TWMS J. App. and Eng. Math. V.15, N.5, 2025, pp. 1100-1114

ESTIMATION AND TESTS WITH BIVARIATE CENSORED DATA

M. BOUKELOUA[1]*, §

Abstract. In this work, we study the asymptotic properties of the maximum likelihood estimator, for bivariate right censored data. We also propose a generalization of the likelihood ratio test and the chi-square test of fit, for this type of data. Finally, we illustrate the performances of our proposed tests through a simulation study and a real data application.

Keywords: Bivariate censored data, maximum likelihood, likelihood ratio test, test of fit, asymptotic distributions.

AMS Subject Classification: 62F12, 62F05, 62N01

1. Introduction

In survival analysis, a phenomenon of censorship often prevents the observation of the lifetime of interest, and provides only a partial information about it. There exist several kinds of censorship, but we focus on bivariate right censored data. Many practical situations require the study of a bivariate lifetime. For instance, the study of the lifetime of twins, the lifetime of spouses having subscribed to a pension contract and the operating time of a system comprised of pairs of components. Concerning matched data, we quote for example, the study of the recovery time for diseases affecting eyes, lungs, ears and kidneys. This field of research is very active, the nonparametric approaches have been developed by many authors. [1] introduced three estimates of the bivariate survival function, under right censoring and they established a law of the iterated logarithm for theses estimates. They also showed that they are asymptotically equivalent in the sense that their difference is $o(n^{-1})$ almost surely (a.s.). [2] established their weak convergence to a centered Gaussian process. Otherwise, [3] used one of these estimates to construct a kernel estimator for the density function. Other estimators of the survival function have been proposed in

[1] Laboratoire de Génie des Procédés pour le Développement Durable et les Produits de Santé (LGPDDPS)- Ecole Nationale Polytechnique de Constantine - Algeria.
Laboratoire de Biostatistique, Bioinformatique et Méthodologie Mathématique Appliquées aux Sciences de la Santé (BIOSTIM)- Faculté de Médecine - Université Salah Boubnider Constantine 3 - Algeria.
e-mail: boukeloua.mohamed@gmail.com, mohamed.boukeloua@enp-constantine.dz; ORCID: https://orcid.org/http://orcid.org/0000-0002-5522-498X.
∗ Corresponding author.
§ Manuscript received: December 28, 2023; accepted: May 30, 2024.
TWMS Journal of Applied and Engineering Mathematics, Vol.15, No.5;

the literature, let us cite for example the works of [4, 5], [6], [7] and [8]. Other authors have been interested in semiparametric models, copula models and statistical tests with bivariate censored data, such as [9], [10], [11] and [12].

In the present paper, we study a parametric model with bivariate censored data. In particular, we are interested in the estimation of the parameter of the model, by the maximum likelihood method. At first, we give the explicit form of the likelihood function for the considered model. Then, we establish the weak consistency and the asymptotic normality of the maximum likelihood estimator. This asymptotic study leads to the generalization of the likelihood ratio test to the case of bivariate censored data. We conclude our theoretical study by proposing a chi-square goodness-of-fit test, for bivariate censored data. For that, we were inspired by the paper of [13], who investigated the same test in the univariate case. The performances of our tests is then assessed using a simulation study and an application on a real dataset.

2. The likelihood function for bivariate censored data

Let $(\Omega, \mathcal{A}, P)$ be a probability space and (X, Y) be a pair of positive random variables, with distribution function F, survival function S and continuous density function $f_{X,Y}$. We consider the case of bivariate censored data, this means that instead of observing (X, Y), one can only observe the vector $(U, V, \Delta_1, \Delta_2)$, where $U = \min(X, C)$, $V = \min(Y, D)$, $\Delta_1 = 1_{\{X \leq C\}}$ and $\Delta_2 = 1_{\{Y \leq D\}}$, ($1_{\{.\}}$ denotes the indicator function). The positive pair (C, D) represents the censoring variables and it is independent from the pair (X, Y). We denote its distribution function by G, its survival function by $\overline{G}$ and its density function (assumed to be continuous) by $f_{C,D}$. We assume that the survival function S belongs to a parametric family $\{S(\cdot, \cdot; \theta)/\theta \in \Theta\}$, where Θ is a compact set of $\mathbb{R}^d$. The function $S(\cdot, \cdot; \theta)$ can be seen as the survival function of the pushforward measure under (X, Y), of a probability measure P_θ defined on the measurable space $(\Omega, \mathcal{A})$. We assume that the model is identifiable, in other words

$$\forall \theta_1, \theta_2 \in \Theta : \theta_1 \neq \theta_2 \Rightarrow S(\cdot, \cdot; \theta_1) \neq S(\cdot, \cdot; \theta_2)$$

and that G does not depend on θ. We denote by θ_T the true value of θ.

In the sequel, for any random variable T, f_T denotes the probability density of T. We also denote by $\xrightarrow{P}$ (resp. $\xrightarrow{\mathcal{D}}$) the convergence in probability (resp. in distribution).

Let $(U_i, V_i, \Delta_{1i}, \Delta_{2i})_{1 \leq i \leq n}$ be a sample of independent copies of the vector $(U, V, \Delta_1, \Delta_2)$. The likelihood function of $(U, V, \Delta_1, \Delta_2)$ is given by

$$L(\theta) = \prod_{i=1}^{n} f_{U,V,\Delta_1,\Delta_2}(u_i, v_i, \delta_{1i}, \delta_{2i}; \theta),$$

where $f_{U,V,\Delta_1,\Delta_2}$ is the density of $(U, V, \Delta_1, \Delta_2)$ with respect to the measure $\lambda^{\otimes 2} \otimes \mu^{\otimes 2}$, λ (resp. μ) being the Lebesgue measure (resp. the counting measure) on $\mathbb{R}$ (resp. on $\{0, 1\}$) and $(u_i, v_i, \delta_{1i}, \delta_{2i})$ is a realization of $(U_i, V_i, \Delta_{1i}, \Delta_{2i})$.

Studying the different possible cases corresponding to the values assumed by the pair

$(\delta_{1i}, \delta_{2i})$, we get

$$\begin{aligned} L(\theta) = \prod_{i=1}^{n} \left(f_{X,Y}(u_i, v_i; \theta)\overline{G}(u_i, v_i)\right)^{\delta_{1i}\delta_{2i}} \left(f_X(u_i;\theta) - \frac{\partial}{\partial u} F(u_i, v_i; \theta)\right)^{\delta_{1i}(1-\delta_{2i})} \\ \times \left(f_D(v_i) - \frac{\partial}{\partial v} G(u_i, v_i)\right)^{\delta_{1i}(1-\delta_{2i})} \left(f_Y(v_i;\theta) - \frac{\partial}{\partial v} F(u_i, v_i; \theta)\right)^{\delta_{2i}(1-\delta_{1i})} \\ \times \left(f_C(u_i) - \frac{\partial}{\partial u} G(u_i, v_i)\right)^{\delta_{2i}(1-\delta_{1i})} \left(S(u_i, v_i; \theta) f_{C,D}(u_i, v_i)\right)^{(1-\delta_{1i})(1-\delta_{2i})}. \end{aligned}$$

Given that we are interested in the estimation of the parameter θ, we will consider only the functions that depend on θ. So, we study the following pseudo-likelihood function

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} h(u_i, v_i, \delta_{1i}, \delta_{2i}; \theta),$$

where

$$\begin{aligned} h(u_i, v_i, \delta_{1i}, \delta_{2i}; \theta) = \left(f_{X,Y}(u_i, v_i; \theta)\right)^{\delta_{1i}\delta_{2i}} \left(f_X(u_i;\theta) - \frac{\partial}{\partial u} F(u_i, v_i; \theta)\right)^{\delta_{1i}(1-\delta_{2i})} \\ \times \left(f_Y(v_i;\theta) - \frac{\partial}{\partial v} F(u_i, v_i; \theta)\right)^{\delta_{2i}(1-\delta_{1i})} S(u_i, v_i; \theta)^{(1-\delta_{1i})(1-\delta_{2i})}, \end{aligned}$$

with support $A = \{(u, v, \delta_1, \delta_2) \in \mathbb{R}^2 \times \{0,1\}^2 / h(u, v, \delta_1, \delta_2; \theta) > 0\}$ assumed to be independent of θ.
We estimate θ by the maximum likelihood estimator, defined by

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} l(\theta),$$

where $l(\theta) = \log(\mathcal{L}(\theta))$.
Note that the maximum likelihood method exhibits sometimes some computational complications. [14] proposed some solutions to overcome these complications for the Dirichlet distribution, which can be used for other distributions. Among these solutions, we cite the methods of the choice of the starting values in the numerical calculations such as the method of moments, the method of Ronning ([15]), the method of Dishon ([16]) and the method of Wicker ([17]). [14] also proposed other solutions such as the method of re-parametrization and a stable algorithm based on the Levenberg-Marquardt algorithm (see [18] and [19]) with a damping parameter.

3. Asymptotic study of the maximum likelihood estimator

In this section, we establish the weak consistency and the asymptotic normality of $\widehat{\theta}_n$, under the following hypotheses.

H1: The function $h(u, v, \delta_1, \delta_2; .)$ is continuous on Θ for all $(u, v, \delta_1, \delta_2) \in A$.

H2: There exists a function $\psi : \mathbb{R}^2 \times \{0,1\}^2 \longrightarrow \mathbb{R}$, such that $\forall \theta \in \Theta$, $|\log h(U, V, \Delta_1, \Delta_2; \theta)| \le \psi(U, V, \Delta_1, \Delta_2)$ $a.s.$ and $\mathbb{E}_{\theta_T}(\psi(U, V, \Delta_1, \Delta_2)) < \infty$, where $\mathbb{E}_{\theta_T}(.)$ denotes the expectation under P_{θ_T}.

H3: There exists a compact neighborhood N of θ_T, included in Θ, such that the function $h(u, v, \delta_1, \delta_2; .)$ is twice continuously differentiable on N, for every $(u, v, \delta_1, \delta_2) \in A$.

H4: $\forall \theta \in N$, $\forall i, j \in \{1, ..., d\}$, we have $\left|\frac{\partial}{\partial \theta_i} \log h(U, V, \Delta_1, \Delta_2; \theta)\right| \le \psi(U, V, \Delta_1, \Delta_2)$ $a.s.$ and $\left|\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log h(U, V, \Delta_1, \Delta_2; \theta)\right| \le \psi(U, V, \Delta_1, \Delta_2)$ $a.s.$

H5: The Fisher information matrix $I(\theta_T) = (I_{i,j}(\theta_T))_{1\leq i,j\leq d}$ exists, where
$I_{i,j}(\theta_T) = \mathbb{E}_{\theta_T}\left(\frac{\partial}{\partial\theta_i} \log h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T} \times \frac{\partial}{\partial\theta_j} \log h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T}\right).$

H6: The matrix $J(\theta_T)$ in nonsingular, where $J(\theta) = (J_{i,j}(\theta))_{1\leq i,j\leq d}$ with
$J_{i,j}(\theta) = \mathbb{E}_{\theta_T}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log h(U,V,\Delta_1,\Delta_2;\theta)\right),\ \forall\theta\in N.$

Before stating our results, we will give some examples of models that satisfy these hypotheses.

Example 3.1. *We assume that X has an exponential distribution with parameter a (denoted by $\mathcal{E}(a)$) and that given $X = x$, Y has an $\mathcal{E}(bx)$ distribution, where $a \in \Theta_1$ and $b \in \Theta_2$, Θ_1 and Θ_2 being two compact sets included in $]0,+\infty[$. The parameter of interest is $\theta = (a,b)^\top \in \Theta = \Theta_1 \times \Theta_2$. The density function of (X,Y) is then given by*

$$f_{X,Y}(x,y;\theta) = abxe^{-ax-bxy}, \forall(x,y) \in]0,+\infty[\times]0,+\infty[. \qquad (1)$$

Concerning the censoring variables, we assume that C has an $\mathcal{E}(a_C)$ distribution and that given $C = x$, D has an $\mathcal{E}(b_{C,D}x)$ distribution, where $a_C \in \Theta_1$ and $b_{C,D} \in \Theta_2$.
We can check that this model satisfies the assumptions **H1–H6** *above.*

Example 3.2. *We assume that X (resp. Y) has an $\mathcal{E}(a)$ (resp. $\mathcal{E}(b)$) distribution and that they are related by a Clayton copula with parameter γ, where $a \in \Theta_1$, $b \in \Theta_2$ and $\gamma \in \Theta_3$, Θ_1, Θ_2 and Θ_3 being three compact sets included in $]0,+\infty[$. Recall that the Clayton copula with parameter γ is defined by*

$$C_\gamma(t_1,t_2) = \left(t_1^{-\gamma} + t_2^{-\gamma} - 1\right)^{-1/\gamma}, \forall(t_1,t_2) \in]0,1[\times]0,1[.$$

So, denoting by $F_X(x;a)$ (resp. $F_Y(y;b)$) the distribution function of X (resp. Y), we get

$$f_{X,Y}(x,y;\theta) = f_X(x;a)f_Y(y;b)c_\gamma(F_X(x;a),F_Y(y;b)), \forall(x,y) \in]0,+\infty[\times]0,+\infty[, \qquad (2)$$

where $\theta = (a,b,\gamma)^\top \in \Theta = \Theta_1 \times \Theta_2 \times \Theta_3$ is the parameter of interest and

$$c_\gamma(t_1,t_2) = \frac{\partial^2 C_\gamma}{\partial t_1 \partial t_2}(t_1,t_2) = (1+\gamma)t_1^{-\gamma-1}t_2^{-\gamma-1}\left(t_1^{-\gamma} + t_2^{-\gamma} - 1\right)^{-1/\gamma-2}$$

is the density of C_γ with respect to the Lebesgue measure on $[0,1]^2$.
Furthermore, we assume that C (resp. D) has an $\mathcal{E}(a_C)$ (resp. $\mathcal{E}(b_D)$) distribution and that they are related by a Clayton copula with parameter $\gamma_{C,D}$, where $a_C \in \Theta_1$, $b_D \in \Theta_2$ and $\gamma_{C,D} \in \Theta_3$.
We can check that this model satisfies the assumptions **H1–H6** *above.*

Now, we are in a position to state the results we seek.

Theorem 3.1. *Under hypotheses* **H1** *and* **H2**, *we have*

$$\widehat{\theta}_n \xrightarrow{P} \theta_T.$$

Proof. Set

$$M(\theta) = \mathbb{E}_{\theta_T}(\log h(U,V,\Delta_1,\Delta_2;\theta))$$

and

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^n \log h(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta).$$

Since $\widehat{\theta}_n$ maximize $M_n(\theta)$, we can apply Theorem 5.7 of [20]. To this end, we have to verify that $\sup_{\theta\in\Theta}|M_n(\theta) - M(\theta)| \xrightarrow{P} 0$ and $\sup\{M(\theta)/\theta \in \Theta, ||\theta - \theta_T|| \geq \varepsilon\} < M(\theta_T),\ \forall\varepsilon > 0$. Under **H1** and **H2**, Lemma 2.4 of [21] implies the first relation as well as the continuity of

$M(\theta)$. It remains to prove the second relation. Let us first show that θ_T maximize $M(\theta)$. Jensen inequality (see [22]) permits to write

$$\begin{aligned}
M(\theta) - M(\theta_T) &= \mathbb{E}_{\theta_T}\left(\log \frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)}\right) \\
&\leq \log \mathbb{E}_{\theta_T}\left(\frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)}\right) \\
&= \log\left(\int \frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)} 1_{\{\Delta_1=1,\Delta_2=1\}} dP_{\theta_T}\right. \\
&+ \int \frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)} 1_{\{\Delta_1=1,\Delta_2=0\}} dP_{\theta_T} \\
&+ \int \frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)} 1_{\{\Delta_1=0,\Delta_2=1\}} dP_{\theta_T} \\
&+ \left.\int \frac{h(U,V,\Delta_1,\Delta_2;\theta)}{h(U,V,\Delta_1,\Delta_2;\theta_T)} 1_{\{\Delta_1=0,\Delta_2=0\}} dP_{\theta_T}\right) \\
&= \log\left(\int\int f_{U,V,\Delta_1,\Delta_2}(u,v,1,1;\theta)dudv\right. \\
&+ \int\int f_{U,V,\Delta_1,\Delta_2}(u,v,1,0;\theta)dudv \\
&+ \int\int f_{U,V,\Delta_1,\Delta_2}(u,v,0,1;\theta)dudv \\
&+ \left.\int\int f_{U,V,\Delta_1,\Delta_2}(u,v,0,0;\theta)dudv\right) \\
&= \log(\varphi_{1,1}(\theta) + \varphi_{1,0}(\theta) + \varphi_{0,1}(\theta) + \varphi_{0,0}(\theta)) \\
&= \log(1) \\
&= 0,
\end{aligned}$$

where $\varphi_{i,j}(\theta) = P_\theta(\Delta_1 = i, \Delta_2 = j)$, $i, j \in \{0,1\}$. Therefore

$$M(\theta) \leq M(\theta_T),\ \forall \theta \in \Theta. \tag{3}$$

Furthermore, for any $\varepsilon > 0$, the continuity of $M(\theta)$ on the compact set $\Gamma = \{\theta \in \Theta / ||\theta - \theta_T|| \geq \varepsilon\}$ ensures the existence of $\tilde{\theta} \in \Gamma$ such that

$$\sup_{\theta\in\Gamma} M(\theta) = M(\tilde{\theta}) < M(\theta_T),$$

in view of the identifiability of the model and the strict concavity of the logarithmic function. □

Let us now move on to the asymptotic normality of $\widehat{\theta}_n$.

Theorem 3.2. *Under hypotheses* **H1–H6**, *we have*

$$\sqrt{n}(\widehat{\theta}_n - \theta_T) \xrightarrow{\mathcal{D}} \mathcal{N}(0, J(\theta_T)^{-1} I(\theta_T) J(\theta_T)^{-1}).$$

Proof. Set for all $\theta \in N$

$$D(U,V,\Delta_1,\Delta_2;\theta) = \left(\frac{\partial}{\partial\theta_1}\log h(U,V,\Delta_1,\Delta_2;\theta), ..., \frac{\partial}{\partial\theta_d}\log h(U,V,\Delta_1,\Delta_2;\theta)\right)^\top. \tag{4}$$

Under **H4**, we have for all $j \in \{1, ..., d\}$

$$\mathbb{E}_{\theta_T}\left(\frac{\partial}{\partial\theta_j}\log h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T}\right) = \frac{\partial}{\partial\theta_j}M(\theta)\Big|_{\theta=\theta_T} = 0,$$

since θ_T maximize $M(\theta)$ (see (3)). Thus, the central limit theorem gives us

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta_T) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\theta_T)). \tag{5}$$

Otherwise, using a Taylor expansion, there exists $\bar{\theta}$ inside the segment that links $\widehat{\theta}_n$ and θ_T such that

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\widehat{\theta}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta_T) \\ &+ J_n(\bar{\theta})\sqrt{n}(\widehat{\theta}_n - \theta_T), \end{aligned} \tag{6}$$

where $J_n(\theta) = (J_n^{j,k}(\theta))_{1\le j,k\le d}$ and

$$J_n^{j,k}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log h(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta).$$

Consider now the following decomposition

$$||J_n(\bar{\theta}) - J(\theta_T)|| \le ||J_n(\bar{\theta}) - J(\bar{\theta})|| + ||J(\bar{\theta}) - J(\theta_T)||.$$

Under **H3** and **H4**, we get in view of Lemma 2.4 of [21]

$$\sup_{\theta\in N}||J_n(\theta) - J(\theta)|| \xrightarrow{P} 0$$

as well as the continuity of $J(\theta)$. This latter combined with Theorem 3.1, gives us

$$||J(\bar{\theta}) - J(\theta_T)|| \xrightarrow{P} 0.$$

Hence $J_n(\bar{\theta}) \xrightarrow{P} J(\theta_T)$, and consequently $J_n(\bar{\theta})$ is nonsingular with probability approaching one. Therefore relation (6) entails

$$\begin{aligned} \sqrt{n}(\widehat{\theta}_n - \theta_T) &= -(J_n(\bar{\theta}))^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta_T) \\ &= -(J(\theta_T))^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\theta_T) + o_p(1) \end{aligned} \tag{7}$$

and the claimed result follows readily from relation (5). □

Remark 1. *Under the hypothesis*
H7 *For all* $i, j \in \{1, \ldots, d\}$

$$\int\int \frac{\partial^2}{\partial\theta_i\partial\theta_j} f_{U,V,\Delta_1,\Delta_2}(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T} d\lambda^{\otimes 2}(u,v)d\mu^{\otimes 2}(\delta_1,\delta_2) = 0,$$

we have

$$I(\theta_T) = -J(\theta_T).$$

Indeed, it suffices to take the expectation under P_{θ_T}*, of both sides of the following expression*

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T} = \frac{\frac{\partial^2}{\partial\theta_i\partial\theta_j} f_{U,V,\Delta_1,\Delta_2}(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T}}{f_{U,V,\Delta_1,\Delta_2}(U,V,\Delta_1,\Delta_2;\theta_T)}$$
$$- \frac{\frac{\partial}{\partial\theta_i} h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T}}{h(U,V,\Delta_1,\Delta_2;\theta_T)}$$
$$\times \frac{\frac{\partial}{\partial\theta_j} h(U,V,\Delta_1,\Delta_2;\theta)\Big|_{\theta=\theta_T}}{h(U,V,\Delta_1,\Delta_2;\theta_T)}.$$

Note that the models given in examples 3.1 and 3.2 above satisfy the hypothesis **H7***.*

4. Hypothesis tests

In the same context developed previously, we study in this section two statistical tests, namely a likelihood ratio test and a chi-square test of fit. In the first one, we test the value of the parameter and in the second one, we test the model.

4.1. Likelihood ratio test.

For a given value $\theta_0 \in \Theta$, consider the test problem of the hypothesis

$$\mathcal{H}_0 : \theta_T = \theta_0 \text{ against } \mathcal{H}_1 : \theta_T \neq \theta_0.$$

By analogy with the usual case of complete data, we propose the following test statistic.

$$\Phi_n = 2\log\left(\frac{\mathcal{L}(\widehat{\theta}_n)}{\mathcal{L}(\theta_0)}\right) = 2(l(\widehat{\theta}_n) - l(\theta_0)) \tag{8}$$

whose limiting distribution is given by the next theorem.

Theorem 4.1. *Under* $\mathcal{H}_0$ *and hypotheses* **H1–H7***, we have*

$$2(l(\widehat{\theta}_n) - l(\theta_0)) \xrightarrow{\mathcal{D}} \chi^2(d).$$

Let α be an asymptotic level. According to this theorem, the critical region of our test is

$$\left\{2(l(\widehat{\theta}_n) - l(\theta_0)) > q_{1-\alpha}\right\},$$

where $q_{1-\alpha}$ is the $(1-\alpha)-$quantile of the $\chi^2(d)$ distribution.

Proof. By a Taylor expansion, there exists $\bar{\theta}$ inside the segment that links $\widehat{\theta}_n$ and θ_0 such that

$$l(\theta_0) = l(\widehat{\theta}_n) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D(U_i,V_i,\Delta_{1i},\Delta_{2i};\widehat{\theta}_n)^\top \sqrt{n}(\theta_0 - \widehat{\theta}_n)$$
$$+ \frac{n}{2}(\theta_0 - \widehat{\theta}_n)^\top J_n(\bar{\theta})(\theta_0 - \widehat{\theta}_n).$$

Taking into account the fact that $J_n(\bar{\theta}) = J(\theta_0) + o_p(1)$, $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ (under $\mathcal{H}_0$) and that $\widehat{\theta}_n$ maximize $l(\theta)$, we obtain under the hypothesis **H7**

$$\begin{aligned} 2(l(\widehat{\theta}_n) - l(\theta_0)) &= -n(\widehat{\theta}_n - \theta_0)^\top J(\theta_0)(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &= n(\widehat{\theta}_n - \theta_0)^\top I(\theta_0)(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top BB^\top \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &= \sqrt{n}(B^\top(\widehat{\theta}_n - \theta_0))^\top \sqrt{n}B^\top(\widehat{\theta}_n - \theta_0) + o_p(1), \end{aligned}$$

where $I(\theta_0) = BB^\top$ is the Cholesky decomposition of $I(\theta_0)$ (see [23]). Combining this with Theorem 3.2, we obtain the claimed result. □

Remark 2. *using this theorem, the set*

$$\left\{\theta \in \Theta : 2(l(\widehat{\theta}_n) - l(\theta)) \leq q_{1-\alpha}\right\}$$

is an asymptotic confidence region for θ_T.

4.2. **Chi-square goodness-of-fit test.** In this subsection, we test the hypothesis

$$\mathcal{H}_0 : S(x,y) \in \{S(x,y;\theta),\ \theta \in \Theta\} \text{ against } \mathcal{H}_1 : S(x,y) \notin \{S(x,y;\theta),\ \theta \in \Theta\}. \quad (9)$$

[13] studied a chi-square test of fit, for right censored data, in the univariate context. In the sequel, we propose a generalization of this test to the bivariate context. For that, we make use of the nonparametric estimator of S, introduced by [1] and given by

$$S_n(s,t) = \begin{cases} \prod_{i=1}^n \left(\frac{N(U_i,0)}{N(U_i,0)+1}\right)^{\alpha_{1i}(s,0)} \prod_{i=1}^n \left(\frac{N(s,V_i)}{N(s,V_i)+1}\right)^{\alpha_{2i}(s,t)} & \text{if } N(s,t) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $N(s,t) = \sum_{i=1}^n 1_{\{U_i>s,V_i>t\}}$, $\alpha_{1i} = 1_{\{U_i\leq s,V_i>t,\Delta_{1i}=1\}}$ and $\alpha_{2i} = 1_{\{U_i>s,V_i\leq t,\Delta_{2i}=1\}}$.
Let T_1 and T_2 be two positive real numbers satisfying $\overline{H}(T_1,T_2) > 0$, where $\overline{H}$ is the survival function of (U,V). We define the empirical process associated with $S_n(s,t)$ on $[0,T_1]\times[0,T_2]$, by

$$Z_n(s,t) = \sqrt{n}(S_n(s,t) - S(s,t;\theta_T))$$

and set

$$\widehat{Z}_n(s,t) = \sqrt{n}(S_n(s,t) - S(s,t;\widehat{\theta}_n)).$$

In order to define the statistic of the studied test, we need the following theorem, that gives the weak convergence of the process $\widehat{Z}_n(s,t)$. This latter results from the fact that the process $Z_n(s,t)$ converges to a centered Gaussian process, denoted by $Z(s,t)$ (see [2]).

Theorem 4.2. *Under hypotheses* **H1– H7**, *the process $\widehat{Z}_n(s,t)$, for $0 < s,s' < T_1$ and $0 < t,t' < T_2$ converges weakly to a centered Gaussian process $\widehat{Z}(s,t)$, with covariance function*

$$cov(\widehat{Z}(s,t),\widehat{Z}(s',t')) = cov(Z(s,t),Z(s',t')) - \nabla_\theta S(s,t;\theta_T)^\top I^{-1}(\theta_T)\nabla_\theta S(s',t';\theta_T),$$

where $\nabla_\theta S(\cdot,\cdot;\theta)$ is the gradient of $S(\cdot,\cdot;\theta)$.

Proof. Applying Theorem 2 of [2], the proof can be carried out in the same way as that of Theorem 1 of [13]. Remark that hypotheses (A.1) and (A.2) of [13] are satisfied in our case, by virtue of hypotheses **H3**, **H6** and **H7** and the continuity of $J(\theta)$ follows from Lemma 2.4 of [21]. As for hypothesis (A.3), it follows from relation (7). □

We are now ready to define the statistic of the test (9). Let $0 < s_1 < s_2 < ... < s_p < T_1$ be a partition of $[0, T_1]$, $0 < t_1 < t_2 < ... < t_q < T_2$ be a partition of $[0, T_2]$ and set

$$\widehat{Z}_n = \Big(\widehat{Z}_n(s_1,t_1), \widehat{Z}_n(s_1,t_2), \ldots, \widehat{Z}_n(s_1,t_q), \widehat{Z}_n(s_2,t_1),\ \ldots,$$
$$\widehat{Z}_n(s_2,t_q), \ldots, \widehat{Z}_n(s_p,t_1), \ldots \widehat{Z}_n(s_p,t_q)\Big)^\top .$$

By Theorem 4.2, the vector $\widehat{Z}_n$ converges in distribution to a centered Gaussian vector $\widehat{Z}$, whose covariance matrix will be denoted by Σ. We define the statistic of our test as follows

$$\widehat{Q}_n = \widehat{Z}_n^\top \widehat{\Sigma}_n^{-1} \widehat{Z}_n,$$

where $\widehat{\Sigma}_n$ is the estimate of Σ, obtained by replacing θ_T by $\widehat{\theta}_n$, S by S_n and in the covariance function of the process $W(s,t)$ defined on page 250 of [2], we replace the different terms by their empirical counterparts.
The statistic $\widehat{Q}_n$ generalizes the modified Pearson statistic, introduced by [13]. Its asymptotic distribution is given by the following theorem.

Theorem 4.3. *Under $\mathcal{H}_0$ and hypotheses* **H1–H7***, and if Σ is nonsingular, we get*

$$\widehat{Q}_n \xrightarrow{\mathcal{D}} \chi^2(pq).$$

Proof. This theorem can be proved in the same way as Theorem 2 of [13]. □

According to this theorem, the critical region of the test (9) is given by

$$\left\{\widehat{Q}_n > q_{1-\alpha}\right\},$$

where $q_{1-\alpha}$ is the $(1-\alpha)-$quantile of the $\chi^2(pq)$ distribution.
It may happen that the matrix $\widehat{\Sigma}_n$ is singular or near-singular. To overcome this problem, we can add, as in the ridge regression, the identity matrix I multiplied by a positive parameter a, so the test statistic becomes

$$\widehat{Q}_n = \widehat{Z}_n^\top \left(\widehat{\Sigma}_n + aI\right)^{-1} \widehat{Z}_n.$$

Remark 3. *To carry out this test, we have used one of the three estimators of S, introduced by* [1]. *The same estimator has been used by* [3] *in the estimation of the density function. Notice that our Theorems 4.2 and 4.3 still hold for the two other estimators of S, since the difference between these estimators is $o(n^{-1})$ a.s., uniformly on (s,t) (see* [1]*).*

5. Simulation study

In this section, we present the results of a simulation study aiming to illustrate the performances of our tests proposed in Section 4 for finite size samples. This study is based on the two models given in examples 3.1 and 3.2 above.
We start by presenting the results for the first model. In this model, X has an $\mathcal{E}(a)$ distribution and given $X = x$, Y has an $\mathcal{E}(bx)$ distribution. The density of the couple (X,Y) is given in relation (1). Moreover, the variable C has an $\mathcal{E}(a_C)$ distribution and given $C = x$, D has an $\mathcal{E}(b_{C,D}x)$ distribution. We take different values of the parameters a, b, a_C and $b_{C,D}$ to get different values of the rate of censoring (RC). First, we deal with the likelihood ratio test of the hypothesis $\mathcal{H}_0 : \theta_T = \theta_0$ against $\mathcal{H}_1 : \theta_T \neq \theta_0$ (introduced in subsection 4.1). We consider two cases for the value of θ_0: $\theta_0 = (7,4)^\top$ and $\theta_0 = (1,3)^\top$. In order to asses the type I error of this test at the asymptotic level $\alpha = 0.05$, we generate 1000 samples of size n of the latent variables from the distribution

characterized by $f_{X,Y}(x, y; \theta_T)$ with $\theta_T = \theta_0$ (i.e., under $\mathcal{H}_0$), then we calculate the values of the test statistic Φ_n (given in relation (8)), corresponding to these 1000 samples. On the basis of the obtained values, we compute the proportion of the rejection of $\mathcal{H}_0$ which is an estimation of the type I error of the test. We take different values of the sample size n to show its influence on the test. We also estimate the type I error of the Wald and the Rao tests (see [24], pages 408-409). Recall that their statistics are respectively defined by

$$W_n = n \left(\widehat{\theta}_n - \theta_0\right)^\top I(\widehat{\theta}_n) \left(\widehat{\theta}_n - \theta_0\right)$$

and

$$R_n = \frac{1}{n} \left(\sum_{i=1}^{n} D(U_i, V_i, \Delta_{1i}, \Delta_{2i}; \theta_0)\right)^\top I(\theta_0)^{-1} \left(\sum_{i=1}^{n} D(U_i, V_i, \Delta_{1i}, \Delta_{2i}; \theta_0)\right),$$

where $D(U, V, \Delta_1, \Delta_2; \theta)$ is defined in relation (4).
The results we obtain are given in Table 1, where the most accurate ones are written in bold. We remark that in all cases, the likelihood ratio test is the best test. Not surprisingly, the performance of the tests increases (resp. decreases) when the sample size (resp. the rate of censoring) increases.

	$\theta_T = \theta_0 = (7, 4)^\top$, $a_C = 3$, $b_{C,D} = 1/2$, $RC \approx 30\%$		$\theta_T = \theta_0 = (1, 3)^\top$, $a_C = 1$, $b_{C,D} = 3$, $RC \approx 50\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Likelihood ratio test	**0.061**	**0.055**	**0.066**	**0.060**
Wald test	0.075	0.070	0.079	0.077
Rao test	0.035	0.039	0.072	0.070

TABLE 1. Type I error of the tests on the parameter for the first model.

To asses the power of the studied tests, we generate 1000 samples of size n of the latent variables from the distribution characterized by $f_{X,Y}(x, y; \theta_T)$ with $\theta_T \neq \theta_0$ (i.e., under $\mathcal{H}_1$), then we calculate the values of the tests statistics Φ_n, W_n and R_n corresponding to these samples and we use them to estimate the power of the tests by the proportion of the rejection of $\mathcal{H}_0$. The obtained results are given in Table 2 below. These results show that the best tests are the Wald test and the likelihood ratio test. The power of the tests increases when the sample size increases.

	$\theta_0 = (7, 4)^\top$, $\theta_T = (8, 5)^\top$, $a_C = 24/7$, $b_{C,D} = 3/5$, $RC \approx 30\%$		$\theta_0 = (7, 4)^\top$, $\theta_T = (10, 8)^\top$, $a_C = 30/7$, $b_{C,D} = 1$, $RC \approx 30\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Likelihood ratio test	0.538	0.889	**0.916**	**0.991**
Wald test	**0.772**	**0.925**	0.911	0.984
Rao test	0.413	0.456	0.478	0.981

TABLE 2. Power of the tests on the parameter for the first model.

Consider now the chi-square test of the hypothesis $\mathcal{H}_0 : S(x, y) \in \{S(x, y; \theta),\ \theta \in \Theta\}$ against $\mathcal{H}_1 : S(x, y) \notin \{S(x, y; \theta),\ \theta \in \Theta\}$ (introduced in subsection 4.2). At the asymptotic level $\alpha = 0.05$, we proceed as previously in order to estimate the type I error of this test. We also estimate the type I error of the Kolmogorov-Smirnov and the Cramér-Von-Mises tests (see [20] page 277). Recall that their statistics are respectively defined by

$$KS_n = \sqrt{n} \sup_{(x,y)\in\mathbb{R}^2} \left| S_n(x,y) - S(x,y;\widehat{\theta}_n) \right|$$

and

$$CVM_n = n \int \left(S_n(x,y) - S(x,y;\widehat{\theta}_n) \right)^2 dS_n(x,y).$$

The results we get are presented in Table 3 below. We remark that in all cases, the best test is the chi-square goodness-of-fit test. The performance of the tests increases (resp. decreases) when the sample size (resp. the rate of censoring) increases.

	$\theta_T = (7,4)^\top$, $a_C = 3$, $b_{C,D} = 1/2$, $RC \approx 30\%$		$\theta_T = (1,3)^\top$, $a_C = 1$, $b_{C,D} = 3$, $RC \approx 50\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Chi-square test	**0.042**	**0.057**	**0.040**	**0.045**
Kolmogorov-Smirnov test	0.040	0.042	0.035	0.041
Cramér-Von-Mises test	0.037	0.038	0.033	0.039

TABLE 3. Type I error of the goodness-of-fit tests for the first model.

To estimate the power of these tests, we proceed as previously, where we generate the latent variables from other models than that defined by $\mathcal{H}_0$. These models and the obtained results are given in Table 4, where $Ray(\sigma)$ denotes the Rayleigh distribution with parameter σ. These results show that the chi-square test is the best test.

	$X \sim \mathcal{E}(7)$, $Y \sim \mathcal{E}(14)$, X and Y independent, $C \sim \mathcal{E}(3)$, $D \sim \mathcal{E}(6)$, C and D independent, $RC = 30\%$		$X \sim Ray(\sqrt{3})$, $Y \sim Ray(\sqrt{6})$, X and Y independent, $C \sim Ray(\sqrt{7})$, $D \sim Ray(\sqrt{14})$, C and D independent, $RC = 30\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Chi-square test	**0.287**	**0.541**	**0.696**	**0.926**
Kolmogorov-Smirnov test	0.115	0.337	0.127	0.579
Cramér-Von-Mises test	0.109	0.310	0.115	0.468

TABLE 4. Power of the goodness-of-fit tests for the first model.

Now, we move on into the second model presented in example 3.2 above. In this example, X has an $\mathcal{E}(a)$ distribution, Y has an $\mathcal{E}(b)$ distribution and X and Y are related by a Clayton copula with parameter γ. The density of (X, Y) is given in relation (2) above. Moreover, C (resp. D) has an $\mathcal{E}(a_C)$ (resp. $\mathcal{E}(b_D)$) distribution and they are related by a Clayton copula with parameter $\gamma_{C,D}$. Following the same steps described above, we obtain the results presented respectively in Tables 5 and 6 for the type I error and the power of the tests on the parameter. Concerning the goodness-of-fit tests, we give our obtained results in Tables 7 and 8 fro the type I error and the power, respectively. From these results, we can deduce similar conclusions to those of the first model.

	$\theta_T = \theta_0 = (7, 14)^\top$, $a_C = 3$, $b_D = 6$, $\gamma = \gamma_{C,D} = 1$, $RC = 30\%$		$\theta_T = \theta_0 = (1, 3)^\top$, $a_C = 1$, $b_D = 3$, $\gamma = \gamma_{C,D} = 1$, $RC = 50\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Likelihood ratio test	**0.059**	**0.057**	**0.072**	**0.060**
Wald test	0.068	0.060	0.075	0.063
Rao test	0.062	0.058	0.073	0.061

Table 5. Type I error of the tests on the parameter for the second model.

	$\theta_0 = (7, 14)^\top$, $\theta_T = (8, 15)^\top$, $a_C = 24/7$, $b_D = 45/7$, $\gamma = \gamma_{C,D} = 1$, $RC = 30\%$		$\theta_0 = (7, 14)^\top$, $\theta_T = (10, 18)^\top$, $a_C = 30/7$, $b_D = 54/7$, $\gamma = \gamma_{C,D} = 1$, $RC = 30\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Likelihood ratio test	**0.258**	**0.814**	**0.513**	**0.907**
Wald test	0.217	0.714	0.452	0.877
Rao test	0.204	0.655	0.446	0.856

Table 6. Power of the tests on the parameter for the second model.

	$\theta_T = (7, 14)^\top$, $a_C = 3$, $b_D = 6$, $\gamma = \gamma_{C,D} = 1$, $RC = 30\%$		$\theta_T = (1, 3)^\top$, $a_C = 1$, $b_D = 3$, $\gamma = \gamma_{C,D} = 1$, $RC = 50\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Chi-square test	0.066	**0.051**	**0.070**	**0.069**
Kolmogorov-Smirnov test	**0.060**	0.053	0.078	0.071
Cramér-Von-Mises test	0.062	0.045	0.072	0.070

Table 7. Type I error of the goodness-of-fit tests for the second model.

	$X \sim \mathcal{E}(7)$, $Y \sim \mathcal{E}(14)$, $C \sim \mathcal{E}(3)$, $D \sim \mathcal{E}(6)$, X and Y (resp. C and D) have a Gumbel copula with parameter 2, $RC = 30\%$		$X \sim Ray(\sqrt{3})$, $Y \sim Ray(\sqrt{6})$, $C \sim Ray(\sqrt{7})$, $D \sim Ray(\sqrt{14})$, X and Y (resp. C and D) have a Calyton copula with parameter 1, $RC = 30\%$	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Chi-square test	**0.517**	**0.795**	**0.557**	**0.822**
Kolmogorov-Smirnov test	0.114	0.441	0.236	0.502
Cramér-Von-Mises test	0.110	0.375	0.219	0.466

Table 8. Power of the goodness-of-fit tests for the second model.

6. Real data application

[25] reported a study on the time to infection after the insertion of a catheter for 38 kidney patients. When an infection occurred, the catheter was removed and the time to infection was recorded. If the catheter was removed by other reasons, then the infection time is right censored. After removing the first catheter, an other one was inserted and the second infection time was observed or censored. So, the variable of interest is (X, Y), where X is the first infection time and Y is the second infection time. In our study, we use our proposed chi-square goodness-of-fit test to fit the bivariate model described in example 3.1 above to this set of data. We also use the Kolmogorov-Smirnov and the Cramér-Von-Mises tests. The infection times being recorded in days, we divide by 365 to treat them in years. All the goodness-of-fit tests accept the theoretical model as a model describing the data with the p-values given in Table 9 below. Having positive results for these tests, we use the method of moments to estimate the parameters of the model. We get values near to $\theta_0 = (2.5, 6)^\top$. So, we perform tests on the value of the parameter using two values: $\theta_0 = (2.5, 6)^\top$ and $\theta_0 = (7, 10)^\top$. As in the simulation study, we use the likelihood ratio, the Wald and the Rao tests. The p-values of these tests are presented in Table 10 below. We remark that the likelihood ratio test accepts the value $(2.5, 6)^\top$ and rejects the value $(7, 10)^\top$ of the parameter. However, the Wald test rejects both values and the Rao test accepts both of them.

Test	p-value
Chi-square test	0.803
Kolmogorov-Smirnov test	0.854
Cramér-Von-Mises test	0.858

Table 9. p-value of the goodness-of-fit tests for the infection times data.

Test	$\theta_0 = (2.5, 6)^\top$	$\theta_0 = (7, 10)^\top$
Likelihood ratio test	0.313	1.796×10^{-13}
Wald test	2.676×10^{-5}	0
Rao test	0.874	0.548

Table 10. p-value of the tests on the parameter for the infection times data.

7. Conclusions

In this work, we have studied parametric statistical models in the presence of bivariate right censored data. First, we studied the asymptotic properties of the maximum likelihood estimator in this context, namely, we have established its weak consistency and asymptotic normality. Then, we have proposed a likelihood ratio test for the value of the parameter as well as a chi-squared goodness-of-fit test. We have determined the asymptotic distributions of the two tests under the null hypothesis. The goodness-of-fit test is based on one of the three estimators of the bivariate survival function proposed by [1]. However, the same results hold immediately for the two other estimators. Comparison between the tests tht result from each estimator, could be developed in further researches.

References

[1] Campbell, G. and Földes, A., (1982), Large sample properties of nonparametric bivariate estimators with censored data, International Colloquium on Nonparametric Statistical Inference, Budapest 1980, Amsterdam: North-Holland, pp. 23-28.

[2] Campbell, G., (1982), Asymptotic properties of several nonparametric multivariate distribution function estimators under random censoring, J. Survival Analysis (Columbus, Ohio, 1981), 2, pp. 243-256.

[3] Wells, M.T. and Yeo, K. P., (1996), Density estimation with bivariate censored data, Journal of the American Statistical Association, 91(436), pp. 1566-1574.

[4] Dabrowska, D., (1988), Kaplan-Meier estimate on the plane, The Annals of Statistics, 16, pp. 1475-1489.

[5] Dabrowska, D., (1989), Kaplan-Meier estimate on the plane: weak convergence, LIL, and the bootstrap, Journal of Multivariate Analysis, 29, pp. 308-325.

[6] Prentice, R. L. and Cai, J., (1992), A covariance function for bivariate survival data and a bivariate survivor function estimate, Biometrika, 79, pp. 495–512.

[7] Pruitt, R., (1993), Small sample comparison of six bivariate survival curve estimators, Journal of Statistical Computation and Simulation, 45, pp. 147-167.

[8] van der Laan, M. J., (1993), Efficient and inefficient estimation in semi-parametric models, Ph.D. Thesis, University of Utrecht.

[9] Ghosh, D., (2006), Semiparametric global cross-ratio models for bivariate censored data,Scandinavian Journal of Statistics, 33, pp. 609-619.

[10] Beaudoin, D., (2007), Estimation de la dépendance et choix de modèles pour des données bivariées sujettes à censure et à troncation , Ph.D. Thesis, Université Laval, Québec.

[11] Degges, R. C., (2011), Testing umbrella alternatives for bivariate censored data, Ph.D. Thesis, North Dakota State University.

[12] Gribkova, S., (2014), Contributions en inférence statistique en présence de censure multivariée, Ph.D. Thesis, Université Pierre et Marie Curie, Paris.

[13] Habib, M. G. and Thomas, D. R., (1986), Chi-square goodness-of-fit tests for randomly censored data, The Annals of Statistics, 14(2), pp. 759-765.

[14] Giordan, M. and Wehrens, R., (2015), A comparison of computational approaches for maximum likelihood estimation of the Dirichlet parameters on high-dimensional data, SORT, 39(1), pp. 109-126.

[15] Ronning, G., (1989), Maximum likelihood estimation of Dirichlet distributions, Journal of Statistical Computation and Simulation, 32, pp. 215-221.

[16] Dishon, M. and Weiss, G., (1980), Small sample comparison of estimation methods for the beta distribution, Journal of Statistical Computation and Simulation, 11, pp. 1-11.

[17] Wicker, N., Muller, J., Kalathur, R. K. R. and Poch, O., (2008), A maximum likelihood approximation method for Dirichlet's parameter estimation, Computational Statistics and Data Analysis, 52, pp. 1315-1322.

[18] Levenberg, K., (1944), A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics, 2, pp. 164-168.

[19] Marquardt, D., (1963), An algorithm for least-squares estimation of nonlinear parameters, SIAM Journal on Applied Mathematics, 11, pp. 431-441.

[20] van der Vaart, A. W., (1998), Asymptotic Statistics, Dover publications, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[21] Newey, W. K. and McFadden, D., (1994), Large sample estimation and hypothesis testing, Handbook of Econometrics, Vol 4, ed. by R. Engle and D. McFadden. New York: North Holland.

[22] Billingsley, P., (1995), Probability and Measure, The University of Chicago, John Wiley & Sons.

[23] Lütkepohl, H., (1996), Handbook of matrices, Humboldt University of Berlin, John Wiley & Sons.

[24] Pardo, L., (2006), Statistical inference based on divergence measures, Chapman & Hall/CRC, Madrid. John Wiley & Sons.

[25] McGilchrist, C. A. and Aisbett, C. W., (1991), Regression with frailty in survival analysis, Biometrics, 142, pp. 461-466.

Mohamed Boukeloua is currently an associate professor of mathematics in the National Polytechnic Institute of Constantine, Algeria. He is a member of Laboratory of Process Engineering for Sustainable Development and Health Products in the same institute. He is also a member of Laboratory of Biostatistics, Bioinformatics and Mathematical Methodology Applied on Health Sciences, Faculty of Medicine, Salah Boubnider University of Constantine, Algeria. He holds a PhD in Mathematical Statistics from Brothers Mentouri University of Constantine, Algeria. His research interests are : Parametric and Non parametric Inference, Censored data, $\phi-$divergences and their applications, Copula models, Kernel estimation, Non parametric regression, Dependent data, Bayesian inference.