsymbol{\xi}_{ij},\ \boldsymbol{\delta}_{ij}, \boldsymbol{\Gamma}_{ij}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\Lambda}_{ij})$ *is the cdf of unified multivariate skew-normal with*

$$\boldsymbol{\xi}_{ij}=\begin{pmatrix} \mathbf{a}^T\boldsymbol{\mu}_{\mathbf{y}}^{(i)} \\ \mathbf{b}^T\boldsymbol{\mu}_{\mathbf{y}}^{(i)} \end{pmatrix},\ \boldsymbol{\delta}_{ij}=\begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{\mu}_{\mathbf{x}}^{(i)} \\ \boldsymbol{\Delta}\boldsymbol{\mu}_{\mathbf{y}}^{(j)} \end{pmatrix},\ \boldsymbol{\Gamma}_{ij}=\begin{pmatrix} \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(ii)}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ij)}\boldsymbol{\Delta}^T \\ & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(jj)}\boldsymbol{\Delta}^T \end{pmatrix},$$

$$\boldsymbol{\Omega}_{ij}=\begin{pmatrix} \mathbf{a}^T\sum_{\mathbf{yy}}^{(ii)}\mathbf{a} & \mathbf{a}^T\sum_{\mathbf{yy}}^{(ij)}\mathbf{b} \\ & \mathbf{b}^T\sum_{\mathbf{yy}}^{(jj)}\mathbf{b} \end{pmatrix} \quad \textit{and} \quad \boldsymbol{\Lambda}_{ij}=\begin{pmatrix} \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ii)}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ij)}\mathbf{b} \\ \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(ij)}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(jj)}\mathbf{b} \end{pmatrix}^T$$

where $\boldsymbol{\mu}_{\mathbf{x}}^{(i)}$ *and* $\boldsymbol{\mu}_{\mathbf{y}}^{(j)}$ *are respectively the mean vectors of the* i*th permutation of the random vector* $\mathbf{X}$ *and the* j*th permutation of the random vector* $\mathbf{Y}$ *and* $\sum_{\mathbf{xx}}^{(ii)}=Var(\mathbf{X}^{(i)})$, $\sum_{\mathbf{yy}}^{(jj)}=Var(\mathbf{Y}^{(j)})$ *and* $\sum_{\mathbf{xy}}^{(ij)}=Cov(\mathbf{X}^{(i)},\ \mathbf{Y}^{(j)})$.

Proof: We have

$$\begin{aligned} F_{\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}}(y_1,\,y_2) &= P\left(\mathbf{a}^T\mathbf{Y}_{(n)}\le y_1,\ \mathbf{b}^T\mathbf{Y}_{[n]}\le y_2\right) \\ &= \sum_{i=1}^{n!}\sum_{j=1}^{n!} P\left(\mathbf{a}^T\mathbf{Y}^{(i)}\le y_1,\ \mathbf{b}^T\mathbf{Y}^{(j)}\le y_2|\mathbf{A}^{(ij)}\right)P\left(\mathbf{A}^{(ij)}\right) \end{aligned}$$

where $\mathbf{A}^{(ij)}=\{\boldsymbol{\Delta}\mathbf{X}^{(i)}\ge\mathbf{0},\ \boldsymbol{\Delta}\mathbf{Y}^{(j)}\ge\mathbf{0}\}$. Since

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X}^{(i)} \\ \boldsymbol{\Delta}\mathbf{Y}^{(j)} \\ \mathbf{a}^T\mathbf{Y}^{(i)} \\ \mathbf{b}^T\mathbf{Y}^{(j)} \end{pmatrix}_{2n\times1} = \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{a}^T \\ \mathbf{0} & \mathbf{b}^T & \mathbf{0} \end{pmatrix}_{2n\times3n} \begin{pmatrix} \mathbf{X}^{(i)} \\ \mathbf{Y}^{(j)} \\ \mathbf{Y}^{(i)} \end{pmatrix}_{3n\times1}, \tag{2.1}$$

the full rank assumption implies nonsingularity of the matrix on the right hand side of (2.1). Furthermore,

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X}^{(i)} \\ \boldsymbol{\Delta}\mathbf{Y}^{(j)} \\ \mathbf{a}^T\mathbf{Y}^{(i)} \\ \mathbf{b}^T\mathbf{Y}^{(j)} \end{pmatrix} \sim N_{2n}\left(\begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{\mu}_{\mathbf{x}}^{(i)} \\ \boldsymbol{\Delta}\boldsymbol{\mu}_{\mathbf{y}}^{(j)} \\ \mathbf{a}^T\boldsymbol{\mu}_{\mathbf{y}}^{(i)} \\ \mathbf{b}^T\boldsymbol{\mu}_{\mathbf{y}}^{(j)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(ii)}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ij)}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ii)}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{xy}}^{(ij)}\mathbf{b} \\ & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(jj)}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(ij)}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(jj)}\mathbf{b} \\ & & \mathbf{a}^T\sum_{\mathbf{yy}}^{(ii)}\mathbf{a} & \mathbf{a}^T\sum_{\mathbf{yy}}^{(ij)}\mathbf{b} \\ & & & \mathbf{b}^T\sum_{\mathbf{yy}}^{(jj)}\mathbf{b} \end{pmatrix}\right).$$

Now, similar to Sheikhi and Jamalizadeh (2011), we immediately conclude that

$$\left(\mathbf{a}^T\mathbf{Y}^{(i)},\,\mathbf{b}^T\mathbf{Y}^{(j)}\right)^T|\,\boldsymbol{\Delta}\mathbf{X}^{(i)}\ge\mathbf{0},\,\boldsymbol{\Delta}\mathbf{Y}^{(j)}\ge\mathbf{0}\sim SUN_{2,\,2(n-1)}(\boldsymbol{\xi}_{ij},\,\boldsymbol{\delta}_{ij},\,\boldsymbol{\Gamma}_{ij},\,\boldsymbol{\Omega}_{ij},\,\boldsymbol{\Lambda}_{ij})\,.$$

This proves the Theorem. $\square$

Remark 2.1. If the rank of the matrix $\left(\boldsymbol{\Delta},\ \mathbf{a}^T, \mathbf{b}^T\right)^T$ is at most $n-1$, the joint distribution of $\left(\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}\right)^T$ is a mixture of a unified skew-normals and a singular unified skew-normals. In this section we assume that the matrix $\left(\boldsymbol{\Delta},\ \mathbf{a}^T, \mathbf{b}^T\right)^T$ is of full rank. A special case will be investigated later in the paper.

Let (X_i, Y_i), $i = 1,\ 2, ...,\ n$ be a random sample of size n from a bivariate normal $N_2\left(\mu_x,\ \mu_y,\ \sigma_x^2,\ \sigma_y^2,\ \rho\right)$, then the model (1.1) reduces to the following:

$$\text{(2.2)} \qquad \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n}\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_x \mathbf{1}_n \\ \mu_y \mathbf{1}_n \end{pmatrix}, \sum = \begin{pmatrix} \sum_{\mathbf{xx}} & \sum_{\mathbf{xy}} \\ & \sum_{\mathbf{yy}} \end{pmatrix}\right)$$

where $\sum_{\mathbf{xx}} = \sigma_x^2 \mathbf{I}_n$, $\sum_{\mathbf{yy}} = \sigma_y^2 \mathbf{I}_n$ and $\sum_{\mathbf{xy}} = \rho\sigma_x\sigma_y \mathbf{1}_n \mathbf{1}_n^T$ where ρ is the correlation coefficient between X and Y.

The following corollary describes the joint distribution of a linear combination of concomitants of order statistics and a linear combination of their order statistics under the independence assumption.

Corollary 2.1. *Suppose the matrix* $\left(\boldsymbol{\Delta},\ \mathbf{a}^T,\ \mathbf{b}^T\right)^T$ *is of full rank. Then under the assumption of model (2.2) the distribution of the random vector* $\left(\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}\right)^T$ *is* $SUN_{2,\ 2(n-1)}\left(\boldsymbol{\xi},\ \mathbf{0}_{2n-2},\ \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda}\right)$ *where*

$$\boldsymbol{\xi} = \begin{pmatrix} \mu_x \mathbf{a}^T \mathbf{1}_n \\ \mu_y \mathbf{b}^T \mathbf{1}_n \end{pmatrix},\quad \boldsymbol{\Omega} = \sigma_y^2 \begin{pmatrix} \mathbf{a}^T\mathbf{a} & \mathbf{a}^T\mathbf{b} \\ & \mathbf{b}^T\mathbf{b} \end{pmatrix},\quad \boldsymbol{\Gamma} = \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}\boldsymbol{\Delta}^T \\ & \sigma_y^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^T \end{pmatrix},$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \rho\sigma_x\sigma_y \boldsymbol{\Delta}\mathbf{a} & \rho\sigma_x\sigma_y \boldsymbol{\Delta}\mathbf{b} \\ \sigma_y^2 \boldsymbol{\Delta}\mathbf{a} & \sigma_y^2 \boldsymbol{\Delta}\mathbf{b} \end{pmatrix}.$$

Proof: We have

$$\begin{aligned} F_{\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}}(y_1,\ y_2) &= P\left(\mathbf{a}^T\mathbf{Y}_{(n)} \le y_1,\ \mathbf{b}^T\mathbf{Y}_{[n]} \le\ y_2\right) \\ &= \sum_{i=1}^{n!}\sum_{j=1}^{n!} P\left(\mathbf{a}^T\mathbf{Y}^{(i)} \le y_1, \mathbf{b}^T\mathbf{Y}^{(j)} \le y_2 | \mathbf{A}^{(ij)}\right) P\left(\mathbf{A}^{(ij)}\right). \end{aligned}$$

Since $P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \ge \mathbf{0}, \boldsymbol{\Delta}\mathbf{Y}^{(j)} \ge \mathbf{0}\right) = \left(\frac{1}{n!}\right)^2$, $i, j = 1, ..., n!$, by independence we have

$$F_{\mathbf{a}^T\mathbf{Y}_{(n)},\ \mathbf{b}^T\mathbf{Y}_{[n]}}(y_1,\ y_2) = P\left(\mathbf{a}^T\mathbf{Y} \le y_1,\ \mathbf{b}^T\mathbf{Y} \le y_2 | \boldsymbol{\Delta}\mathbf{X} \ge \mathbf{0},\ \boldsymbol{\Delta}\mathbf{Y} \ge \mathbf{0}\right).$$

Moreover, $\left(\boldsymbol{\Delta}\mathbf{X},\ \boldsymbol{\Delta}\mathbf{Y},\ \mathbf{a}^T\mathbf{Y},\ \mathbf{b}^T\mathbf{Y}\right)^T$ follows an $2n$ dimensional multivariate normal distribution with $\boldsymbol{\mu} = \left(\mathbf{0}_{n-1},\ \mathbf{0}_{n-1},\ \mu_y \mathbf{a}^{\mathbf{T}}\mathbf{1}_n,\ \mu_y \mathbf{b}^{\mathbf{T}}\mathbf{1}_n\right)^T$ and

$$\sum = \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}\boldsymbol{\Delta}^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}\mathbf{a} & \rho\sigma_x\sigma_y \boldsymbol{\Delta}\mathbf{b} \\ & \sigma_y^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^T & \sigma_y^2 \boldsymbol{\Delta}\mathbf{a} & \sigma_y^2 \boldsymbol{\Delta}\mathbf{b} \\ & & \sigma_y^2 \mathbf{a}^T\mathbf{a} & \sigma_y^2 \mathbf{a}^T\mathbf{b} \\ & & & \sigma_y^2 \mathbf{b}^T\mathbf{b} \end{pmatrix}.$$

So, as in Theorem 2.1 the proof is completed. □

We easily find that $\mathbf{\Gamma} = [\gamma_{ij}]$, where

$$\gamma_{ij} = \begin{cases} 2\sigma_x^2 & |i-j| = 0 \\ -\sigma_x^2 & |i-j| = 1 \\ 0 & |i-j| = 2, ..., 2(n-1) \end{cases}$$

and $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{bmatrix}$ with $\mathbf{\Lambda}_1 = \left(\lambda_{11}, ..., \ \lambda_{(n-1)1}\right)^T$ where $\lambda_{k1} = \sigma_x^2 \left\{a_{k+1} - a_k\right\}$, $k = 1, ..., n-1$ and $\mathbf{\Lambda}_2 = \left(\lambda_{12}, ..., \ \lambda_{(n-1)2}\right)^T$ where $\lambda_{k2} = \sigma_y^2 \left\{b_{k+1} - b_k\right\}$, $k = 1, ..., n-1$.

Let the difference matrix $\mathbf{\Delta}_1$ of dimension $n-1 \times n$ be such that its first $i-1$ rows are $\mathbf{e}_{n,1}^T - \mathbf{e}_{n,k}^T$, $k = 2, 3, ..., i-1$ and the last $n-i$ rows are $\mathbf{e}_{n,k}^T - \mathbf{e}_{n,1}^T$, $k = i, \ ... \ n-1$. Also, let the matrix $\mathbf{\Delta}_2$ of dimension $n-1 \times n$ be such that its first $j-1$ rows are $\mathbf{e}_{n,1}^T - \mathbf{e}_{n,k}^T$, $k = 2, 3, ..., j-1$ and the last $n-j$ rows are $\mathbf{e}_{n,j}^T - \mathbf{e}_{n,1}^T$, $k = j, \ ... \ n-1$ and $\mathbf{1}_{n,i}$ be a $n-1$ dimensional vector with the first i elements equal to 1 and the rest -1. Further, let $\mathbf{X}_i$ be a permutation of the random vector $\mathbf{X}$, such that its ith element is located in the first place.

Theorem 2.2. *For a random sample of size n from a bivariate normal random vector (X, Y), the joint distribution of $Y_{i:n}$, $Y_{[j:n]}$ is*

$$\begin{aligned} F_{Y_{i:n},\ Y_{[j:n]}}(y_1,\, y_2) \;=\; & k_1 F_{SUN}(\min(y_1,\, y_2)\,;\ \mu_y,\ \mathbf{0}_{2(n-1)},\ \sigma_y^2,\ \mathbf{\Gamma},\ \mathbf{\Lambda}) \\ & + k_2 F_{SSUN}(y_1,\, y_2\,;\ \mu_y \mathbf{1}_2\ , \mathbf{0}_{2(n-1)},\ \sigma_y^2 \mathbf{I}_2,\ \mathbf{\Gamma},\ \mathbf{\Lambda}) \end{aligned}$$

where $F_{SUN}(.\,;\ \mu_y,\ \mathbf{0}_{2(n-1)},\ \sigma_y^2,\ \mathbf{\Gamma},\ \mathbf{\Lambda})$ is the cdf of a non-singular unified multivariate skew-normal distribution $SUN_{1,\ 2n-2}\left(\mu_y,\ \mathbf{0},\ \sigma_y^2,\ \mathbf{\Gamma},\ \mathbf{\Lambda}\right)$ with

$$\mathbf{\Gamma} = \begin{pmatrix} \sigma_x^2 \mathbf{\Delta}_1 \mathbf{\Delta}_1^T & \rho\sigma_x\sigma_y \mathbf{\Delta}_1 \mathbf{\Delta}_1^T \\ & \sigma_y^2 \mathbf{\Delta}_1 \mathbf{\Delta}_1^T \end{pmatrix}, \ \mathbf{\Lambda} = \begin{pmatrix} \rho\sigma_x\sigma_y \mathbf{1}_{n,i} \\ \sigma_y^2 \mathbf{1}_{n,i} \end{pmatrix}$$

and $F_{SSUN}(.\,;\ \mu_y \mathbf{1}_2,\ \mathbf{0}_{2(n-1)},\ \sigma_y^2 \mathbf{I}_2,\ \mathbf{\Gamma},\ \mathbf{\Lambda})$ is the cdf of a singular unified multivariate skew-normal distribution $SSUN_{2,\ 2n-2}(\mu_y \mathbf{1}_2\ , \mathbf{0}_{2(n-1)}, \sigma_y^2 \mathbf{I}_2, \mathbf{\Gamma}, \mathbf{\Lambda})$ where $\mathbf{I}_2$ is an identity matrix of dimension 2 and

$$\mathbf{\Gamma} = \begin{pmatrix} \sigma_x^2 \mathbf{\Delta}_2 \mathbf{\Delta}_2^T & \rho\sigma_x\sigma_y \mathbf{\Delta}_2 \mathbf{\Delta}_1^T \\ & \sigma_y^2 \mathbf{\Delta}_1 \mathbf{\Delta}_1^T \end{pmatrix}, \ \mathbf{\Lambda} = \begin{pmatrix} -\rho\sigma_x\sigma_y \mathbf{J}_{n-1} & \rho\sigma_x\sigma_y \mathbf{1}_{n,j} \\ \sigma_y^2 \mathbf{1}_{n,i} & -\sigma_y^2 \mathbf{J}_{n-1} \end{pmatrix},$$

$k_1 = n!(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(-2\rho))^n$ *and* $k_2 = n(n-1)((n-1)!)^2(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(-2\rho))^n$.

Proof: Let B_{ij} denote the event that Y_i is the ith order statistic among $\{Y_1,\ Y_2, ...,\ Y_n\}$ and X_j is the jth order statistic among $\{X_1,\ X_2, ...,\ X_n\}$. So,

$B_{ij} = \{\boldsymbol{\Delta}_1 \mathbf{Y}_i > \mathbf{0},\ \boldsymbol{\Delta}_2 \mathbf{X}_j > \mathbf{0}\}$ and we have

$$\begin{aligned}
& F_{Y_{i:n},\ Y_{[j:n]}}(u,\ v) \\
=\ & P\left(Y_{i:n} \le u,\ Y_{[j:n]} \le v\right) \\
=\ & \sum_{i=1}^{n}\sum_{j=1}^{n} P\left(Y_i \le u,\ Y_j \le v | B_{ij}\right) P\left(B_{ij}\right) \\
=\ & \sum_{i=1}^{n} P\left(Y_i \le u,\ Y_i \le v | B_{ii}\right) P\left(B_{ii}\right) + \sum_{\substack{i=1 \\ i \ne j}}^{n}\sum_{j=1}^{n} P\left(Y_i \le u,\ Y_j \le v | B_{ij}\right) P\left(B_{ij}\right) \\
=\ & n! P\left(Y_1 \le \min(u,v) | B_{11}\right) P\left(B_{11}\right) \\
& +(n^2-n)((n-1)!)^2 P\left(Y_1 \le u, Y_2 \le v | B_{12}\right) P\left(B_{12}\right).
\end{aligned}$$

The last equality holds by the independence assumption. Since the distribution of $Y_1|B_{11}$ is identical to the distribution of $Y_1|\{\boldsymbol{\Delta}_1 \mathbf{Y}_1 > \mathbf{0},\ \boldsymbol{\Delta}_1 \mathbf{X}_1 > \mathbf{0}\}$, we have

$$\begin{pmatrix} \boldsymbol{\Delta}_1 \mathbf{X}_1 \\ \boldsymbol{\Delta}_1 \mathbf{Y}_1 \\ Y_1 \end{pmatrix} \sim N_{2n-1}\left(\begin{pmatrix} \mathbf{0}_{n-1} \\ \mathbf{0}_{n-1} \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T & \rho\sigma_x\sigma_y \mathbf{1}_{n,i} \\ & \sigma_y^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T & \sigma_y^2 \mathbf{1}_{n,i} \\ & & \sigma_y^2 \end{pmatrix} \right).$$

So, $Y_1|B_{11} \sim SUN_{1,\ 2n-2}\left(\mu_y,\ \mathbf{0}_{2(n-1)},\ \sigma_y^2,\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda}\right)$ where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T \\ & \sigma_y^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda} = \begin{pmatrix} \rho\sigma_x\sigma_y \mathbf{1}_{n,i} \\ \sigma_y^2 \mathbf{1}_{n,i} \end{pmatrix}.$$

Also, the conditional distribution of Y_1 and Y_2 given B_{12} is the same as the distribution of $(Y_1,\ Y_2)^T|\{\boldsymbol{\Delta}_2 \mathbf{X}_2 > \mathbf{0},\ \boldsymbol{\Delta}_1 \mathbf{Y}_1 > \mathbf{0}\}$. Moreover, $(\boldsymbol{\Delta}_2 \mathbf{X}_2,\ \boldsymbol{\Delta}_1 \mathbf{Y}_1,\ Y_1,\ Y_2)^T$ follows a $2n$ multivariate singular normal distribution with rank $2n-1$, $\boldsymbol{\mu} = \left(\mathbf{0}_{n-1},\ \mathbf{0}_{n-1},\ \mu_y \mathbf{1}_2\right)^T$ and

$$\sum = \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}_2 \boldsymbol{\Delta}_2^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}_2 \boldsymbol{\Delta}_1^T & -\rho\sigma_x\sigma_y \mathbf{J}_{n-1} & \rho\sigma_x\sigma_y \mathbf{1}_{n,j} \\ & \sigma_y^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T & \sigma_y^2 \mathbf{1}_{n,i} & -\sigma_y^2 \mathbf{J}_{n-1} \\ & & \sigma_y^2 & 0 \\ & & & \sigma_y^2 \end{pmatrix}$$

where $\mathbf{J}_{n-1} = (1,\ \mathbf{0}_{n-2})^T$.

We note that the matrix $(\boldsymbol{\Delta}_2 \mathbf{X}_2,\ \boldsymbol{\Delta}_1 \mathbf{Y}_1)^T$ is of full rank but $(\boldsymbol{\Delta}_2 \mathbf{X}_2,\ \boldsymbol{\Delta}_1 \mathbf{Y}_1,\ Y_1,\ Y_2)^T$ is not. Hence, according to the case (3) of Arellano-Valle and Azzalini (2006) we conclude that $(Y_1,\ Y_2)^T|\{\boldsymbol{\Delta}_2 \mathbf{X}_2,\ \boldsymbol{\Delta}_1 \mathbf{Y}_1 > \mathbf{0}\} \sim SSUN_{2,2n-2}(\mu_y \mathbf{1}_2,\ \mathbf{0}_{2(n-1)},\ \sigma_y^2 \mathbf{I}_2,\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$ where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \sigma_x^2 \boldsymbol{\Delta}_2 \boldsymbol{\Delta}_2^T & \rho\sigma_x\sigma_y \boldsymbol{\Delta}_2 \boldsymbol{\Delta}_1^T \\ & \sigma_y^2 \boldsymbol{\Delta}_1 \boldsymbol{\Delta}_1^T \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda} = \begin{pmatrix} -\rho\sigma_x\sigma_y \mathbf{J}_{n-1} & \rho\sigma_x\sigma_y \mathbf{1}_{n,j} \\ \sigma_y^2 \mathbf{1}_{n,i} & -\sigma_y^2 \mathbf{J}_{n-1} \end{pmatrix}.$$

On the other hand, using the orthant probabilities (e.g. Kotz *et al.* 2000) we easily obtain

$$\begin{aligned} P\,(B_{11}) &= P(X_2 > X_1,\ X_3 > X_1, ...,\ X_3 > X_1,\ Y_2 > Y_1,\ Y_3 > Y_1, ...,\ Y_n > Y_1) \\ &= (P(X_2 > X_1,\ Y_2 > Y_1))^n \\ &= (\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(-2\rho))^n\,. \end{aligned}$$

So, $k_1 = n!(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(-2\rho))^n$. Similarly, $k_2 = n(n-1)((n-1)!)^2(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(-2\rho))^n$.

This completes the proof. □

Remark 2.2. As a special case, we assume $n = 2$, $(X,\ Y)^T \sim BN(0,\ 0,\ 1,\ 1,\ \rho)$, $i = 1$ and $j = 2$. Then the joint pdf of $Y_{1:2}$ and $Y_{[2:2]}$ is obtained as

$$F_{Y_{1:2},\ Y_{[2:2]}}(\,y_1,\, y_2) = k_1 F_{SUN}(\,\min(y_1,\, y_2)\,) + k_2 F_{SSUN}(\,y_1,\, y_2\,)$$

where k_1 and k_2 are as in Theorem 2.2 with $n = 2$ and $F_{SUN}(.)$ and $F_{SSUN}(.,.)$ are the cdfs of

$$\varphi((\,\min(y_1,\, y_2))\frac{\Phi_2\left((\rho,\ -1)^T \min(y_1,\, y_2);\ \ \mathbf{M}_1\right)}{\Phi_2\left((0,0)^T;\ \mathbf{M}_2\right)}$$

and

$$\varphi(y_1)\varphi(\,y_2)\frac{\Phi_2\left((\rho,\ 1)^T\,(y_1 -\ y_2);\ \ \mathbf{M}_3\right)}{\Phi_2\left((0,0)^T;\ \mathbf{M}_4\right)}$$

respectively where

$$\mathbf{M}_1 = \begin{pmatrix} 2-\rho^2 & \rho \\ \rho & 1 \end{pmatrix}, \qquad \mathbf{M}_2 = 2\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

$$\mathbf{M}_3 = \begin{pmatrix} 2-\rho^2 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{M}_4 = 2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and their joint pdf is
(2.3)

$$f_{Y_{1:2},\ Y_{[2:2]}}(\,y_1,\, y_2) = \begin{cases} \varphi(y)\frac{\Phi_2\left((\rho,\ -1)^T y; \mathbf{M}_1\right)}{\Phi_2\left((0,0)^T;\ \mathbf{M}_2\right)} & if\ \ y_1 = y_2 = y \\ \varphi(y_1)\varphi(y_2)\frac{\Phi_2(\rho(y_2-y_1),\ y_2-y_1;\ \ \mathbf{M}_3)}{\Phi_2\left((0,0)^T;\ \mathbf{M}_4\right)} & if\ \ y_1 <\ y_2\,. \end{cases}$$

Remark 2.3. When X and Y are independent, the joint density (2.3) becomes

$$f_{Y_{1:2},\ Y_{[2:2]}}(y_1,\ y_2) = \begin{cases} 2\varphi(y)(1-\Phi(y)) & if \quad y_1 = y_2 = y \\ 2\varphi(y_1)\varphi(y_2) & if \quad y_1 < y_2 \end{cases}$$

which is the same as the joint distribution (8) of He and Nagaraja (2009) under these assumptions (see e.g. He, 2007, p. 35).

Furthermore, He and Nagaraja (2009) discussed some relations between $Y_{i:n}$ and $Y_{[j:n]}$ in a bivariate setting. In particular, they showed that $Corr(Y_{i:n}, Y_{[j:n]}) = Corr(Y_{n-i+1:n},\ Y_{[n-j+1:n]})$. The following remark shows that, in addition, the joint distribution of $Y_{i:n},\ Y_{[j:n]}$ and $Y_{n-i+1:n},\ Y_{[n-j+1:n]}$ belong to a same family and differ only in one parameter. The relation (24) of He and Nagaraja (2009) is a direct consequence.

Remark 2.4. Let B'_{ij} denote the event that Y_i is the $(n-i+1)$th order statistic among $\{Y_1,\ Y_2, ...,\ Y_n\}$ and X_j is the $(n-j+1)$th order statistic among $\{X_1,\ X_2, ...,\ X_n\}$. Then $B'_{ij} = \{\mathbf{\Delta}_1\mathbf{Y}_i < \mathbf{0},\ \mathbf{\Delta}_2\mathbf{X}_j < \mathbf{0}\} = \{-\mathbf{\Delta}_1\mathbf{Y}_i > \mathbf{0},\ -\mathbf{\Delta}_2\mathbf{X}_j > \mathbf{0}\}$. Hence, the joint distribution of $Y_{n-i+1:n},\ Y_{[n-j+1:n]}$ is

$$\begin{aligned} F_{Y_{n-i+1:n},\ Y_{[n-j+1:n]}}(y_1,\ y_2) \ = \ & k_1 F_{SUN}(\min(y_1,\ y_2)\,;\mu_y,\ \mathbf{0},\ \sigma_y^2,\ \mathbf{\Gamma},\ \mathbf{\Lambda}') \\ & + k_2 F_{SSUN}(y_1,\ y_2\,;\mu_y \mathbf{1}_2\ ,\mathbf{0},\ \sigma_y^2 \mathbf{I}_2,\ \mathbf{\Gamma},\ \mathbf{\Lambda}') \end{aligned}$$

where $\mathbf{\Lambda}' = -\mathbf{\Lambda}$ and the parameters as in Theorem 2.2.

3. NUMERICAL EXAMPLE

Loperfido (2008b), with the assumption of exchangeability, have estimated the distribution of extreme values of vision of left eye (Y_1) and vision of right eye (Y_2) and the conditional distribution of age (X), given these extreme values as a skew-normal family. Johnson and Wichern (2002, p.24) provide data consisting of mineral content measurements of three bones (radius, humerus, ulna) in two arms (dominant and non dominant) for each of 25 old women. We consider the following variables:

X_1 : Dominant radius

X_2 : Non dominant radius

Y_1 : Dominant ulna

Y_2 : Non dominant ulna

The sample data is presented in Table 1. We apply model (1.1) to this data and obtain the unbiased estimates of the parameters of these models as

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \begin{bmatrix} 0.8438 \\ 0.8191 \end{bmatrix}, \; \hat{\boldsymbol{\mu}}_{\mathbf{y}} = \begin{bmatrix} 0.7044 \\ 0.6938 \end{bmatrix}, \; \hat{\boldsymbol{\Sigma}}^2_{\mathbf{x}} = \begin{bmatrix} 0.0130 & 0.0103 \\ 0.0103 & 0.0114 \end{bmatrix},$$

$$\hat{\boldsymbol{\Sigma}}^2_{\mathbf{y}} = \begin{bmatrix} 0.0115 & 0.0088 \\ 0.0088 & 0.0105 \end{bmatrix} \text{ and } \hat{\boldsymbol{\Sigma}}_{\mathbf{xy}} = \begin{bmatrix} 0.0091 & 0.0085 \\ 0.0085 & 0.0105 \end{bmatrix}.$$

Table 1: Data of measurements of two bones in 25 old women.

Dominant radius	Non dominant radius	Dominant ulna	Non dominant ulna
1.103	1.052	0.873	0.872
0.842	0.859	0.590	0.744
0.925	0.873	0.767	0.713
0.857	0.744	0.706	0.674
0.795	0.809	0.549	0.654
0.787	0.799	0.782	0.571
0.933	0.880	0.737	0.803
0.799	0.851	0.618	0.682
0.945	0.876	0.853	0.777
0.921	0.906	0.823	0.765
0.792	0.825	0.686	0.668
0.815	0.751	0.678	0.546
0.755	0.724	0.662	0.595
0.880	0.866	0.810	0.819
0.900	0.838	0.723	0.677
0.764	0.757	0.586	0.541
0.733	0.748	0.672	0.752
0.932	0.898	0.836	0.805
0.856	0.786	0.578	0.610
0.890	0.950	0.758	0.718
0.688	0.532	0.533	0.482
0.940	0.850	0.757	0.731
0.493	0.616	0.546	0.615
0.835	0.752	0.618	0.664
0.915	0.936	0.869	0.868

Yang (1981) has considered general linear functions of the form

$$L = \frac{1}{n} \sum_{i=1}^{n} J(\frac{i}{n}) Y_{[i:n]}$$

where J is a smooth function. He has established that L is asymptotically normal and may be used to construct consistent estimator of various conditional quantities such as $E(Y|X = x)$, $P(Y \in A|X = x)$ and $Var(Y|X = x)$. We assume that J is a quadratic function and estimate the joint distribution of L and the *sample midrange* of $\mathbf{Y}$, i.e. $T = \frac{1}{2} \sum_{i=1}^{2} Y_{i:n}$. The joint distribution of T and L is as in Theorem 2.1 with $\boldsymbol{\Delta} = \begin{pmatrix} -1 & 1 \end{pmatrix}$, $\mathbf{a} = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix}^T$ and $\mathbf{b} = \begin{pmatrix} 1/8 & 1/2 \end{pmatrix}^T$.

In particular,

$$\boldsymbol{\xi}_{11}=\boldsymbol{\xi}_{21}=\begin{pmatrix}0.6991\\0.4345\end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\xi}_{12}=\boldsymbol{\xi}_{22}=\begin{pmatrix}0.6991\\0.4389\end{pmatrix}.$$

Also, if

$$M_n = n^{-1}\sum_{i=1}^{n} h(n)^{-1}K\Big(\frac{(i/n)-F_n(x)}{h(n)}\Big)Y_{[i:n]}$$

where $F_n(x)$ is the proportion of the X_i less than or equal to x, $K(x)$ is some pdf on real line and $h(n)\to 0$ as $n\to\infty$, then M_n is a mean square consistent estimator of the regression function $E(Y|X=x)$. We assume that $K(x)$ is the pdf of the normal distribution with mean 0.8314 and variance 0.0108, i.e. $K(x)$ is the pdf of the radius. Moreover, we set $h(n)=\frac{1}{n-1}$. At $x=0.8$, we obtain $M_2 = 0.012Y_{[1:2]} + 0.515Y_{[2:2]}$. Again, the joint distribution of T and M_2 is as in Theorem 2.1 with $\boldsymbol{\Delta}=\begin{pmatrix}-1 & 1\end{pmatrix}$, $\mathbf{a}=\begin{pmatrix}1/2 & 1/2\end{pmatrix}^T$ and $\mathbf{b}=\begin{pmatrix}0.012 & 0.515\end{pmatrix}^T$.

4. CONCLUSION

In this paper we model the joint distribution of a linear combination of concomitants of order statistics and linear combinations of their order statistics as a unified skew-normal family assuming a multivariate normal distribution. However, there are many interesting further work which may be carried out. Viana and Lee (2006) have studied the covariance structure of two random vectors $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{[n]}$ in the presence of a random variable Z. We may generalize their work by extending our results in the presence of one or more covariates. The results of this paper may be extended to elliptical distributions or using exchangeability assumption. Other results such as the regression analysis of concomitants using their order statistics are also of interest.

ACKNOWLEDGMENTS

The authors would like to thank the referees for their helpful comments and suggestions.

REFERENCES

[1] Aigner, D.J.; Lovell, C.A.K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model, *J. Econometrics*, **12**, 21–37.

[2] Andel, J.; Netuka, I. and Zvara, K. (1984). On threshold autoregressive processes, *Kybernetika*, **20**, 89–106.

[3] Arellano-Valle, R.B. and Genton, M.G. (2007). On the exact distribution of linear combinations of order statistics from dependent random variables, *J. Mult. Anal.*, **98**, 1876–1894.

[4] Arellano-Valle, R.B. and Azzalini, A. (2006). On the Unification of Families of Skew-normal Distributions, *Scand. J. Statics.*, **33**, 561–574.

[5] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scand. J. Statics.*, **12**, 171–178.

[6] Crocetta, C. and Loperfido, N. (2005). The exact sampling distribution of L-statistics, *Metron*, **63**, 213–223.

[7] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, third ed., John Wiley and Sons, New York.

[8] Goel, P.K. and Hall, P. (1994). On the average difference between concomitants and order statistics, *Ann. Probab.*, **22**, 126–144.

[9] He, Q. (2007). Inference on correlation from incomplete bivariate samples, Ph.D. Dissertation, Ohio State University, Columbus, OH.

[10] He, Q. and Nagaraja, H.N. (2009). Distribution of concomitants of order statistics and their order statistics, *J. Statist. Plann. Infer.*, **139**, 2643–2655.

[11] Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*, 5th Ed., Prentice-Hall.

[12] Loperfido, N. (2008a). A note on skew-elliptical distributions and linear functions of order statistics, *Statist. Probab. Lett.*, **78**, 3184–3186.

[13] Loperfido, N. (2008b). Modeling maxima of longitudinal contralateral observations, *Test*, **17**, 370–380.

[14] Nagaraja, H.N. (1982). A note on linear functions of ordered correlated normal random variables, *Biometrika*, **69**, 284–285.

[15] Roberts, C.D. (1966). A correlation model useful in the study of twins, *J. Am. Statist. Assoc.*, **61**, 1184–1190.

[16] Sheikhi, A. and Jamalizadeh, A. (2011). Regression analysis using order statistics, *Stat. Paper*, **52**, 885–892.

[17] Tsukibayashi, S. (1998). The joint distribution and moments of an extreme of the dependent variable and the concomitant of an extreme of the independent variable, *Comm. Statist. Theory Methods*, **27**, 1639–1651.

[18] Viana, M.A.G. and Lee, H.M. (2006). Correlation analysis of ordered symmetrically dependent observations and their concomitants of order statistics, *The Canadian Journal of Statistics*, **34**(2), 327–340.

[19] YANG, S.S. (1981). Linear functions of concomitants of order statistics with application to non-parametric estimation of a regression function, *J. Amer. Stat. Assoc.*, **76**, 658–662.

REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 135–168

ON THE MAXIMUM LIKELIHOOD ESTIMATOR FOR IRREGULARLY OBSERVED TIME SERIES DATA FROM COGARCH(1,1) MODELS

Authors: Moosup Kim
– Department of Statistics, Seoul National University, Republic of Korea
moosupkim@gmail.com

Sangyeol Lee
– Department of Statistics, Seoul National University, Republic of Korea
sylee@stats.snu.ac.kr

Received: December 2011 Revised: August 2012 Accepted: August 2012

Abstract:

- In this paper, we study the asymptotic properties of the maximum likelihood estimator (MLE) in COGARCH(1,1) models driven by Lévy processes as proposed by Maller *et al.* ([13]). We show that the MLE is consistent and asymptotically normal under some conditions relevant to the moments of the driving Lévy process and the sampling scheme.

Key-Words:

- *COGARCH(1,1) models; maximum likelihood estimation; consistency; asymptotic normality; sampling scheme; irregular time spaces.*

AMS Subject Classification:

- 62F12, 62M86.

1. INTRODUCTION

GARCH models are prominent stochastic models in finance, designed to capture the time-varying conditional volatilities and heavy tail phenomenon of financial time series. We refer to Bollerslev ([5]); Bougerol and Picard ([6]); Nelson ([16]); Basrak *et al.* ([1]); Berkes *et al.* ([3]), and for its estimation, to Hall and Yao ([9]); Berkes and Horváth ([2]); Francq and Zakoïan ([8]). In ordinary discrete time GARCH models, time series are assumed to be equally spaced. However, in some situations, time series are often observed irregularly. This phenomenon happens, for instance, in tick-by-tick data and daily data which is not observed on weekends and holidays. To accommodate the irregularity of time spaces, several authors have made efforts to extend the discrete time GARCH model to a continuous time counterpart. Nelson ([15]) demonstrated that the discrete time GARCH process with Gaussian innovations is a finite approximation of a bivariate diffusion process. Therein, the limiting diffusion process is driven by two independent Brownian motions, which unfortunately undermines the spirit of GARCH processes since they are originally designed to have a single innovation sequence. Later, Klüppelberg *et al.* ([12]) proposed a continuous time GARCH (COGARCH) process driven by a Lévy process, which can be seen as an analogue of discrete time GARCH process. Also, Maller *et al.* ([13]) demonstrated that the discrete time GARCH process embeds in COGARCH processes and further, the embedded GARCH process converges in a strong sense to the original COGARCH process that embeds it as the discrete grid used for obtaining the embedded process gets finer (cf. Theorem 2.1 of [13]). For more details, we refer to Kallsen and Vesenmayer ([11]).

Concerning the estimation of COGARCH parameters, Haug *et al.* ([10]) considered a method of moment estimator which is suitable for equally spaced time series and verified its consistency and asymptotic normality under some regularity conditions, which, however, is not directly applicable to irregularly spaced time series. On the other hand, Müller ([14]) proposed an MCMC-based estimation for COGARCH(1,1) models driven by a compound Poisson process, which is suitable for irregularly spaced time series, which, however, has a defect that computation is somewhat intensive. Maller *et al.* ([13]) proposed using a Gaussian maximum likelihood estimator (MLE) in COGARCH(1,1) models but its asymptotic properties such as consistency and asymptotic normality has not been thoroughly investigated yet in the literature. Motivated by this, we are led to study the asymptotic behavior of the MLE in COGARCH(1,1) models. Since some empirical study to evaluate finite sample performance has been already implemented by [13], here we focus on the rigorous verification of the asymptotic properties of the MLE.

The organization of this paper as follows. In Section 2.1, we give a brief review for COGARCH(1,1) processes. In Section 2.2, we present the main result of this paper. In Section 3, we provide the proof for the result presented in Section 2.2.

2. THE COGARCH(1,1) MODEL AND ESTIMATION

2.1. COGARCH(1,1) Processes

In this subsection, we summarize the COGARCH(1,1) process. Let $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t : t \geq 0\})$ be a filtered probability space satisfying the usual conditions:

- $\mathcal{F}_0$ has all the measurable sets of P-measure 0,
- each $\mathcal{F}_t$ is right continuous, i.e., $\mathcal{F}_t = \bigcap_{t<s} \mathcal{F}_s$.

Let $L := \{L_t, \mathcal{F}_t : t \geq 0\}$ be a càdlàg Lévy process with characteristic triplet (γ, ϕ, Π) satisfying $\int_{\mathbb{R}} \min\{1, x^2\}\Pi(dx) < \infty$. The characteristic function of L_t is given by

$$u \mapsto \mathrm{E}e^{\mathrm{i}uL_t} = \exp\left\{\mathrm{i}t\gamma u - \frac{t\phi^2u^2}{2} + t\int_{\mathbb{R}}\{e^{\mathrm{i}ux} - 1 - \mathrm{i}ux1_{(|x|\leq 1)}\}\Pi(dx)\right\}.$$

which is called Lévy-Khintchine's representation (cf. Theorem 43 of Chapter I of Protter ([17])). In this paper, we assume $\phi = 0$.

Let $\eta^\circ > 0$, $\varphi^\circ > 0$, and $\beta^\circ > 0$ satisfying $\eta^\circ > \varphi^\circ$. Define $\Delta L_s := L_s - L_{s-}$ and

$$X_t := \eta^\circ t - \sum_{0<s\leq t} \log\left(1 + \varphi^\circ(\Delta L_s)^2\right),$$

which is a càdlàg process. Let σ_0^2 be an integrable random variable which is independent of $\{L_t\}$. Define

$$\sigma_t^2 := \left(\beta^\circ \int_0^t e^{X_s}ds + \sigma_0^2\right)e^{-X_{t-}},$$

which is a càglàd process. According to Proposition 3.2 of Klüppelberg *et al.* ([12]), the process $\{\sigma_t^2\}$ satisfies the stochastic integral equation

$$\sigma_t^2 - \sigma_0^2 = \int_0^t (\beta^\circ - \eta^\circ\sigma_s^2)ds + \varphi^\circ \sum_{0<s<t} \sigma_s^2(\Delta L_s)^2. \tag{2.1}$$

Note that due to $\phi = 0$, L is a quadratic pure jump, i.e., $[L, L]_t - [L, L]_0 = \sum_{0<s\leq t}(\Delta L_s)^2$ (cf. p. 71 of [17]) and (2.1) is rewritten as

$$\sigma_t^2 - \sigma_0^2 = \int_0^t (\beta^\circ - \eta^\circ\sigma_s^2)ds + \varphi^\circ \int_{(0,t)} \sigma_s^2 d[L, L]_s, \tag{2.2}$$

i.e., $\{\sigma_{t+}^2\}$ is the almost surely unique and càdlàg solution of the stochastic differential equation

$$d\sigma_{t+}^2 = (\beta^\circ - \eta^\circ \sigma_t^2)dt + \varphi^\circ \sigma_t^2 d[L,L]_t.$$

Later, we take σ_0^2 so that the solution is strictly stationary (see (3.1)). Finally, we define the integrated COGARCH(1,1) process as

$$G_t := \int_{(0,t]} \sigma_s dL_s, \quad t \geq 0.$$

2.2. Gaussian ML Estimation

In this subsection, we consider the maximum likelihood estimation method as proposed by Maller *et al.* ([13]) and study its asymptotic properties. Particularly, we consider the situation in which $\{G_t : t \geq 0\}$ is observed discretely with irregular time spaces. For each $n \in \mathbb{N}$, we set $N = N_n \in \mathbb{N}$,

$$0 = t_0 < t_1 < \cdots < t_N < \infty, \quad \Delta t_k := t_k - t_{k-1},$$

and

$$Y_{nk} := G_{t_k} - G_{t_{k-1}},$$

where $\{\Delta t_k\}$ are allowed to be nonidentical. By putting $\Delta := \Delta_n := \max\{\Delta t_1, ..., \Delta t_N\}$, we assume that $\Delta \to 0$ and $t_N \to \infty$ as $n \to \infty$.

Let $\theta^\circ = (\beta^\circ, \varphi^\circ, \eta^\circ)'$ be the vector of (unknown) true parameters. Let $\theta = (\beta, \eta, \varphi)'$ and

$$\Theta := \{\theta = (\beta, \eta, \varphi) : \beta_* \leq \beta \leq \beta^*, \eta_* \leq \eta \leq \eta^*, \varphi_* \leq \varphi \leq \varphi^*, \eta - \varphi \geq c_*\},$$

where $0 < \beta_* < \beta^* < \infty$, $0 < \eta_* < \eta^* < \infty$, $0 < \varphi_* < \varphi^* < \infty$, and $0 < c_* < \infty$. We assume that $\theta^\circ \in \Theta$.

Following [13], we set $\tilde{\sigma}_{n,k}^2(\theta)$ $(k = 0, 1, 2, ..., N)$ to be the solution of the recursion formula:

$$\begin{aligned}
\tilde{\sigma}_{n0}^2(\theta) &:= \frac{\beta}{\eta - \varphi}, \\
\tilde{\sigma}_{nk}^2(\theta) &:= \beta \Delta t_k + e^{-\eta \Delta t_k} \tilde{\sigma}_{n,k-1}^2(\theta) + \varphi e^{-\eta \Delta t_k} Y_{nk}^2 \quad \text{for } k = 1, 2, ..., N.
\end{aligned}$$

More precisely,

$$\tilde{\sigma}_{nk}^2(\theta) = \beta \sum_{i=0}^{k-1} \Delta t_{k-i} e^{-\eta(t_k - t_{k-i})} + e^{-\eta t_k} \tilde{\sigma}_{n0}^2(\theta) + \varphi e^{-\eta t_k} \sum_{i=1}^{k} e^{\eta t_{k-i}} Y_{n,k-i+1}^2,$$

which can be viewed as an estimate of $\sigma^2_{t_k}$ when $\theta = \theta^\circ$. By observing the argument:

$$\mathrm{E}\left\{(G_{t+h} - G_t)^2 | \mathcal{F}_t\right\} = \left(\sigma_t^2 - \frac{\beta^\circ}{\eta^\circ - \varphi^\circ}\right)\left(\frac{\exp\{(\eta^\circ - \varphi^\circ)h\} - 1}{\eta^\circ - \varphi^\circ}\right) + \frac{\beta^\circ h}{\eta^\circ - \varphi^\circ},$$

provided that $\mathrm{E}\{L_1^2\} = 1$ and $\{\sigma_t^2\}$ is strictly stationary (see the proof of Proposition 5.1 of [12]), we use the terms:

$$\tilde{\rho}^2_{nk}(\theta) := \left(\tilde{\sigma}^2_{n,k-1}(\theta) - \frac{\beta}{\eta - \varphi}\right)\left(\frac{\exp\{(\eta - \varphi)\Delta t_k\} - 1}{\eta - \varphi}\right) + \frac{\beta \Delta t_k}{\eta - \varphi}$$

as estimates of conditional variances of Y_{nk} when $\theta = \theta^\circ$.

Let $m = m_n$ be a positive integer. Then we define a Gaussian log-likelihood function of $\theta = (\beta, \varphi, \eta)$ as

$$\mathcal{L}_N(\theta) := \sum_{k=m}^{N} \mathrm{l}_{nk}(\theta)\Delta t_k, \quad \mathrm{l}_{nk}(\theta) = -\left(\frac{Y^2_{nk}}{\tilde{\rho}^2_{nk}(\theta)} + \log \frac{\tilde{\rho}^2_{nk}(\theta)}{\Delta t_k}\right),$$

which is slightly different from that of [13] in which Δt_k does not appear. Below, we show that $\hat{\theta}_n$, a measurable maximum point of $\mathcal{L}_N$, i.e.,

$$\mathcal{L}_N(\hat{\theta}_n) = \max_{\theta \in \Theta} \mathcal{L}_N(\theta)$$

is consistent and asymptotically normal under some regularity conditions such as

- **C1**: $\theta^\circ \in \Theta$. $\Delta \to 0$ and $t_N \to \infty$. $t_m = o(t_N)$ and $e^{-\eta_* t_m} = O(\Delta^{1/2})$.
- **C2**: $\phi = 0$, i.e., $\{L_t : t \geq 0\}$ is a quadratic pure jump.
- **C3**: $\mathrm{E}\{L_1\} = 0$, $\mathrm{E}\{L_1^2\} = 1$, and $\mathrm{E}\{L_1^4\} < \infty$; $\Psi(2) < 0$, where $\Psi(z) := \log \mathrm{E} e^{-zX_1}$.
- **C4**: θ° is an interior point of Θ; $t_N \Delta \to 0$; $\mathrm{E}|G_h|^{4+\delta} = O(h)$ for some $\delta > 0$; $\int_{\mathbb{R}} x^3 d\Pi(x) = 0$.

The following is the main result of this paper, the proof of which is presented in the next section.

Theorem 2.1. *Under* **C1-C3**,

$$\hat{\theta}_n \xrightarrow{P} \theta^\circ. \tag{2.3}$$

Suppose that **C4** *also holds. Then,*

$$\sqrt{t_N}(\hat{\theta}_n - \theta^\circ) \Rightarrow N(\mathbf{0}, \tau \Sigma^{-1}), \tag{2.4}$$

where

$$\tau := \int_{\mathbb{R}} x^4 \Pi(dx) = \lim_{h \downarrow 0} \frac{h\mathrm{E}\left\{(G_h - G_0)^4 | \mathcal{F}_0\right\}}{\left\{\mathrm{E}\left\{(G_h - G_0)^2 | \mathcal{F}_0\right\}\right\}^2}$$

and Σ *is a positive definite matrix presented in Proposition 3.3.*

3. PROOFS

In what follows, K denotes a generic constant. We begin with the existence of a strictly stationary solution of (2.2). Let $\{L_t^*\}$ be an independent copy of $\{L_t : 0 \le t < \infty\}$. We extend the time domain of $\{L_t\}$ and $\{X_t\}$ to $\mathbb{R}$ by letting

$$L_t := -L^*_{(-t)-} \quad \text{for } -\infty < t < 0$$

and

$$X_t := \eta^\circ t + \sum_{t<s\le 0} \log\left(1 + \varphi^\circ(\Delta L_s)^2\right) \qquad \text{for } -\infty < t < 0.$$

Note that $\{L_t : t \in \mathbb{R}\}$ and $\{X_t : t \in \mathbb{R}\}$ are càdlàg processes and still have independent and strictly stationary increments. We define

$$\sigma_u^2 := \beta^\circ \int_{-\infty}^{u} e^{X_v - X_{u-}} dv \qquad \text{for } u \le 0. \tag{3.1}$$

Lemma 3.1. *Suppose that* **C3** *holds. Then, σ_u^2 is square integrable.*

Proof: Note that

$$\mathrm{E}\left\{\int_{-\infty}^{u} e^{X_v - X_{u-}} dv\right\}^2 = \lim_{h\to\infty} \mathrm{E}\left\{\int_{u-h}^{u} e^{X_v - X_{u-}} dv\right\}^2 < \infty,$$

(cf. the proof of Proposition 4.1 of [12]). This completes the proof. □

It can be easily checked that $\{\sigma_u^2\}$ with $\sigma_0^2 = \int_{-\infty}^0 e^{X_v - X_{0-}} dv$ is the almost surely unique strictly stationary solution of (2.2).

3.1. The Proof of Consistency

In this subsection, we assume that **C1-C3** hold. Note that

$$\sigma_0^2(\theta) := \beta/\eta + \varphi \int_{(-\infty,0)} e^{\eta u} \sigma_u^2 d[L,L]_u$$

is integrable, since

$$\mathrm{E}\int_{(-\infty,0]} e^{\eta u} \sigma_u^2 d[L,L]_u = \lim_{h\to\infty} \mathrm{E}\int_{(-h,0]} e^{\eta u} \sigma_u^2 d[L,L]_u = \mathrm{E}\sigma_0^2 \int_0^\infty e^{-\eta u} du < \infty.$$

We set

$$\sigma_t^2(\theta) := \beta/\eta + (\sigma_0^2(\theta) - \beta/\eta)e^{-\eta t} + \varphi e^{-\eta t} \int_{(0,t)} e^{\eta s} \sigma_s^2 d[L,L]_s, \quad (t > 0)$$

which is a càglàd process.

Lemma 3.2. *$\{\sigma_t^2(\theta)\}$ is strictly stationary and satisfies the stochastic differential equation*

$$d\sigma_{t+}^2(\theta) = (\beta - \eta\sigma_t^2(\theta))dt + \varphi\sigma_t^2 d[L,L]_t. \tag{3.2}$$

Especially, $\sigma_t^2(\theta) \geq \beta/\eta$ and $\mathrm{E}\sigma_t^4(\theta) < \infty$.

Proof: Note that

$$\sigma_{t+}^2(\theta) - \sigma_{0+}^2(\theta) = \left(\sigma_0^2(\theta) - \beta/\eta\right)\left(e^{-\eta t} - 1\right) + \varphi e^{-\eta t}\int_{(0,t]} e^{\eta s}\sigma_s^2 d[L,L]_s.$$

By using Fubini's theorem, we can see that

$$\begin{aligned}
&\int_0^t (\beta - \eta\sigma_s^2(\theta))ds = \int_0^t (\beta - \eta\sigma_{s+}^2(\theta))ds \\
&= \int_0^t \left\{ -\eta(\sigma_0^2(\theta) - \beta/\eta)e^{-\eta s} - \varphi\eta e^{-\eta s}\int_{(0,s]} e^{\eta u}\sigma_u^2 d[L,L]_u \right\} ds \\
&= \left(\sigma_0^2(\theta) - \beta/\eta\right)\left(e^{-\eta t} - 1\right) - \varphi\int_0^t \eta e^{-\eta s}\int_{(0,s]} e^{\eta u}\sigma_u^2 d[L,L]_u ds \\
&= \left(\sigma_0^2(\theta) - \beta/\eta\right)\left(e^{-\eta t} - 1\right) - \varphi\int_{(0,t]} \left\{\int_u^t \eta e^{-\eta s} ds\right\} e^{\eta u}\sigma_u^2 d[L,L]_u \\
&= \left(\sigma_0^2(\theta) - \beta/\eta\right)\left(e^{-\eta t} - 1\right) - \varphi\int_{(0,t]} \left(e^{-\eta u} - e^{-\eta t}\right) e^{\eta u}\sigma_u^2 d[L,L]_u \\
&= \left(\sigma_0^2(\theta) - \beta/\eta\right)\left(e^{-\eta t} - 1\right) - \varphi\int_{(0,t]} \sigma_u^2 d[L,L]_u + \varphi e^{-\eta t}\int_{(0,t]} e^{\eta u}\sigma_u^2 d[L,L]_u,
\end{aligned}$$

and which implies (3.2). Now that the strict stationarity can be easily checked, $\sigma_0^2(\theta) \geq \beta/\eta$ obviously implies $\sigma_t^2(\theta) \geq \beta/\eta$. Moreover,

$$\sigma_0^2(\theta) \leq \frac{\beta}{\eta} + \sum_{j=0}^{\infty} e^{-\eta j}\int_{(-j-1,-j]} \sigma_u^2 d[L,L]_u,$$

which indicates the square integrability since $\mathrm{E}\{\int_{(0,1]}\sigma_u^2 d[L,L]_u\}^2 < \infty$ due to **C3**. This gives the lemma. □

Lemma 3.3. $\sigma_0^2(\theta^\circ) = \sigma_0^2$ *a.s. Hence, $\sigma_t^2(\theta^\circ) = \sigma_t^2$ a.s. for every $t \geq 0$ and $\sigma_t^2 \geq \beta^\circ/\eta^\circ$.*

Proof: By using Fubini's theorem, we obtain

$$\begin{aligned}
\sigma_0^2(\theta^\circ) &= \beta^\circ/\eta^\circ + \varphi^\circ \int_{(-\infty,0)} e^{\eta^\circ u}\sigma_u^2 d[L,L]_u \\
&= \beta^\circ/\eta^\circ + \varphi^\circ \int_{(-\infty,0)} e^{\eta^\circ u}\left(\beta^\circ \int_{-\infty}^{u} e^{X_v - X_{u-}} dv\right) d[L,L]_u \\
&= \beta^\circ/\eta^\circ + \beta^\circ\varphi^\circ \int_{-\infty}^{0}\left(\int_{(v,0)} e^{\eta^\circ u - X_{u-}} d[L,L]_u\right) e^{X_v} dv \\
&= \beta^\circ \int_{-\infty}^{0} e^{\eta^\circ v} dv + \beta^\circ\varphi^\circ \int_{-\infty}^{0}\left(\int_{(v,0)} e^{\eta^\circ u - X_{u-}} d[L,L]_u\right) e^{X_v} dv \\
&= \beta^\circ \int_{-\infty}^{0}\left(e^{\eta^\circ v - X_v} + \varphi^\circ \int_{(v,0)} e^{\eta^\circ u - X_{u-}} d[L,L]_u\right) e^{X_v} dv.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
e^{\eta^\circ v - X_v} &= \exp\left\{-\sum_{v<s\le 0}\log(1+\varphi^\circ(\Delta L_s)^2)\right\} \\
&= \sum_{v<w\le 0}\left[\exp\left\{-\sum_{w\le s\le 0}\log(1+\varphi^\circ(\Delta L_s)^2)\right\}\right. \\
&\quad \left. - \exp\left\{-\sum_{w<s\le 0}\log(1+\varphi^\circ(\Delta L_s)^2)\right\}\right] + 1 \\
&= -\varphi^\circ \sum_{v<w\le 0}\exp\left\{-\sum_{w\le s\le 0}\log(1+\varphi^\circ(\Delta L_s)^2)\right\}(\Delta L_w)^2 + 1 \\
&= -\varphi^\circ \int_{(v,0]} e^{\eta^\circ w - X_{w-}} d[L,L]_w + 1,
\end{aligned}$$

and thus,

$$\sigma_0^2(\theta^\circ) = \beta^\circ \int_{-\infty}^{0}\left\{1 - \varphi^\circ e^{-X_{0-}}(\Delta L_0)^2\right\} e^{X_v} dv.$$

Since $(\Delta L_0)^2 = 0 = X_{0-}$ a.s., we obtain

$$\sigma_0^2(\theta^\circ) = \beta^\circ \int_{-\infty}^{0} e^{X_v - X_{0-}} dv = \sigma_0^2 \quad \text{a.s.}$$

This verifies the uniqueness of the solution of (3.2) and completes the proof. □

The following proposition plays a key role in proving the consistency.

Proposition 3.1. *If $\sigma_0^2(\theta) = \sigma_0^2$ a.s., then $\theta = \theta^\circ$. Hence,*

$$\Upsilon(\theta) := -\mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)} + \log\sigma_0^2(\theta)\right\}, \quad \theta \in \Theta$$

has the unique maximum at $\theta = \theta^\circ$. Moreover, $\Upsilon(\theta)$ is uniformly continuous in $\theta \in \Theta$.

Proof: Suppose that $\sigma_0^2(\theta) = \sigma_0^2$, a.s. Then, we have

$$\int_{(-\infty,0)} \left\{ \varphi e^{\eta u} - \varphi^\circ e^{\eta^\circ u} \right\} \sigma_u^2 d[L,L]_u = \beta^\circ/\eta^\circ - \beta/\eta, \quad \text{a.s.}$$

Also, by the strict stationarity, for every t,

$$\int_{(-\infty,t)} \left\{ \varphi e^{\eta(t-u)} - \varphi^\circ e^{\eta^\circ (t-u)} \right\} \sigma_u^2 d[L,L]_u = \beta^\circ/\eta^\circ - \beta/\eta, \quad \text{a.s.}$$

which implies $\sigma_t^2(\theta) = \sigma_t^2$, a.s. Moreover, both the processes are càglàd processes and so are indistinguishable. Thus, we have

$$\int_0^t (\beta - \eta\sigma_s^2) ds + \varphi \int_{(0,t)} \sigma_s^2 d[L,L]_s = \int_0^t (\beta^\circ - \eta^\circ \sigma_s^2) ds + \varphi^\circ \int_{(0,t)} \sigma_s^2 d[L,L]_s.$$

Suppose that $\varphi^\circ \neq \varphi$. Then,

$$\int_{(0,t)} \sigma_s^2 d[L,L]_s = \frac{1}{\varphi - \varphi^\circ} \int_0^t \left\{ \beta^\circ - \beta + (\eta - \eta^\circ)\sigma_s^2 \right\} ds,$$

which implies that there exist constants α, γ such that

$$\sigma_t^2 = \sigma_0^2 + \int_0^t (\alpha + \gamma\sigma_s^2) ds.$$

If $\gamma \neq 0$, $\sigma_t^2 = \gamma^{-1}\{(\alpha + \gamma\sigma_0^2)e^{\gamma t} - \alpha\}$, which contradicts the strictly stationarity of $\{\sigma_t^2\}$. On the other hand, if $\gamma = 0$, $\sigma_t^2 = \sigma_0^2 + \alpha t$. In this case, $\alpha \neq 0$ contradicts the strictly stationarity of $\{\sigma_t^2\}$ as well. Thus, $\alpha = \gamma = 0$, which in turn produces $\sigma_t^2 = \sigma_0^2$ a.s. for every $t > 0$. Then, we should have

$$0 = \beta^\circ - \eta^\circ \sigma_0^2 + \varphi^\circ \sigma_0^2 \{[L,L]_1 - [L,L]_0\} \quad \text{a.s.}$$

However, the above is also false since $[L,L]_1 - [L,L]_0$ is independent of σ_0^2. Therefore, $\varphi = \varphi^\circ$. If $\eta \neq \eta^\circ$, then $\sigma_t^2 = c$ for some constant c. Thus from the same reasoning, we conclude that $\eta = \eta^\circ$, and $\beta^\circ = \beta$.

Now, we have that for $h > 0$, η_1, η_2 satisfying $\eta_* \leq \eta_2 < \eta_1 \leq \eta^*$,

$$\begin{aligned}
&\left\| \int_{(-\infty,0)} |e^{\eta_1 u} - e^{\eta_2 u}| \sigma_u^2 d[L,L]_u \right\|_2 = \left\| \int_{(-\infty,0)} e^{\eta_2 u} |e^{(\eta_1 - \eta_2)u} - 1| \sigma_u^2 d[L,L]_u \right\|_2 \\
\leq &\left\| \sup_{-h<u<0} |e^{(\eta_1-\eta_2)u} - 1| \int_{(-\infty,0)} e^{\eta_2 u} \sigma_u^2 d[L,L]_u \right. \\
&\left. + e^{-\eta_2 h} \int_{(-\infty,-h]} e^{\eta_2 (u+h)} \sigma_u^2 d[L,L]_u \right\|_2 \\
\leq &\left(\sup_{-h<u<0} |e^{(\eta_1-\eta_2)u} - 1| + e^{-\eta_2 h} \right) \left\| \int_{(-\infty,0]} e^{\eta_2 u} \sigma_u^2 d[L,L]_u \right\|_2,
\end{aligned}$$

which implies

$$\lim_{\delta\to 0}\sup_{|\eta_1-\eta_2|<\delta}\left\|\int_{(-\infty,0)} e^{\eta_1 u}\sigma_u^2 d[L,L]_u - \int_{(-\infty,0)} e^{\eta_2 u}\sigma_u^2 d[L,L]_u\right\|_2 = 0.$$

This in turn implies that Υ is continuous. So the proposition is established. □

The proof of the consistency is based on the uniform convergence of the likelihood function, which can be obtained from the ergodic theorem and smoothness condition on the likelihood function.

Lemma 3.4. *Let $\sigma^2_{n,k-1}(\theta) := \sigma^2_{t_{k-1}}(\theta)$. Then,*

$$\frac{1}{t_N}\sum_{k=m}^{N}\left\{\frac{\sigma^2_{n,k-1}}{\sigma^2_{n,k-1}(\theta)} + \log\sigma^2_{n,k-1}(\theta)\right\}\Delta t_k \xrightarrow{P} \mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)} + \log\sigma_0^2(\theta)\right\}.$$

Proof: Since

$$\begin{aligned}&\sup_{t_{k-1}<u\le t_k}|\sigma^2_{n,k-1}(\theta)-\sigma^2_u(\theta)|\\ &\le \varphi e^{\eta\Delta t_k}\int_{t_{k-1}}^{t_k}\sigma_t^2 d[L,L]_t + \left(1-e^{-\eta\Delta t_k}\right)\varphi\int_{-\infty}^{t_{k-1}} e^{-\eta(t_{k-1}-t)}\sigma_t^2 d[L,L]_t\end{aligned}$$

and

$$\sup_{t_{k-1}<s\le t_k}\sigma_s^2(\theta)\le\sigma^2_{t_{k-1}}(\theta)+\varphi\int_{t_{k-1}}^{t_k}\sigma_t^2 d[L,L]_t,$$

we have that $\mathrm{E}\sup_{t_{k-1}<s\le t_k}\sigma_s^4(\theta)<\infty$ and

$$\max_{m\le k\le N}\left\|\sup_{t_{k-1}<u\le t_k}|\sigma^2_{n,k-1}(\theta)-\sigma^2_u(\theta)|\right\|_2 = o(1).$$

Thus, we have

$$\begin{aligned}&\mathrm{E}\left|\frac{1}{t_N}\sum_{k=m}^{N}\left\{\frac{\sigma^2_{n,k-1}}{\sigma^2_{n,k-1}(\theta)}+\log\sigma^2_{n,k-1}(\theta)\right\}\Delta t_k - \frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta)}+\log\sigma_s^2(\theta)\right\}ds\right|\\ &\le\frac{1}{t_N}\sum_{k=m}^{N}\mathrm{E}\sup_{t_{k-1}<s\le t_k}\left\{\left|\frac{\sigma^2_{n,k-1}}{\sigma^2_{n,k-1}(\theta)}-\frac{\sigma_s^2}{\sigma_s^2(\theta)}\right|+\left|\log\sigma^2_{n,k-1}(\theta)-\log\sigma_s^2(\theta)\right|\right\}\Delta t_k\to 0.\end{aligned}$$

On the other hand, by the ergodic theorem,

$$\frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta)}+\log\sigma_s^2(\theta)\right\}ds\xrightarrow{P}\mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)}+\log\sigma_0^2(\theta)\right\},$$

(cf. Lemma A.1). Hence, the lemma is validated. □

Lemma 3.5. *There exists a constant $c > 0$ such that for all large n,*

$$\min_{m\le k\le N}\inf_{\theta\in\Theta}\tilde{\sigma}^2_{nk}(\theta)\wedge\min_{m\le k\le N}\inf_{\theta\in\Theta}\frac{\tilde{\rho}^2_{nk}(\theta)}{\Delta t_k} > c \quad \text{a.s.}$$

Proof: Since

$$\tilde{\sigma}^2_{nk}(\theta)\ge\beta\sum_{i=0}^{k-1}\Delta t_{k-i}e^{-\eta(t_k-t_{k-i})} \quad \text{a.s.},$$

we have that for all large n,

$$\min_{m\le k\le N}\inf_{\theta\in\Theta}\tilde{\sigma}^2_{nk}(\theta)\ge\frac{\beta_*}{2\eta^*}>0 \quad \text{a.s.}$$

and

$$\min_{m\le k\le N}\left\{\inf_{\theta\in\Theta}\tilde{\sigma}^2_{n,k-1}(\theta)-\sup_{\theta\in\Theta}\frac{\beta}{\eta-\varphi}\left\{\frac{e^{(\eta-\varphi)\Delta t_k}-1}{(\eta-\varphi)\Delta t_k}-1\right\}\right\}\ge\frac{\beta_*}{4\eta^*}>0 \quad \text{a.s.}$$

This completes the proof. □

Now, we prove that $\{\tilde{\sigma}^2_{nk}(\theta)\}$ approximates to $\{\sigma^2_{nk}(\theta)\}$ well.

Lemma 3.6.

$$\mathrm{E}\left(\int_{(0,h]}G_{s-}\sigma_s dL_s\right)^2 = O(h^2) \quad \text{as } h\to 0.$$

Proof: From Corollary 4.1 of [12], we obtain

$$\mathrm{E}|\sigma_t^2-\sigma_0^2|^2 = 2\{\mathrm{Var}(\sigma_0^2)-\mathrm{Cov}(\sigma_t^2,\sigma_0^2)\} = 2\,\mathrm{Var}(\sigma_0^2)\{1-e^{t\Psi(1)}\},$$

i.e., $\mathrm{E}|\sigma_t^2-\sigma_0^2|^2 = O(t)$ as $t\to 0$. Further, for $h>0$,

$$\begin{aligned}\mathrm{E}\left(\int_{(0,h]}G_{s-}\sigma_s dL_s\right)^2 &= \int_{(0,h]}\mathrm{E}\{G^2_{s-}\sigma^2_s\}ds\\ &= \int_{(0,h]}\mathrm{E}G^2_{s-}\{\sigma^2_s-\sigma^2_0\}ds+\int_{(0,h]}\mathrm{E}G^2_{s-}\sigma^2_0 ds.\end{aligned}$$

Since

$$\left|\mathrm{E}G^2_{s-}\{\sigma^2_s-\sigma^2_0\}\right|\le\mathrm{E}^{1/2}G^4_{s-}\mathrm{E}^{1/2}(\sigma^2_s-\sigma^2_0)^2 = O(s) \quad \text{as } s\to 0$$

and

$$\mathrm{E}\{G^2_{s-}\sigma^2_0\} = \mathrm{E}\left\{\mathrm{E}\left(G^2_{s-}|\mathcal{F}_0\right)\sigma^2_0\right\} = O(s) \quad \text{as } s\to 0,$$

the lemma is established. □

Lemma 3.7. *Suppose that $e^{-\eta_* t_m} = O(\Delta^{1/2})$. Then,*

$$\max_{m \le k \le N} \|\sigma_{nk}^2(\theta) - \tilde{\sigma}_{nk}^2(\theta)\|_2 = O(\Delta^{1/2}).$$

Proof: Since

$$\tilde{\sigma}_{nk}^2(\theta) = \beta \sum_{i=0}^{k} \Delta t_{k-i} e^{-\eta(t_k - t_{k-i})} + e^{-\eta t_k} \tilde{\sigma}_{n0}^2(\theta) + \varphi e^{-\eta t_k} \sum_{i=1}^{k} e^{\eta t_{k-i}} Y_{n,k-i+1}^2$$

and

$$\begin{aligned}
&\sum_{i=1}^{k} e^{-\eta(t_k - t_{k-i})} Y_{n,k-i+1}^2 \\
&= \sum_{i=1}^{k} e^{-\eta(t_k - t_{k-i})} \left\{ [G,G]_{t_{k-i+1}} - [G,G]_{t_{k-i}} + 2 \int_{(t_{k-i}, t_{k-i+1}]} (G_{u-} - G_{t_{k-i}}) dG_u \right\} \\
&= \sum_{i=1}^{k} e^{-\eta(t_k - t_{k-i})} \left\{ \int_{(t_{k-i}, t_{k-i+1}]} \sigma_u^2 d[L,L]_u + 2 \int_{(t_{k-i}, t_{k-i+1}]} (G_{u-} - G_{t_{k-i}}) \sigma_u dL_u \right\},
\end{aligned}$$

we only have to deal with

$$e^{-\eta t_k} \sum_{0 < s \le t_k} e^{\eta s} \sigma_s^2 d[L,L]_s - e^{-\eta t_k} \sum_{i=1}^{k} e^{\eta t_{k-i}} \int_{(t_{k-i}, t_{k-i+1}]} \sigma_u^2 d[L,L]_u \tag{3.3}$$

and

$$\sum_{i=1}^{k} e^{-\eta(t_k - t_{k-i})} \int_{(t_{k-i}, t_{k-i+1}]} (G_{u-} - G_{t_{k-i}}) \sigma_u dL_u. \tag{3.4}$$

Note that (3.3) is bounded by

$$\begin{aligned}
&\sum_{i=1}^{k} (e^{\eta \Delta t_{k-i+1}} - 1) e^{-\eta(t_k - t_{k-i})} \int_{(t_{k-i}, t_{k-i+1}]} \sigma_u^2 d[L,L]_u \\
&\quad = \sum_{i=1}^{k} (e^{\eta \Delta t_{k-i+1}} - 1) e^{-\eta(t_k - t_{k-i})} \int_{(t_{k-i}, t_{k-i+1}]} \sigma_u^2 du \\
&\qquad + \sum_{i=1}^{k} (e^{\eta \Delta t_{k-i+1}} - 1) e^{-\eta(t_k - t_{k-i})} \int_{(t_{k-i}, t_{k-i+1}]} \sigma_u^2 d\left\{ [L,L]_u - u \right\},
\end{aligned}$$

where the second term is a sum of martingale differences. Thus, the L^2-norm of (3.3) is $O(\Delta^{1/2})$ uniformly in $m \le k \le N$, since

$$\begin{aligned}
\mathrm{E}\left(\int_{(s,t]} \sigma_u^2 d\left\{ [L,L]_u - u \right\} \right)^2 &= \mathrm{E}\left(\int_{(0,t-s]} \sigma_u^2 d\left\{ [L,L]_u - u \right\} \right)^2 \\
&= \mathrm{E}\sigma_0^4 \cdot \mathrm{E}[L,L]_1^2 \cdot (t-s).
\end{aligned}$$

Moreover, since (3.4) is also a sum of martingale differences, the L^2-norm of (3.4) is $O(\Delta^{1/2})$ due to Lemma 3.6. Hence, the proof is completed. □

For vector $\mathbf{x} = (x_1, x_2, x_3)'$, we denote $|\mathbf{x}| := \sqrt{\mathbf{x}'\mathbf{x}}$.

Lemma 3.8.

$$\max_{m \le k \le N} \left\| \sup_{\theta \in \Theta} \frac{1}{\tilde{\rho}^2_{nk}(\theta)} \left| \frac{\partial}{\partial \theta} \tilde{\rho}^2_{nk}(\theta) \right| \right\|_2 < \infty.$$

Proof: Due to Lemma 3.5, we have

$$\begin{aligned} \sup_{\theta \in \Theta} \frac{1}{\tilde{\rho}^2_{nk}(\theta)} \left| \frac{\partial}{\partial \theta} \tilde{\rho}^2_{nk}(\theta) \right| &\le K \sup_{\theta \in \Theta} \left\{ \left| \frac{\partial}{\partial \theta} \tilde{\sigma}^2_{n,k-1}(\theta) \right| + O(\Delta t_k) \tilde{\sigma}^2_{n,k-1}(\theta) \right\} \\ &\le K \left\{ 1 + \sum_{i=1}^{k-1} e^{-(\eta_*/2)(t_{k-1} - t_{k-i-1})} Y^2_{n,k-i-1} \right\}. \end{aligned}$$

Further, according to the proof of Lemma 3.7,

$$\max_{m \le k \le N} \left\| \sum_{i=1}^{k-1} e^{-(\eta_*/2)(t_{k-1} - t_{k-i-1})} Y^2_{n,k-i-1} - \int_{(0,t_{k-1}]} e^{-(\eta_*/2)(t_{k-1}-s)} \sigma^2_s d[L,L]_s \right\|_2 \to 0.$$

Since

$$\max_{m \le k \le N} \left\| \int_{(0,t_{k-1}]} e^{-(\eta_*/2)(t_{k-1}-s)} \sigma^2_s d[L,L]_s \right\|_2 \le \left\| \int_{(-\infty,0]} e^{-(\eta_*/2)s} \sigma^2_s d[L,L]_s \right\|_2 < \infty,$$

the lemma is validated. □

In fact, Lemma 3.10 below shows a more general result. However, Lemma 3.8 is sufficient to verify the consistency. Finally, we verify the uniform convergence of the likelihood function. In what follows, we denote

$$\rho^2_{nk}(\theta) := \left(\sigma^2_{n,k-1}(\theta) - \frac{\beta}{\eta - \varphi} \right) \left(\frac{\exp\{(\eta - \varphi)\Delta t_k\} - 1}{\eta - \varphi} \right) + \frac{\beta \Delta t_k}{\eta - \varphi}$$

and $\rho^2_{nk} := \rho^2_{nk}(\theta^\circ)$.

Proposition 3.2.

$$\sup_{\theta \in \Theta} \left| \frac{1}{t_N} \mathcal{L}_N(\theta) - \Upsilon(\theta) \right| = o_P(1).$$

Proof: We have

$$\begin{aligned} &\frac{1}{t_N} \sum_{k=m}^{N} \mathrm{l}_{nk}(\theta) \Delta t_k \\ &= \frac{1}{t_N} \sum_{k=m}^{N} \{ \mathrm{l}_{nk}(\theta) - \mathrm{E}(\mathrm{l}_{nk}(\theta) | \mathcal{F}_{n,k-1}) \} \Delta t_k + \frac{1}{t_N} \sum_{k=m}^{N} \mathrm{E}(\mathrm{l}_{nk}(\theta) | \mathcal{F}_{n,k-1}) \Delta t_k \\ &= \frac{1}{t_N} \sum_{k=m}^{N} \{ \mathrm{l}_{nk}(\theta) - \mathrm{E}(\mathrm{l}_{nk}(\theta) | \mathcal{F}_{n,k-1}) \} \Delta t_k - \frac{1}{t_N} \sum_{k=m}^{N} \left\{ \frac{\rho^2_{nk}}{\tilde{\rho}^2_{nk}(\theta)} + \log \frac{\tilde{\rho}^2_{nk}(\theta)}{\Delta t_k} \right\} \Delta t_k, \end{aligned}$$

where the first term is a sum of martingale differences, which converges to 0 in probability. Then, we obtain from Lemmas 3.5 and 3.7 that

$$\begin{aligned}
&\max_{m\le k\le N}\left\|\frac{\rho_{nk}^2}{\tilde\rho_{nk}^2(\theta)}-\frac{\rho_{nk}^2}{\rho_{nk}^2(\theta)}+\log\tilde\rho_{nk}^2(\theta)-\log\rho_{nk}^2(\theta)\right\|_1\\
&\le K\max_{m\le k\le N}\left(\left\|\sigma_{n,k-1}^2\right\|_2+1\right)\left\|\tilde\sigma_{n,k-1}^2(\theta)-\sigma_{n,k-1}^2(\theta)\right\|_2\longrightarrow 0
\end{aligned}$$

and

$$\max_{m\le k\le N}\sup_{\theta\in\Theta}\left|\frac{\rho_{nk}^2(\theta)}{\Delta t_k\sigma_{n,k-1}^2(\theta)}-1\right|=O(\Delta),$$

which implies

$$\begin{aligned}
\frac{1}{t_N}\sum_{k=m}^N\left\{\frac{\rho_{nk}^2}{\tilde\rho_{nk}^2(\theta)}+\log\frac{\tilde\rho_{nk}^2(\theta)}{\Delta t_k}\right\}\Delta t_k-\frac{1}{t_N}\sum_{k=m}^N\left\{\frac{\sigma_{n,k-1}^2}{\sigma_{n,k-1}^2(\theta)}+\log\sigma_{n,k-1}^2(\theta)\right\}\Delta t_k&\\
=o_P(1).&
\end{aligned}$$

Note that due to Lemma 3.4, the pointwise convergence holds:

$$\frac{1}{t_N}\sum_{k=m}^N \mathrm{l}_{nk}(\theta)\Delta t_k\xrightarrow{P}\Upsilon(\theta),\quad \text{for each }\theta\in\Theta. \tag{3.5}$$

Below, we verify the uniform convergence. Letting $\dot{\mathrm{l}}_{nk}(\theta):=\frac{\partial}{\partial\theta}\mathrm{l}_{nk}(\theta)$, we have

$$\frac{1}{t_N}\sup_{|\theta_1-\theta_2|<h}|\mathcal{L}_N(\theta_1)-\mathcal{L}_N(\theta_2)|\le\frac{1}{t_N}\sum_{k=m}^N\sup_{\theta\in\Theta}\left|\dot{\mathrm{l}}_{nk}(\theta)\right|\Delta t_k h$$

since $\theta_1+\lambda(\theta_2-\theta_1)\in\Theta$ for any $\lambda\in(0,1)$. Moreover, due to Lemmas 3.1, 3.5, and 3.8,

$$\begin{aligned}
\max_{m\le k\le N}\mathrm{E}\sup_{\theta\in\Theta}\left|\dot{\mathrm{l}}_{nk}(\theta)\right| &\le \max_{m\le k\le N}\mathrm{E}\sup_{\theta\in\Theta}\left\{\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta)}+1\right\}\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta)\right|\\
&\le K\max_{m\le k\le N}\mathrm{E}\left\{\frac{Y_{nk}^2}{\Delta t_k}+1\right\}\sup_{\theta\in\Theta}\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta)\right|\\
&= \max_{m\le k\le N}\mathrm{E}\left\{\frac{\mathrm{E}(Y_{nk}^2|\mathcal{F}_{t_{k-1}})}{\Delta t_k}+1\right\}\sup_{\theta\in\Theta}\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta)\right|\\
&\le K\max_{m\le k\le N}\mathrm{E}\left\{\sigma_{t_{k-1}}^2+1\right\}\sup_{\theta\in\Theta}\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta)\right|\\
&\le K\max_{m\le k\le N}\left\|\sup_{\theta\in\Theta}\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta)\right|\right\|_2<\infty.
\end{aligned}$$

Therefore, we obtain

$$\lim_{h\to0}\limsup_{n\to\infty}\mathrm{E}\frac{1}{t_N}\sup_{|\theta_1-\theta_2|<h}|\mathcal{L}_N(\theta_1)-\mathcal{L}_N(\theta_2)|=0. \tag{3.6}$$

Now, for given $h > 0$, take finitely many open balls $B_h(\theta_i) := \{\theta \in \Theta : |\theta - \theta_i| < h\}$ with $\theta_i \in \Theta$ such that $\Theta \subset \bigcup_i B_h(\theta_i)$. Then,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{1}{t_N}\mathcal{L}_N(\theta) - \Upsilon(\theta)\right| \\
&\quad \le \max_i \sup_{\theta\in B_h(\theta_i)}\left|\frac{1}{t_N}\mathcal{L}_N(\theta) - \frac{1}{t_N}\mathcal{L}_N(\theta_i)\right| + \max_i \left|\frac{1}{t_N}\mathcal{L}_N(\theta_i) - \Upsilon(\theta_i)\right| \\
&\qquad + \max_i \sup_{\theta\in B_h(\theta_i)} |\Upsilon(\theta_i) - \Upsilon(\theta)|.
\end{aligned}$$

Thus, we obtain from (3.5) that for every $\epsilon > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\sup_{\theta\in\Theta}\left|\frac{1}{t_N}\mathcal{L}_N(\theta) - \Upsilon(\theta)\right| > \epsilon\right) \\
&\quad \le \limsup_{n\to\infty} P\left(\sup_{|\theta_1-\theta_2|<h}\left|\frac{1}{t_N}\mathcal{L}_N(\theta_1) - \frac{1}{t_N}\mathcal{L}_N(\theta_2)\right| > \frac{\epsilon}{3}\right) \\
&\qquad + P\left(\sup_{|\theta_1-\theta_2|<h}|\Upsilon(\theta_1) - \Upsilon(\theta_2)| > \frac{\epsilon}{3}\right),
\end{aligned}$$

so that the uniform convergence is achieved by letting $h \to 0$ thanks to (3.6) and Proposition 3.1. □

The Proof of Consistency. Let $\epsilon > 0$ and $B_\epsilon(\theta^\circ) := \{\theta \in \Theta : |\theta - \theta^\circ| < \epsilon\}$. Then, $\Theta - B_\epsilon(\theta^\circ)$ is compact, since Θ is taken as a compact subset in $\mathbb{R}^3$.

$$H_n = \left\{\theta \in \Theta : \Upsilon(\theta) < \Upsilon(\theta^\circ) - \frac{1}{n}\right\}, \quad n \in \mathbb{N}$$

constitute a collection of open subsets relative to Θ, which covers $\Theta - B_\epsilon(\theta^\circ)$ since $\Upsilon(\theta) < \Upsilon(\theta^\circ)$ for each $\theta \in \Theta - B_\epsilon(\theta^\circ)$ (cf. Proposition 3.1). By virtue of compactness, there is $n_0 \in \mathbb{N}$ such that $\Theta - B_\epsilon(\theta^\circ) \subset H_{n_0}$, i.e.,

$$\sup\{\Upsilon(\theta) : \theta \in \Theta - B_\epsilon(\theta^\circ)\} \le \Upsilon(\theta^\circ) - \frac{1}{n_0}.$$

Therefore, by Proposition 3.2, we have that with probability tending to 1,

$$\sup\left\{\frac{1}{t_N}\mathcal{L}_N(\theta) : \theta \in \Theta - B_\epsilon(\theta^\circ)\right\} \le \Upsilon(\theta^\circ) - \frac{1}{2n_0}.$$

On the other hand,

$$\frac{1}{t_N}\mathcal{L}_N(\theta^\circ) \xrightarrow{P} \Upsilon(\theta^\circ), \quad \frac{1}{t_N}\mathcal{L}_N(\theta^\circ) \le \frac{1}{t_N}\mathcal{L}_N(\hat{\theta}_n).$$

Hence, $\lim_{n\to\infty} P\left(\hat{\theta}_n \in B_\epsilon(\theta^\circ)\right) = 1$. □

3.2. The Proof of Asymptotic Normality

In this subsection, we assume that **C1-C4** hold. By Taylor's theorem, we have

$$
\begin{aligned}
(3.7)\quad \mathbf{0} &= \frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\dot{\mathrm{l}}_{nk}(\hat{\theta}_n)\Delta t_k\\
&= \frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\dot{\mathrm{l}}_{nk}(\theta^\circ)\Delta t_k+\left\{\frac{1}{t_N}\sum_{k=m}^{N}\ddot{\mathrm{l}}_{nk}(\theta_n^*)\Delta t_k\right\}\cdot\sqrt{t_N}(\hat{\theta}_n-\theta^\circ),
\end{aligned}
$$

where

$$
\dot{\mathrm{l}}_{nk}(\theta^\circ)=\frac{\partial}{\partial\theta}\mathrm{l}_{nk}(\theta^\circ)=\left(\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\tilde{\rho}_{nk}^2(\theta^\circ),
$$

$$
\begin{aligned}
\ddot{\mathrm{l}}_{nk}(\theta)=\frac{\partial}{\partial\theta\partial\theta'}\mathrm{l}_{nk}(\theta) \;&=\; \left(1-2\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta)}\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta)}\frac{\partial}{\partial\theta}\tilde{\rho}_{nk}^2(\theta)\frac{1}{\tilde{\rho}_{nk}^2(\theta)}\frac{\partial}{\partial\theta'}\tilde{\rho}_{nk}^2(\theta)\\
&\quad+\left(\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta)}-1\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta)}\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\rho}_{nk}^2(\theta).
\end{aligned}
$$

More precisely,

$$
\ddot{\mathrm{l}}_{nk}(\theta_n^*)=\begin{pmatrix}\frac{\partial}{\partial\theta'}\frac{\partial}{\partial\beta}\mathrm{l}_{nk}(\theta^\circ+\lambda_1(\hat{\theta}_n-\theta^\circ))\\ \frac{\partial}{\partial\theta'}\frac{\partial}{\partial\eta}\mathrm{l}_{nk}(\theta^\circ+\lambda_2(\hat{\theta}_n-\theta^\circ))\\ \frac{\partial}{\partial\theta'}\frac{\partial}{\partial\varphi}\mathrm{l}_{nk}(\theta^\circ+\lambda_3(\hat{\theta}_n-\theta^\circ))\end{pmatrix}
$$

for some $\lambda_1,\lambda_2,\lambda_3\in(0,1)$.

Let $\frac{\partial^q}{\partial\beta^i\partial\eta^j\partial\varphi^l}$ be a differential operator of order q, where q,i,j,l are non-negative integers with $i+j+l=q$. Observe that

$$
\begin{aligned}
(3.8)\,&\frac{1}{\Delta t_k}\frac{\partial^q}{\partial\beta^i\partial\eta^j\partial\varphi^l}\tilde{\rho}_{nk}^2(\theta)\\
&=\;\frac{\partial^q}{\partial\beta^i\partial\eta^j\partial\varphi^l}\tilde{\sigma}_{n,k-1}^2(\theta)+O(\Delta t_k)\left\{\sum_{p\le q}\sum_{a,b,c}\frac{\partial^p}{\partial\beta^a\partial\eta^b\partial\varphi^c}\tilde{\sigma}_{n,k-1}^2(\theta)+1\right\}
\end{aligned}
$$

uniformly in $\theta \in \Theta$, and

$$\begin{aligned}(3.9)\quad & \frac{\partial^q}{\partial\beta^a\partial\eta^b\partial\varphi^c}\tilde{\sigma}^2_{nk}(\theta)\\
&= \frac{\partial^q}{\partial\beta^a\partial\eta^b\partial\varphi^c}\left\{\beta\sum_{i=0}^{k-1}\Delta t_{k-i}e^{-\eta(t_k-t_{k-i})}\right\}\\
&\quad+(-1)^q 1(a=0,c=0)\varphi\sum_{i=1}^{k}(t_k-t_{k-i})^q e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}\\
&\quad+(-1)^{q-1}1(a=0,c=1)\sum_{i=1}^{k}(t_k-t_{k-i})^{q-1}e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}\\
&\quad+\frac{\partial^q}{\partial\beta^a\partial\eta^b\partial\varphi^c}\frac{\beta e^{-\eta t_k}}{\eta-\varphi}.\end{aligned}$$

Define

$$\begin{aligned}\frac{\partial^q}{\partial\beta^a\partial\eta^b\partial\varphi^c}\sigma^2_t(\theta) := & \frac{\partial^q}{\partial\beta^a\partial\eta^b\partial\varphi^c}\frac{\beta}{\eta}\\
&+(-1)^q 1(a=0,c=0)\varphi\int_{-\infty<s<t}(t-s)^q e^{-\eta(t-s)}\sigma^2_s d[L,L]_s\\
&+(-1)^{q-1}1(a=0,c=1)\int_{-\infty<s<t}(t-s)^{q-1}e^{-\eta(t-s)}\sigma^2_s d[L,L]_s.\end{aligned}$$

Below, we show that a nice approximation to $\left\{\frac{\partial}{\partial\theta}\sigma^2_{nk}(\theta)\right\}$ is achievable similarly to Lemma 3.7. For a random vector $\mathbf{X}=(X_1,X_2,X_3)'$, we denote $\|\mathbf{X}\|_2 := \sqrt{\mathrm{E}\mathbf{X}'\mathbf{X}}$.

Lemma 3.9.

$$\max_{m\le k\le N}\left\|\frac{\partial}{\partial\theta}\tilde{\sigma}^2_{nk}(\theta)-\frac{\partial}{\partial\theta}\sigma^2_{nk}(\theta)\right\|_2 = O(\Delta^{1/2}).$$

Proof: We can express

$$\begin{aligned}&\sum_{i=1}^{k}(t_k-t_{k-i})\,e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}\\
&= \sum_{i=1}^{k}(t_k-t_{k-i})\,e^{-\eta(t_k-t_{k-i})}\left\{\int_{(t_{k-i},t_{k-i+1}]}\sigma^2_u\,d[L,L]_u + 2\int_{(t_{k-i},t_{k-i+1}]}(G_{u-}-G_{t_{k-i}})\sigma_u\,dL_u\right\}\end{aligned}$$

and

$$\begin{aligned}&\left|\sum_{i=1}^{k}(t_k-t_{k-i})\,e^{-\eta(t_k-t_{k-i})}\int_{(t_{k-i},t_{k-i+1}]}\sigma^2_u d[L,L]_u-\int_0^{t_k}(t_k-u)e^{-\eta(t_k-u)}\sigma^2_u d[L,L]_u\right|\\
&\le \sum_{i=1}^{k}\sup_{u\in(t_{k-i},t_{k-i+1}]}\left|(t_k-t_{k-i})\,e^{-\eta(t_k-t_{k-i})}-(t_k-u)e^{-\eta(t_k-u)}\right|\int_{(t_{k-i},t_{k-i+1}]}\sigma^2_u d[L,L]_u\\
&\le K\sum_{i=1}^{k}\Delta t_{k-i+1}e^{-\eta(t_k-t_{k-i})}(t_k-t_{k-i}+1)\int_{(t_{k-i},t_{k-i+1}]}\sigma^2_u d[L,L]_u.\end{aligned}$$

By virtue of the above facts, the lemma can be proven in the same fashion to prove Lemma 3.7. □

Lemma 3.10. *For any $p > 0$ and any nonnegative integer q,*

$$\max_{m\le k\le N} \mathrm{E}\sup_\theta \left|\frac{1}{\tilde\sigma^2_{n,k-1}(\theta)}\frac{\partial^q\tilde\sigma^2_{n,k-1}(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|^p \vee \mathrm{E}\sup_\theta\left|\frac{1}{\sigma^2_0(\theta)}\frac{\partial^q\sigma^2_0(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|^p < \infty \tag{3.10}$$

and

$$\max_{m\le k\le N} \mathrm{E}\sup_\theta \left|\frac{1}{\tilde\rho^2_{nk}(\theta)}\frac{\partial^q\tilde\rho^2_{nk}(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|^p \vee \mathrm{E}\sup_\theta\left|\frac{1}{\rho^2_{nk}(\theta)}\frac{\partial^q\rho^2_{nk}(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|^p < \infty.$$

Proof: Assume that $p > 1$. In view of (3.9), we have

$$\left|\frac{1}{\tilde\sigma^2_{n,k-1}(\theta)}\frac{\partial^q\tilde\sigma^2_{n,k-1}(\theta)}{\partial\beta^a\partial\eta^b\partial\varphi^c}\right| \le K\frac{1+\sum_{l=0}^1\sum_{i=1}^k (t_k-t_{k-i})^{q-l}\, e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}}{\beta/\eta+\varphi\sum_{i=1}^k e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}}.$$

Since $\frac{x}{c+x}\le x^{1/p}$ holds for every $x>0$ and $c>0$, letting $B_j := \{i : j\le t_i < j+1, i\le k\}$, we have

$$\begin{aligned}
&\mathrm{E}\sup_\theta\left|\frac{\sum_{i=1}^k (t_k-t_{k-i})^q\, e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}}{\beta/\eta+\varphi\sum_{i=1}^k e^{-\eta(t_k-t_{k-i})}Y^2_{n,k-i+1}}\right|^p \\
&\le \mathrm{E}\sup_\theta\left|\sum_{j=0}^{[t_k]}([t_k]-j+1)^q\,\frac{\sum_{i\in B_j} e^{-\eta(t_k-t_i)}Y^2_{n,i+1}}{\beta/\eta+\varphi\sum_{i\in B_j} e^{-\eta(t_k-t_i)}Y^2_{n,i+1}}\right|^p \\
&\le \mathrm{E}\sup_\theta\left|\frac{1}{\varphi}\sum_{j=0}^{[t_k]}([t_k]-j+1)^q\left(\sum_{i\in B_j} e^{-\eta(t_k-t_i)}Y^2_{n,i+1}\right)^{1/p}\right|^p \\
&\le K\mathrm{E}\left|\sum_{j=0}^{[t_k]}([t_k]-j+1)^q\, e^{-\eta_*/p([t_k]-j)}\left(\sum_{i\in B_j}Y^2_{n,i+1}\right)^{1/p}\right|^p \\
&\le K\left|\sum_{j=0}^{[t_k]}([t_k]-j+1)^q\, e^{-\eta_*/p([t_k]-j)}\left(\mathrm{E}\sum_{i\in B_j}Y^2_{n,i+1}\right)^{1/p}\right|^p < \infty
\end{aligned}$$

uniformly in $m\le k\le N$. Similarly, we also have

$$\frac{1}{\sigma^2_0(\theta)}\frac{\partial^q\sigma^2_0(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l} \le K\frac{1+\sum_{l=0}^1\int_{-\infty}^0 (-u)^{q-l}\, e^{\eta u}\sigma^2_u d[L,L]_u}{\beta/\eta+\varphi\int_{-\infty}^0 e^{\eta u}\sigma^2_u d[L,L]_u}$$

and

$$
\begin{aligned}
&\mathrm{E}\sup_{\theta}\left|\frac{\int_{-\infty}^{0}(-u)^q\,e^{\eta u}\sigma_u^2 d[L,L]_u}{\beta/\eta+\varphi\int_{-\infty}^{0}e^{\eta u}\sigma_u^2 d[L,L]_u}\right|^p\\
&\le\ \mathrm{E}\sup_{\theta}\left|\sum_{j=0}^{\infty}(j+1)^q\,\frac{\int_{-j-1}^{-j}e^{\eta u}\sigma_u^2 d[L,L]_u}{\beta/\eta+\varphi\int_{-j-1}^{-j}e^{\eta u}\sigma_u^2 d[L,L]_u}\right|^p\\
&\le\ \mathrm{E}\sup_{\theta}\left|\frac{1}{\varphi}\sum_{j=0}^{\infty}(j+1)^q\left(\int_{-j-1}^{-j}e^{\eta u}\sigma_u^2 d[L,L]_u\right)^{1/p}\right|^p\\
&\le\ K\mathrm{E}\left|\sum_{j=0}^{\infty}(j+1)^q\,e^{-(\eta_*/p)j}\left(\int_{-j-1}^{-j}\sigma_u^2 d[L,L]_u\right)^{1/p}\right|^p\\
&\le\ K\left|\sum_{j=0}^{\infty}(j+1)^q\,e^{-(\eta_*/p)j}\left(\mathrm{E}\int_{-j-1}^{-j}\sigma_u^2 d[L,L]_u\right)^{1/p}\right|^p<\infty.
\end{aligned}
$$

Therefore, we obtain (3.10).

Now, due to (3.8),

$$
\begin{aligned}
\left|\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial^q\tilde\rho_{nk}^2(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|\ \le\ &K\left|\frac{1}{\sigma_{n,k-1}^2(\theta)}\frac{\partial^q\tilde\sigma_{n,k-1}^2(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|\\
&+O(\Delta t_k)\left\{\sum_{p\le q}\sum_{a,b,c}\left|\frac{1}{\sigma_{n,k-1}^2(\theta)}\frac{\partial^p\tilde\sigma_{n,k-1}^2(\theta)}{\partial\beta^a\partial\eta^b\partial\varphi^c}\right|+1\right\}
\end{aligned}
$$

uniformly in $\theta\in\Theta$. Further, a similar argument can be applied to $\sup_\theta\left|\frac{1}{\rho_{nk}^2(\theta)}\frac{\partial^q\rho_{nk}^2(\theta)}{\partial\beta^i\partial\eta^j\partial\varphi^l}\right|$. Hence, the lemmas is proved. □

Below, we establish a central limit theorem for the asymptotic normality. To this end, we show that the score function can be approximated by a sum of square integrable martingale differences. For vector $\mathbf{x}=(x_1,x_2,x_3)'$, we denote $|\mathbf{x}|:=\sqrt{\mathbf{x}'\mathbf{x}}$. And for random vector $\mathbf{X}=(X_1,X_2,X_3)'$, we denote $\|\mathbf{X}\|_1:=\mathrm{E}|\mathbf{X}|$.

Lemma 3.11. *Suppose that $t_N\Delta\to 0$. Then,*

$$
\begin{aligned}
&\frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\tilde\rho_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\tilde\rho_{nk}^2(\theta^\circ)\Delta t_k\\
&=\frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\rho_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\rho_{nk}^2(\theta^\circ)\Delta t_k+o_P(1) \qquad (3.11)\\
&=\frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\sigma_{n,k-1}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\Delta t_k+o_P(1). \qquad (3.12)
\end{aligned}
$$

Proof: Due to Lemmas 3.5, 3.7, 3.9, and 3.10, we have

$$
\begin{aligned}
&\left\| \left(\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta^\circ)} - \frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)} \right) \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\|_1 \\
&= \mathrm{E}\left\{ Y_{nk}^2 \left| \left(\frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} - \frac{1}{\rho_{nk}^2(\theta^\circ)} \right) \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right| \right\} \\
&= \mathrm{E}\left\{ \mathrm{E}\left\{ Y_{nk}^2 | \mathcal{F}_{t_{k-1}} \right\} \left| \left(\frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} - \frac{1}{\rho_{nk}^2(\theta^\circ)} \right) \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right| \right\} \\
&= \mathrm{E}\left| \left(\frac{\rho_{nk}^2(\theta^\circ)}{\tilde\rho_{nk}^2(\theta^\circ)} - 1 \right) \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right| \\
&\le \left\| \frac{\rho_{nk}^2(\theta^\circ) - \tilde\rho_{nk}^2(\theta^\circ)}{\tilde\rho_{nk}^2(\theta^\circ)} \right\|_2 \left\| \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\|_2 \\
&\le K \left\| \sigma_{n,k-1}^2(\theta^\circ) - \tilde\sigma_{n,k-1}^2(\theta^\circ) \right\|_2 \left\| \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\|_2 = O(\Delta^{1/2})
\end{aligned}
$$

uniformly in $m \le k \le N$, and

$$
\begin{aligned}
&\left\| \frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)} \left\{ \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} - \frac{1}{\rho_{nk}^2(\theta^\circ)} \frac{\partial \rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\} \right\|_1 \\
&= \mathrm{E}\left\{ \mathrm{E}(Y_{nk}^2 | \mathcal{F}_{t_{k-1}}) \frac{1}{\rho_{nk}^2(\theta^\circ)} \left| \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} - \frac{1}{\rho_{nk}^2(\theta^\circ)} \frac{\partial \rho_{nk}^2(\theta^\circ)}{\partial\theta} \right| \right\} \\
&= \left\| \frac{1}{\tilde\rho_{nk}^2(\theta^\circ)} \frac{\partial \tilde\rho_{nk}^2(\theta^\circ)}{\partial\theta} - \frac{1}{\rho_{nk}^2(\theta^\circ)} \frac{\partial \rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\|_1 \\
&\le K \left\{ \left\| \frac{\partial \tilde\sigma_{n,k-1}^2(\theta^\circ)}{\partial\theta} - \frac{\partial \sigma_{n,k-1}^2(\theta^\circ)}{\partial\theta} \right\|_1 \right. \\
&\qquad \left. + \left\| \tilde\sigma_{n,k-1}^2(\theta^\circ) - \sigma_{n,k-1}^2(\theta^\circ) \right\|_2 \left\| \frac{1}{\rho_{n,k}^2(\theta^\circ)} \frac{\partial \rho_{nk}^2(\theta^\circ)}{\partial\theta} \right\|_2 \right\} \\
&= O(\Delta^{1/2}),
\end{aligned}
$$

uniformly in $m \le k \le N$. Thus, (3.11) follows.

Further, note that

$$
\begin{aligned}
\frac{1}{\rho_{nk}^2(\theta^\circ)} \frac{\partial}{\partial\theta} \rho_{nk}^2(\theta^\circ) &= \frac{1}{\sigma_{n,k-1}^2(\theta^\circ)} \frac{\partial}{\partial\theta} \sigma_{n,k-1}^2(\theta^\circ) \\
&\quad + O(\Delta) \left\{ \left| \frac{1}{\sigma_{n,k-1}^2(\theta^\circ)} \frac{\partial}{\partial\theta} \sigma_{n,k-1}^2(\theta^\circ) \right| + 1 \right\}
\end{aligned}
$$

uniformly in $m \le k \le N$ and

$$\begin{aligned}
&\mathrm{E}\left|\frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\sigma_{n,k-1}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\Delta t_k\right| \\
&\le \frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}\mathrm{E}\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}+1\right)\left|\frac{1}{\sigma_{n,k-1}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\right|\Delta t_k \\
&\le \frac{1}{\sqrt{t_N}}\sum_{k=m}^{N}2\mathrm{E}\left|\frac{1}{\sigma_{n,k-1}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\right|\Delta t_k = O(t_N^{1/2}),
\end{aligned}$$

so that (3.12) holds. This completes the proof. □

The following proposition and lemma are concerned with the stability of the sum of conditional variances of the score function.

Proposition 3.3. *Suppose that* **C2-C3** *hold and* $\int_{\mathbb{R}} x^3 d\Pi(x) = 0$. *Then, as* $h \downarrow 0$,

$$\mathrm{E}\left\{(G_{t+h}-G_t)^4|\mathcal{F}_t\right\} = h\left(\int_{\mathbb{R}} x^4\Pi(dx)+o(1)\right)\sigma_t^4, \tag{3.13}$$

$$\mathrm{E}\left\{(G_{t+h}-G_t)^2|\mathcal{F}_t\right\} = h(1+o(1))\sigma_t^2 \tag{3.14}$$

uniformly in $t \ge 0$, *and therefore,*

$$\tau := \int_{\mathbb{R}} x^4\Pi(dx) = \lim_{h\downarrow 0}\frac{h\mathrm{E}\left\{(G_{t+h}-G_t)^4|\mathcal{F}_t\right\}}{\left\{\mathrm{E}\left\{(G_{t+h}-G_t)^2|\mathcal{F}_t\right\}\right\}^2} \quad \text{for every } t \ge 0.$$

Further,

$$\Sigma := \mathrm{E}\frac{1}{\sigma_0^4(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_0^2(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma_0^2(\theta^\circ)$$

is positive definite.

Proof: We defer the proof of (3.13) and (3.14) to Lemma A.2. Since $\mathrm{E}\frac{1}{\sigma_0^4(\theta^\circ)}\left|\frac{\partial}{\partial\theta}\sigma_0^2(\theta^\circ)\right|^2 < \infty$ (cf. Lemma 3.10), Σ is well defined and symmetric. Moreover, since we have that for $\lambda \in \mathbb{R}^3$, $\lambda'\frac{\partial}{\partial\theta}\sigma_0^2(\theta^\circ) = 0$ a.s. if and only if $\lambda = \mathbf{0}$, Σ is positive definite. □

For a matrix $A = (a_{ij})_{i,j=1,2,3}$, we denote $|A| := \left(\sum_{ij}|a_{ij}|^2\right)^{1/2}$.

Lemma 3.12.

$$\frac{1}{t_N}\sum_{k=m}^{N}\frac{1}{\sigma_{n,k-1}^4(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma_{n,k-1}^2(\theta^\circ)\Delta t_k \xrightarrow{P} \Sigma.$$

Proof: Notice

$$
\begin{aligned}
&\sup_{u\in(t_{k-1},t_k]}\left|\frac{1}{\sigma^4_{n,k-1}(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)-\frac{1}{\sigma^4_u(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma^2_u(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma^2_u(\theta^\circ)\right| \\
&\le \sup_{u\in(t_{k-1},t_k]}\left|\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}-\frac{1}{\sigma^2_u(\theta^\circ)}\right|\left|\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)\right| \\
&\quad+\sup_{u\in(t_{k-1},t_k]}\frac{1}{\sigma^2_u(\theta^\circ)}\left|\left(\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)-\frac{\partial}{\partial\theta}\sigma^2_u(\theta^\circ)\right)\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)\right| \\
&\quad+\sup_{u\in(t_{k-1},t_k]}\left|\frac{1}{\sigma^2_u(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma^2_u(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)\right|\left|\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}-\frac{1}{\sigma^2_u(\theta^\circ)}\right| \\
&\quad+\sup_{u\in(t_{k-1},t_k]}\left|\frac{1}{\sigma^2_u(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma^2_u(\theta^\circ)\frac{1}{\sigma^2_u(\theta^\circ)}\left(\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)-\frac{\partial}{\partial\theta'}\sigma^2_u(\theta^\circ)\right)\right|.
\end{aligned}
$$

We concentrate on the third term since the other terms can be treated similarly.

Note that

$$
\begin{aligned}
&\sup_{u\in(t_{k-1},t_k]}\left|\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}-\frac{1}{\sigma^2_u(\theta^\circ)}\right|\left|\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}\frac{\partial}{\partial\theta'}\sigma^2_{n,k-1}(\theta^\circ)\right| \\
&\le K\sup_{u\in(t_{k-1},t_k]}|\sigma^2_{n,k-1}(\theta^\circ)-\sigma^2_u(\theta^\circ)|\left|\frac{1}{\sigma^2_{n,k-1}(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)\frac{1}{\sigma^2_u(\theta^\circ)}\frac{\partial}{\partial\theta'}\sigma^2_u(\theta^\circ)\right|.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\sup_{u\in(t_{k-1},t_k]}|\sigma^2_{n,k-1}(\theta^\circ)-\sigma^2_u(\theta^\circ)| \\
&\le \varphi^\circ e^{\eta^\circ\Delta t_k}\int_{t_{k-1}}^{t_k}\sigma^2_t d[L,L]_t+\varphi^\circ\left(1-e^{-\eta^\circ\Delta t_k}\right)\int_{-\infty}^{t_{k-1}}e^{-\eta^\circ(t_{k-1}-t)}\sigma^2_t d[L,L]_t,
\end{aligned}
$$

we can have

$$\max_{m\le k\le N}\left\|\sup_{u\in(t_{k-1},t_k]}|\sigma^2_{n,k-1}(\theta^\circ)-\sigma^2_u(\theta^\circ)|\right\|_2=o(1).$$

Similarly,

$$\max_{m\le k\le N}\left\|\sup_{u\in(t_{k-1},t_k]}\left|\frac{\partial}{\partial\theta}\sigma^2_{n,k-1}(\theta^\circ)-\frac{\partial}{\partial\theta}\sigma^2_u(\theta^\circ)\right|\right\|_2=o(1).$$

Moreover, we have that for $h > 0$ and $p > 1, q > 0$,

$$\begin{aligned}
&\left| \frac{\int_{-\infty}^{h} (h-u)^q \, e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u}{\beta/\eta + \varphi \int_{-\infty}^{h} e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u} \right|^p \\
&\leq \left| \frac{\int_0^h e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u}{\beta/\eta + \varphi \int_0^h e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u} + \sum_{j=0}^{\infty} (j+1)^q \frac{\int_{-j-1}^{-j} e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u}{\beta/\eta + \varphi \int_{-j-1}^{-j} e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u} \right|^p \\
&\leq \left| \left(\int_0^h e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u \right)^{1/p} + \frac{1}{\varphi} \sum_{j=0}^{\infty} (j+1)^q \left(\int_{-j-1}^{-j} e^{-\eta(h-u)} \sigma_u^2 d[L,L]_u \right)^{1/p} \right|^p \\
&\leq \left| \left(\int_0^h \sigma_u^2 d[L,L]_u \right)^{1/p} + \frac{1}{\varphi} \sum_{j=0}^{\infty} (j+1)^q \, e^{-\eta(h+j)} \left(\int_{-j-1}^{-j} \sigma_u^2 d[L,L]_u \right)^{1/p} \right|^p .
\end{aligned}$$

Thus, it can be seen that

$$\left\| \sup_{u \in (t_{k-1}, t_k]} \left| \frac{1}{\sigma_u^2(\theta^\circ)} \frac{\partial}{\partial \theta'} \sigma_u^2(\theta^\circ) \right| \right\|_4 < \infty$$

and

$$\begin{aligned}
&\left\| \sup_{u \in (t_{k-1}, t_k]} \left| \frac{1}{\sigma_{n,k-1}^2(\theta^\circ)} \frac{\partial}{\partial \theta} \sigma_{n,k-1}^2(\theta^\circ) \frac{1}{\sigma_u^2(\theta^\circ)} \frac{\partial}{\partial \theta'} \sigma_u^2(\theta^\circ) \right| \right\|_2 \\
&\leq \left\| \left| \frac{1}{\sigma_{n,k-1}^2(\theta^\circ)} \frac{\partial}{\partial \theta} \sigma_{n,k-1}^2(\theta^\circ) \right| \right\|_4 \left\| \sup_{u \in (t_{k-1}, t_k]} \left| \frac{1}{\sigma_u^2(\theta^\circ)} \frac{\partial}{\partial \theta'} \sigma_u^2(\theta^\circ) \right| \right\|_4 < \infty.
\end{aligned}$$

Therefore,

$$\mathrm{E} \sup_{u \in (t_{k-1}, t_k]} \left| \frac{1}{\sigma_{n,k-1}^4(\theta^\circ)} \frac{\partial}{\partial \theta} \sigma_{n,k-1}^2(\theta^\circ) \frac{\partial}{\partial \theta'} \sigma_{n,k-1}^2(\theta^\circ) - \frac{1}{\sigma_u^4(\theta^\circ)} \frac{\partial}{\partial \theta} \sigma_u^2(\theta^\circ) \frac{\partial}{\partial \theta'} \sigma_u^2(\theta^\circ) \right| \to 0$$

uniformly in $m \leq k \leq N$. Then the lemma is validated by the ergodic theorem (cf. Lemma A.1). □

Now, we establish the asymptotical normality of the score function.

Proposition 3.4. *Suppose that there exists $\delta > 0$ such that $\mathrm{E}|G_h|^{4+\delta} = O(h)$. Then,*

$$\frac{1}{\sqrt{t_N}} \sum_{k=m}^{N} \dot{\mathrm{l}}_{nk}(\theta^\circ) \Delta t_k \Rightarrow N(0, \tau\Sigma).$$

Proof: Due to Lemma 3.11, it suffices to show that

$$\frac{1}{\sqrt{t_N}} \sum_{k=m}^{N} \left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)} - 1 \right) \frac{1}{\sigma_{n,k-1}^2(\theta^\circ)} \frac{\partial}{\partial \theta} \sigma_{n,k-1}^2(\theta^\circ) \Delta t_k \Rightarrow N(0, \tau\Sigma).$$

Let λ be any vector in $\mathbb{R}^3$ and

$$\xi_{n,k-1} := \frac{1}{\sigma^4_{n,k-1}(\theta^\circ)} \left\{ \lambda' \frac{\partial}{\partial\theta} \sigma^2_{n,k-1}(\theta^\circ) \right\}^2 .$$

Note that

$$\frac{\Delta t_k}{\sqrt{t_N}} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right) \frac{1}{\sigma^2_{n,k-1}(\theta^\circ)} \lambda' \frac{\partial}{\partial\theta} \sigma^2_{n,k-1}(\theta^\circ), \quad k = m, ..., N$$

are row-wise martingale differences with respect to $\{\mathcal{F}_{t_k} : k = m, ..., N\}$. Since due to Proposition 3.3 and Lemma 3.12

$$\frac{1}{t_N} \sum_{k=m}^{N} \left\{ \Delta t_k \left(\frac{\mathrm{E}\left\{ Y^4_{nk} | \mathcal{F}_{t_{k-1}} \right\}}{\rho^4_{nk}(\theta^\circ)} - 1 \right) \right\} \xi_{n,k-1} \Delta t_k \xrightarrow{P} \tau \lambda' \Sigma \lambda,$$

it suffices to verify Lindeberg's condition for martingale differences (cf. Theorem 35.12 of Billingsley ([4])). For $\epsilon > 0$ and $A > 0$, we have

$$\begin{aligned}
&\sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 \xi_{n,k-1} \frac{\Delta t^2_k}{t_N} I \left\{ \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 \xi_{n,k-1} \frac{\Delta t^2_k}{t_N} > \epsilon \right\} \\
&\le \sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 A \frac{\Delta t^2_k}{t_N} I \left\{ \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 A \frac{\Delta t^2_k}{t_N} > \epsilon \right\} \\
&\quad + \sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 \xi_{n,k-1} \frac{\Delta t^2_k}{t_N} I \left\{ \xi_{n,k-1} > A \right\},
\end{aligned}$$

and further, due to Lemma 3.5 and the fact $\mathrm{E}|G_h|^{4+\delta} = O(h)$,

$$\begin{aligned}
&\sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 A \frac{\Delta t^2_k}{t_N} I \left\{ \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 A \frac{\Delta t^2_k}{t_N} > \epsilon \right\} \\
&\le \sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 A \frac{\Delta t^2_k}{t_N} \left| \frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right|^{\delta/2} \left(\frac{A}{\epsilon} \right)^{\delta/4} \frac{\Delta t^{\delta/2}_k}{t^{\delta/4}_N} \\
&\le K \sum_{k=m}^{N} \mathrm{E} \left| \frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right|^{2+\delta/2} \frac{\Delta t^{2+\delta/2}_k}{t^{1+\delta/4}_N} \le K \sum_{k=m}^{N} \left\{ \left(\frac{\mathrm{E}|Y_{nk}|^{4+\delta}}{\Delta t^{2+\delta/2}_k} + 1 \right) \right\} \frac{\Delta t^{2+\delta/2}_k}{t^{1+\delta/4}_N} \\
&\le K \sum_{k=m}^{N} \left\{ \left(\frac{O(\Delta t_k)}{\Delta t^{2+\delta/2}_k} + 1 \right) \right\} \frac{\Delta t^{2+\delta/2}_k}{t^{1+\delta/4}_N} \to 0,
\end{aligned}$$

and due to Proposition 3.3,

$$\begin{aligned}
&\sum_{k=m}^{N} \mathrm{E} \left(\frac{Y^2_{nk}}{\rho^2_{nk}(\theta^\circ)} - 1 \right)^2 \xi_{n,k-1} \frac{\Delta t^2_k}{t_N} I \left\{ \xi_{n,k-1} > A \right\} \\
&= \sum_{k=m}^{N} \mathrm{E} \left\{ \Delta t_k \left(\frac{\mathrm{E}\{ Y^4_{nk} | \mathcal{F}_{t_{k-1}} \}}{\rho^4_{nk}(\theta^\circ)} - 1 \right) \xi_{n,k-1} \frac{\Delta t_k}{t_N} I \left\{ \xi_{n,k-1} > A \right\} \right\} \\
&\le K \mathrm{E} \left\{ \xi_0 I \left\{ \xi_0 > A \right\} \right\}.
\end{aligned}$$

Then by letting $A \to \infty$, we establish the proposition. □

Note that every component of $\frac{1}{t_N}\sum_{k=m}^{N} \ddot{\mathrm{l}}_{nk}(\theta_n^*)\Delta t_k$ is expressed as

$$\begin{aligned}\frac{1}{t_N}\sum_{k=m}^{N}\frac{\partial^2}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\theta_n^*)\Delta t_k &= \frac{1}{t_N}\sum_{k=m}^{N}\frac{\partial^2}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\theta^\circ)\Delta t_k \\ &\quad +\frac{1}{t_N}\sum_{k=m}^{N}\frac{\partial}{\partial\theta'}\frac{\partial^2}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\delta\theta_n^*+(1-\delta)\theta^\circ)(\theta_n^*-\theta^\circ)\Delta t_k\end{aligned}$$

with $\delta\in(0,1)$. The rest of this subsection is devoted to verifying the convergence of $\frac{1}{t_N}\sum_{k=m}^{N}\ddot{\mathrm{l}}_{nk}(\theta_n^*)\Delta t_k$.

Lemma 3.13.

$$\frac{1}{t_N}\sum_{k=m}^{N}\left(1-2\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\tilde{\rho}_{nk}^2(\theta^\circ)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta'}\tilde{\rho}_{nk}^2(\theta^\circ)\Delta t_k \xrightarrow{P} -\Sigma$$

and

$$\frac{1}{t_N}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\rho}_{nk}^2(\theta^\circ)\Delta t_k \xrightarrow{P} \mathbf{0}.$$

Hence,

$$\frac{1}{t_N}\sum_{k=m}^{N}\ddot{\mathrm{l}}_{nk}(\theta^\circ)\Delta t_k \xrightarrow{P} -\Sigma.$$

Proof: For convenience, we set $\partial_1 := \frac{\partial}{\partial\beta^i\partial\eta^j\partial\varphi^l}$ and $\partial_2 := \frac{\partial}{\partial\beta^a\partial\eta^b\partial\varphi^c}$ to denote any differential operators of the first order. Due to Lemmas 3.5 and 3.7, we have

$$\begin{aligned}&\mathrm{E}\left\{Y_{nk}^2\left|\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}-\frac{1}{\rho_{nk}^2(\theta^\circ)}\right|\left|\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right|\right\}\\ &= \mathrm{E}\left\{\rho_{nk}^2\left|\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}-\frac{1}{\rho_{nk}^2(\theta^\circ)}\right|\left|\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right|\right\}\\ &= \mathrm{E}\left\{\left|\frac{\rho_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right|\left|\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right|\right\}\\ &\le K\left\|\tilde{\sigma}_{n,k-1}^2(\theta^\circ)-\sigma_{n,k-1}^2(\theta^\circ)\right\|_2\left\|\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right\|_2\\ &\longrightarrow 0,\end{aligned}$$

uniformly in $m\le k\le N$, since

$$\max_{m\le k\le N}\left\|\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right\|_2<\infty$$

(cf. Lemma 3.10). Moreover, due to Proposition 3.3,

$$\begin{aligned}
&\frac{1}{t_N^2}\sum_{k=m}^{N}\mathrm{E}\left\{\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}-1\right)^2\left(\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right)^2\right\}\Delta t_k^2\\
&= \frac{1}{t_N^2}\sum_{k=1}^{N}\mathrm{E}\left\{\Delta t_k\left(\frac{\mathrm{E}\left(Y_{nk}^4|\mathcal{F}_{t_{k-1}}\right)}{\rho_{nk}^4(\theta^\circ)}-1\right)\left(\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right)^2\right\}\Delta t_k\\
&= \frac{K}{t_N^2}\sum_{k=1}^{N}\mathrm{E}\left(\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right)^2\Delta t_k\longrightarrow 0,
\end{aligned}$$

so that we get

$$\frac{1}{t_N}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\rho_{nk}^2(\theta^\circ)}-1\right)\left(\frac{\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right)\Delta t_k = o_P(1),$$

i.e.,

$$\frac{1}{t_N}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta}\tilde{\rho}_{nk}^2(\theta^\circ)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial}{\partial\theta'}\tilde{\rho}_{nk}^2(\theta^\circ)\Delta t_k\xrightarrow{P}\mathbf{0}.$$

Similarly, we can see that

$$\frac{1}{t_N}\sum_{k=m}^{N}\left(\frac{Y_{nk}^2}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right)\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\rho}_{nk}^2(\theta^\circ)\Delta t_k\xrightarrow{P}\mathbf{0}.$$

On the other hand, we have

$$\begin{aligned}
&\mathrm{E}\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}\left|\partial_1\tilde{\rho}_{nk}^2(\theta^\circ)-\partial_1\rho_{nk}^2(\theta^\circ)\right|\left|\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right|\\
&\le K\left\|\partial_1\tilde{\sigma}_{n,k-1}^2(\theta^\circ)-\partial_1\sigma_{n,k-1}^2(\theta^\circ)\right\|_2\left\|\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right\|_2\longrightarrow 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left|\frac{1}{\tilde{\rho}_{nk}^2(\theta^\circ)}-\frac{1}{\rho_{nk}^2(\theta^\circ)}\right|\left|\partial_1\rho_{nk}^2(\theta^\circ)\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right| &= \mathrm{E}\left|\frac{\rho_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}-1\right|\left|\frac{\partial_1\rho_{nk}^2(\theta^\circ)}{\rho_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right|\\
&\le K\left\|\tilde{\sigma}_{n,k-1}^2(\theta^\circ)-\sigma_{n,k-1}^2(\theta^\circ)\right\|_2\left\|\frac{\partial_1\rho_{nk}^2(\theta^\circ)}{\rho_{nk}^2(\theta^\circ)}\frac{\partial_2\tilde{\rho}_{nk}^2(\theta^\circ)}{\tilde{\rho}_{nk}^2(\theta^\circ)}\right\|_2\longrightarrow 0
\end{aligned}$$

uniformly in $m \le k \le N$. Therefore,

$$\frac{1}{t_N}\sum_{k=m}^{N}\ddot{\mathrm{l}}_{nk}(\theta^\circ)\Delta t_k = -\frac{1}{t_N}\sum_{k=m}^{N}\frac{1}{\sigma_{n,k-1}^4(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_{n,k-1}^2(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma_{n,k-1}^2(\theta^\circ)\Delta t_k+o_P(1).$$

Henceforth, the lemma is validated by Lemma 3.12. $\square$

Lemma 3.14. *We have*

$$\max_{m\le k\le N} \mathrm{E}\sup_{\theta\in\Theta}\left|\frac{\partial^3}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\theta)\right| < \infty.$$

Hence,

$$\frac{1}{t_N}\sum_{k=m}^{N}\frac{\partial}{\partial\theta'}\frac{\partial^2}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\delta\theta_n^* + (1-\delta)\theta^\circ)(\theta_n^* - \theta^\circ)\Delta t_k \xrightarrow{P} 0.$$

Proof: Observe that $\frac{\partial^3}{\partial\beta^i\partial\eta^j\partial\varphi^l}\mathrm{l}_{nk}(\theta)$ is a finite sum of the terms:

$$\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta)}\prod_{i=1}^{3}\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)},\quad \frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta)}\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial}{\partial\beta^a\partial\eta^b\partial\varphi^c}\tilde\rho_{nk}^2(\theta)\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial^2}{\partial\beta^i\partial\eta^j\partial\varphi^l}\tilde\rho_{nk}^2(\theta),$$
$$\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta)}\frac{1}{\tilde\rho_{nk}^2(\theta)}\frac{\partial^3}{\partial\beta^a\partial\eta^b\partial\varphi^c}\tilde\rho_{nk}^2(\theta),$$

where $\partial_i, (i=1,2,3)$ are differential operators of the first order. Now, by Lemmas 3.5 and 3.10,

$$\begin{aligned}
&\mathrm{E}\sup_{\theta}\frac{Y_{nk}^2}{\tilde\rho_{nk}^2(\theta)}\prod_{i=1}^{3}\left|\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right| \le K\mathrm{E}\frac{Y_{nk}^2}{\Delta t_k}\prod_{i=1}^{3}\sup_{\theta}\left|\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\\
&\le K\mathrm{E}\frac{\rho_{nk}^2}{\Delta t_k}\prod_{i=1}^{3}\sup_{\theta}\left|\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right| \le K\mathrm{E}\left(\sigma_{n,k-1}^2 + O(\Delta)\right)\prod_{i=1}^{3}\sup_{\theta}\left|\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\\
&\le K\|\sigma_{n,k-1}^2 + O(\Delta)\|_2\left\|\prod_{i=1}^{3}\sup_{\theta}\left|\frac{\partial_i\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\right\|_2\\
&\le K\|\sigma_{n,k-1}^2 + O(\Delta)\|_2\left\|\sup_{\theta}\left|\frac{\partial_1\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\right\|_4\left\|\sup_{\theta}\left|\frac{\partial_2\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\right\|_8\left\|\sup_{\theta}\left|\frac{\partial_3\tilde\rho_{nk}^2(\theta)}{\tilde\rho_{nk}^2(\theta)}\right|\right\|_8 < \infty
\end{aligned}$$

uniformly in $m \le k \le N$. The other terms can be treated in essentially the same fashion. Hence, the lemmas is asserted. □

The following proposition is due to Lemmas 3.13-3.14:

Proposition 3.5.

$$\frac{1}{t_N}\sum_{k=m}^{N}\ddot{\mathrm{l}}_{nk}(\theta_n^*)\Delta t_k \xrightarrow{P} -\Sigma.$$

The Proof of Asymptotic Normality. (2.4) can be proven by using standard arguments (cf. the proof of Theorem 2.2 in Francq and Zakoïan ([8])) and the results in (3.7) and Propositions 3.3-3.5. □

APPENDIX

Lemma A.1. *Suppose that* **C3** *holds. Then,*

$$\frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta)}+\log\sigma_s^2(\theta)\right\}ds \xrightarrow{P} \mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)}+\log\sigma_0^2(\theta)\right\}, \tag{A.1}$$

$$\frac{1}{t_N}\int_{t_m}^{t_N}\frac{1}{\sigma_s^4(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_s^2(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma_s^2(\theta^\circ)ds \xrightarrow{P} \mathrm{E}\frac{1}{\sigma_0^4(\theta^\circ)}\frac{\partial}{\partial\theta}\sigma_0^2(\theta^\circ)\frac{\partial}{\partial\theta'}\sigma_0^2(\theta^\circ). \tag{A.2}$$

Proof: We only verify (A.1) since (A.2) can be proved similarly. Let $h > 0$ and

$$\sigma_s^2(\theta,h) := \beta/\eta + \varphi\int_{(s-h,s)} e^{-\eta(s-u)}\sigma_u^2 d[L,L]_u.$$

Then we have

$$\begin{aligned}
&\mathrm{E}\left|\frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta)}+\log\sigma_s^2(\theta)\right\}ds - \frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta,h)}+\log\sigma_s^2(\theta,h)\right\}ds\right| \\
&\le \mathrm{E}\left|\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)}+\log\sigma_0^2(\theta)\right\}-\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta,h)}+\log\sigma_0^2(\theta,h)\right\}\right| \\
&\le \mathrm{E}K\left\{\sigma_0^2|\sigma_0^2(\theta)-\sigma_0^2(\theta,h)|\right\} \le K\|\sigma_0^2\|_2\|\sigma_0^2(\theta)-\sigma_0^2(\theta,h)\|_2 \le Ke^{-\eta h}.
\end{aligned} \tag{A.3}$$

Note that

$$\sigma_s^2 = \beta^\circ\int_{s-h}^{s} e^{X_u - X_{s-}}du + \sigma_{s-h}^2 e^{-X_{s-h}-X_{s-}}$$

and thus,

$$\frac{\sigma_s^2}{\sigma_s^2(\theta,h)}+\log\sigma_s^2(\theta,h) \in \mathcal{G}_{s-h}^s,$$

where $\mathcal{G}_s^t := \sigma\{\sigma_u, L_u - L_s : s < u < t\}$. Let

$$\alpha(v) := \sup_{0\le t<\infty}\sup\left\{P(A\cap B)-P(A)P(B) : A\in\mathcal{G}_{-\infty}^t, B\in\mathcal{G}_{t+v}^\infty\right\},$$

and define α^* in the same way with replacing $\mathcal{G}_s^t$ by $\sigma\{\sigma_u^2 : s < u < t\}$. According to the proof of Theorem 3.5 of Haug *et al.* ([10]), we can have

$$0 \le \alpha(v) \le 6\alpha^*(v) \to 0 \quad \text{as } v\to\infty,$$

(cf. Fasen ([7])), which implies that $\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta,h)}+\log\sigma_s^2(\theta,h) : s\ge 0\right\}$ is càglàd, strictly stationary and strong mixing. Thus,

$$Z_i := \int_{i-1}^{i}\left(\frac{\sigma_s^2}{\sigma_s^2(\theta,h)}+\log\sigma_s^2(\theta,h)\right)ds \in \mathcal{G}_{i-h-1}^i, \quad i = 1,2,...$$

is strictly stationary and ergodic. Then, since by the ergodic theorem,

$$\begin{aligned}\frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta,h)}+\log\sigma_s^2(\theta,h)\right\}ds &= \frac{1}{t_N}\sum_{i=[t_m]+1}^{[t_N]} Z_i + o_P(1)\\ &\xrightarrow{P} \mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta,h)}+\log\sigma_0^2(\theta,h)\right\}\end{aligned}$$

owing to (A.3), by letting $h\to\infty$, we get

$$\frac{1}{t_N}\int_{t_m}^{t_N}\left\{\frac{\sigma_s^2}{\sigma_s^2(\theta)}+\log\sigma_s^2(\theta)\right\}ds \xrightarrow{P} \mathrm{E}\left\{\frac{\sigma_0^2}{\sigma_0^2(\theta)}+\log\sigma_0^2(\theta)\right\}.$$

This completes the proof. □

Lemma A.2. *Suppose that* **C2-C3** *hold and* $\int_{\mathbb{R}} x^3 d\Pi(x)=0$. *Then, as* $h\downarrow 0$,

$$\begin{aligned}\mathrm{E}\left\{(G_{t+h}-G_t)^4|\mathcal{F}_t\right\} &= h\left(\int_{\mathbb{R}} x^4\Pi(dx)+o(1)\right)\sigma_t^4,\\ \mathrm{E}\left\{(G_{t+h}-G_t)^2|\mathcal{F}_t\right\} &= h(1+o(1))\sigma_t^2\end{aligned}$$

uniformly in $t\geq 0$.

Proof: By the strict stationarity, it suffices to consider the case $t=0$. For $h>0$,

$$\begin{aligned}(G_h-G_0)^2 &= 2\int_{(0,h]} G_{u-}dG_u+[G,G]_h = 2\int_{(0,h]} G_{u-}\sigma_u dL_u+\int_{(0,h]}\sigma_u^2 d[L,L]_u,\\ (G_h-G_0)^4 &= 2\int_{(0,h]} G_{s-}^2 dG_s^2+[G^2,G^2]_h\\ &= 4\int_{(0,h]} G_{s-}^3\sigma_s dL_s+2\int_{(0,h]} G_{s-}^2\sigma_s^2 d[L,L]_s\\ &\quad +4\int_{(0,h]} G_{s-}^2\sigma_s^2 d[L,L]_s+\int_{(0,h]}\sigma_s^4 d[[L,L],[L,L]]_s\\ &\quad +4\int_{(0,h]} G_{s-}\sigma_s^3 d[[L,L],L]_s,\end{aligned}$$

where

$$\mathrm{E}\left\{\int_{(0,h]} G_{s-}^3\sigma_s dL_s\,\middle|\,\mathcal{F}_0\right\}=0.$$

Since $\int_{\mathbb{R}} x^3 d\Pi(x)=0$,

$$\mathrm{E}\left\{\int_{(0,h]} G_{s-}\sigma_s^3 d[[L,L],L]_s\,\middle|\,\mathcal{F}_0\right\}=0.$$

Thus, we have

$$\mathrm{E}\left\{(G_h - G_0)^4|\mathcal{F}_0\right\} = 6\int_{(0,h]}\mathrm{E}\left\{G_{s-}^2\sigma_s^2|\mathcal{F}_0\right\}ds + \int_{\mathbb{R}}x^4\Pi(dx)\int_{(0,h]}\mathrm{E}\left\{\sigma_s^4|\mathcal{F}_0\right\}ds.$$

Let $Z_s = \int_{(0,s]} G_{u-}\sigma_u dL_u$. By the integration by parts and associativity (cf. [17]), we can write

$$\begin{aligned} Z_s\sigma_{s+}^2 \;&=\; \beta^\circ\int_{(0,s]}Z_{u-}du \;-\; \eta^\circ\int_{(0,s]}Z_{u-}\sigma_u^2du \;+\; \varphi^\circ\int_{(0,s]}Z_{s-}\sigma_u^2d[L,L]_u + \int_{(0,s]}\sigma_u^3G_{u-}dL_u \\ &\quad+\left[\int_{(0,\cdot]}(\beta^\circ - \eta^\circ\sigma_u^2)du \;+\; \varphi^\circ\int_{(0,\cdot]}\sigma_u^2d[L,L]_u\,,\;\int_{(0,\cdot]}G_{u-}\sigma_u dL_u\right]_s. \end{aligned}$$

Note that for $F \in \mathcal{F}_0$,

$$\begin{aligned} \mathrm{E}\left\{\int_{(0,s]}Z_{u-}du\cdot 1_F\right\} \;&=\; \int_{(0,s]}\mathrm{E}\left\{Z_{u-}1_F\right\}du = \int_{(0,s]}\mathrm{E}\left\{\mathrm{E}(Z_{u-}|\mathcal{F}_0)1_F\right\}du = 0, \\ \mathrm{E}\left\{\int_{(0,s]}Z_{u-}\sigma_u^2du\cdot 1_F\right\} \;&=\; \int_{(0,s]}\mathrm{E}\left\{Z_u\sigma_u^21_F\right\}du. \end{aligned}$$

Since

$$\begin{aligned} &\left[\int_{(0,\cdot]}(\beta^\circ - \eta^\circ\sigma_u^2)du + \varphi^\circ\int_{(0,\cdot]}\sigma_u^2d[L,L]_u, \int_{(0,\cdot]}G_{u-}\sigma_u dL_u\right] \\ &= \left[\int_{(0,\cdot]}(\beta^\circ - \eta^\circ\sigma_u^2)du, \int_{(0,\cdot]}G_{u-}\sigma_u dL_u\right] + \left[\varphi^\circ\int_{(0,\cdot]}\sigma_u^2d[L,L]_u, \int_{(0,\cdot]}G_{u-}\sigma_u dL_u\right] \\ &= \varphi^\circ\int_{(0,\cdot]}G_{u-}\sigma_u^3d[[L,L],L]_u, \end{aligned}$$

we have

$$\mathrm{E}\left\{\left[\int_{(0,\cdot]}(\beta^\circ - \eta^\circ\sigma_u^2)du + \varphi^\circ\int_{(0,\cdot]}\sigma_u^2d[L,L]_u, \int_{(0,\cdot]}G_{u-}\sigma_u dL_u\right]1_F\right\} = 0$$

due to $\int_{\mathbb{R}}x^3d\Pi(x) = 0$. Thus,

$$\mathrm{E}\left\{Z_s\sigma_s^21_F\right\} = (\varphi^\circ - \eta^\circ)\int_0^s\mathrm{E}\left\{Z_u\sigma_u^21_F\right\}du, \quad \mathrm{E}\left\{Z_0\sigma_0^21_F\right\} = 0,$$

which implies $\mathrm{E}\left\{Z_s\sigma_s^21_F\right\} = 0$ for each $F \in \mathcal{F}_0$. This in turn implies $\mathrm{E}\left\{Z_s\sigma_s^2|\mathcal{F}_0\right\} = 0$ and

$$\begin{aligned} \mathrm{E}\left\{G_{s-}^2\sigma_s^2|\mathcal{F}_0\right\} \;&=\; \mathrm{E}\left\{G_s^2\sigma_s^2|\mathcal{F}_0\right\} = \mathrm{E}\left\{\sigma_s^2\int_{(0,s]}\sigma_u^2d[L,L]_u + 2\sigma_s^2\int_{(0,s]}G_{u-}\sigma_u dL_u\,\middle|\,\mathcal{F}_0\right\} \\ &=\; \varphi_\circ^{-1}\mathrm{E}\left\{\sigma_s^2\left\{\sigma_{s+}^2 - \sigma_{0+}^2 - \beta^\circ s + \eta^\circ\int_0^s\sigma_u^2du\right\}\middle|\,\mathcal{F}_0\right\} \\ &=\; \varphi_\circ^{-1}\left\{\mathrm{E}\{\sigma_s^4|\mathcal{F}_0\} - (\sigma_0^2 + \beta^\circ s)\mathrm{E}\{\sigma_s^2|\mathcal{F}_0\} + \eta^\circ\int_0^s\mathrm{E}\{\sigma_s^2\sigma_u^2|\mathcal{F}_0\}du\right\}. \end{aligned}$$

Since $\sigma_s^2 = \beta^\circ \int_0^s e^{-(X_{s-}-X_u)}du + \sigma_0^2 e^{-X_{s-}}$, we can have

$$\begin{aligned}
\mathrm{E}\{\sigma_s^2|\mathcal{F}_0\} &= \beta^\circ \int_0^s \mathrm{E}e^{-X_{s-u}}du + \sigma_0^2\mathrm{E}e^{-X_s} = \beta^\circ \int_0^s \mathrm{E}e^{-X_u}du + \sigma_0^2\mathrm{E}e^{-X_s} \\
&= \beta^\circ \int_0^s \mathrm{E}e^{-X_u}du + \sigma_0^2\mathrm{E}e^{-X_s} \\
&= \frac{\beta^\circ(e^{s\Psi(1)}-1)}{\Psi(1)} + e^{s\Psi(1)}\sigma_0^2.
\end{aligned}$$

Then, observing

$$\sigma_s^4 = \beta_\circ^2\left\{\int_0^s e^{-(X_{s-}-X_u)}du\right\}^2 + 2\sigma_0^2 e^{-X_{s-}}\beta^\circ \int_0^s e^{-(X_{s-}-X_u)}du + \sigma_0^4 e^{-2X_{s-}},$$

we obtain

$$\begin{aligned}
\mathrm{E}\left\{\sigma_s^4|\mathcal{F}_0\right\} &= \beta_\circ^2\mathrm{E}\left\{\int_0^s e^{-X_u}du\right\}^2 + 2\beta^\circ\sigma_0^2\mathrm{E}\left\{\int_0^s e^{X_u-2X_s}du\right\} + \sigma_0^4\mathrm{E}e^{-2X_s} \\
&= \beta_\circ^2\left\{\frac{2}{\Psi(1)\Psi(2)} + \frac{2}{\Psi(2)-\Psi(1)}\left(\frac{e^{s\Psi(2)}}{\Psi(2)} - \frac{e^{s\Psi(1)}}{\Psi(1)}\right)\right\} \\
&\quad +2\beta^\circ\sigma_0^2\frac{e^{s\Psi(2)}-e^{s\Psi(1)}}{\Psi(2)-\Psi(1)} + \sigma_0^4 e^{s\Psi(2)}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left\{\sigma_s^2\sigma_u^2|\mathcal{F}_0\right\} &= \mathrm{E}\left\{\sigma_u^2\mathrm{E}\left\{\sigma_s^2|\mathcal{F}_u\right\}\middle|\,\mathcal{F}_0\right\} \\
&= \mathrm{E}\left\{\sigma_u^2\frac{\beta^\circ(e^{(s-u)\Psi(1)}-1)}{\Psi(1)} + e^{(s-u)\Psi(1)}\sigma_u^4\middle|\,\mathcal{F}_0\right\} \\
&= \mathrm{E}\{\sigma_u^2|\mathcal{F}_0\}\frac{\beta^\circ\{e^{(s-u)\Psi(1)}-1\}}{\Psi(1)} + e^{(s-u)\Psi(1)}\mathrm{E}\left\{\sigma_u^4|\mathcal{F}_0\right\} \\
&= \left\{\frac{\beta^\circ(e^{u\Psi(1)}-1)}{\Psi(1)} + e^{u\Psi(1)}\sigma_0^2\right\}\frac{\beta^\circ\{e^{(s-u)\Psi(1)}-1\}}{\Psi(1)} \\
&\quad +e^{(s-u)\Psi(1)}\beta_\circ^2\left\{\frac{2}{\Psi(1)\Psi(2)} + \frac{2}{\Psi(2)-\Psi(1)}\left(\frac{e^{u\Psi(2)}}{\Psi(2)} - \frac{e^{u\Psi(1)}}{\Psi(1)}\right)\right\} \\
&\quad +2e^{(s-u)\Psi(1)}\beta^\circ\sigma_0^2\frac{e^{u\Psi(2)}-e^{u\Psi(1)}}{\Psi(2)-\Psi(1)} + e^{(s-u)\Psi(1)}\sigma_0^4 e^{u\Psi(2)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{E}\left\{(G_h-G_0)^4|\mathcal{F}_0\right\} &= 6\int_{(0,h]}\mathrm{E}\left\{G_{s-}^2\sigma_s^2|\mathcal{F}_0\right\}ds + \int_{\mathbb{R}} x^4\Pi(dx)\int_{(0,h]}\mathrm{E}\left\{\sigma_s^4|\mathcal{F}_0\right\}ds \\
&= h\left(\int_{\mathbb{R}} x^4\Pi(dx) + o(1)\right)\sigma_0^4, \\
\mathrm{E}\left\{(G_h-G_0)^2|\mathcal{F}_0\right\} &= h(1+o(1))\sigma_0^2.
\end{aligned}$$

This completes the proof. □

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2011-0010936) grant number:2011-0010936 (S. Lee), and was also supported by the National Research Foundation of Korea Grant funded by the Korean Government (Ministry of Education, Science and Technology) [NRF-2010-355-C00010] (M. Kim).

REFERENCES

[1] Basrak, B.; Davis, R.A. and Mikosch, T. (2002). Regular variation of GARCH processes, *Stochastic Process. Appl.*, **99**, 95–115.

[2] Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes, *Ann. Statist.*, **32**, 633–655.

[3] Berkes, I.; Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–227.

[4] Billingsley, P. (1995). *Probability and measure*, Wiley Series in Probability and Mathematical Statistics, 3rd ed., John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327.

[6] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series, *J. Econometrics*, **52**, 115–127.

[7] Fasen, V. (2010). Asymptotic results for sample autocovariance functions and extremes of integrated generalized Ornstein-Uhlenbeck processes, *Bernoulli*, **16**, 51–79.

[8] Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli*, **10**, 605–637.

[9] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors, *Econometrica*, **71**, 285–317.

[10] Haug, S.; Klüppelberg, C.; Lindner, A. and Zapp, M. (2007). Method of moment estimation in the COGARCH$(1,1)$ model, *Econom. J.*, **10**, 320–341.

[11] Kallsen, J. and Vesenmayer, B. (2009). COGARCH as a continuous-time limit of GARCH$(1,1)$, *Stochastic Process. Appl.*, **119**, 74–98.

[12] Klüppelberg, C.; Lindner, A. and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour, *J. Appl. Probab.*, **41**, 601–622.

[13] Maller, R.A.; Müller, G. and Szimayer, A. (2008). GARCH modelling in continuous time for irregularly spaced time series data, *Bernoulli*, **14**, 519–542.

[14] Müller, G. (2010). MCMC estimation of the COGARCH(1, 1) model, *Journal of Financial Econometrics*, **8**, 481–510.

[15] Nelson, D.B. (1990a). ARCH models as diffusion approximations, *J. Econometrics*, **45**, 7–38.

[16] Nelson, D.B. (1990b). Stationarity and persistence in the GARCH(1, 1) model, *Econometric Theory*, **6**, 318–334.

[17] Protter, P.E. (2004). *Stochastic integration and differential equations*, vol. 21 of Applications of Mathematics (New York), 2nd ed., Springer-Verlag, Berlin. Stochastic Modelling and Applied Probability.

REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 169–190

BOUNDARY KERNELS FOR DISTRIBUTION FUNCTION ESTIMATION

Author: Carlos Tenreiro
– CMUC, Department of Mathematics, University of Coimbra, Apartado 3008, 3001–501 Coimbra, Portugal
tenreiro@mat.uc.pt

Received: January 2012 Accepted: August 2012

Abstract:

- Boundary effects for kernel estimators of curves with compact supports are well known in regression and density estimation frameworks. In this paper we address the use of boundary kernels for distribution function estimation. We establish the Chung-Smirnov law of iterated logarithm and an asymptotic expansion for the mean integrated squared error of the proposed estimator. These results show the superior theoretical performance of the boundary modified kernel estimator over the classical kernel estimator for distribution functions that are not smooth at the extreme points of the distribution support. The automatic selection of the bandwidth is also briefly discussed in this paper. Beta reference distribution and cross-validation bandwidth selectors are considered. Simulations suggest that the cross-validation bandwidth performs well, although the simpler reference distribution bandwidth is quite effective for the generality of test distributions.

Key-Words:

- *kernel distribution function estimation; boundary kernels; Chung-Smirnov property; MISE expansion; bandwidth selection.*

AMS Subject Classification:

- 62G05, 62G20.

1. INTRODUCTION

Given $X_1, ..., X_n$ independent copies of an absolutely continuous real random variable with unknown density and distribution functions f and F, respectively, a kernel estimator of F is introduced by authors such as Tiago de Oliveira [33], Nadaraya [20] or Watson and Leadbetter [35]. Such an estimator arises as an integral of the Parzen-Rosenblatt kernel density estimator (see Rosenblatt [25] and Parzen [21]) and is defined, for $x \in \mathbb{R}$, by

$$\bar{F}_{nh}(x) = \frac{1}{n}\sum_{i=1}^{n} \bar{K}\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where, for $u \in \mathbb{R}$,

$$\bar{K}(u) = \int_{]-\infty,u]} K(v)dv,$$

with K a kernel on $\mathbb{R}$, that is, a bounded and symmetric probability density function with support $[-1, 1]$ and $h = h_n$ a sequence of strictly positive real numbers converging to zero when n goes to infinity. Theoretical properties of this estimator, including bandwidth selection, have been investigated by several authors. Classical and more recent references, showing a continued interest in the subject, are, among others, Winter [36, 37], Yamato [38], Falk [7], Singh, Gasser and Prasad [28], Swanepoel [30], Jones [13], Shirahata and Chu [27], Sarda [26], Altman and Léger [1], Bowman, Hall and Prvan [2], Tenreiro [31, 32], Liu and Yang [16], Giné and Nickl [11], Mason and Swanepoel [18] and Chacón and Rodrígues-Casal [3].

If the support of f is known to be the finite interval $[a, b]$, from the continuity of F it is well known that the kernel estimator (1.1) is an asymptotically unbiased estimator of F if and only if $h \to 0$ as n goes to infinity (see Yamato [38], Lemma 1). However, if F is not smooth enough at the extreme points of the distribution support, the bias of $\bar{F}_{nh}$ does not achieve the standard h^2 order of convergence on the left and right boundary regions. In fact, assuming that the restriction of F to the interval $[a, b]$ is twice continuously differentiable, for $x = a + \alpha h$ with, $\alpha \in [0, 1]$, we have

$$\mathrm{E}\bar{F}_{nh}(x) - F(x) = hF'_+(a)\varphi_1(\alpha) + \frac{h^2}{2}F''_+(a)\varphi_2(\alpha) + o(h^2),$$

uniformly in α, with

$$\varphi_1(\alpha) = \alpha(\mu_{0,\alpha}(K) - 1) - \mu_{1,\alpha}(K), \tag{1.2}$$

$\varphi_2(\alpha) = \alpha^2(\mu_{0,\alpha}(K) - 1) - 2\alpha\mu_{1,\alpha}(K) + \mu_{2,\alpha}(K)$ and $\mu_{\ell,\alpha}(K) = \int_{-1}^{\alpha} z^\ell K(z)dz$. A similar expansion is valid for x in the right boundary region. As noticed by Gasser and Müller [9] in a regression context, this local behaviour dominates the global behaviour of the estimator which implies an inferior global order of convergence for the kernel estimator (1.1) which can be confirmed by examining the asymptotic behaviour of widely used measures of the quality of kernel estimators such as the maximum absolute deviation or the mean integrated squared error.

This type of boundary effect for kernel estimators of curves with compact supports is well-known in regression and density function estimation frameworks and several modified estimators have been proposed in the literature (see Müller [19], Karunamuni and Alberts [14], and Karunamuni and Zhang [15], and references therein). In order to improve the theoretical performance of the standard kernel distribution function estimator when the underlying distribution function F is not smooth enough at the extreme points of the distribution support, the use of the so-called boundary kernels, suggested for regression and density kernel estimators by Gasser and Müller [9], Rice [24], Gasser, Müller and Mammitzsch [10] and Müller [19], is addressed in this paper, which is organised as follows.

In Section 2, we introduce the boundary modified kernel distribution function estimator and some families of boundary kernels are presented, one of them leading to proper distribution function estimators. Contrary to the boundary modified kernel density estimators which possibly assume negative values, in a distribution function estimation framework the theoretical advantage of using boundary kernels is compatible with the natural property of obtaining a proper distribution function estimate. In Section 3 we show that the Chung-Smirnov theorem, that gives the supremum norm convergence rate of the empirical distribution function estimator, is also valid for the boundary kernel distribution function estimator. In Section 4 we present an asymptotic expansion for the mean integrated squared error of the estimator. This result illustrates the superior theoretical performance of the boundary kernel distribution function estimator over the classical kernel estimator whenever the underlying distribution function is not smooth enough at the extreme points of the distribution support. The automatic selection of the bandwidth is addressed in Section 5 where beta reference distribution and cross-validation bandwidth selectors are considered. Simulations suggest that the cross-validation bandwidth performs well, although the simpler reference distribution bandwidth is quite effective for the generality of test distributions. All the proofs can be found in Section 6. The simulations and plots in this paper were carried out using the R software [23].

2. KERNEL ESTIMATOR WITH BOUNDARY KERNELS

In order to deal with the boundary effects that occur in nonparametric regression and density function estimation, the use of boundary kernels is proposed and studied by authors such as Gasser and Müller [9], Rice [24], Gasser, Müller and Mammitzsch [10] and Müller [19]. Next we extend this approach to a distribution function estimation framework, where we assume that the support of the underlying distribution is known to be the finite interval $[a, b]$.

We consider the boundary modified kernel distribution function estimator given by

$$\tilde{F}_{nh}(x) = \frac{1}{n}\sum_{i=1}^{n} \bar{K}_{x,h}\left(\frac{x - X_i}{h}\right), \tag{2.1}$$

for $x \in]a, b[$ and $0 < h \leq (b-a)/2$, where

$$\bar{K}_{x,h}(u) = \int_{]-\infty,u]} K_{x,h}(v)dv,$$

and $K_{x,h}$ takes the form

$$K_{x,h}(u) = \begin{cases} K^L(u;(x-a)/h), & a < x < a+h \\ K(u), & a+h \leq x \leq b-h \\ K^R(u;(b-x)/h), & b-h < x < b, \end{cases}$$

where K is a bounded and symmetric probability density function with support $[-1, 1]$, and $K^L(\cdot;\alpha)$ and $K^R(\cdot;\alpha)$ are second order (left and right) boundary kernels for $\alpha \in]0, 1[$. Therefore, $K^L(\cdot;\alpha)$ and $K^R(\cdot;\alpha)$ are such that theirs supports are contained in the intervals $[-1, \alpha]$ and $[-\alpha, 1]$, respectively, and

$$\int K^\ell(u;\alpha)du = 1, \int uK^\ell(u;\alpha)du = 0 \text{ and } \int u^2K^\ell(u;\alpha)du \neq 0,$$

for all $\alpha \in]0, 1[$, with $\ell = R, L$. Additionally we define $\tilde{F}_{nh}(x) = 0$ for $x \leq a$ and $\tilde{F}_{nh}(x) = 1$ for $x \geq b$.

If we write

$$\bar{K}^\ell(u;\alpha) = \int_{]-\infty,u]} K^\ell(v;\alpha)dv,$$

for $\ell = L, R$, the kernel $\bar{K}_{x,h}$ can be written as

$$\bar{K}_{x,h}(u) = \begin{cases} \bar{K}^L(u;(x-a)/h), & a < x < a+h \\ \bar{K}(u), & a+h \leq x \leq b-h \\ \bar{K}^R(u;(b-x)/h), & b-h < x < b. \end{cases}$$

In the following examples we present three families of boundary kernels. We will assume that $K^R(u;\alpha) = K^L(-u;\alpha)$. In this case, we have $\bar{K}^R(u;\alpha) = 1 - \bar{K}^L(-u;\alpha)$.

Example 2.1. In a density estimation setting the standard choice for K^L is

$$K^L(u;\alpha) = (A_\alpha(K) + B_\alpha(K)u)K(u)I(-1 \le u \le \alpha),$$

where $A_\alpha(K) = \mu_{2,\alpha}(K)/D_\alpha(K)$, $B_\alpha(K) = -\mu_{1,\alpha}(K)/D_\alpha(K)$ and $D_\alpha(K) = \mu_{0,\alpha}(K)\mu_{2,\alpha}(K) - \mu_{1,\alpha}(K)^2$. Despite being negative for small values of α, this type of boundary kernels is suitable for density estimation. Contrary to nonnegative boundary kernels, they allow the control of the variability of the estimator near the support distribution boundary (see Gasser and Müller [9]). In this case, we get

$$\bar{K}^L(u;\alpha) = (A_\alpha(K)\bar{K}(u) + B_\alpha(K)\mu_{1,u}(K))I(-1 \le u \le \alpha) + I(u > \alpha).$$

A local behaviour analysis of the modified kernel distribution function estimator near the end points of the distribution support reveals that this class of boundary kernels may not be especially appropriate for the estimation of a distribution function. Restricting our analysis to the left-sided boundary region, and assuming the continuity of the second derivative of F in $]a, a+h[$, for $x = a + \alpha h$, with $\alpha \in]0,1[$, we have

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = \frac{h^2}{2} F''(x)\mu(\alpha) + o(h^2) \tag{2.2}$$

and

$$\operatorname{Var} \tilde{F}_{nh}(x) = \frac{F(x)(1-F(x))}{n} - \frac{h}{n} F'(x)\nu(\alpha) + O(n^{-1}h^2), \tag{2.3}$$

where

$$\mu(\alpha) = \int_{-1}^{\alpha} z^2 K^L(z;\alpha)\, dz$$

and

$$\nu(\alpha) = \int_{-1}^{\alpha} z B^L(z;\alpha)\, dz,$$

with $B^L(u;\alpha) = 2\bar{K}^L(u;\alpha)\ K^L(u;\alpha)$ (see expansions (6.4) and (6.5) in Section 6).

For the previous class of kernels the quantity $\nu(\alpha)$ can be negative for small values of α, which leads to an estimator whose local variability is larger than the empirical distribution function one. Additionally, as $\mu(\alpha)$ converges to a strictly negative value, when α tends to zero, a local bias can occur for small values of α (at the order of convergence h^2). In the next examples we take for $K^L(\cdot;\alpha)$ a symmetric probability density function with support $[-\alpha, \alpha]$. In this case, F_{nh} is nonnegative and $\nu(\alpha) > 0$, for $\alpha \in]0,1[$. Therefore, the boundary kernel estimator has a local variability inferior to the empirical distribution function one. Additionally, $\mu(\alpha)$ converges to zero, as α approaches zero (for the boundary kernels of Example 2.2, this is true whenever K is continuous on a neighbourhood of the origin with $K(0) > 0$).

Example 2.2. If K is such that $\int_0^\alpha K(u)du > 0$ for all $\alpha > 0$, for

$$K^L(u;\alpha) = (2\bar{K}(\alpha)-1)^{-1}K(u)I(-\alpha \le u \le \alpha),$$

we have

$$\bar{K}^L(u;\alpha) = (2\bar{K}(\alpha)-1)^{-1}(\bar{K}(u)-\bar{K}(-\alpha))I(-\alpha \le u \le \alpha) + I(u > \alpha).$$

Example 2.3. If we take

$$K^L(u;\alpha) = K(u/\alpha)/\alpha$$

we get

$$\bar{K}^L(u;\alpha) = \bar{K}(u/\alpha).$$

Finally, note that, for these two last classes of boundary kernels, $\tilde{F}_{nh}$ is, with probability one, a continuous probability distribution function. Therefore, in a distribution function estimation framework, the theoretical advantage of using boundary kernels, which we establish in the following sections, is compatible with the natural property of obtaining proper distribution function estimates.

3. UNIFORM CONVERGENCE RESULTS

The almost sure (or complete) uniform convergence of the classical kernel distribution function estimator $\bar{F}_{nh}$ to F was established by Nadaraya [20], Winter [36] and Yamato [38], whereas Winter [37] proved that, under certain regularity conditions, $\bar{F}_{nh}$ has the Chung-Smirnov law of iterated logarithm property (see also Degenhardt [5] and Chacón and Rodrígues-Casal [3]). In this section we show that these results are also valid for the boundary kernel distribution function estimator (2.1). For that, we will need the following lemma that gives upper bounds for $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$ and $||\mathrm{E}\tilde{F}_{nh} - F||$, where $||\cdot||$ denotes the supremum norm.

Lemma 3.1. *For all* $0 < h \le (b-a)/2$, *we have*

$$||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}|| \le C_K ||\, F_n - F|| \tag{3.1}$$

and

$$||\mathrm{E}\tilde{F}_{nh} - F|| \le C_K \sup_{x,y\in[a,b]:\,|x-y|\le h} |F(x)-F(y)|, \tag{3.2}$$

where F_n *is the empirical distribution function and*

$$C_K = \max\left(1, \max_{\ell=L,R} \sup_{\alpha\in\,]0,1[} \int |K^\ell(u;\alpha)|\,du\right).$$

Moreover, if the derivative F' is continuous on $[a,b]$, then

$$||\mathrm{E}\tilde{F}_{nh} - F|| \leq h\, C_K \sup_{x,y\in[a,b]:\,|x-y|\leq h} |F'(x) - F'(y)|. \tag{3.3}$$

The next results follow straightforwardly from Lemma 3.1 after separating the difference $\tilde{F}_{nh} - F$ into a stochastic component $\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}$ and a non-stochastic bias component $\mathrm{E}\tilde{F}_{nh} - F$. The first one is a consequence of a well-known exponential inequality due to Dvoretzky, Kiefer and Wolfowitz [6], which gives a bound on the tail probabilities of $||F_n - F||$, and the second one follows from the law of iterated logarithm for the empirical distribution function estimator due to Smirnov [29] and Chung [4] (see also van der Vaart [34], p. 268, and references therein). Also note that the condition imposed on the boundary kernels is trivially satisfied by nonnegative boundary kernels such as those of the Examples 2.2 and 2.3. It is also fulfilled by the boundary kernels of Example 2.1.

Theorem 3.1. *For $\ell = L, R$, let K^ℓ be such that*

$$\sup_{\alpha\in\,]0,1[} \int |K^\ell(u;\alpha)|\, du < \infty.$$

If $h \to 0$, then

$$||\tilde{F}_{nh} - F|| \to 0 \text{ almost completely.}$$

Theorem 3.2. *Under the conditions of Theorem 3.1, if F is Lipschitz and $(n/\log\log n)^{1/2}h \to 0$, then $\tilde{F}_{nh}$ has the Chung-Smirnov property, i.e.,*

$$\limsup_{n\to+\infty} (2n/\log\log n)^{1/2} ||\tilde{F}_{nh} - F|| \leq 1 \text{ almost surely.}$$

Moreover, the same is true whenever F' is Lipschitz on $[a,b]$ and h satisfies the less restrictive condition $(n/\log\log n)^{1/2}h^2 \to 0$.

Remark 3.1. If F is Lipschitz and the bandwidth fulfills the more restrictive condition $n^{1/2}h \to 0$, the Chung-Smirnov property can be deduced from the strong approximation property $\sqrt{n}\,||\tilde{F}_{nh} - F_n|| = o(1)$ *almost surely*, that can be derived by adapting the approach by Fernholz [8]. In this case, $\sqrt{n}\,||\tilde{F}_{nh} - F||$ and the Kolmogorov statistic $\sqrt{n}\,||F_n - F||$ have the same asymptotic distribution.

Remark 3.2. When F' is Lipschitz on $[a,b]$ and $(n/\log\log n)^{1/2}h^2 \to 0$, F_{nh} has the Chung-Smirnov property without assuming the continuity of F' at $x = a$ or $x = b$. This shows that F_{nh} improves on $\bar{F}_{nh}$ for distribution functions which are not smooth enough at the extreme points of the distribution support (cf. Winter [37], Theorem 3.2).

Remark 3.3. If F is the uniform distribution on $[a,b]$, from inequality (3.3) we deduce that $||\mathrm{E}\tilde{F}_{nh} - F|| = 0$, for all $0 < h \le (b-a)/2$. Therefore,

$$||\tilde{F}_{nh} - F|| = ||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}|| \le C_K ||F_n - F||,$$

and $\tilde{F}_{nh}$ has the Chung-Smirnov property even when h does not converge to zero as n goes to infinity.

Remark 3.4. In practice the bandwidth h is usually chosen on the basis of the data, that is, $h = \hat{h}(X_1, ..., X_n)$. From the proof of Lemma 3.1 we easily conclude that the so-called automatic boundary kernel estimator defined by (2.1) with $h = \hat{h}$ satisfies the inequalities

$$||\tilde{F}_{n\hat{h}} - F|| \le C_K \Big\{ ||F_n - F|| + \sup_{x,y\in[a,b]\colon |x-y|\le\hat{h}} |F(x) - F(y)| \Big\},$$

for any F, and

$$||\tilde{F}_{n\hat{h}} - F|| \le C_K \Big\{ ||F_n - F|| + \hat{h} \sup_{x,y\in[a,b]\colon |x-y|\le\hat{h}} |F'(x) - F'(y)| \Big\},$$

whenever F' is continuous on $[a,b]$. Therefore, under the conditions of Theorems 3.1 and 3.2, if the assumptions on h are replaced by their almost sure counterparts, we conclude that the automatic boundary kernel estimator, $\tilde{F}_{n\hat{h}}$, is an almost sure uniform convergent estimator of F that enjoys the Chung-Smirnov property.

4. MISE ASYMPTOTIC EXPANSION

A widely used measure of the quality of the kernel estimator is the mean integrated squared error given by

$$\begin{aligned}\mathrm{MISE}(F;h) &= \mathrm{E}\int \{\tilde{F}_{nh}(x) - F(x)\}^2 dx \\ &= \int \mathrm{Var}\,\tilde{F}_{nh}(x) dx + \int \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2 dx \\ &=: \tilde{\mathbf{V}}(F;h) + \tilde{\mathbf{B}}(F;h),\end{aligned}$$

where the integrals are over $\mathbb{R}$. Denoting by $\bar{\mathbf{V}}(F;h)$ and $\bar{\mathbf{B}}(F;h)$ the corresponding variance and bias terms for the classical kernel distribution function estimator (1.1), the approach followed by Swanepoel [30] leads to the following expansions whenever the restriction of F to the interval $[a,b]$ is twice continuously differentiable:

$$\bar{\mathbf{V}}(F;h) = \frac{1}{n}\int F(x)(1-F(x))dx - \frac{h}{n}\int uB(u)du + O\left(n^{-1}h^2\right) \tag{4.1}$$

where

$$B(u) = 2\bar{K}(u)K(u),$$

for $u \in \mathbb{R}$, and

$$\text{(4.2)} \quad \bar{\mathbf{B}}(F;h) = h^3 \left(F'_+(a)^2 + F'_-(b)^2\right) \int_0^1 \varphi_1(\alpha)^2 d\alpha$$
$$+ h^4 \big(F'_+(a)F''_+(a) - F'_-(b)F''_-(b)\big) \int u^2 K(u)\, du \int_0^1 \varphi_1(\alpha) d\alpha$$
$$+ \frac{h^4}{4} \left(\int u^2 K(u)\, du\right)^2 ||F''||_2^2 + o(h^4),$$

where φ_1 is given by (1.2) and $||\cdot||_2$ is the L_2 distance in $[a,b]$.

Depending on the smoothness of F on $\mathbb{R}$, we see that two different orders of convergence to zero for the mean integrated square error can be obtained. In the smooth case, that is, when $F'_+(a) = F'_-(b) = 0$, the previous expansions agree with the classical ones (cf. Jones [13]). However, in the non-smooth case an inferior global order of convergence occurs and a different order of convergence for the optimal bandwidth, in the sense of minimising the asymptotic MISE, takes place.

Next we show that, even when F is not smooth at the extreme points of the distribution support, the leading terms of the MISE expansion of the boundary kernel estimator agree with those given in Jones [13] for the classical kernel distribution function estimator. This shows the theoretical advantage of the boundary kernel distribution function estimator over the classical kernel estimator. Next define $B^\ell(u;\alpha) = 2\bar{K}^\ell(u;\alpha)K^\ell(u;\alpha)$, for $u \in \mathbb{R}$, $\alpha \in]0,1[$ and $\ell = L, R$.

Theorem 4.1. *For $\ell = L, R$, let K^ℓ be such that*

$$\int_0^1 \left(\int |K^\ell(u;\alpha)| du\right)^2 d\alpha < \infty,$$

and assume that the restriction of F to the interval $[a,b]$ is twice continuously differentiable. We have

$$\tilde{\mathbf{V}}(F;h) = \frac{1}{n}\int F(x)(1-F(x))dx - \frac{h}{n}\int uB(u)du + O\left(n^{-1}h^2\right)$$

and

$$\tilde{\mathbf{B}}(F;h) = \frac{h^4}{4}\left(\int u^2 K(u) du\right)^2 ||F''||_2^2 + o\left(h^4\right).$$

Note that the previous assumptions on the boundary kernels are trivially satisfied by nonnegative boundary kernels such as those of Examples 2.2 and 2.3, and also by the boundary kernels of Example 2.1. Next we give the asymptotically optimal choice for the bandwidth in the sense of minimising the leading terms in the expansion of the MISE.

Theorem 4.2. *Under the conditions of Theorem 4.1, let us assume that* $C_B > 0$ *where*

$$C_B = 2\int uB(u)\,du - \int_0^1\!\!\int u\left(B^L(u;\alpha) + B^R(u;\alpha)\right)\,du d\alpha.$$

Then the asymptotically optimal bandwidth is given by

$$h_0 = \min\left(\delta(K)||F''||_2^{-2/3}n^{-1/3}\,,\; \frac{b-a}{2}\min\left(1\,,\int uB(u)du/C_B\right)\right), \tag{4.3}$$

where

$$\delta(K) = \left(\int uB(u)\,du\right)^{1/3}\left(\int u^2K(u)du\right)^{-2/3}.$$

Remark 4.1. Following the approach by Marron and Jones [17], and taking into account the results of Swanepoel [30] and Jones [13], we conclude that the uniform density on $[-1,1]$ is the optimal kernel in the sense of minimising the asymptotic MISE. However, as noticed by Jones [13], other suboptimal kernels, such as the Epanechnikov kernel on $[-1,1]$, have a performance very close to the optimal one.

Remark 4.2. For the boundary kernels of Example 2.3, we have $C_B = \int uB(u)du > 0$ and the asymptotically optimal bandwidth is simply given by $h_0 = \min\left(\delta(K)||F''||_2^{-2/3}n^{-1/3}, (b-a)/2\right)$.

5. BANDWIDTH SELECTION

In a kernel estimation setting the bandwidth is usually chosen on the basis of the data. For the classical kernel distribution function estimator (1.1) and assuming that f is a smooth function over the whole real line, two main approaches for the automatic selection of h can be found in the literature. Cross-validation methods are discussed in Sarda [26], Altman and Léger [1] and Bowman, Hall and Prvan [2], and direct plug-in methods, including normal reference distribution methods, are proposed by Altman and Léger [1], Polansky and Baker [22] and Tenreiro [32]. In the following subsections we consider two fully automatic bandwidth selectors for the boundary kernel distribution function estimator. The first one is a reference distribution method based on the beta distribution family. The second one is a cross-validation bandwidth selector inspired in the approach of Bowman, Hall and Prvan [2].

5.1. A reference distribution method

A commonly used quick and simple method for choosing the bandwidth involves using the asymptotically optimal bandwidth for a fixed reference distribution having the same mean and scale as that estimated for the underlying distribution. In what follows a beta distribution over the interval $[a, b]$ with both shape parameters greater than or equal to 2 is taken as reference distribution. The restriction on the shape parameters values takes into account the assumptions on F imposed in Theorem 4.1. If X has a beta distribution over the interval $[a, b]$ with shape parameters p and q, the expected value of X is given by

$$\mathrm{E}(X) = a + (b-a)\frac{p}{p+q}$$

and the variance of X by

$$\mathrm{Var}(X) = (b-a)^2 \frac{pq}{(p+q)^2(p+q+1)}$$

(see Johnson, Kotz and Balakrishnan [12], p. 222). Taking the sample mean $\bar{X}$ and the sample variance S^2 as estimators of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$, respectively, the method of moments estimators for the parameters p and q are given by

$$\tilde{p} = \tilde{X}(\tilde{X}(1-\tilde{X})\tilde{S}^{-2} - 1) \quad \text{and} \quad \tilde{q} = (1-\tilde{X})(\tilde{X}(1-\tilde{X})\tilde{S}^{-2} - 1),$$

where $\tilde{X} = (\bar{X} - a)/(b-a)$ and $\tilde{S}^2 = S^2/(b-a)^2$. Thus, denoting by $\hat{F}$ the beta distribution over the interval $[a, b]$ with shape parameters $\hat{p} = \max(2, \tilde{p})$ and $\hat{q} = \max(2, \tilde{q})$, the considered beta optimal bandwidth, which we denote by $\hat{h}_{\mathrm{BR}}$, is defined by (4.3) with $||\hat{F}''||_2$ in place of $||F''||_2$ where

$$||\hat{F}''||_2^2 = \frac{(\hat{p}-1)(\hat{q}-1)\mathrm{B}(2\hat{p}-3, 2\hat{q}-3)}{(b-a)(2(\hat{p}+\hat{q})-5)\mathrm{B}(\hat{p},\hat{q})^2},$$

and $\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the beta function.

5.2. A cross-validation method

An alternative approach for bandwidth selection can be based on the cross-validation ideas of Bowman, Hall and Prvan [2]. The cross-validation function proposed by these authors is a mean over all the observations of the integrated squared error between the indicator function $I(X_i \le x)$ associated to the observation X_i, and the boundary kernel estimator constructed from the data with observation X_i omitted, that is,

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \int \{I(X_i \le x) - \tilde{F}_{-ih}(x)\}^2 dx,$$

where

$$\tilde{F}_{-ih}(x) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} \bar{K}_{x,h}\left(\frac{x-X_j}{h}\right).$$

The cross-validation bandwidth, which we denote by $\hat{h}_{\mathrm{CV}}$, is the minimiser of $\mathrm{CV}(h)$. The main motivation for this method comes from the equality

$$\mathrm{E}\left(\mathrm{CV}(h) - \frac{1}{n}\sum_{i=1}^{n}\int\{I(X_i \leq x) - F(x)\}^2 dx\right) = \mathrm{E}\int\{\tilde{F}_{n-1,h}(x) - F(x)\}^2 dx,$$

which shows that the criterion function $\mathrm{CV}(h)$ provides an unbiased estimator of $\mathrm{MISE}(F;h)$ for a sample size $n-1$, shifted vertically by an unknown term which is independent of h. Although the asymptotic behaviour of the cross-validation bandwidth is not discussed in this paper, it will be of interest to know whether $\hat{h}_{\mathrm{CV}}$ is asymptotically equivalent to the asymptotically optimal bandwidth h_0. As shown in Bowman, Hall and Prvan [2], this property is valid for the standard kernel distribution function estimator.

5.3. A simulation study

In order to analyse the finite sample performance of the bandwidth selectors $\hat{h}_{\mathrm{BR}}$ and $\hat{h}_{\mathrm{CV}}$, a simulation study was carried out for a set of beta mixture distributions with support $[0,1]$ that represents different shapes and boundary behaviours. Their weights and shape parameters are given in Table 1 and the corresponding probability density and cumulative distribution functions are shown in Figure 1.

Table 1: Beta mixture test distributions.

	Beta mixture distribution $\sum_i w_i B(p_i,q_i)$		
	Weights w	1st shape parameters p	2nd shape parameters q
#1	$(1/4,3/4)$	$(1,6)$	$(6,1)$
#2	$(1/10,7/10,2/10)$	$(1,2,3)$	$(2,2,1)$
#3	$(1/10,7/10,2/10)$	$(1,2,6)$	$(2,6,1)$
#4	$(5/16,5/16,3/16,2/16,1/16)$	$(1,25,160,320,800)$	$(10,60,100,80,90)$

From each distribution we generated 500 samples of sizes $n = 25, 50, 100$ and 200, and we calculated the integrated squared error $\mathrm{ISE}(F;h) = \int\{\tilde{F}_{nh}(x) - F(x)\}^2 dx$ for $h = \hat{h}_{\mathrm{BR}}$ and $h = \hat{h}_{\mathrm{CV}}$ as a measure of the performance of each bandwidth selector. The integrated squared error associated to the asymptotically optimal bandwidth h_0 was also evaluated for the sake of comparison.

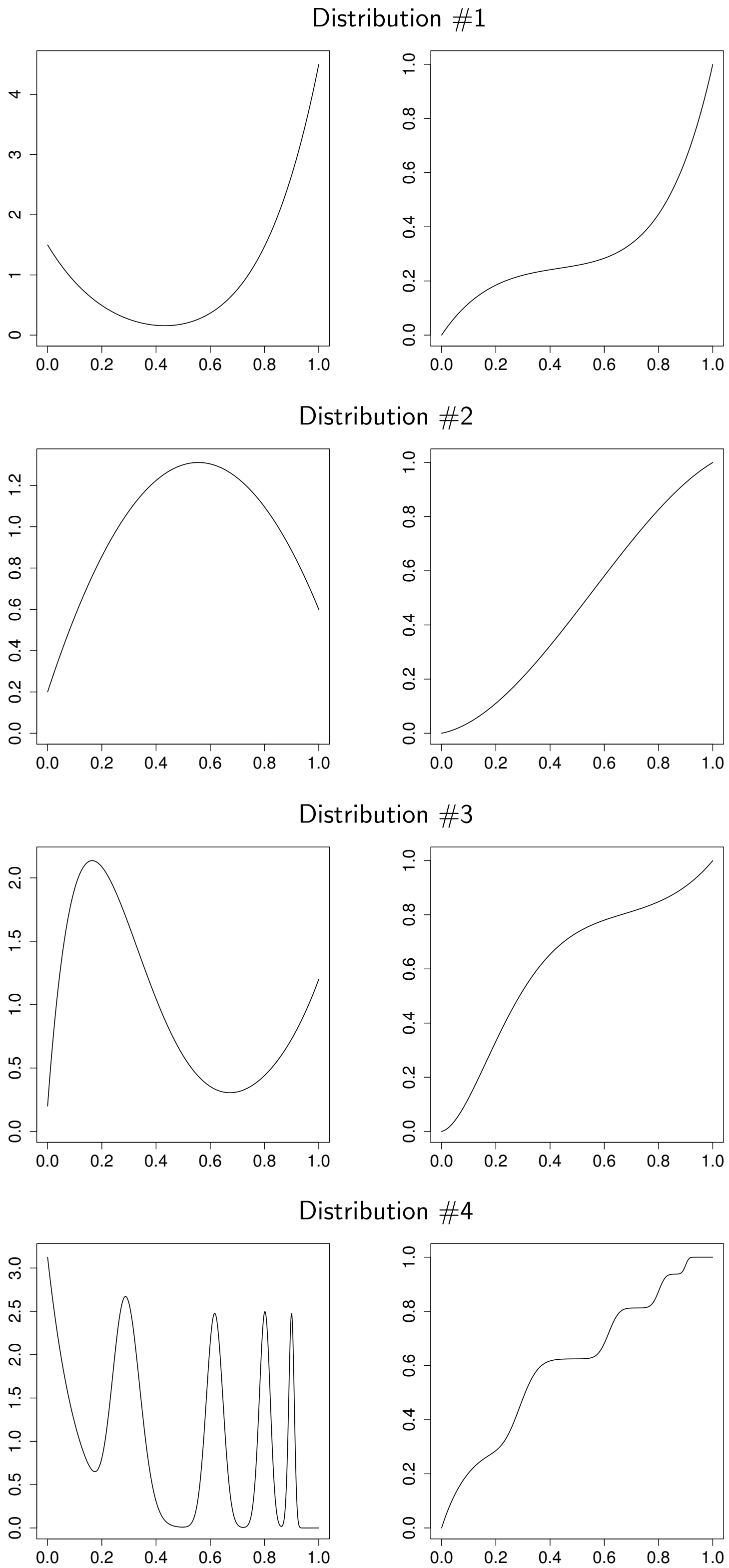

Figure 1: Beta mixture test density and cumulative distribution functions.

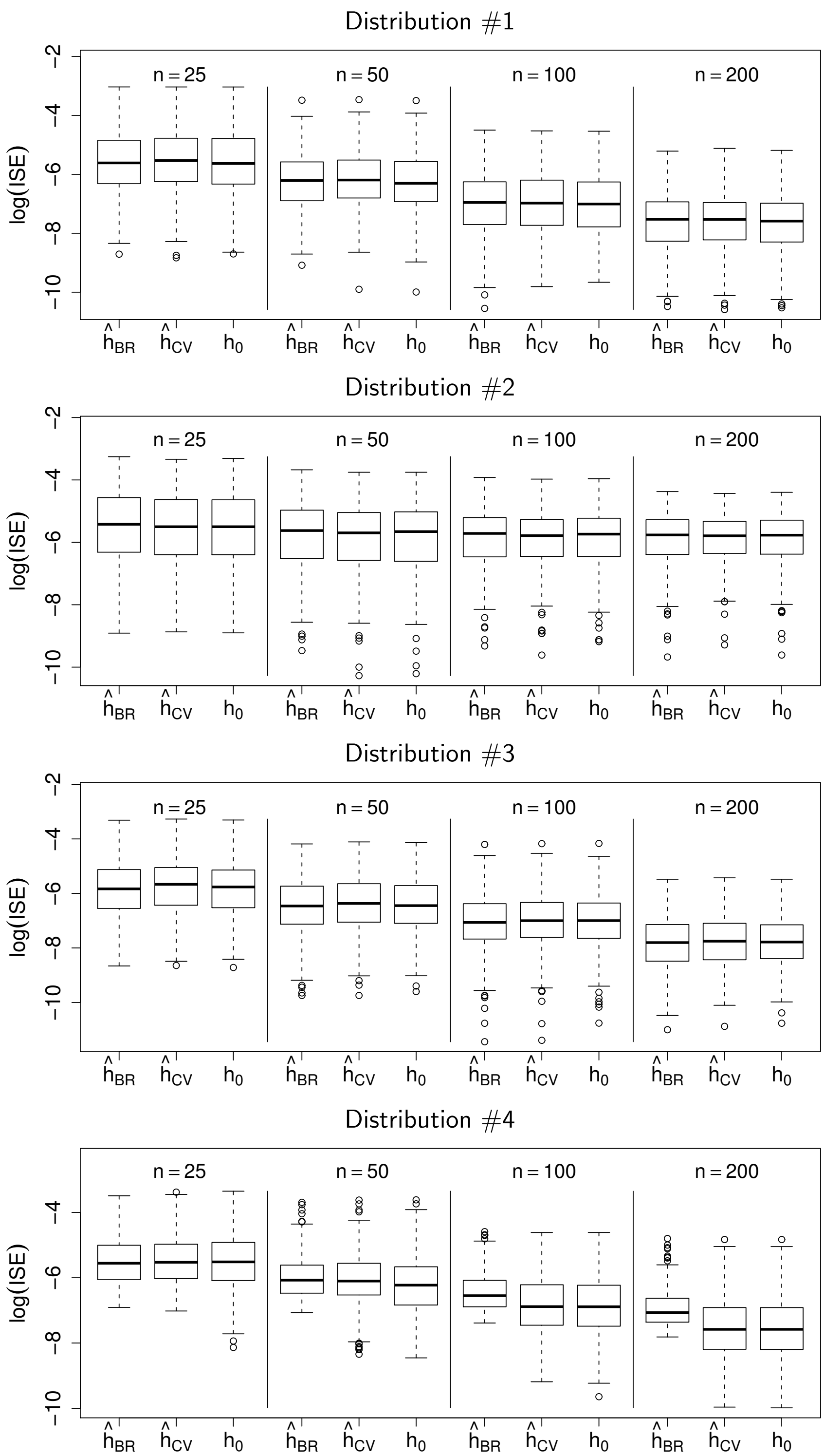

Figure 2: Integrated squared error results for the smoothing parameters $h = \hat{h}_{\rm BR}$, $h = \hat{h}_{\rm CV}$ and $h = h_0$ and sample sizes $n = 25, 50, 100$ and 200. K is the Epanechnikov density function. The number of replications for each case is 500.

In the implementation of cross-validation method the minimisation of CV(h) was confined to the interval $[\hat{h}_{\mathrm{BR}}/10, 1/2]$. The previous integrals have been numerically evaluated using the composite Simpson's rule. The Epanechnikov density $K(t) = \frac{3}{4}(1-t^2)I(|t| \leq 1)$ was taken as kernel function and we restrict our attention to the boundary kernels defined by $K^{\ell}(u,\alpha) = K(u/\alpha)/\alpha$ for $\ell = L, R$ (see Example 2.3). The integrated squared error empirical distributions (log scale) are presented in Figure 2.

For all the considered test distributions, Figure 2 suggests that the cross-validation bandwidth performs quite well showing a performance close to that one of the oracle estimator with bandwidth h_0. Additionally, for distributions #1, #2 and #3 there is no indication of significant differences between the bandwidths $\hat{h}_{\mathrm{CV}}$ and $\hat{h}_{\mathrm{BR}}$. This can be seen as an evidence of the well-known fact that smoothing has only a second order effect in kernel distribution function estimation. For the beta mixture #4 the cross-validation approach is clearly more effective than the beta optimal smoothing for large sample sizes. This distribution presents features that are not revealed until the sample size is above some threshold which explains the fact that both methods performed similarly for small sample sizes but not for large ones. In this latter case the cross-validation method is able to adapt to distributional shape while the beta distribution reference method does not reveal such a property.

In conclusion, we can say that the cross-validation bandwidth reveals a very good performance, although the simpler and less time consuming beta reference distribution bandwidth shows it self to be quite effective for the generality of test distributions.

6. PROOFS

Proof of Lemma 3.1: We start by the analysis of the stochastic component $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$. For that we follow the approach by Winter [37]. In order to deal with kernels that could have negative values, we need the following version of the integration by parts result presented by Winter [37, Lemma 2.1].

Lemma 6.1. *If Φ is a probability distribution function and*

$$\Psi(u) = \int_{]-\infty,u]} \psi(v)dv,$$

where ψ is a Lebesgue integrable function with $\int \psi(v)dv = 1$, then

$$\int \Phi d\Psi + \int \Psi d\Phi = 1.$$

Proof: Denoting by μ_Φ and μ_Ψ the finite signed measures defined by $\mu_\Phi(]-\infty,x]) = \Phi(x)$ and $\mu_\Psi(]-\infty,x]) = \Psi(x)$, for all $x \in \mathbb{R}$, it is enough to apply Fubini's theorem to the indicator function $(s,t) \to I(s>t)$ which is integrable with respect to the product measure $\mu_\Phi \otimes \mu_\Psi$. □

Returning to the proof of Lemma 3.1, for $x \in]a, a+h[$, we have

$$\tilde{F}_{nh}(x) = \int \bar{K}^L((x-y)/h;(x-a)/h)dF_n(y) = 1 - \int \Psi_{x,h}(y)dF_n(y),$$

$$\mathrm{E}\tilde{F}_{nh}(x) = \int \bar{K}^L((x-y)/h;(x-a)/h)dF(y) = 1 - \int \Psi_{x,h}(y)dF(y),$$

and

$$\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x) = \int \Psi_{x,h}(y)(dF(y) - dF_n(y)),$$

where $\Psi_{x,h}(u) = \int_{]-\infty,u]} \psi_{x,h}(v)dv$ with $\psi_{x,h}(v) = K^L((x-v)/h;(x-a)/h)/h$.

From Lemma 6.1 we get

$$\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x) = \int \{F_n(y) - F(y)\}d\Psi_{x,h}(y),$$

and therefore

$$\sup_{x\in]a,a+h[} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F|| \sup_{\alpha \in]0,1[} \int |K^L(u;\alpha)|\, du \tag{6.1}$$

because

$$\sup_{x\in]a,a+h[} \int d|\Psi_{x,h}|(y) = \sup_{x\in]a,a+h[} \int |\psi_{x,h}(u)|\, du \leq \sup_{\alpha\in]0,1[} \int |K^L(u;\alpha)|\, du.$$

Similarly, we get

$$\sup_{x\in]b-h,b[} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F|| \sup_{\alpha \in]0,1[} \int |K^R(u;\alpha)|\, du, \tag{6.2}$$

and the standard approach (see Winter [37]) can be used for $x \in [a+h, b-h]$, in order to obtain

$$\sup_{x\in [a+h,b-h]} |\tilde{F}_{nh}(x) - \mathrm{E}\tilde{F}_{nh}(x)| \leq ||\, F_n - F||. \tag{6.3}$$

Finally, from (6.1), (6.2) and (6.3) we obtain the upper bound (3.1) for $||\tilde{F}_{nh} - \mathrm{E}\tilde{F}_{nh}||$.

In the analysis of the bias component $||\mathrm{E}\tilde{F}_{nh} - F||$, we first note that, for $x \in \,]a, a+h[$, the expectation of $\tilde{F}_{nh}(x)$ is given by

$$\begin{aligned}
\mathrm{E}\tilde{F}_{nh}(x) &= \int \bar{K}^L((x-y)/h;(x-a)/h)f(y)\,dy \\
&= \iint K^L(u;(x-a)/h)f(y)I(y \le x-uh)\,dudy \\
&= \int F(x-uh)K^L(u;(x-a)/h)\,du.
\end{aligned}$$

Therefore,

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = \int \{F(x-uh) - F(x)\}K^L(u;(x-a)/h)\,du, \tag{6.4}$$

which leads to

$$\sup_{x \in \,]a,a+h[} |\mathrm{E}\tilde{F}_{nh}(x) - F(x)| \le \sup_{x,y \in [a,b]:\,|x-y| \le h} |F(x) - F(y)| \sup_{\alpha \in \,]0,1[} \int |K^L(u;\alpha)|\,du.$$

Additionally, if F' is continuous on $[a,b]$, from the Taylor formula we have

$$F(x-uh) - F(x) = -uhF'(x) - uh\int_0^1 \{F'(x-tuh) - F'(x)\}\,dt.$$

Using the fact that $\int uK^L(u;\alpha)du = 0$, for all $\alpha \in \,]0,1[$, from (6.4) we get

$$\mathrm{E}\tilde{F}_{nh}(x) - F(x) = -h\int \{F'(x-tuh) - F'(x)\}uK^L(u;(x-a)/h)\,du$$

which leads to

$$\sup_{x \in \,]a,a+h[} |\mathrm{E}\tilde{F}_{nh}(x) - F(x)| \le h \sup_{x,y \in [a,b]:\,|x-y| \le h} |F'(x) - F'(y)| \sup_{\alpha \in \,]0,1[} \int |K^L(u;\alpha)|\,du.$$

A similar analysis can be carried out for the cases $x \in [a+h, b-h]$ and $x \in \,]b-h, b[$, leading to the bounds (3.2) and (3.3) for the bias term $||\mathrm{E}\tilde{F}_{nh} - F||$. $\square$

Proof of Theorem 4.1: We start by the analysis of the bias term $\tilde{\mathbf{B}}(F;h) = \int \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2 dx$. By using the continuity of the second derivative of F and the Taylor expansion

$$F(x-uh) - F(x) = -uhF'(x) + u^2h^2\int_0^1 (1-t)F''(x-tuh)\,dt,$$

from (6.4) we get

$$\begin{aligned}
&\int_a^{a+h} \{\mathrm{E}\tilde{F}_{nh}(x) - F(x)\}^2\,dx \\
&= h^5 \int_0^1 \left(\iint_0^1 (1-t)F''(a+\alpha h - tuh)u^2K^L(u;\alpha)\,dtdu\right)^2 d\alpha \\
&\le h^5\,||F''||^2 \int_0^1 \left(\int |K^L(u;\alpha)|du\right)^2 d\alpha = O\left(h^5\right).
\end{aligned}$$

A similar upper bound can be obtained for the term $\int_{b-h}^{b}\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2dx$. The stated expansion for $\tilde{\mathbf{B}}(F;h)$ follows now from the dominated convergence theorem:

$$
\begin{aligned}
&\int_{a+h}^{b-h}\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2dx\\
&=\int_{a+h}^{b-h}\left(\int\{F(x-uh)-F(x)\}K(u)\,du\right)^2dx\\
&=h^4\int_{a+h}^{b-h}\left(\int\!\!\int_0^1(1-t)F''(x-tuh)u^2K(u)\,dtdu\right)^2dx\\
&=\frac{h^4}{4}\left(\int u^2K(u)\,du\right)^2||F''||_2^2+o\left(h^4\right).
\end{aligned}
$$

The analysis of the variance term, $\tilde{\mathbf{V}}(F;h)=\int\mathrm{Var}\,\tilde{F}_{nh}(x)dx$, can be made easy by considering the expansion

$$
\begin{aligned}
(6.5)\qquad n\,\mathrm{Var}\,\tilde{F}_{nh}(x)&=F(x)(1-F(x))+\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,du\\
&\quad-\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}^2-2\{\mathrm{E}\tilde{F}_{nh}(x)-F(x)\}F(x),
\end{aligned}
$$

where $B_{x,h}$ is defined as $K_{x,h}$ with K replaced by B. In fact, from the first part of the proof we conclude that the integral over $[a,b]$ of the last two terms is of order $O(h^2)$, and from standard arguments we get

$$\int_{a+h}^{b-h}\!\!\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,dudx=-h\int uB(u)\,du+O\left(h^2\right)$$

and

$$
\begin{aligned}
&\left|\int_a^{a+h}\!\!\int\{F(x-uh)-F(x)\}B_{x,h}(u)\,dudx\right|\\
&\qquad\le h^2\,||F'||\int_0^1\!\!\int|u||B^L(u;\alpha)|\,dud\alpha\\
&\qquad\le h^2||F'||\int_0^1\left(\int|K^L(u;\alpha)|du\right)^2d\alpha=O\left(h^2\right).
\end{aligned}
$$

Taking into account that the same order of convergence can be obtained for the term $\int_{b-h}^{b}\int\{F(x-uh)-F(x)\}B_{x,h}(u)dudx$, we finally get the stated expansion for $\tilde{\mathbf{V}}(F;h)$. □

Proof of Theorem 4.2: We shall restrict our attention to the case where F is the uniform distribution function on the interval $[a,b]$. From Remark 3.3 and equality (6.5) we get

$$\mathrm{MISE}(F;h)=\frac{b-a}{6n}-\frac{h}{n}\left(\int uB(u)\,du-h\,\frac{C_B}{b-a}\right),$$

for $0 < h \leq (b-a)/2$. It is now easy to conclude that

$$h_0 = \frac{b-a}{2} \min\left(1\,, \int uB(u)\,du/C_B\right)$$

is the minimiser of $\mathrm{MISE}(F;h)$, for $0 < h \leq (b-a)/2$. $\square$

ACKNOWLEDGMENTS

Research partially supported by the Centre of Mathematics of the University of Coimbra funded by the European Regional Development Fund through the program COMPETE and by the Portuguese Government through the Fundação para a Ciência e Tecnologia under the project PEst-C/MAT/UI0324/2011.

REFERENCES

[1] Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation, *Journal of Statistical Planning and Inference*, **46**, 195–214.

[2] Bowman, A.; Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions, *Biometrika*, **85**, 799–808.

[3] Chacón, J.E. and Rodríguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator, *Statistics & Probability Letters*, **80**, 1414–1419.

[4] Chung, K.-L. (1949). An estimate concerning the Kolmogoroff limit distribution, *Transactions of the American Mathematical Society*, **67**, 36–50.

[5] Degenhardt, H.J.A. (1993). Chung-Smirnov property for perturbed empirical distribution functions, *Statistics & Probability Letters*, **16**, 97–101.

[6] Dvoretzky, A.; Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Annals of Mathematical Statistics*, **27**, 642–669.

[7] Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statistica Neerlandica*, **37**, 73–83.

[8] Fernholz, L.T. (1991). Almost sure convergence of smoothed empirical distribution functions, *Scandinavian Journal of Statistics*, **18**, 255–262.

[9] Gasser, T. and Müller, H.-G. (1979). *Kernel estimation of regression functions*. In "Smoothing techniques for curve estimation" (T. Gasser and M. Rosenblatt, Eds.), Lecture Notes in Mathematics 757, 23–68.

[10] Gasser, T.; Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society. Series B. Methodological*, **47**, 238–252.

[11] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation, *Probability Theory and Related Fields*, **143**, 569–596.

[12] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions 2*, Wiley, New York.

[13] Jones, M.C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statistics & Probability Letters*, **9**, 129–132.

[14] Karunamuni, R.J. and Alberts, T. (2005). On boundary correction in kernel density estimation, *Statistical Methodology*, **2**, 191–212.

[15] Karunamuni, R.J. and Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator, *Statistics & Probability Letters*, **78**, 499–507.

[16] Liu, R. and Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function, *Journal of Nonparametric Statistics*, **20**, 661–677.

[17] Marron, J.S. and Nolan, D. (1989). Canonical kernels for density estimation, *Statistics & Probability Letters*, **7**, 191–195.

[18] Mason, D.M. and Swanepoel, J.W.H. (2011). A general result on the uniform in bandwidth consistency of kernel-type function estimators, *TEST*, **20**, 72–94.

[19] Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, **78**, 521–530.

[20] Nadaraya, E.A. (1964). Some new estimates for distribution functions, *Theory of Probability and Its Applications*, **9**, 497–500.

[21] Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics*, **33**, 1065–1076.

[22] Polansky, A.M. and Baker, E.R. (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation and Simulation*, **65**, 63–80.

[23] R Development Core Team (2011). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`

[24] Rice, J. (1984). Boundary modification for kernel regression, *Communications in Statistics – Theory and Methods*, **13**, 893–900.

[25] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Annals of Mathematical Statistics*, **27**, 832–837.

[26] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference*, **35**, 65–75.

[27] Shirahata, S. and Chu, I-S. (1992). Integrated squared error of kernel-type estimator of distribution function, *Annals of the Institute of Statistical Mathematics*, **44**, 579–591.

[28] Singh, R.S.; Gasser, T. and Prasad, B. (1983). Nonparametric estimates of distribution function, *Communications in Statistics – Theory and Methods*, **12**, 2095–2108.

[29] Smirnov, N.V. (1944). Approximate laws of distribution of random variables from empirical data (in Russian), *Uspehi Matematičeskih Nauk*, **10**, 179–206.

[30] Swanepoel, J.W.H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function, *Communications in Statistics – Theory and Methods*, **17**, 3785–3799.

[31] Tenreiro, C. (2003). On the asymptotic behaviour of the ISE for automatic kernel distribution estimators, *Journal of Nonparametric Statistics*, **15**, 485–504.

[32] Tenreiro, C. (2006). Asymptotic behaviour of multistage plug-in bandwidth selectors for kernel distribution function estimators, *Journal of Nonparametric Statistics*, **18**, 101–116.

[33] Tiago de Oliveira, J. (1963). Estatística de densidades: resultados assintóticos, *Revista da Faculdade de Ciências de Lisboa*, **9**, 111–206.

[34] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[35] Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis II, *Sankhyā Ser. A*, **26**, 101–116.

[36] Winter, B.B. (1973). Strong uniform consistency of integrals of density estimators, *Canadian Journal of Statistics*, **1**, 247–253.

[37] Winter, B.B. (1979). Convergence rate of perturbed empirical distribution functions, *Journal of Applied Probability*, **16**, 163–173.

[38] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, *Bulletin of Mathematical Statistics*, **15**, 69–78.

REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 191–210

AN EXPONENTIAL-NEGATIVE BINOMIAL DISTRIBUTION

Authors: Mahtab Hajebi
– Department of Statistics, Amirkabir University of Technology, Tehran, IRAN

Sadegh Rezaei
– Department of Statistics, Amirkabir University of Technology, Tehran, IRAN
srezaei@aut.ac.ir

Saralees Nadarajah
– School of Mathematics, University of Manchester, Manchester M13 9PL, UK
Saralees.Nadarajah@manchester.ac.uk

Received: July 2012 Revised: August 2012 Accepted: August 2012

Abstract:

- A new three-parameter distribution is proposed for modeling lifetime data. It is advocated as most reasonable among the many exponential mixture type distributions proposed in recent years. An account of the mathematical properties of the new distribution including estimation issues is presented. Two real data applications are described to show superior performance versus at least four of the known lifetime models.

Key-Words:

- *exponential distribution; maximum likelihood estimation; negative binomial distribution.*

AMS Subject Classification:

- Primary: 62E15; Secondary: 62E20.

1. INTRODUCTION

The exponential distribution is the first and most popular model for failure times. In recent years, many authors have proposed generalizations of the exponential distribution. The generalizations are based on a "failure of a system" framework.

Suppose a series system is made up of Z unknown independent components. (The variable Z could be determined by such factors as economy, man power, and customer demand.) Let $Y_1, Y_2, ..., Y_Z$ denote the failure times of the Z components, assumed to be independent of Z. Then the system lifetime is $X = \min(Y_1, Y_2, ..., Y_Z)$. It is reasonable to assume that Y_js are exponential random variables, so the cumulative distribution function (cdf) and the probability density function (pdf) of X are

$$F_X(x) = 1 - \sum_{n=0}^{\infty} \exp(-n\beta x) \Pr(Z = n) \tag{1.1}$$

and

$$f_X(x) = \beta \sum_{n=0}^{\infty} n \exp(-n\beta x) \Pr(Z = n), \tag{1.2}$$

respectively, for $x > 0$ and $\beta > 0$.

Several authors have constructed models for (1.1) and (1.2) by taking Z to follow different distributions. Models with Z belonging to the *Panjer class* (Panjer [15]) have widespread applications in risk theory. The Panjer class includes the geometric, Poisson, negative binomial and other distributions. Panjer [15]'s paper was a breakthrough on the iterative computation of the distribution of aggregate claims, see, for example, Rolski *et al.* [18]. Extended versions of the Panjer class have been introduced by Sundt and Jewell [19], Hess *et al.* [7] and Pestana and Velosa [16]. Panjer class is also used in other contexts, see, for example, Katz [9].

Adamidis and Loukas [1] take Z to be a geometric random variable with parameter p, so yielding

$$f_X(x) = \frac{\beta(1-p)\exp(-\beta x)}{[1 - p\exp(-\beta x)]^2} \tag{1.3}$$

for $x > 0$, $0 < p < 1$ and $\beta > 0$. The case of Z being geometric has been considered much earlier by Rényi [17] in the context of rarefection and by Gnedenko and Korolev [5] and Kovalenko [10] with applications to reliability. We shall refer to (1.3) as the EG distribution. Kus [11] and Hemmati *et al.* [6] take Z to be a Poisson random variable with parameter λ, so yielding

$$f(x) = \frac{\lambda\beta}{1 - \exp(-\lambda)} \exp\{-\lambda - \beta x + \lambda \exp(-\beta x)\} \tag{1.4}$$

for $x > 0$, $\lambda > 0$ and $\beta > 0$. We shall refer to this as the EP distribution. Tahmasbi and Rezaei [20] take Z to be a logarithmic random variable with parameter p, so yielding

$$f(x) = -\frac{1}{\log p} \frac{\beta(1-p)\exp(-\beta x)}{1-(1-p)\exp(-\beta x)} \tag{1.5}$$

for $x > 0$, $0 < p < 1$ and $\beta > 0$. We shall refer to this as the EL distribution. We are not aware of any other model for (1.1) and (1.2) considered in the literature.

In this paper, we propose a new model for (1.1) and (1.2). We take Z to be a negative binomial random variable given by the probability mass function (pmf)

$$f_Z(z) = \binom{z-1}{k-1}(1-p)^k p^{z-k} \tag{1.6}$$

for $z = k, k+1, \ldots$. Geometric pmf is a particular case of (1.6). Poisson pmf is a limiting case of (1.6). Then (1.1) and (1.2) reduce to

$$F_X(x) = 1 - \frac{(1-p)^k \exp(-k\beta x)}{[1-p\exp(-\beta x)]^k} \tag{1.7}$$

and

$$f_X(x) = \frac{k\beta(1-p)^k \exp(-k\beta x)}{[1-p\exp(-\beta x)]^{k+1}}, \tag{1.8}$$

respectively, for $x > 0$, $k > 0$, $0 < p < 1$ and $\beta > 0$. The corresponding hazard rate function (hrf) is

$$h_X(x) = \frac{k\beta}{1-p\exp(-\beta x)} \tag{1.9}$$

for $x > 0$, $k > 0$, $0 < p < 1$ and $\beta > 0$. The corresponding quantile function is

$$F^{-1}(u) = \frac{1}{\beta}\log\left[p + \frac{1-p}{(1-u)^{1/k}}\right] \tag{1.10}$$

for $0 < u < 1$. We shall refer to the distribution given by (1.7) and (1.8) as the exponential negative binomial (ENB) distribution. The exponential distribution arises as the particular case for $k = 1$ and $p = 0$. The EG distribution of Adamidis and Loukas [1] arises as the particular case for $k = 1$.

Note that $d\log f(x)/dx < 0$ for all $x > 0$, so $f(x)$ is a monotonically decreasing function all the time. Note also that $f(0) = k\beta/(1-p)$, $f(\infty) = 0$ and $f(x) \sim k\beta(1-p)^k \exp(-k\beta x)$ as $x \to \infty$. So, the pdf takes a finite value at $x = 0$ and has an exponentially decaying upper tail. Clearly, the hrf given by (1.9) is also a monotonically decreasing function with $h(0) = k\beta/(1-p)$ and $h(\infty) = k\beta$.

Figure 1 illustrates possible shapes of (1.8) for selected parameter values. Figure 2 illustrates possible shapes of (1.9) for selected parameter values. The upper tails of (1.8) become lighter with increasing p and with increasing k. The upper tails of (1.9) become heavier with increasing p and become lighter with increasing k.

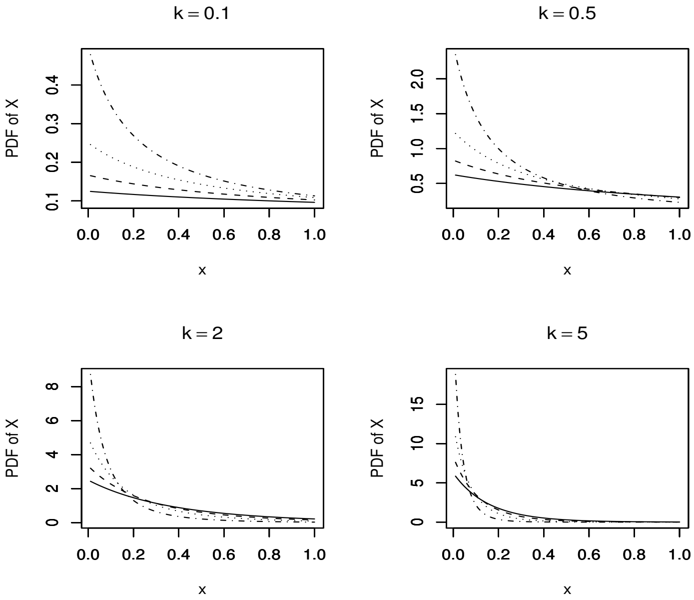

Figure 1: Plots (1.8) versus x for $\beta = 1$, $k = 0.1, 0.5, 2, 5$, $p = 0.2$ (solid curve), $p = 0.4$ (curve of dashes), $p = 0.6$ (curve of dots) and $p = 0.8$ (curve of dots and dashes).

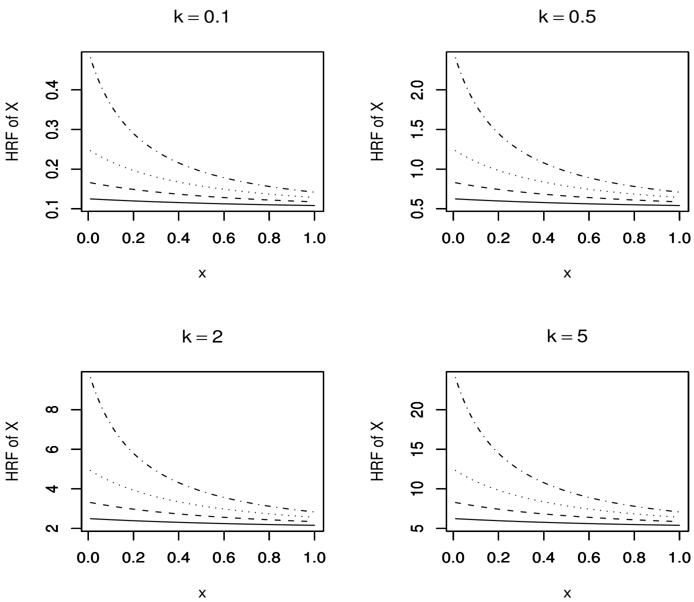

Figure 2: Plots (1.9) versus x for $\beta = 1$, $k = 0.1, 0.5, 2, 5$, $p = 0.2$ (solid curve), $p = 0.4$ (curve of dashes), $p = 0.6$ (curve of dots) and $p = 0.8$ (curve of dots and dashes).

The new distribution given by (1.7) and (1.8) can be motivated in several different ways. Firstly, the negative binomial distribution is a generalization of the geometric and Poisson distributions (Poisson is a limiting particular case). The negative binomial distribution with support over the set of all non-negative integers is also a generalization of the Poisson distribution in the sense that it can deduced as a hierarchical model if $X \sim$ Poisson (Λ) with Λ being a gamma random variable, see, for example, Casella and Berger [3].

So, (1.8) can be considered a generalization of (1.3) and (1.4). The logarithmic distribution is used to construct (1.5). The logarithmic distribution is widely used in population studies, iteration, fractality and chaos. But it is not a well known model for counts as the geometric, Poisson and negative binomial distributions are.

Secondly, using the series expansion

$$(1-a)^{-k-1} = \sum_{i=0}^{\infty} \binom{-k-1}{i} (-a)^i,$$

we can rewrite (1.8) as

$$f_X(x) = k\beta(1-p)^k \sum_{i=0}^{\infty} \binom{-k-1}{i} (-p)^i \exp\left\{-(k+i)\beta x\right\}. \tag{1.11}$$

Integrating (1.11), we can rewrite (1.7) as

$$F_X(x) = 1 - k(1-p)^k \sum_{i=0}^{\infty} \binom{-k-1}{i} \frac{(-p)^i}{k+i} \exp\left\{-(k+i)\beta x\right\}. \tag{1.12}$$

It follows from (1.11) and (1.12) that the ENB distribution is a mixture of the exponential distribution, the earliest and the best known model for failure times.

Our third motivation is simulation based. We shall see later (see Section 6) that the ENB distribution provides significantly better fits than the EG, EP and EL distributions, the only known competing distributions under the framework of (1.1) and (1.2), for more than tens of thousands of simulated samples. This is the case even when the samples are simulated from the EG, EP and EL distributions.

Our fourth and final motivation is real data based. We shall see later (see Section 7) that the proposed distribution outperforms the EP and EL distributions as well as the two-parameter Weibull distribution and the three-parameter Weibull Poisson distribution (Hemmati *et al.* [6]) with respect to at least two real data sets.

The contents of this paper are organized as follows. An account of mathematical properties of the new distribution is provided in Sections 2 to 4. The properties studied include: raw moments, order statistics and their moments, and asymptotic distribution of the extreme values. Estimation by the methods of moments and maximum likelihood is presented in Section 5. A simulation study to compare the performance of the proposed distribution versus the EG, EP and EL distributions is presented in Section 6. Finally, Section 7 illustrates an application by using two real data sets.

2. MOMENTS

Let X denote a random variable with the pdf (1.8). It follows from Lemma A.1 in the Appendix that

$$\begin{aligned} E\left(X^n\right) &= \frac{n!(1-p)^k}{\beta^n k^n} {}_{n+2}F_{n+1}\left(1+k, k, ..., k; k+1, ..., k+1; p\right) \\ &= \frac{n!(1-p)^k}{\beta^n k^n} {}_{n+1}F_{n}\left(k, ..., k; k+1, ..., k+1; p\right), \end{aligned}$$

where ${}_pF_q(a_1, a_2, ..., a_p; b_1, b_2, ..., b_q; x)$ denotes the generalized hypergeometric function defined by

$${}_pF_q\left(a_1, a_2, ..., a_p; b_1, b_2, ..., b_q; x\right) = \sum_{k=0}^{\infty} \frac{\left(a_1\right)_k \left(a_2\right)_k \cdots \left(a_p\right)_k}{\left(b_1\right)_k \left(b_2\right)_k \cdots \left(b_q\right)_k} \frac{x^k}{k!},$$

where $(e)_k = e(e+1)\cdots(e+k-1)$ denotes the ascending factorial. In particular, the first four moments of X are

$$E\left(X\right) = \frac{(1-p)^k}{\beta k} {}_2F_1\left(k, k; k+1; p\right),$$

$$E\left(X^2\right) = \frac{2(1-p)^k}{\beta^2 k^2} {}_3F_2\left(k, k, k; k+1, k+1; p\right),$$

$$E\left(X^3\right) = \frac{6(1-p)^k}{\beta^3 k^3} {}_4F_3\left(k, k, k, k; k+1, k+1, k+1; p\right)$$

and

$$E\left(X^4\right) = \frac{24(1-p)^k}{\beta^4 k^4} {}_5F_4\left(k, k, k, k, k; k+1, k+1, k+1, k+1; p\right).$$

The variance, skewness and kurtosis of X can be obtained using the relationships $Var(X) = E(X^2) - (E(X))^2$, $Skewness(X) = E(X - E(X))^3/(Var(X))^{3/2}$ and $Kurtosis(X) = E(X - E(X))^4/(Var(X))^2$. The variations of E (X), Var (X), Skewness (X) and Kurtosis (X) versus k and p for $\beta = 1$ are illustrated in Figure 3. It appears that $E(X)$ and $Var(X)$ are decreasing functions with respect to both k and p. Skewness (X) and Kurtosis (X) appear to increase with respect to a. With respect to p, they initially increase before decreasing.

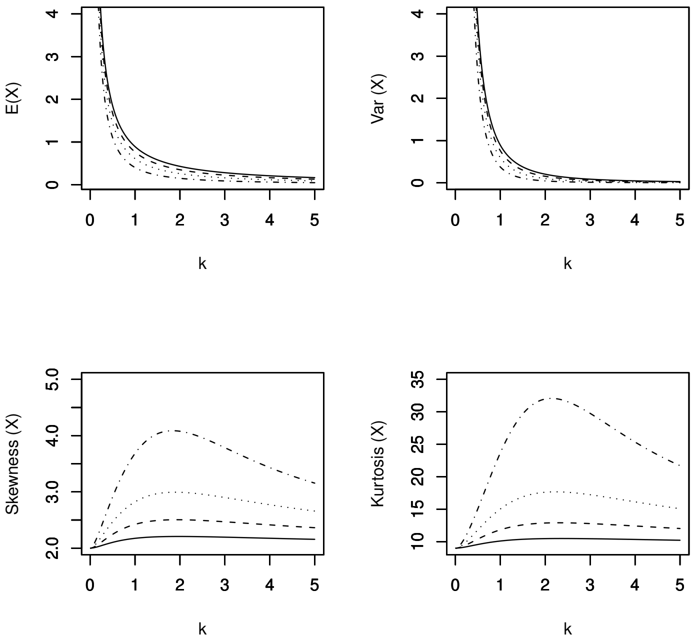

Figure 3: Mean, variance, skewness and kurtosis for (1.8) versus k for $p = 0.2$ (solid curve), $p = 0.4$ (curve of dashes), $p = 0.6$ (curve of dots) and $p = 0.8$ (curve of dots and dashes).

3. ORDER STATISTICS

Suppose $X_1, X_2, ..., X_n$ is a random sample from (1.8). Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the corresponding order statistics. It is well known that the pdf and the cdf of the rth order statistic, say $Y = X_{r:n}$, are given by

$$\begin{aligned} f_Y(y) &= \frac{n!}{(r-1)!(n-r)!} F_X^{r-1}(y) \left\{1 - F_X(y)\right\}^{n-r} f_X(y) \\ &= \frac{n!}{(r-1)!(n-r)!} \sum_{\ell=0}^{n-r} \binom{n-r}{\ell} (-1)^{\ell} F_X^{r-1+\ell}(y) f_X(y) \end{aligned}$$

and

$$F_Y(y) = \sum_{j=r}^{n} \binom{n}{j} F_X^{j}(y) \left\{1 - F_X(y)\right\}^{n-j} = \sum_{j=r}^{n} \sum_{\ell=0}^{n-j} \binom{n}{j} \binom{n-j}{\ell} (-1)^{\ell} F_X^{j+\ell}(y),$$

respectively, for $r = 1, 2, ..., n$. It follows from (1.8) and (1.7) that

$$f_Y(y) = \frac{k\beta n!}{(r-1)!(n-r)!} \sum_{\ell=0}^{n-r} \binom{n-r}{\ell} (-1)^{\ell} \frac{(1-p)^{(r+\ell)k} \exp\left[-(r+\ell)k\beta y\right]}{\left[1 - p\exp(-\beta y)\right]^{(r+\ell)k+1}}$$

and

$$F_Y(y) = \sum_{j=r}^{n} \sum_{\ell=0}^{n-j} \binom{n}{j} \binom{n-j}{\ell} (-1)^\ell \frac{(1-p)^{(j+\ell)k} \exp\left[-(j+\ell)k\beta y\right]}{\left[1-p\exp(-\beta y)\right]^{(j+\ell)k}}.$$

Using Lemma A.1 in the Appendix, the qth moment of Y can be expressed as

$$E\left(Y^q\right) = \frac{q!n!}{\beta^q k^q (r-1)!(n-r)!} \sum_{\ell=0}^{n-r} \binom{n-r}{\ell} \frac{(-1)^\ell (1-p)^{(r+\ell)k}}{(r+\ell)^{q+1}} G(\ell)$$

for $q \geq 1$, where $G(\ell) = {}_{q+1}F_q(\ell k + rk, ..., \ell k + rk; 1 + \ell k + rk, ..., 1 + \ell k + rk; p)$.

4. EXTREME VALUES

If $\overline{X} = (X_1 + \cdots + X_n)/n$ denotes the sample mean then by the usual central limit theorem $\sqrt{n}(\overline{X} - E(X))/\sqrt{Var(X)}$ approaches the standard normal distribution as $n \to \infty$. Sometimes one would be interested in the asymptotics of the extreme values $M_n = \max(X_1, ..., X_n)$ and $m_n = \min(X_1, ..., X_n)$.

Let $g(t) = 1/(k\beta)$. Take the cdf and the pdf as specified by (1.7) and (1.8), respectively. Since $f(x) \sim k\beta(1-p)^k \exp(-k\beta x)$ as $x \to \infty$,

$$\lim_{t\to\infty} \frac{1 - F(t + xg(t))}{1 - F(t)} = \lim_{t\to\infty} \frac{f(t + x/(k\beta))}{f(t)} = \exp(-x).$$

Since $f(0) = k\beta/(1-p)$,

$$\lim_{t\to 0} \frac{F(tx)}{F(t)} = \lim_{t\to\infty} \frac{xf(tx)}{f(t)} = x.$$

Hence, it follows from Theorem 1.6.2 in Leadbetter *et al.* [12] that there must be norming constants $a_n > 0$, b_n, $c_n > 0$ and d_n such that

$$\Pr\left\{a_n\left(M_n - b_n\right) \leq x\right\} \to \exp\left\{-\exp(-x)\right\}$$

and

$$\Pr\left\{c_n\left(m_n - d_n\right) \leq x\right\} \to 1 - \exp\left(-x\right)$$

as $n \to \infty$. The form of the norming constants can also be determined. For instance, using Corollary 1.6.3 in Leadbetter *et al.* [12], one can see that $b_n = F^{-1}(1 - 1/n)$ and $a_n = k\beta$, where $F^{-1}(\cdot)$ is given by (1.10).

5. ESTIMATION

Here, we consider estimation by the methods of moments and maximum likelihood and provide expressions for the associated Fisher information matrix.

Suppose $x_1, ..., x_n$ is a random sample from (1.8). For moments estimation, let $m_1 = (1/n) \sum_{j=1}^{n} x_j$, $m_2 = (1/n) \sum_{j=1}^{n} x_j^2$ and $m_3 = (1/n) \sum_{j=1}^{n} x_j^3$. By equating the theoretical moments of (1.8) with the sample moments, we obtain the equations:

$$\frac{(1-p)^k}{\beta k} {}_2F_1 \left(k, k; k+1; p\right) = m_1,$$

$$\frac{2(1-p)^k}{\beta^2 k^2} {}_3F_2 \left(k, k, k; k+1, k+1; p\right) = m_2,$$

and

$$\frac{6(1-p)^k}{\beta^3 k^3} {}_4F_3 \left(k, k, k, k; k+1, k+1, k+1; p\right) = m_3.$$

The method of moments estimators (mmes), say $\widetilde{p}$, $\widetilde{k}$ and $\widetilde{\beta}$, are the simultaneous solutions of these three equations.

Now consider estimation by the method of maximum likelihood. The log likelihood function of the three parameters is:

$$\begin{aligned} \log L(p, k, \beta) \; &= \; n \log(k\beta) + nk \log(1-p) - k\beta \sum_{i=1}^{n} x_i \\ & \quad -(k+1) \sum_{i=1}^{n} \log\left[1 - p \exp\left(-\beta x_i\right)\right]. \end{aligned} \tag{5.1}$$

It follows that the maximum likelihood estimators (mles), say $\widehat{p}$, $\widehat{k}$ and $\widehat{\beta}$, are the simultaneous solutions of the equations:

$$\frac{n}{k} + n \log(1-p) = \beta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log\left[1 - p \exp\left(-\beta x_i\right)\right],$$

$$\frac{n}{\beta} = k \sum_{i=1}^{n} x_i + p(k+1) \sum_{i=1}^{n} \frac{x_i \exp\left(-\beta x_i\right)}{1 - p \exp\left(-\beta x_i\right)},$$

and

$$\frac{nk}{1-p} = (k+1) \sum_{i=1}^{n} \frac{\exp\left(-\beta x_i\right)}{1 - p \exp\left(-\beta x_i\right)}.$$

For interval estimation of (p, k, β) and tests of hypothesis, one requires the Fisher information matrix:

$$\mathbf{I} = \begin{pmatrix} E\left(-\frac{\partial^2 \log L}{\partial p^2}\right) & E\left(-\frac{\partial^2 \log L}{\partial p \partial k}\right) & E\left(-\frac{\partial^2 \log L}{\partial p \partial \beta}\right) \\ E\left(-\frac{\partial^2 \log L}{\partial k \partial p}\right) & E\left(-\frac{\partial^2 \log L}{\partial k^2}\right) & E\left(-\frac{\partial^2 \log L}{\partial k \partial \beta}\right) \\ E\left(-\frac{\partial^2 \log L}{\partial \beta \partial p}\right) & E\left(-\frac{\partial^2 \log L}{\partial \beta \partial k}\right) & E\left(-\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{pmatrix}.$$

Using Lemma A.1 in the Appendix, the elements of this matrix for (5.1) can be worked out as:

$$I_{11} = \frac{n}{k^2},$$

$$\begin{aligned} I_{12} = I_{21} &= \frac{n(1-p)^k}{\beta k} {}_2F_1\left(k, k; k+1; p\right) \\ &+ \frac{npk(1-p)^k}{\beta(k+1)^2} {}_3F_2\left(k+3, k+1, k+1; k+2, k+2; p\right), \end{aligned}$$

$$I_{13} = I_{31} = \frac{n}{1-p} - \frac{nk}{(k+1)(1-p)},$$

$$I_{22} = \frac{n}{\beta^2} - \frac{npk(1-p)^k}{\beta^2(k+1)^2} {}_3F_2\left(k+3, k+1, k+1; k+2, k+2; p\right),$$

$$\begin{aligned} I_{23} = I_{32} &= \frac{nk(1-p)^k}{\beta(k+1)} {}_3F_2\left(k+3, k+1, k+1; k+2, k+2; p\right) \\ &+ \frac{npk(k+1)(1-p)^k}{\beta(k+2)^2} {}_2F_1\left(k+2, k+2; k+3; p\right), \end{aligned}$$

and

$$I_{33} = \frac{nk}{(1-p)^2} - \frac{nk(k+1)}{(k+2)(1-p)^2}.$$

Under regularity conditions, the asymptotic distribution of $(\widehat{p}, \widehat{k}, \widehat{\beta})$ as $n \to \infty$ is trivariate normal with zero means and variance co-variance matrix $\mathbf{I}^{-1}$. So, $\mathrm{Var}\left(\widehat{p}\right) = (I_{33}I_{22} - I_{32}I_{23})/\Delta$, $\mathrm{Cov}\left(\widehat{p}, \widehat{k}\right) = -(I_{33}I_{12} - I_{32}I_{13})/\Delta$, $\mathrm{Cov}\left(\widehat{p}, \widehat{\beta}\right) = (I_{23}I_{12} - I_{22}I_{13})/\Delta$, $\mathrm{Var}\left(\widehat{k}\right) = (I_{33}I_{11} - I_{31}I_{13})/\Delta$, $\mathrm{Cov}\left(\widehat{k}, \widehat{\beta}\right) = -(I_{23}I_{11} - I_{21}I_{13})/\Delta$ and $\mathrm{Var}\left(\widehat{\beta}\right) = (I_{22}I_{11} - I_{21}I_{12})/\Delta$, where $\Delta = I_{11}(I_{33}I_{22} - I_{32}I_{23}) - I_{21}(I_{33}I_{12} - I_{32}I_{13}) + I_{31}(I_{23}I_{12} - I_{22}I_{13})$.

6. A SIMULATION STUDY

Here, we perform a simulation study to compare the performance of the proposed distribution versus those given by (1.3), (1.4) and (1.5); that is, the EG, EP and EL distributions, the only known competing distributions under the framework of (1.1) and (1.2). We use the following scheme:

1. Generate ten thousand samples of size n from (1.8);

2. For each sample, fit the models given by (1.8), (1.3), (1.4) and (1.5);

3. Let ℓ_{1i}, ℓ_{2i}, ℓ_{3i} and ℓ_{4i}, $i = 1, 2, ..., 10000$ denote the maximized log-likelihoods for (1.8), (1.3), (1.4) and (1.5) for the ten thousand samples;

4. Draw the box plots of $2(\ell_{1i} - \ell_{2i})$, $2(\ell_{1i} - \ell_{3i})$ and $2(\ell_{1i} - \ell_{4i})$, $i = 1, 2, ..., 10000$.

This scheme compares the fits of the four distributions when simulated samples are from the proposed distribution. For completeness, we repeated the above scheme with simulated samples coming from the EG, EP and EL distributions.

The resulting box plots are shown in Figure 4 for $n = 25$ and $(\beta, \lambda, k, p) = (1, 1, 2, 0.5)$. The figure shows that proposed distribution provides the best fit wherever the sample comes from. The relative performances of the EG, EP and EL distributions with respect to the proposed one appear similar. The four distributions are not nested, so the likelihood ratio test may not be used to discriminate between them. But the differences in the log-likelihood are so large that they are significant even with respect to the AIC and BIC criteria.

For reasons of space, we have presented results for only one value for n and the parameters. But the conclusions of Figure 4 hold also for larger sample sizes and other parameter values.

The results are not surprising because, as explained in Section 1, the proposed distribution is flexible enough to contain the EG and EP distributions as particular cases. The logarithmic distribution used to construct the EL distribution is not flexible and is certainly not widely used.

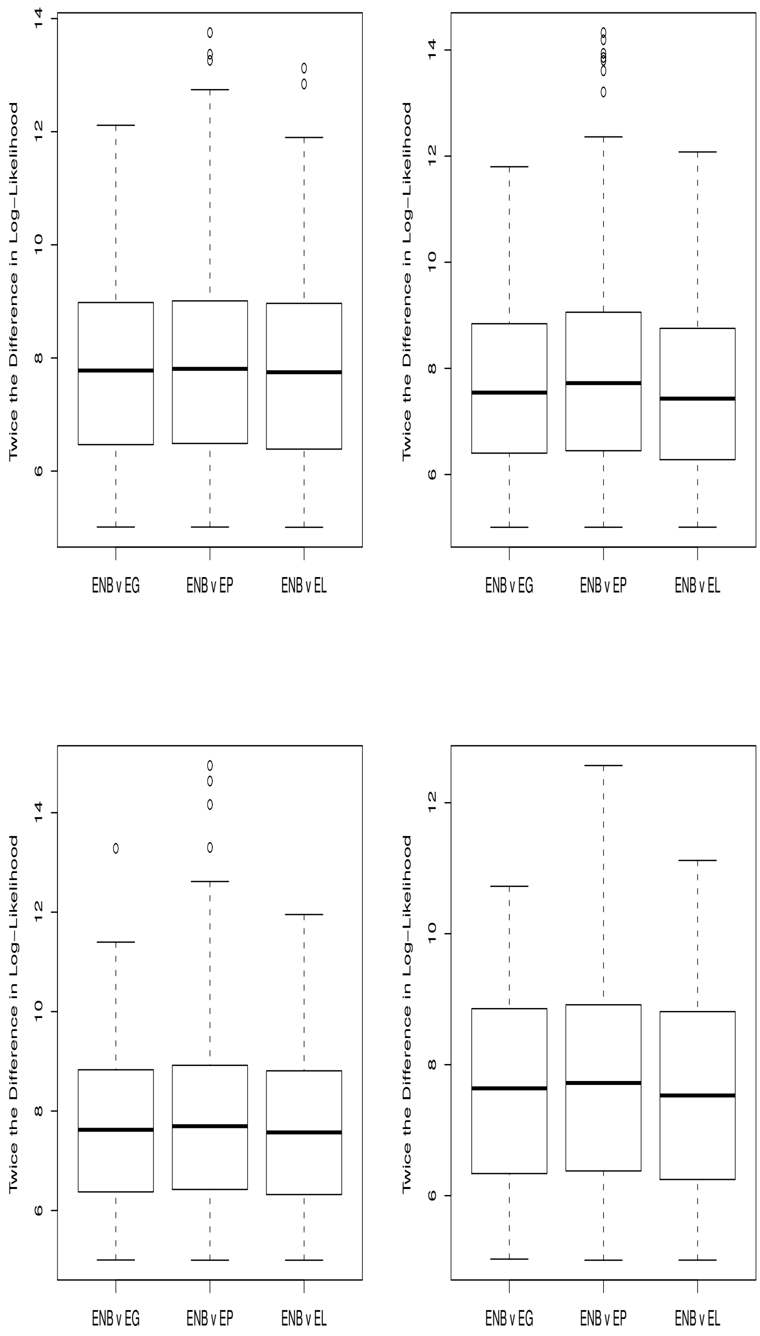

Figure 4: Box plots of $2(\ell_{1i} - \ell_{2i})$, $2(\ell_{1i} - \ell_{3i})$ and $2(\ell_{1i} - \ell_{4i})$ when the simulated samples are from the proposed distribution (top left), the EG distribution (top right), the EP distribution (bottom left) and the EL distribution (bottom right).

7. APPLICATIONS

Here, we illustrate applicability of the ENB distribution using two real data sets. The first data set contains intervals in days between successive failures of a piece of software. See Jelinski and Moranda [8] and Linda [13]. The second data set consists of lifetimes of pressure vessels. See Pal *et al.* [14].

We compare the fit of the ENB distribution with those of the EP and EL distributions as well as those of the Weibull distribution given by the pdf

$$f(x) = \beta\lambda^{\beta}x^{\beta-1}\exp\left\{-(\lambda x)^{\beta}\right\}$$

(for $x > 0$, $\lambda > 0$ and $\beta > 0$) and, the Weibull Poisson distribution (Hemmati *et al.* [6]) given by the pdf

$$f(x) = \frac{\theta\alpha\beta^{\alpha}}{\exp(\theta)-1}\exp\left\{-(\beta x)^{\alpha}\right\}\exp\left\{-\theta\exp\left[-(\beta x)^{\alpha}\right]\right\}$$

for $x > 0$, $\theta > 0$, $\alpha > 0$ and $\beta > 0$. The parameters of the ENB distribution are estimated by the method of maximum likelihood, see Section 5. The parameters of other distributions are also estimated by the method of maximum likelihood.

The mles and the corresponding log-likelihood value, the Kolmogorov Smirnov statistic, its p value, the AIC value and the BIC value are shown in Tables 1 and 2. We can see that the largest log-likelihood value, the largest p value, the smallest AIC value and the smallest BIC value are obtained for the ENB distribution.

Table 1: Fitted estimates for data set 1.

Model	Parameter estimates	Log likelihood	K-S statistic	p-value	AIC	BIC
Weibull	(16.7835, 0.6460)	-131.6366	0.2046	0.1092	267.2732	270.2662
EL	(0.0300, 0.0162)	-129.6636	0.2147	0.0818	263.3273	266.3203
EP	(0.0191, 3.9168)	-131.2939	0.1967	0.1358	266.5878	269.5808
WP	(0.0182, 0.8072, 3.3587)	-129.5968	0.1634	0.3070	265.1936	269.6831
ENB	(0.0076, 0.9491, 0.9462)	-127.7312	0.1372	0.5189	261.4624	265.9519

Table 2: Fitted estimates for data set 2.

Model	Parameter estimates	Log likelihood	K-S statistic	p-value	AIC	BIC
Weibull	(488.1066, 0.7162)	-145.3353	0.1519	0.6904	294.6705	296.6620
EL	(0.1239, 0.0011)	-146.5781	0.1700	0.5531	297.1562	299.1477
EP	(0.0015, 0.6978)	-146.9594	0.1534	0.6787	297.9189	299.9104
WP	(0.0020, 0.7162, 0.0001)	-145.3353	0.1519	0.6904	296.6705	299.6577
ENB	(0.0342, 0.0434, 0.9748)	-143.4332	0.1309	0.8400	292.8665	295.8537

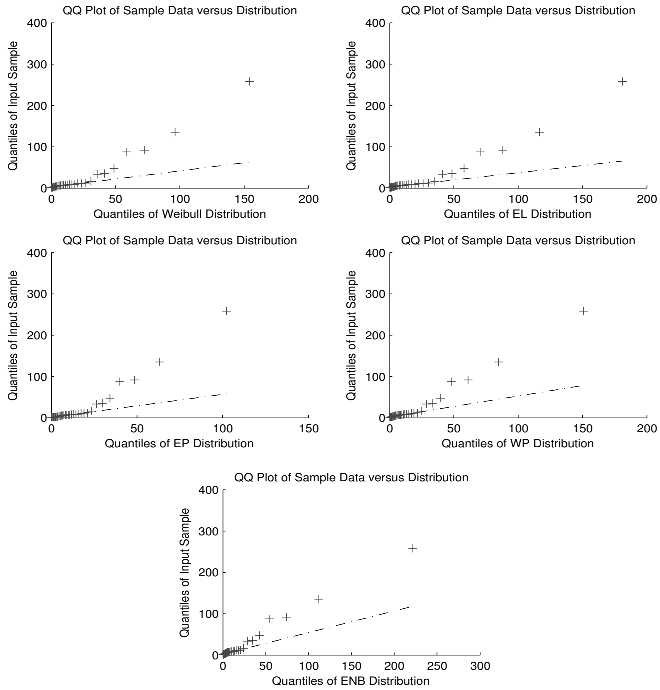

Figure 5: Quantile-quantile plots for the fitted models for the first data set.

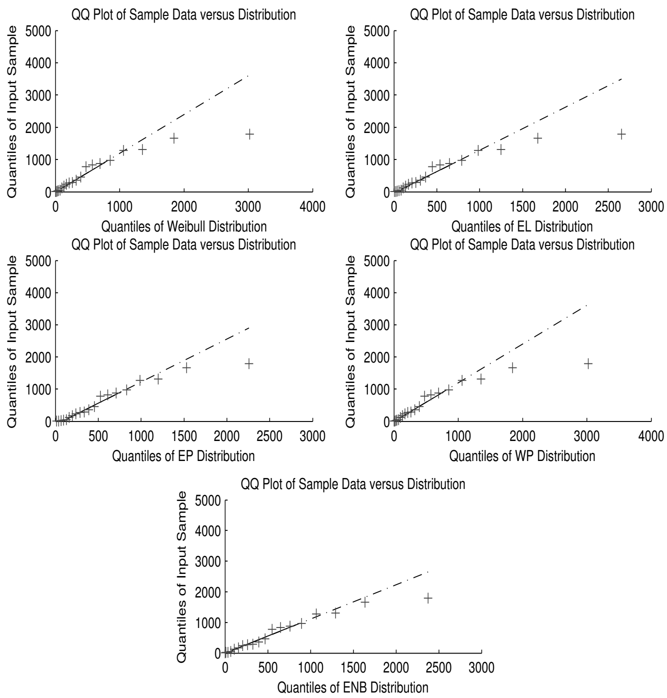

Figure 6: Quantile-quantile plots for the fitted models for the second data set.

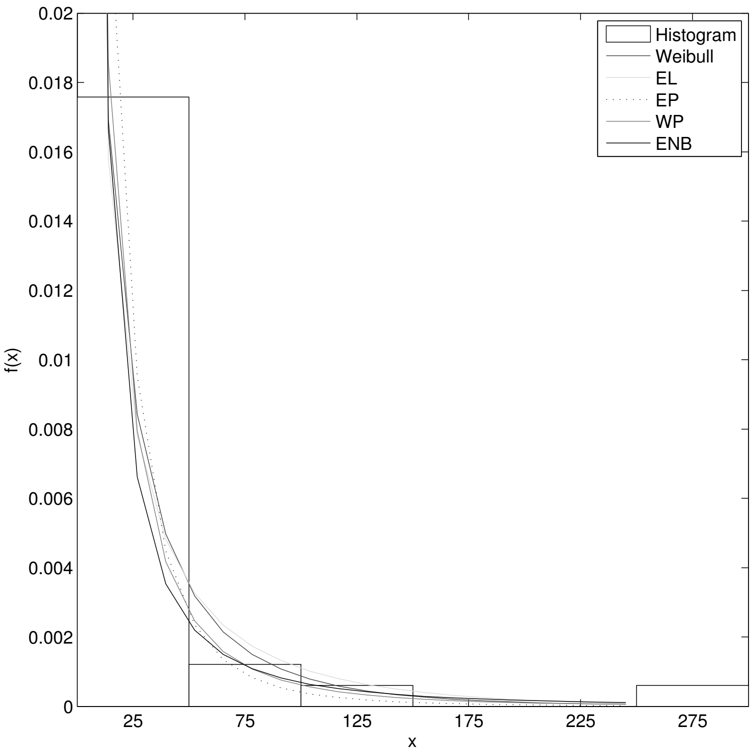

Figure 7: Fitted pdfs and the observed histogram for the first data set.

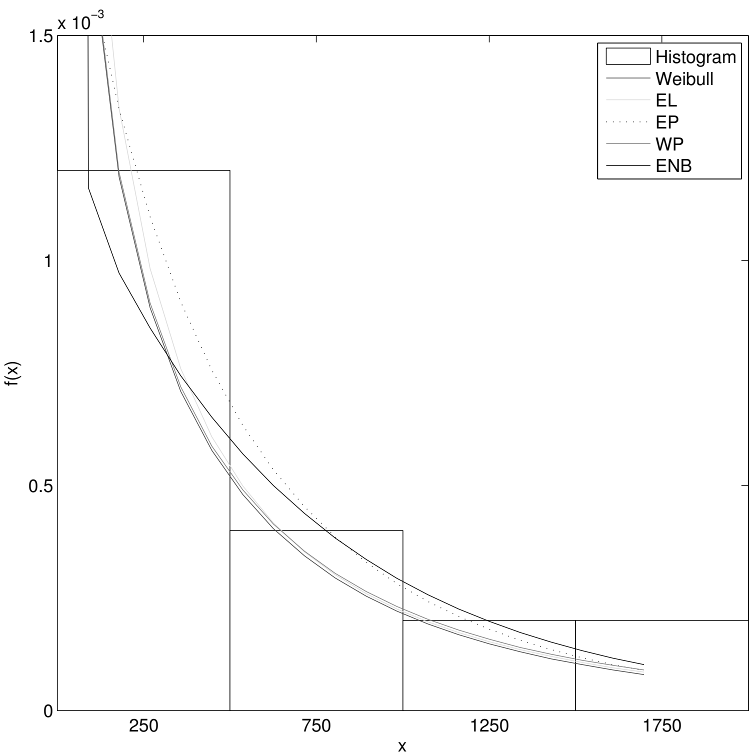

Figure 8: Fitted pdfs and the observed histogram for the second data set.

The conclusion based on Tables 1 and 2 can be verified by means of quantile-quantile plots and density plots. A quantile-quantile plot consists of plots of the observed quantiles against quantiles predicted by the fitted model. For example, for the model based on the ENB distribution, $x_{(j)}$ was plotted versus $F^{-1}((j - 0.375)/(n + 0.25))$, $j = 1, 2, ..., n$, as recommended by Blom [2] and Chambers *et al.* [4], where $F^{-1}(\cdot)$ is given by (1.10), $x_{(j)}$ are the sorted values of the observed data in the ascending order and n is the number of observations. The quantile-quantile plots for the five fitted models and for each data set are shown in Figures 5 and 6. We can see that the model based on the ENB distribution has the points closer to the diagonal line for each data set.

A density plot compares the fitted pdfs of the models with the empirical histogram of the observed data. The density plots for the two data sets are shown in Figures 7 and 8. Again the fitted pdfs for the ENB distribution appear to capture the general pattern of the empirical histograms better.

APPENDIX

We need the following lemma.

Lemma A.1. *Let*

$$I(a,b,c) = k\beta(1-p)^k \int_0^\infty \frac{x^a \exp\left[-(k+b)\beta x\right]}{\left[1-p\,\exp(-\beta x)\right]^{k+1+c}} dx.$$

Then

$$I(a,b,c) = \frac{a!k(1-p)^k{}_{a+2}F_{a+1}\left(1+k+c, k+b, ..., k+b;\ k+b+1, ..., k+b+1;\ p\right)}{\beta^a (k+b)^{a+1}}.$$

Proof: Using the series expansion

$$(1-a)^{-k-1-c} = \sum_{i=0}^{\infty} \binom{-k-1-c}{i} (-a)^i,$$

we can write

$$\begin{aligned} I(a,b,c) &= k\beta(1-p)^k \sum_{i=0}^{\infty} \binom{-k-1-c}{i} (-p)^i \int_0^\infty x^a \exp\left[-(k+b+i)\beta x)\right] dx \\ &= a!k\beta^{-a}(1-p)^k \sum_{i=0}^{\infty} \binom{-k-1-c}{i} \frac{(-p)^i}{(k+b+i)^{a+1}} \\ &= a!k\beta^{-a}(1-p)^k (k+b)^{-a-1} \sum_{i=0}^{\infty} \frac{(k+1+c)_i (k+b)_i \cdots (k+b)_i}{(k+b+1)_i \cdots (k+b+1)_i} \frac{p^i}{i!}. \end{aligned}$$

The result now follows from the definition of hypergeometric functions. □

ACKNOWLEDGMENTS

The authors would like to thank the Editor, the Associate Editor, and the referee for careful reading and for comments which greatly improved the paper.

REFERENCES

[1] Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate, *Statistics and Probability Letters*, **39**, 35–42.

[2] Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*, John Wiley and Sons, New York.

[3] Casella, G. and Berger, R.L. (2001). *Statistical Inference*, second edition, Duxbury Press, California.

[4] Chambers, J.; Cleveland, W.; Kleiner, B. and Tukey, P. (1983). *Graphical Methods for Data Analysis*, Chapman and Hall, London.

[5] Gnedenko, B.V. and Korolev, V.Yu. (1996). *Random Summation: Limit Theorems and Applications*, CRC Press, Bocca Raton.

[6] Hemmati, F.; Khorram, E. and Rezakhah, S. (2011). A new three-parameter ageing distribution, *Journal of Statistical Planning and Inference*, **141**, 2266–2275.

[7] Hess, K.T.; Liewald, A. and Schmidt, K.D. (2002). An extension of Panjer's recursion, *ASTIN Bulletin*, **32**, 283–297.

[8] Jelinski, Z. and Moranda, P.B. (1972). *Software reliability research*, In: "Statistical Performance Evaluation" (W. Freiberger, Ed.), Academic Press, New York, pp. 485–502.

[9] Katz, L. (1965). *Unified treatment of a broad class of discrete probability distributions*, In: "Classical and Contagious Discrete Distributions", Pergamon Press, Oxford, pp. 175–182.

[10] Kovalenko, I.N. (1971). *On the class of limit distributions for thinning streams of homogeneous events*, In: "Selected Translations in Mathematical Statistics and Probability", volume 9, Providence, Rhode Island, pp. 75–81.

[11] Kus, C. (2007). A new lifetime distribution, *Computational Statistics and Data Analysis*, **51**, 4497–4509.

[12] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1987). *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York.

[13] Linda, C.W. (1999). *Reliability Modelling: A Statistical Approach*, Chapman and Hall/CRC, New York.

[14] Pal, N.; Chun, J. and Wooi, K.L. (2006). *Handbook of Exponential and Related Distributions for Engineers and Scientists*, Chapman and Hall, London.

[15] Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions, *ASTIN Bulletin*, **12**, 22–26.

[16] Pestana, D.D. and Velosa, S.F. (2004). Extensions of Katz–Panjer families of discrete distributions, *Revstat*, **2**, 145–162.

[17] Rényi, A. (1976). *A characterization of the Poisson process*, In: "Selected Papers of Alfred Rényi", volume 1, Akademiai Kiado, Budapest, pp. 519–527.

[18] Rolski, T.; Schmidli, H.; Schmidt, V. and Teugels, J. (1999). *Stochastic Processes for Insurance and Finance*, John Wiley and Sons, New York.

[19] Sundt, B. and Jewell, W.S. (1981). Further results on recursive evaluation of compound distributions, *ASTIN Bulletin*, **12**, 27–39.

[20] Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate, *Computational Statistics and Data Analysis*, **52**, 3889–3901.

REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 211–230

NONPARAMETRIC ESTIMATION FOR FUNCTIONAL DATA BY WAVELET THRESHOLDING

Authors: Christophe Chesneau
– Université de Caen, LMNO, Campus II, Science 3,
14032, Caen, France
christophe.chesneau@unicaen.fr

Maher Kachour
– École supérieure de commerce IDRAC,
47, rue Sergent Michel Berthet CP 607, 69258 Lyon Cedex 09, France
maher.kachour@idraclyon.com

Bertrand Maillot
– Université de Caen, LMNO, Campus II, Science 3,
14032, Caen, France
Bertrand.Maillot@unicaen.fr

Received: February 2012 Revised: October 2012 Accepted: December 2012

Abstract:

- This paper deals with density and regression estimation problems for functional data. Using wavelet bases for Hilbert spaces of functions, we develop a new adaptive procedure based on a term-by-term selection of the wavelet coefficients estimators. We study its asymptotic performances by considering the mean integrated squared error over adapted decomposition spaces.

Key-Words:

- *functional data; density estimation; nonparametric regression; wavelets; hard thresholding.*

AMS Subject Classification:

- 62G07, 60B11.

1. INTRODUCTION

Due to technological progress, in particular the enlarged capacity of computer memory and the increasing efficiency of data collection devices, there is a growing number of applied sciences (biometrics, chemometrics, meteorology, medical sciences...) where collected data are curves which require appropriate statistical tools. Because of this, functional data analysis has known a quite important development in the last fifteen years (see, e.g., [26], [27], [14], [7], [16], [17] and [18] for monographs and collective books on this specific subject). However, whereas there has been substantial work on the nonparametric estimation of the probability density function for univariate and multivariate random variables since the papers of [22] and [28], much less attention has been paid to the infinite-dimensional case. The extension of the results from the multivariate framework to the infinite dimensional one is not direct since there is no equivalent of the Lebesgue measure on an infinite dimensional Hilbert space. In fact, the only locally finite and translation invariant measure on an infinite dimensional Hilbert space is the null measure and any locally finite measure μ is even very irregular: denoting by $\mathcal{B}(x,r)$ the ball of center x and radius r, we have that, for any point x, any arbitrary large M and any arbitrary small r such that $\mu(\mathcal{B}(x,r)) < \infty$, there exist $(x_1,x_2) \in \mathcal{B}(x,r)^2$ such that $\mu(\mathcal{B}(x_1,r/4)) < M \times \mu(\mathcal{B}(x_2,r/4))$. For a coverage of the theme of measures on infinite dimension spaces, we refer to [33], [34], [8] and [31].

The first consistency result for a kernel estimator of the density function for infinite dimensional random variables has been obtained in [4] where a rate is given in the special case when the kernel is an indicator function and the density is defined with respect to the Wiener measure. Later, different estimators of the density, based on orthogonal series (see [5]), delta sequences (see [25]) or wavelets (see [24]), have been proposed but none of them is adaptive. Note that the estimation of the density probability function is nonetheless itself of intrinsic interest but it also has a key role in mode estimation and curve clustering (see [6]).

Contrary to the chronology of studies in the multivariate case, in the functional framework, estimators of the regression function have been proposed before those of the density. Ferraty and Vieu introduced the first fully nonparametric estimator of the regression function, at first under the hypothesis that the underlying measure has a fractal dimension in [12] and then using only probabilities of small balls in [13]. However, since these pioneering works, no adaptive estimator has been proposed.

Considering the density estimation problem from functional data, [24] has recently developed a new procedure based on the multiresolution approach on a separable Hilbert space introduced by [19]. This procedure belongs to the family of the linear wavelet estimators. As proved in [24, Theorem 3.1], it enjoys powerful asymptotic properties. However, such a linear wavelet estimator has two drawbacks: it is not adaptive (i.e., its performances are deeply associated to the smoothness of the unknown function) and it is not efficient to estimate functions with complex singularities (the sparsity nature of the wavelet decomposition of the unknown function is not captured). For these reasons, [24, Page 2 lines 14-16] states *"it would be interesting to investigate the advantage of these wavelet estimators for functional data by using wavelet thresholding suggested by [11]"*. This perspective motivates our study.

Adopting the multiresolution approach on a separable Hilbert space H of [19], we construct an adaptive wavelet procedure extending the hard thresholding rule introduced by [11] to a general nonparametric estimation context for functional data. In order to study its asymptotic properties, we introduce two different kinds of decomposition spaces expressed in terms of wavelet coefficients via the new basis (see Section 2). They are related to the maxiset approach introduced by [3] and of interest as they contain a wide variety of unknown functions, complex or not. Exploring the density model and the regression model for functional data, we determine the rates of convergence attained by our estimator under the mean integrated squared error on H and over the intersection of the two considering decomposition spaces. To the best of our knowledge, this study is the first one developing a wavelet-based adaptive estimator in the context of functional data (and studying it theoretically). Let us mention that the new findings includes several obtained results for $H = \mathbb{L}_p([a, b])$.

The paper is structured as follows. In Section 2, we briefly describe the wavelet bases on H and we define some decomposition spaces. The density estimation problem for functional data via wavelet thresholding is considered in Section 3. The regression one is developed in Section 4. The proofs are gathered in Section 5.

2. WAVELET BASES ON H AND DECOMPOSITION SPACES

2.1. Wavelet bases on H

Let us briefly describe the construction of wavelet bases on H introduced by [19]. Let H be a separable Hilbert space of real- or complex-valued functions defined on a complete separable metric space or a normed vector space S. Since H is separable, it has an orthonormal basis $\mathcal{E} = \{e_j;\ j \in \Lambda\}$ for some countable index set Λ. As usual, we denote by $\langle .,. \rangle$ and $||.||$ the inner product and corresponding norm that H is equipped with.

Let $\{\mathcal{I}_k;\ k \geq 0\}$ be an increasing sequence of finite subsets of Λ such that $\bigcup_{k\geq 0} \mathcal{I}_k = \Lambda$ and, for any $k \geq 0$, $\mathcal{J}_k = \mathcal{I}_{k+1}/\mathcal{I}_k$. For any $k \geq 0$, we suppose that there exist $\zeta_{k,\ell} \in S$, $\ell \in \mathcal{I}_k$ and $\eta_{k,\ell} \in S$, $\ell \in \mathcal{J}_k$, such that the two matrices

$$A_k = (e_j(\zeta_{k,\ell}))_{(j,\ell)\in\mathcal{I}_k^2}, \qquad B_k = (e_j(\eta_{k,\ell}))_{(j,\ell)\in\mathcal{J}_k^2},$$

satisfy one of the two following conditions:

(**A1**) $A_k^* A_k = \mathrm{diag}(c_{k,\ell})_{\ell\in\mathcal{I}_k}$ and $B_k^* B_k = \mathrm{diag}(s_{k,\ell})_{\ell\in\mathcal{J}_k}$, where $c_{k,\ell}$, $s_{k,\ell'}$, for $\ell \in \mathcal{I}_k$ and $\ell' \in \mathcal{J}_k$, are positive constants,

(**A2**) $A_k A_k^* = \mathrm{diag}(d_{k,j})_{j\in\mathcal{I}_k}$ and $B_k B_k^* = \mathrm{diag}(t_{k,j})_{j\in\mathcal{J}_k}$, where $d_{k,j}$, $t_{k,j'}$ for $j \in \mathcal{I}_k$ and $j' \in \mathcal{J}_k$, are positive constants.

For any $x \in S$, we set

$$\begin{cases} \phi_k(x;\zeta_{k,\ell}) = \displaystyle\sum_{j\in\mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} \overline{e_j(\zeta_{k,\ell})} e_j(x), \\ \psi_k(x;\eta_{k,\ell}) = \displaystyle\sum_{j\in\mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} \overline{e_j(\eta_{k,\ell})} e_j(x), \end{cases}$$

where

$$g_{j,k,\ell} = \begin{cases} c_{k,\ell} & \text{if } \textbf{(A1)}, \\ d_{k,j} & \text{if } \textbf{(A2)}, \end{cases} \qquad h_{j,k,\ell} = \begin{cases} s_{k,\ell} & \text{if } \textbf{(A1)}, \\ t_{k,j} & \text{if } \textbf{(A2)}. \end{cases}$$

Then the collection

$$\mathcal{B} = \{\phi_0(x;\zeta_{0,\ell}),\ \ell \in \mathcal{I}_0;\ \ \psi_k(x;\eta_{k,\ell}),\ k \geq 0, \ell \in \mathcal{J}_k\}$$

is an orthonormal basis for H (see [19, Theorem 2 (a)]).

Consequently, any $f \in H$ can be expressed on $\mathcal{B}$ as

$$f(x) = \sum_{\ell\in\mathcal{I}_0} \alpha_{0,\ell}\phi_0(x;\zeta_{0,\ell}) + \sum_{k\geq 0}\sum_{\ell\in\mathcal{J}_k} \beta_{k,\ell}\psi_k(x;\eta_{k,\ell}), \qquad x \in S,$$

where

$$\alpha_{0,\ell} = \langle f, \phi_0(.;\zeta_{0,\ell})\rangle, \qquad \beta_{k,\ell} = \langle f, \psi_k(.;\eta_{k,\ell})\rangle. \tag{2.1}$$

We formulate the two following assumptions on $\mathcal{E}$:

- There exists a constant $C_1 > 0$ such that, for any integer $k \geq 0$,

$$\sum_{j\in\mathcal{I}_k} \frac{1}{g_{j,k,\ell}} |e_j(\zeta_{k,\ell})|^2 \leq C_1, \qquad \sum_{j\in\mathcal{J}_k} \frac{1}{h_{j,k,\ell}} |e_j(\eta_{k,\ell})|^2 \leq C_1. \tag{2.2}$$

  This assumption is obviously satisfied under (**A1**) with $C_1 = 1$. Remark also that the second example in [19, Section 4] satisfies both **(A2)** and (2.2).

- There exists a constant $C_2 > 0$ such that, for any integer $k \geq 0$,

$$\sup_{x \in S} \sum_{j \in \mathcal{J}_k} |e_j(x)|^2 \leq C_2 |\mathcal{J}_k|. \tag{2.3}$$

  This assumption is satisfied by the three examples in [19] (we have $\sup_{x \in S} \sup_{j \in \mathcal{J}_k} |e_j(x)| \leq 1$). Remark that it contains [24, (3.16)].

2.2. Decomposition spaces

Let $s > 0$ and $r > 0$. From the wavelet coefficients (2.1) of a function $f \in H$, we define the Besov spaces $\mathcal{B}^s_\infty(H)$ by

$$\mathcal{B}^s_\infty(H) = \left\{ f \in H; \quad \sup_{m \geq 0} |\mathcal{J}_m|^{2s} \sum_{k \geq m} \sum_{\ell \in \mathcal{J}_k} |\beta_{k,\ell}|^2 < \infty \right\} \tag{2.4}$$

and the “weak Besov spaces” $\mathcal{W}^r(H)$ by

$$\mathcal{W}^r(H) = \left\{ f \in H; \quad \sup_{\lambda > 0} \lambda^r \sum_{k \geq 0} \sum_{\ell \in \mathcal{J}_k} \mathbb{I}_{\{|\beta_{k,\ell}| \geq \lambda\}} < \infty \right\}, \tag{2.5}$$

where $\mathbb{I}_{\mathcal{A}}$ is the indicator function on $\mathcal{A}$.

Such kinds of function spaces are extensively used in approximation theory for the study of non linear procedures such as thresholding and greedy algorithms. See, e.g., [10] and [30]. From a statistical point of view, they are connected to the maxiset approach. See, e.g., [3], [21] and [1].

3. DENSITY ESTIMATION FOR FUNCTIONAL DATA

3.1. Problem statement

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $\{X_i;\ i \geq 1\}$ be i.i.d. random variables defined on $\{\Omega, \mathcal{F}, P\}$ and taking values in a complete separable metric space or a Hilbert space S associated with the corresponding Borel σ-algebra $\mathcal{B}$. Let P_X be the probability measure induced by X_1 on $(S, \mathcal{B})$. Suppose that there exists a σ-finite measure ν on the measurable space $(S, \mathcal{B})$ such that P_X is dominated by ν. The Radon-Nikodym theorem ensures the existence of a nonnegative measurable function f such that

$$P_X(B) = \int_B f(x)\nu(dx), \qquad B \in \mathcal{B}.$$

In this context, we aim to estimate f based on n observed functional data $X_1, ..., X_n$.

We suppose that $f \in H$, where H is a separable Hilbert space of real-valued functions defined on S and square integrable with respect to the σ-finite measure ν.

Moreover, we suppose that there exists a known constant $C_f > 0$ such that

$$\sup_{x \in S} f(x) \leq C_f. \tag{3.1}$$

The estimation of the density function for functional data has been first addressed in [4], and the consistency in L^2-norm has been established in [5] for a projection estimator. More recently, [24] established the convergence in mean square -with rate- of a non adaptive wavelets based estimator. We refer to these papers for details and applications of the model.

3.2. Estimator

Following the procedure of [11] and adopting the notation of Section 2, we define the wavelet hard thresholding estimator $\hat{f}$ by

$$\hat{f}(x) = \sum_{\ell \in \mathcal{I}_0} \hat{\alpha}_{0,\ell} \phi_0(x; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \hat{\beta}_{k,\ell} \mathbb{1}_{\left\{ |\hat{\beta}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}} \right\}} \psi_k(x; \eta_{k,\ell}), \tag{3.2}$$

$x \in S$, where

$$\hat{\alpha}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \phi_k(X_i; \zeta_{k,\ell}), \qquad \hat{\beta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \psi_k(X_i; \eta_{k,\ell}), \tag{3.3}$$

κ is a large enough constant and m_n is the integer satisfying

$$\frac{1}{2} \frac{n}{\ln n} < |\mathcal{J}_{m_n}| \leq \frac{n}{\ln n}.$$

The construction of $\hat{f}$ consists in three steps: firstly, we estimate the unknown wavelet coefficients (2.1) of f by (3.3), secondly, we select only the "greatest" $\hat{\beta}_{k,\ell}$ via a hard thresholding and thirdly we reconstruct the selected elements of the initial wavelet basis. The choices of the threshold $\kappa(\ln n/n)^{1/2}$ (corresponding to the "universal threshold") and the definition of m_n are based on theoretical considerations (see Theorem 3.1 below).

Note that $\hat{f}$ is adaptive, i.e., it does not depend on the knowledge of the smoothness of f. It can be viewed as an adaptive and thresholded version of the linear wavelet estimator proposed by [24]

Details on the wavelet hard thresholding estimator for $H = \mathbb{L}_p([a,b])$ and the standard nonparametric models can be found in [11], [9], [20] and [32].

3.3. Results

Theorem 3.1 below evaluates the performance of $\hat{f}$ assuming that f belongs to the decomposition spaces described in Subsection 2.2.

Theorem 3.1. *Consider the density estimation problem described in Subsection 3.1. Suppose that $\mathcal{E}$ satisfies (2.2) and (2.3). Let $\hat{f}$ be given by (3.2). Suppose that f satisfies (3.1) and, for any $\theta \in (0,1)$, $f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$, where $\mathcal{B}_\infty^{\theta/2}(H)$ is (2.4) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$ (2.5) with $r = 2(1-\theta)$. Then there exists a constant $C > 0$ such that*

$$E(||\hat{f} - f||^2) \le C \left(\frac{\ln n}{n}\right)^\theta$$

for n large enough.

An immediate consequence is the following upper bound result: if $f \in \mathcal{B}_\infty^{s/(2s+1)}(H) \cap \mathcal{W}^{2/(2s+1)}(H)$ for $s > 0$, then there exists a constant $C > 0$ such that

$$E(||\hat{f} - f||^2) \le C \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

This rate of convergence corresponds to the near optimal one in the "standard" minimax setting (see, e.g., [20]).

Moreover, applying [21, Theorem 3.2], one can prove that $\mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$ is the "maxiset" associated to $\hat{f}$ at the rate of convergence $(\ln n/n)^\theta$, i.e.,

$$\lim_{n\to\infty} \left(\frac{n}{\ln n}\right)^\theta E(||\hat{f} - f||^2) < \infty \Leftrightarrow f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H).$$

These new theoretical results complete the work of [24] in the sense that our wavelet-based procedure is adaptive thanks to its term-by-term selection of the $\hat{\beta}_{k,\ell}$ and we prove that it achieves a suitable rate of convergence over a wide class of functions well adapted to our setting.

The next section considers another statistical problem: the the regression estimation for functional data. We show how adapt our wavelet methodology to this problem.

4. REGRESSION ESTIMATION FOR FUNCTIONAL DATA

4.1. Problem statement

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $\{(X_i, Y_i);\ i \geq 1\}$ be i.i.d. replica of a couple of random variables (X, Y) defined on $\{\Omega, \mathcal{F}, P\}$, where Y is real valued and X takes values in a complete separable metric space or a Hilbert space S associated with the corresponding Borel σ-algebra $\mathcal{B}$ such that

$$Y = f(X) + \epsilon, \tag{4.1}$$

f denotes an unknown regresion function and ϵ is a random variable independent of X with $\epsilon \sim \mathcal{N}(0, 1)$. We suppose that $f \in H$ where H is a separable Hilbert space of real-valued functions defined on S. Let P_X be the probability measure induced by X_1 on $(S, \mathcal{B})$. Suppose that there exists a σ-finite measure ν on the measurable space $(S, \mathcal{B})$ such that P_X is dominated by ν. As a consequence of the Radon-Nikodym theorem, there exists a nonnegative measurable function g such that

$$P_X(B) = \int_B g(x)\nu(dx), \qquad B \in \mathcal{B}.$$

We suppose that g is known.

In this context, we want to estimate f from $(X_1, Y_1), ..., (X_n, Y_n)$.

The kernel estimator of the regression function for functional data has been proposed by [13] and the convergence in mean square of that estimator has been established by [15] with the rate $\mathcal{O}(h^2 + (nP(X \in \mathcal{B}(x, h))^{-1})$ where h is the bandwidth. Note that the optimal choice of h depends on the underlying unknown distribution.

We shall suppose that there exist two known constants $C_f > 0$ and $c_g > 0$ such that

$$\sup_{x \in S} f(x) \leq C_f, \qquad \inf_{x \in S} g(x) \geq c_g. \tag{4.2}$$

4.2. Results

Theorem 4.1 below explores the performance of $\hat{f}$ assuming that f belongs to the decomposition spaces described in Subsection 2.2.

Theorem 4.1. *Consider the regression estimation problem described above. Suppose that $\mathcal{E}$ satisfies (2.2) and (2.3). Let $\hat{f}$ be as in (3.2) with*

$$\hat{\alpha}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\phi_k(X_i;\zeta_{k,\ell}), \qquad \hat{\beta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell}),$$

κ is a large enough constant and m_n is the integer satisfying

$$\frac{1}{2}\frac{n}{(\ln n)^2} < |\mathcal{J}_{m_n}| \leq \frac{n}{(\ln n)^2}.$$

Suppose that f and g satisfy (4.2) and, for any $\theta \in (0,1)$, $f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$, where $\mathcal{B}_\infty^{\theta/2}(H)$ is (2.4) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$ (2.5) with $r = 2(1-\theta)$. Then there exists a constant $C > 0$ such that

$$E(||\hat{f} - f||^2) \leq C\left(\frac{(\ln n)^2}{n}\right)^{\theta}$$

for n large enough.

Again, note that, if $f \in \mathcal{B}_\infty^{s/(2s+1)}(H) \cap \mathcal{W}^{2/(2s+1)}(H)$ for $s > 0$, then there exists a constant $C > 0$ such that

$$E(||\hat{f} - f||^2) \leq C\left(\frac{(\ln n)^2}{n}\right)^{2s/(2s+1)}.$$

This rate of convergence corresponds to the near optimal one in the "standard" minimax setting (see, e.g., [20]) up to an extra logarithmic term. To the best of our knowledge, Theorem 4.1 is first one studying an adaptive wavelet-based estimator for functional data in the nonparametric regression context.

CONCLUSION AND PERSPECTIVES

We construct an efficient and new adaptive estimator for an unknown function f belonging to a separable Hilbert space H. To reach this goal, we combine several existing techniques: the wavelet basis on H developed by [19], the hard thresholding rule introduced by [11] and some elements related to the maxiset approach proposed by [3]. Rates of convergence are determined under the mean integrated squared error on H over $\mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$. Perspectives of this work are

- To determine the optimal lower bounds over the considered spaces,
- To remove the logarithmic term by perhaps considering other thresholding techniques. Thanks to its performances in numerous *i.i.d.* nonparametric models, the block thresholding introduced by [2] is a good candidate.

- Consider the regression model (4.1) with an unknown g.

These aspects require further investigations that we leave for a future work.

5. PROOFS

In this section, C denotes any constant that does not depend on j, k and n. Its value may change from one term to another and may depends on ϕ or ψ.

Proof of Theorem 3.1: The proof of Theorem 3.1 is a consequence of [21, Theorem 3.1] with $c(n) = (\ln n/n)^{1/2}$, $\sigma_i = 1$, $r = 2$ and the following proposition.

Proposition 5.1. *For any $k \in \{0, ..., m_n\}$ and any $\ell \in \mathcal{I}_k$ or $\ell \in \mathcal{J}_k$, let $\alpha_{k,\ell}$ and $\beta_{k,\ell}$ be given by (2.1), and $\hat{\alpha}_{k,\ell}$ and $\hat{\beta}_{k,\ell}$ be given by (3.3). Then*

(i) *There exists a constant $C > 0$ such that*

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) \leq C\frac{\ln n}{n}.$$

(ii) *There exists a constant $C > 0$ such that*

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) \leq C\left(\frac{\ln n}{n}\right)^2.$$

(iii) *For $\kappa > 0$ large enough, there exists a constant $C > 0$ such that*

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq 2\left(\frac{\ln n}{n}\right)^2.$$

Let us now prove **(i)**, **(ii)** and **(iii)** of Proposition 5.1 (which corresponds to [21, (3.1) and (3.2) of Theorem 3.1]).

(i) We have

$$E(\hat{\alpha}_{k,\ell}) = E(\phi_k(X_1; \zeta_{k,\ell})) = \int_S f(x)\phi_k(x; \zeta_{k,\ell})\nu(dx) = \alpha_{k,\ell}. \tag{5.1}$$

So

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) = V(\hat{\alpha}_{k,\ell}) = \frac{1}{n}V(\phi_k(X_1; \zeta_{k,\ell})) \leq \frac{1}{n}E\left(|\phi_k(X_1; \zeta_{k,\ell})|^2\right).$$

It follows from (3.1), the fact that $\mathcal{E}$ is an orthonormal basis of H and (2.2) that

$$\begin{aligned} E\left(|\phi_k(X_1;\zeta_{k,\ell})|^2\right) &= \int_S |\phi_k(x;\zeta_{k,\ell})|^2 f(x)\nu(dx) \\ &\le C_f \int_S |\phi_k(x;\zeta_{k,\ell})|^2 \nu(dx) \\ &= C_f \int_S \left| \sum_{j\in\mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} e_j(\zeta_{k,\ell}) e_j(x) \right|^2 \nu(dx) \\ &= C_f \sum_{j\in\mathcal{I}_k} \frac{1}{g_{j,k,\ell}} |e_j(\zeta_{k,\ell})|^2 \le C_f C_1. \end{aligned} \tag{5.2}$$

Therefore there exists a constant $C > 0$ such that

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) \le C\frac{1}{n} \le C\frac{\ln n}{n}.$$

(ii) Proceeding as in (5.1), we show that $E\left(\psi_k(X_i;\eta_{k,\ell})\right) = \beta_{k,\ell}$. Hence

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) = \frac{1}{n^4} E\left(\left| \sum_{i=1}^{n} U_{i,k,\ell} \right|^4 \right), \tag{5.3}$$

where

$$U_{i,k,\ell} = \psi_k(X_i;\eta_{k,\ell}) - E(\psi_k(X_i;\eta_{k,\ell})).$$

We will bound this last term via the Rosenthal inequality (recalled in the Appendix).

We have $E(U_{1,k,\ell}) = 0$.

By the Hölder inequality and (5.2) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we have

$$E(|U_{1,k,\ell}|^2) \le CE\left(|\psi_k(X_1;\eta_{k,\ell})|^2\right) \le C. \tag{5.4}$$

Let us now investigate the bound of $E(|U_{1,k,\ell}|^4)$. Observe that, thanks to the Cauchy-Schwarz inequality, (2.2) and (2.3), we have

$$\begin{aligned} \sup_{x\in S} |\psi_k(x;\eta_{k,\ell})| &\le \sup_{x\in S} \sum_{j\in\mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} |e_j(\eta_{k,\ell})||e_j(x)| \\ &\le \left(\sum_{j\in\mathcal{J}_k} \frac{1}{h_{j,k,\ell}} |e_j(\eta_{k,\ell})|^2 \right)^{1/2} \left(\sup_{x\in S} \sum_{j\in\mathcal{J}_k} |e_j(x)|^2 \right)^{1/2} \\ &\le C_1^{1/2} C_2^{1/2} \sqrt{|\mathcal{J}_k|} \le C\sqrt{|\mathcal{J}_{m_n}|} \le C\sqrt{\frac{n}{\ln n}}. \end{aligned} \tag{5.5}$$

The Hölder inequality, (5.5) and (5.4) yield

$$E(|U_{1,k,\ell}|^4) \leq CE\left(|\psi_k(X_1;\eta_{k,\ell})|^4\right) \leq CnE\left(|\psi_k(X_1;\eta_{k,\ell})|^2\right) \leq Cn. \tag{5.6}$$

It follows from the Rosenthal inequality, (5.4) and (5.6) that

$$\begin{aligned} \frac{1}{n^4}E\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right|^4\right) &\leq C\frac{1}{n^4}\max\left(nE\left(|U_{1,k,\ell}|^4\right), \left(nE\left(|U_{1,k,\ell}|^2\right)\right)^2\right) \\ &\leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2. \end{aligned} \tag{5.7}$$

By (5.3) and (5.7), we prove the existence of a constant $C > 0$ such that

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) \leq C\left(\frac{\ln n}{n}\right)^2.$$

(iii) We adopt the same notation as in **(ii)**. Observe that

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) = P\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right| \geq n\frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right). \tag{5.8}$$

We will bound this probability via the Bernstein inequality (recalled in the Appendix).

We have $E(U_{1,k,\ell}) = 0$.

By (5.5),

$$|U_{1,k,\ell}| \leq C\sup_{x\in S}|\psi_k(x;\eta_{k,\ell})| \leq C\sqrt{\frac{n}{\ln n}}.$$

Applying (5.2) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we obtain $E(|U_{1,k,\ell}|^2) \leq C$.

It follows from the Bernstein inequality that

$$P\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right| \geq n\frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq 2\exp\left(-\frac{Cn^2\kappa^2\frac{\ln n}{n}}{n + n\kappa\sqrt{\frac{\ln n}{n}}\sqrt{\frac{n}{\ln n}}}\right) \leq 2n^{-w(\kappa)}, \tag{5.9}$$

where

$$w(\kappa) = \frac{C\kappa^2}{1+\kappa}.$$

Since $\lim_{\kappa\to\infty} w(\kappa) = \infty$, combining (5.17) and (5.19), and taking κ such that $w(\kappa) = 2$, we have

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2.$$

The points **(i)**, **(ii)** and **(iii)** of Proposition 5.1 are proved. The proof of Theorem 3.1 is complete. □

Proof of Theorem 4.1: As in the proof of Theorem 3.1, we only need to prove **(i)**, **(ii)** and **(iii)** of Proposition 5.1.

(i) Since X_1 and ϵ_1 are independent and $E(\epsilon_1) = 0$, we have

$$
\begin{aligned}
E(\hat{\alpha}_{k,\ell}) &= E\left(\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) = E\left(\frac{f(X_1)}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) \\
&= \int_S \frac{f(x)}{g(x)}\phi_k(x;\zeta_{k,\ell})g(x)\nu(dx) = \alpha_{k,\ell}. \qquad (5.10)
\end{aligned}
$$

So

$$
\begin{aligned}
E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) &= V\left(\hat{\alpha}_{k,\ell}\right) = \frac{1}{n}V\left(\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) \\
&\leq \frac{1}{n}E\left(\left|\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right|^2\right).
\end{aligned}
$$

It follows from (4.2), $|Y_1| \leq C_f + |\epsilon_1|$, $g(X_1) \geq c_g$, the independence between X_1 and ϵ_1, $E(\epsilon_1^2) = 1$, the fact that $\mathcal{E}$ is an orthonormal basis of H and (2.2) that

$$
\begin{aligned}
E\left(\left|\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right|^2\right) &\leq (C_f^2+1)\frac{1}{c_g}E\left(|\phi_k(X_1;\zeta_{k,\ell})|^2\frac{1}{g(X_1)}\right) \\
&= (C_f^2+1)\frac{1}{c_g}\int_S |\phi_k(x;\zeta_{k,\ell})|^2\frac{1}{g(x)}g(x)\nu(dx) \\
&= C\int_S |\phi_k(x;\zeta_{k,\ell})|^2\nu(dx) \\
&= C\int_S \left|\sum_{j\in\mathcal{I}_k}\frac{1}{\sqrt{g_{j,k,\ell}}}e_j(\zeta_{k,\ell})e_j(x)\right|^2\nu(dx) \\
&= C\sum_{j\in\mathcal{I}_k}\frac{1}{g_{j,k,\ell}}|e_j(\zeta_{k,\ell})|^2 \leq C. \qquad (5.11)
\end{aligned}
$$

Therefore there exists a constant $C > 0$ such that

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) \leq C\frac{1}{n} \leq C\frac{\ln n}{n}.$$

(ii) Proceeding as in (5.10), we show that $E\left(Y_i\psi_k(X_i;\eta_{k,\ell})/g(X_i)\right) = \beta_{k,\ell}$. Set

$$U_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell}) - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\right).$$

and observe that

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) = \frac{1}{n^4}E\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right|^4\right). \qquad (5.12)$$

We will bound this last term via the Rosenthal inequality (recalled in the Appendix).

We have $E(U_{1,k,\ell}) = 0$.

By the Hölder inequality and (5.11) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we obtain

$$E(|U_{1,k,\ell}|^2) \leq CE\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right|^2\right) \leq C. \tag{5.13}$$

Let us now investigate the bound of $E(|U_{1,k,\ell}|^4)$. Observe that, thanks to the Cauchy-Schwarz inequality, (2.2) and (2.3), we have

$$\begin{aligned}
\sup_{x\in S}|\psi_k(x;\eta_{k,\ell})| &\leq \sup_{x\in S}\sum_{j\in\mathcal{J}_k}\frac{1}{\sqrt{h_{j,k,\ell}}}|e_j(\eta_{k,\ell})||e_j(x)| \\
&\leq \left(\sum_{j\in\mathcal{J}_k}\frac{1}{h_{j,k,\ell}}|e_j(\eta_{k,\ell})|^2\right)^{1/2}\left(\sup_{x\in S}\sum_{j\in\mathcal{J}_k}|e_j(x)|^2\right)^{1/2} \\
&\leq C_1^{1/2}C_2^{1/2}\sqrt{|\mathcal{J}_k|} \leq C\sqrt{|\mathcal{J}_{m_n}|} \leq C\sqrt{\frac{n}{(\ln n)^2}}. \qquad (5.14)
\end{aligned}$$

The Hölder inequality, (5.14) and (5.13) yield

$$\begin{aligned}
E(|U_{1,k,\ell}|^4) &\leq CE\left(|\psi_k(X_1;\eta_{k,\ell})|^4\right) \leq CnE\left(|\psi_k(X_1;\eta_{k,\ell})|^2\right) \\
&\leq Cn. \qquad (5.15)
\end{aligned}$$

It follows from the Rosenthal inequality, (5.13) and (5.15) that

$$\begin{aligned}
\frac{1}{n^4}E\left(\left|\sum_{i=1}^{n}U_{i,k,\ell}\right|^4\right) &\leq C\frac{1}{n^4}\max\left(nE\left(|U_{1,k,\ell}|^4\right), \left(nE\left(|U_{1,k,\ell}|^2\right)\right)^2\right) \\
&\leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2. \qquad (5.16)
\end{aligned}$$

By (5.12) and (5.16), we prove the existence of a constant $C > 0$ such that

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) \leq C\left(\frac{\ln n}{n}\right)^2.$$

(iii) We adopt the same notation as in **(ii)**. Since $E\left(Y_i\psi_k(X_i;\eta_{k,\ell})/g(X_i)\right) = \beta_{k,\ell}$, we can write

$$U_{i,k,\ell} = V_{i,k,\ell} + W_{i,k,\ell},$$

where

$$V_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i} - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i}\right),$$

$$W_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i^c} - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i^c}\right),$$

$$\mathcal{A}_i = \left\{|\epsilon_i| \geq c_*\sqrt{\ln n}\right\}$$

and c_* denotes a constant which will be chosen later.

We have

$$\begin{aligned} P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) &= P\left(\left|\sum_{i=1}^{n} U_{i,k,\ell}\right| \geq n\frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \\ &\leq I_1 + I_2, \end{aligned} \tag{5.17}$$

where

$$I_1 = P\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right| \geq \frac{\kappa}{4}\sqrt{n\ln n}\right)$$

and

$$I_2 = P\left(\left|\sum_{i=1}^{n} W_{i,k,\ell}\right| \geq \frac{\kappa}{4}\sqrt{n\ln n}\right).$$

Let us now bound I_1 and I_2.

Upper bound for I_1. The Markov inequality and the Cauchy-Schwarz inequality yield

$$\begin{aligned} I_1 &\leq \frac{4}{\kappa\sqrt{n\ln n}} E\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right|\right) \leq C\sqrt{\frac{n}{\ln n}} E(|V_{1,k,\ell}|) \\ &\leq C\sqrt{\frac{n}{\ln n}} E\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right| \mathbb{1}_{\mathcal{A}_1}\right) \\ &\leq C\sqrt{\frac{n}{\ln n}} \left(E\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right|^2\right)\right)^{1/2} (P(\mathcal{A}_1))^{1/2}. \end{aligned}$$

Using (5.13), an elementary Gaussian inequality and taking c_* large enough, we obtain

$$I_1 \leq C\sqrt{\frac{n}{\ln n}} e^{-c_*^2 \ln n/4} \leq C\frac{1}{n^2}. \tag{5.18}$$

Upper bound for I_2. We will bound this probability via the Bernstein inequality (recalled in the Appendix).

We have $E(W_{1,k,\ell}) = 0$.

Using (4.2) which implies $|Y_1 \mathbb{1}_{\mathcal{A}_1^c}| \leq C_f + c_*\sqrt{\ln n} \leq C\sqrt{\ln n}$ and $g(X_1) \geq c_g$, and (5.14), we obtain

$$|W_{i,k,\ell}| \leq C\sqrt{\ln n} \sup_{x\in S} |\psi_k(x;\eta_{k,\ell})| \leq C\sqrt{\ln n}\sqrt{\frac{n}{(\ln n)^2}} = C\sqrt{\frac{n}{\ln n}}.$$

Applying (5.11) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we obtain $E(|W_{1,k,\ell}|^2) \leq C$.

It follows from the Bernstein inequality that

$$I_2 \leq 2\exp\left(-\frac{Cn^2\kappa^2\frac{\ln n}{n}}{n + n\kappa\sqrt{\frac{\ln n}{n}}\sqrt{\frac{n}{\ln n}}}\right) \leq 2n^{-w(\kappa)}, \tag{5.19}$$

where

$$w(\kappa) = \frac{C\kappa^2}{1+\kappa}.$$

Since $\lim_{\kappa\to\infty} w(\kappa) = \infty$, taking κ such that $w(\kappa) = 2$, we have

$$I_2 \leq 2\frac{1}{n^2}.$$

It follows from (5.17), (5.18) and (5.19) that

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2.$$

Hence the points **(i)**, **(ii)** and **(iii)** of Proposition 5.1 are satisfied by our estimators. The proof of Theorem 4.1 is complete. □

APPENDIX

Here we state the two inequalities that have been used for proving the results in earlier section.

Lemma A.1 ([29]). *Let n be a positive integer, $p \geq 2$ and $V_1, ..., V_n$ be n zero mean i.i.d. random variables such that $E(|V_1|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$E\left(\left|\sum_{i=1}^{n} V_i\right|^p\right) \leq C \max\left(nE(|V_1|^p), n^{p/2}\left(E(V_1^2)\right)^{p/2}\right).$$

Lemma A.2 ([23]). *Let n be a positive integer and $V_1, ..., V_n$ be n i.i.d. zero mean random variables such that there exists a constant $M > 0$ satisfying $|V_1| \leq M < \infty$. Then, for any $\upsilon > 0$,*

$$P\left(\left|\sum_{i=1}^{n} V_i\right| \geq \upsilon\right) \leq 2\exp\left(-\frac{\upsilon^2}{2\left(nE(V_1^2) + \upsilon M/3\right)}\right).$$

ACKNOWLEDGMENTS

The authors thank two anonymous referees for their thorough and useful comments.

REFERENCES

[1] Autin, F. (2004). *Point de vue Maxiset en estimation non paramétrique*, PhD thesis, Université Paris 7. (tel-00008542).

[2] Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898–924.

[3] Cohen, A.; De Vore, R.; Kerkyacharian, G. and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms, *Appl. Comput. Harmon. Anal.*, **11**, 167–191.

[4] Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie: application aux diffusions, *C. R. Math. Acad. Sci. Paris*, **334**(3), 213–216.

[5] Dabo-Niang, S. (2004). Density estimation by orthogonal series in an infinite dimensional space: application to processes of diffusion type I, *J. Nonparametr. Stat.*, **16**(1–2), 171–186.

[6] Dabo-Niang, S. (2006). Mode estimation for functional random variable and its application for curves classification, *Far East J. Theor. Stat.*, **18**(1), 93–119.

[7] Dabo-Niang, S. and Ferraty, F. (Eds) (2008). *Functional and Operatorial Statistics*, Contributions to Statistics, Physica-Verlag, Heidelberg.

[8] Dalecky, Yu.L. and Fomin, S.V. (1991). *Measures and Differential Equations in Infinite-Dimensional Space*, Dordrecht, Kluwer Academic Publishers Group.

[9] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, **3**, 215–228.

[10] DeVore, R. (1998). Non linear approximation, *Acta Numerica*, **7**, 51–150.

[11] Donoho, D.; Johnstone, I.; Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, **24**(2), 508–539.

[12] Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, *C. R. Acad. Sci. Paris Sér. I Math.*, **330**(2), 139–142.

[13] Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination, *J. Nonparametr. Stat.*, **16**(1–2), 111–125.

[14] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice*, Springer-Verlag, New York.

[15] Ferraty, F.; Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects, *Aust. N. Z. J. Stat.*, **49**, 267–286.

[16] Ferraty, F. (Ed) (2010). Special Issue: Statistical Methods and Problems in Infinite-dimensional Spaces, 1st International Workshop on Functional and Operatorial Statistics (IWFOS'2008), *J. Multivariate Anal.*, **101**(2).

[17] Ferraty, F. and Romain, Y. (2010). *The Oxford Handbook of Functional Data Analysis*, Oxford University Press.

[18] Ferraty, F. (Ed) (2011). *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, Physica-Verlag, Heidelberg.

[19] Goh, S.S. (2007). Wavelet bases for Hilbert spaces of functions, *Complex Variables and Elliptic Equations*, **52**, 245–260.

[20] Härdle, W.; Kerkyacharian, G.; Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, 129, Springer-Verlag, New York.

[21] Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, **9**(2), 283–345.

[22] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065Ű-1076.

[23] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

[24] Prakasa Rao, B.L.S. (2010a). Nonparametric density estimation for functional data via wavelets, *Communications in Statistics — Theory and Methods*, **39**(8–9), 1608–1618.

[25] Prakasa Rao, B.L.S. (2010b). Nonparametric density estimation for functional data by delta sequences, *Brazilian Journal of Probability and Statistics*, **24**(3), 468–478.

[26] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*, New York, Springer.

[27] Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, New York, Springer.

[28] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832–837.

[29] Rosenthal, H.P. (1970). On the subspaces of $\mathbb{L}^p$ $(p \geq 2)$ spanned by sequences of independent random variables, *Israel Journal of Mathematics*, **8**, 273–303.

[30] Temlyakov, V.N. (1998). The best m-term approximation and greedy algorithms, *Adv. Comput. Math.*, **8**, 249–265.

[31] Uglanov, A.V. (2000). *Integration on Infinite-Dimensional Surfaces and its Applications*, Dordrecht, Kluwer Academic Publishers.

[32] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

[33] Xia, D.X. (1972). *Measure and Integration Theory on Infinite-Dimensional Spaces. Abstract Harmonic Analysis*, New York, Academic Press.

[34] Yamasaki, Y. (1985). *Measures on Infinite-Dimensional Spaces*, Series in Pure Mathematics 5, World Scientific, Singapore.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, I.P., in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal's website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.