Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Reliability of a soccer player based on the bivariate Rayleigh distribution with right censored and ignorable missing data

Fayyaz Bahari, Safar Parsi & Mojtaba Ganjali

To cite this article: Fayyaz Bahari, Safar Parsi & Mojtaba Ganjali (2020): Reliability of a soccer player based on the bivariate Rayleigh distribution with right censored and ignorable missing data, Journal of Applied Statistics, DOI: 10.1080/02664763.2020.1723504

To link to this article: https://doi.org/10.1080/02664763.2020.1723504

Published online: 04 Feb 2020.

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2020.1723504

Reliability of a soccer player based on the bivariate Rayleigh distribution with right censored and ignorable missing data

Fayyaz Bahari[a], Safar Parsi[a] and Mojtaba Ganjali[b]

[a]Department of Statistics, Faculty of Mathematical Sciences, University of Mohaghegh Ardabil, Ardabil, Iran; [b]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

ABSTRACT

In this paper, we study the performance of a soccer player based on analysing an incomplete data set. To achieve this aim, we fit the bivariate Rayleigh distribution to the soccer dataset by the maximum likelihood method. In this way, the missing data and right censoring problems, that usually happen in such studies, are considered. Our aim is to inference about the performance of a soccer player by considering the stress and strength components. The first goal of the player of interest in a match is assumed as the stress component and the second goal of the match is assumed as the strength component. We propose some methods to overcome incomplete data problem and we use these methods to inference about the performance of a soccer player.

ARTICLE HISTORY
Received 13 August 2019
Accepted 24 January 2020

KEYWORDS
Bivariate Rayleigh distribution; maximum likelihood estimator; soccer player; right censored data; missing data; stress-strength component

1. Introduction

One of the most important criteria in distinguishing the quality of a soccer forward player is his ability to score goals in any match. Also, obtaining the first goal is vital in the result of a soccer match. Therefore, we can talk about the quality of a forward player based on the time that takes for him/her to achieve the first goal of the match. To reach this aim, we consider the stress-strength criterion in statistical studies. Suppose, we have a system with stress and strength components and we are seeking the probability of exceeding the stress component, Y, to the strength component, X, say $R = P(Y < X)$. Therefore, we can consider the time to the first goal of the assumed player in any match as the stress component and we can consider the time to the second goal of any match by any player, as the strength component.

In stress-strength studies such as those of Tong [15] and Constantine *et al.* [2], it is usually assumed that X and Y are independent. But in most studies, X and Y are dependent such as those of Kundu and Gupta [9] and Kakade *et al.* [8] where the properties of R based on the generalized exponential distribution and Gumbel distribution, respectively, are studied.

Meintanis [11] and Pak *et al.* [12] have considered the full data set of the Union of European Football Associations (UEFA) to inference about the performance of a soccer player. They have fitted the Bivariate Rayleigh (BVR) distribution to data based on the maximum

CONTACT Safar Parsi parsi@uma.ac.ir

likelihood method. They have ignored most of the matches that affect the inferences. For example, Pak *et al.* [12] have ignored all the matches without any goal and the matches without penalty scored goals. Consequently, they have lost some useful information about the matches. In statistical studies, these ignored data are known as missing information. Indeed in this paper, our aim is to consider the effects of missing data on studies. In soccer case, the matches that our assumed player has been absent is considered as the missing data because before studying experiments (matches), we lose the experimental unit (assumed player). Furthermore, the matches that the player has been substituted and he/she left the match without goals, is considered as a right censored data. Also, the matches, he/she finished with no goal are considered as the right censored data, because the time of the experiment is over.

We propose some models that contain missing data in the estimating equations to obtain maximum likelihood estimators. We have applied our proposed method for the BVR distribution. In a similar way, it can be used to derive such estimates for other distributions. Moreover, we will consider effects of missing data and censored data together in the estimating equations. See Jana [6], Al-khedhairi *et al.* [1] and Dina [3] for more details about bivariate Rayleigh distribution and its parameters maximum likelihood estimation. Also, see Jeevanand [7] for Bayesian inferences with Rayleigh distribution. The probability density function of BVR distribution is:

$$f(x,y;\boldsymbol{\lambda}) = \begin{cases} 4\lambda_1(\lambda_2+\lambda_3)xy\,\mathrm{e}^{-\lambda_1 x^2-(\lambda_2+\lambda_3)y^2} & 0<x<y \\ 4\lambda_2(\lambda_1+\lambda_3)xy\,\mathrm{e}^{-\lambda_2 y^2-(\lambda_1+\lambda_3)x^2} & x>y>0. \\ \lambda_3\,\mathrm{e}^{-(\lambda_1+\lambda_2+\lambda_3)z^2} & x=y=z \end{cases} \tag{1}$$

where, $\boldsymbol{\lambda}^{\mathrm{T}} = (\lambda_1, \lambda_2, \lambda_3)$ is the unknown vector of parameters of the BVR distribution and $z = \max(x, y)$.

Rest of this paper organized as follow. In Section 2, we will construct appropriate estimating equations based on the log-likelihood function by considering the soccer data problem in our cases. In Section 3, we will derive the asymptotic properties of the estimators. In Section 4, we will use some Monte Carlo simulation study to compare the maximum likelihood estimates with the different incomplete dataset. In Section 5, some applications of the proposed methods are discussed and in the next section, real data are studied by the proposed methods. Also, proof of the theorems are given in the appendices.

2. Maximum likelihood estimator of stress-strength parameter

Suppose X and Y be the random variables of strength and stress components of a system, respectively, where their joint distribution follows from the BVR distribution. It is easy to prove that,

$$R = P(Y < X) = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}. \tag{2}$$

Therefore, based on the invariance property of the maximum likelihood method, the MLE of R will be:

$$\hat{R} = \frac{\hat{\lambda}_2}{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3}, \tag{3}$$

where $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ are the maximum likelihood estimates of parameters of BVR distribution. Accordingly, it is necessary to determine maximum likelihood estimates of parameters. To achieve this aim, we consider the full data case, censored data case, missing data case and censored and missing data case together, say the mixed case, to find MLEs.

2.1. Full data case

If in a soccer competition, all the matches have at least two goals, our assumed player be present in any matches, and he scores at least one goal in any matches, we will have fully observed data. Therefore, in this case, the likelihood function will be:

$$L_F(\boldsymbol{\lambda}) = \prod_{i=1}^{n} f(x_i, y_i; \boldsymbol{\lambda}) = \prod_{I_1} f_1(x_i, y_i; \boldsymbol{\lambda}) \prod_{I_2} f_2(x_i, y_i; \boldsymbol{\lambda}) \prod_{I_3} f_3(z_i; \boldsymbol{\lambda}), \tag{4}$$

where, n is the sample size (the number of games), $I_1 = \{(x_i, y_i) : x_i < y_i,\ i = 1, \ldots, n\}$, $I_2 = \{(x_i, y_i) : x_i > y_i,\ i = 1, \ldots, n\}$ and $I_3 = \{(x_i, y_i) : x_i = y_i,\ i = 1, \ldots, n\}$. Therefore, if in full data case, we consider the log-likelihood function with $l_F(\lambda)$, then the derivatives of the log-likelihood function with respect to λ_1, λ_2 and λ_3 will be as follow, respectively:

$$\begin{aligned}
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_1} &= \frac{n_2}{\lambda_1 + \lambda_3} + \frac{n_1}{\lambda_1} - \sum_{i=1}^{n} x_i^2 \\
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_2} &= \frac{n_1}{\lambda_2 + \lambda_3} + \frac{n_2}{\lambda_2} - \sum_{i=1}^{n} y_i^2. \\
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_3} &= \frac{n_1}{\lambda_2 + \lambda_3} + \frac{n_2}{\lambda_1 + \lambda_3} + \frac{n_3}{\lambda_3} - \sum_{i=1}^{n} \max(x_i, y_i)^2
\end{aligned} \tag{5}$$

In Equation (2.4), n_1 is the number of observations with $x_i < y_i$, n_2 is the number of observations with $x_i > y_i$ and n_3 is the number of observations with $x_i = y_i$. Also, we can rewrite the above equations as:

$$\begin{aligned}
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_1} &= \sum_{i=1}^{n} \left\{ \frac{I(x_i > y_i)}{\lambda_1 + \lambda_3} + \frac{I(x_i < y_i)}{\lambda_1} - x_i^2 \right\} \\
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_2} &= \sum_{i=1}^{n} \left\{ \frac{I(x_i < y_i)}{\lambda_2 + \lambda_3} + \frac{I(x_i > y_i)}{\lambda_2} - y_i^2 \right\}. \\
\frac{\partial l_F(\boldsymbol{\lambda})}{\partial \lambda_3} &= \sum_{i=1}^{n} \left\{ \frac{I(x_i < y_i)}{\lambda_2 + \lambda_3} + \frac{I(x_i > y_i)}{\lambda_1 + \lambda_3} + \frac{I(x_i = y_i)}{\lambda_3} - \max(x_i, y_i)^2 \right\}
\end{aligned} \tag{6}$$

We denote components of the above summations in Equation (6) by $\boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i)$ which is a vector function with dimension 3 and it is called the score function. Moreover, $I(\cdot)$ is the indicator function. Therefore, MLEs can be obtained by solving the following estimating equations:

$$\sum_{i=1}^{n} \boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}. \tag{7}$$

To solve the above equations with respect to parameters, it is enough to use recursive numeric methods such as the Newton-Raphson algorithm or Fisher's score method.

2.2. Censored data case

Assumptions in the last subsection rarely happen on soccer matches. Therefore, in this subsection, we consider some minor assumption on the matches. In this case, we just assume that our player of interest to be present in any matches. Therefore, we will have only the right censored data problem. Suppose some individuals of variables X and Y to be right censored. The relationship between X and Y will be as in Table 1. It is easy to prove the following statements which are necessary to construct the likelihood functions in the censored data case,

$$P(X = x, Y > y_c) = 2\lambda_1 x\, \mathrm{e}^{-\lambda_1 x^2 - (\lambda_2+\lambda_3)y_c^2}, \tag{8}$$

$$P(X > x_c, Y = y) = 2\lambda_2 y\, \mathrm{e}^{-(\lambda_1+\lambda_3)x_c^2 - \lambda_2 y^2}. \tag{9}$$

Therefore, the likelihood function based on the right censored data will be:

$$\begin{aligned} L_C(\boldsymbol{\lambda}) &= \prod_{I_x \cap I_y} f(x_i, y_i; \boldsymbol{\lambda}) \prod_{I_x \cap I_y^c} P(X = x_i,\ Y > y_{c_i}) \prod_{I_x^c \cap I_y} P(X > x_{c_i},\ Y = y_i) \prod_{I_x^c \cap I_y^c} S(x_{c_i}, y_{c_i}), \\ &= \prod_{i=1}^{n} \{[f(x_i, y_i;\ \boldsymbol{\lambda})]^{c_{x_i} c_{y_i}} [P(X = x_i,\ Y > y_{c_i})]^{c_{x_i}(1-c_{y_i})} \\ &\quad \times [P(X > x_{c_i},\ Y = y_i)]^{(1-c_{x_i})c_{y_i}} [S(x_{c_i}, y_{c_i})]^{(1-c_{x_i})(1-c_{y_i})}\}, \end{aligned} \tag{10}$$

where $c_{x_i}(c_{y_i})$ determines the censoring indicator for X_i (Y_i). It is equal to 1 if the ith value of variable $X(Y)$ is observed and it is equal to 0 if the ith individual of variable $X(Y)$ to be censored. Therefore, if we consider log-likelihood function, $l_C(\boldsymbol{\lambda})$, in censored case, then the derivatives of the log-likelihood function with respect to λ_1, λ_2 and λ_3 will be as follow,

Table 1. The relationship between random variables X and Y.

Case	X	Y	Domain of variables
1	*observed* (I_X)	*observed* (I_Y)	$x > 0, y > 0$ ($I_X \cap I_Y$)
2	*observed* (I_X)	*censored* (I_Y^c)	$0 < x < y,\ y > y_c$ ($I_X \cap I_Y^c$)
3	*censored* (I_X^c)	*observed* (I_Y)	$x > y > 0,\ x > x_c$ ($I_X^c \cap I_Y$)
4	*censored* (I_X^c)	*censored* (I_Y^c)	$x > x_c,\ y > y_c$ ($I_X^c \cap I_Y^c$)

Notes: Where some observations maybe censored from the right. The set of fully observed and censored observations of $X(Y)$ are denoted by $I_X(I_Y)$ and $I_X^c(I_Y^c)$, respectively. Also, X_c and Y_c show the censoring time of the variables.

respectively:

$$
\begin{aligned}
\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_1} &= \sum_{I_1} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1} - x_i^2 \right\} + \sum_{I_2} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1 + \lambda_3} - x_i^2 \right\} \\
&\quad + \sum_{I_0} c_{x_i} c_{y_i} \{-z_i^2\} + \sum_{i=1}^{n} c_{x_i} (1 - c_{y_i}) \left\{ \frac{1}{\lambda_1} - x_i^2 \right\} \\
&\quad + \sum_{i=1}^{n} c_{y_i} (1 - c_{x_i}) \{-x_{c_i}^2\} + \sum_{i=1}^{n} (1 - c_{x_i})(1 - c_{y_i}) \{-x_{c_i}^2\} \\
\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_2} &= \sum_{I_1} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2 + \lambda_3} - y_i^2 \right\} + \sum_{I_2} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2} - x_i^2 \right\} \\
&\quad + \sum_{I_0} c_{x_i} c_{y_i} \{-z_i^2\} + \sum_{i=1}^{n} c_{x_i} (1 - c_{y_i}) \left\{ \frac{1}{\lambda_1} - y_{c_i}^2 \right\}. \\
&\quad + \sum_{i=1}^{n} c_{y_i} (1 - c_{x_i}) \{-y_i^2\} + \sum_{i=1}^{n} (1 - c_{x_i})(1 - c_{y_i}) \{-y_{c_i}^2\} \\
\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_3} &= \sum_{I_1} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2 + \lambda_3} - y_i^2 \right\} + \sum_{I_2} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1 + \lambda_3} - x_i^2 \right\} \\
&\quad + \sum_{I_0} c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_3} - z_i^2 \right\} + \sum_{i=1}^{n} c_{x_i} (1 - c_{y_i}) \{-y_i^2\} \\
&\quad + \sum_{i=1}^{n} c_{y_i} (1 - c_{x_i}) \{-x_{c_i}^2\} + \sum_{i=1}^{n} (1 - c_{x_i})(1 - c_{y_i}) \{-z_{c_i}^2\}
\end{aligned}
\tag{11}
$$

We can rewrite the above equations as follow:

$$
\begin{aligned}
\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_1} &= \sum_{i=1}^{n} \Bigg[I(x_i > y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1} - x_i^2 \right\} + I(x_i < y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1 + \lambda_3} - x_i^2 \right\} \\
&\quad + I(x_i = y_i) c_{x_i} c_{y_i} \{-z_i^2\} + c_{x_i} (1 - c_{y_i}) \left\{ \frac{1}{\lambda_1} - x_i^2 \right\} \\
&\quad + c_{y_i} (1 - c_{x_i}) \{-x_{c_i}^2\} + (1 - c_{x_i})(1 - c_{y_i}) \{-x_{c_i}^2\} \Bigg] \\
\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_2} &= \sum_{i=1}^{n} \Bigg[I(x_i > y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2 + \lambda_3} - y_i^2 \right\} + I(x_i < y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2} - x_i^2 \right\} \\
&\quad + I(x_i = y_i) c_{x_i} c_{y_i} \{-z_i^2\} + c_{x_i} (1 - c_{y_i}) \left\{ \frac{1}{\lambda_1} - y_{c_i}^2 \right\}. \\
&\quad + c_{y_i} (1 - c_{x_i}) \{-y_i^2\} + (1 - c_{x_i})(1 - c_{y_i}) \{-y_{c_i}^2\} \Bigg]
\end{aligned}
$$

$$\frac{\partial l_C(\boldsymbol{\lambda})}{\partial \lambda_3} = \sum_{i=1}^{n} \Bigg[I(x_i > y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_2 + \lambda_3} - y_i^2 \right\} + I(x_i < y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_1 + \lambda_3} - x_i^2 \right\}$$
$$+ I(x_i = y_i) c_{x_i} c_{y_i} \left\{ \frac{1}{\lambda_3} - z_i^2 \right\} + c_{x_i}(1 - c_{y_i})\{-y_i^2\}$$
$$+ c_{y_i}(1 - c_{x_i})\{-x_{c_i}^2\} + (1 - c_{x_i})(1 - c_{y_i})\{-z_{c_i}^2\} \Bigg] \tag{12}$$

As in the full data case, we denote components of the above summations in equation (12) by $\boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i)$ which is a vector function with dimension 3. In this case, MLEs are obtained by solving the following estimating equations:

$$\sum_{i=1}^{n} \boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}. \tag{13}$$

2.3. Missing data case

We consider the other problem of the soccer data set i.e. the occurence of missing data. Suppose our assumed player is absent in some matches which causes some information of data to be missing. Also in this case, it is assumed that any match contains at least two goals and if our assumed player is present, he can score at least one goal. By the above assumption, we will just encounter the missing data problem. Therefore, we consider the case that some values in recording the random variable Y are missing with an ignorable mechanism, based on the taxonomy of Rubin [13] and Little and Rubin [10], which means that the data are Missing at Random (MAR) or are Missing Completely at Random (MCAR). We say missing mechanism is MCAR if missingness of the data does not depend on the data and we say missing mechanism is MAR if missingness of the data just depends on the observed data. Moreover, in this case, the variable X is fully observed.

To estimate parameters of interest, we may remove missing data from the data and estimate the parameters by the complete observed dataset which is called the Complete Case (CC) method. The CC method usually gives us biased estimates. To solve this problem, we use the Inverse Probability Weights (IPW) method of Horvitz and Thompson [5] which gives unbiased estimates. Therefore, in this case, we will use two methods, say the CC method and the IPW method to estimate parameters. The estimating equations of CC method and IPW method are, respectively:

$$\sum_{i=1}^{n} \delta_i \boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}, \tag{14}$$

and

$$\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(x_i)} \boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}, \tag{15}$$

where δ_i is the missing indicator and is equal to 1 if the ith individual is observed and is equal to 0 if the ith individual has missing value. Also, $\pi(x_i)$ is the probability of observing the ith individual response of data which can be estimated by the logistic regression method or some nonparametric methods based on the observed variable X as follow:

$$\hat{\pi}(x_i) = \frac{\sum_{j=1}^{n} \delta_j K_h(x_j - x_i)}{\sum_{j=1}^{n} K_h(x_j - x_i)}, \tag{16}$$

where $K_h(\cdot)$ is a kernel function with smoothing parameter h. The parameters can be obtained from Equations (14) and (15) in a similar way to the full data case. To find more details about the IPW method see Zhao and Lipsitz [17].

2.4. Mixed case

In this case, we consider soccer competitions with no restrictions. It means that we consider a case that the dataset may contain missing and right censored observations. Therefore, some values of the variable Y can be missed and some values of this variable can be right censored and also the variable X can contain the right censored data. We consider the *CC* and *IPW* methods again similar to the missing data case. Therefore, the estimating equations based on the log-likelihood function will be as follow for *CC* and *IPW* methods, respectively:

$$\sum_{i=1}^{n} \delta_i \boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}, \tag{17}$$

$$\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(x_i)} \boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i) = \mathbf{0}. \tag{18}$$

3. Asymptotic distribution

To derive the asymptotic distribution of $\hat{R}$, it is necessary to derive the asymptotic distribution of $\hat{\boldsymbol{\lambda}}$. The Fisher information using the full data is:

$$I_F(\boldsymbol{\lambda}) = \begin{bmatrix} -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1^2}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_2}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_3}\right) \\ -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_2}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_2^2}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_2 \partial \lambda_3}\right) \\ -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_3}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_2 \partial \lambda_3}\right) & -E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_3^3}\right) \end{bmatrix}, \tag{19}$$

where it is easy to prove that:

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1^2}\right) = -\frac{n\lambda_2}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_1+\lambda_3)^2} - \frac{n}{\lambda_1^2(\lambda_1+\lambda_2+\lambda_3)},$$

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_2}\right) = 0,$$

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_3}\right) = -\frac{n\lambda_2}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_1+\lambda_3)^2},$$

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial\lambda_2^2}\right) = -\frac{n\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_2+\lambda_3)^2} - \frac{n}{\lambda_2^2(\lambda_1+\lambda_2+\lambda_3)},$$

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial\lambda_2\partial\lambda_3}\right) = -\frac{n\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_2+\lambda_3)^2},$$

$$E\left(\frac{\partial^2 l_F(\boldsymbol{\lambda})}{\partial^2\lambda_3}\right) = -\frac{n\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_2+\lambda_3)^2} - \frac{n\lambda_2}{(\lambda_1+\lambda_2+\lambda_3)(\lambda_1+\lambda_3)^2} - \frac{n}{\lambda_3^2(\lambda_1+\lambda_2+\lambda_3)}.$$

Moreover, in a similar way, we can find the Fisher information for censored data, missing cases with *CC* and *IPW* method and mixed cases with *CC* and *IPW* methods which are indicated by I_C, I_{CC}, I_{IPW}, I_{CCC} and I_{CIPW}, respectively. The Fisher information using the censored data is as follow:

$$I_C(\boldsymbol{\lambda}) = \begin{bmatrix} -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1^2}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_2}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_3}\right) \\ -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_2}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_2^2}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_2\partial\lambda_3}\right) \\ -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_3}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_2\partial\lambda_3}\right) & -E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_3^2}\right) \end{bmatrix}, \tag{20}$$

where it is easy to prove that:

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1^2}\right) = -\frac{nE(C_x)E(C_y)\lambda_2}{(\lambda_1+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)} - \frac{nE(C_x)E(1-C_y)}{\lambda_1^2},$$

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_2}\right) = 0,$$

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_1\partial\lambda_3}\right) = -\frac{nE(C_x)E(C_y)\lambda_2}{(\lambda_1+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)},$$

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_2^2}\right) = -\frac{nE(C_x)E(C_y)\lambda_1}{(\lambda_2+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)} - \frac{nE(1-C_x)E(C_y)}{\lambda_2^2},$$

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial\lambda_2\partial\lambda_3}\right) = -\frac{nE(C_x)E(C_y)\lambda_1}{(\lambda_2+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)},$$

$$E\left(\frac{\partial^2 l_C(\boldsymbol{\lambda})}{\partial^2\lambda_3}\right) = -\frac{nE(C_x)E(C_y)\lambda_2}{(\lambda_1+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)} - \frac{nE(C_x)E(C_y)\lambda_1}{(\lambda_2+\lambda_3)^2(\lambda_1+\lambda_2+\lambda_3)} - \frac{nE(C_x)E(C_y)}{(\lambda_3(\lambda_1+\lambda_2+\lambda_3)}.$$

Moreover, it is easy to prove that $I_{CC} = E(\delta)I_F$, $I_{IPW} = I_F$, $I_{CCC} = E(\delta)I_C$ and $I_{CIPW} = I_C$.

One can distinguish that all of the mentioned Fisher informations are positive definite. Now, we can construct the following theorems:

Theorem 3.1: *For large n,*

(a) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_F - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_F^{-1}(\boldsymbol{\lambda}))$,
(b) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_C - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_C^{-1}(\boldsymbol{\lambda}))$,
(c) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_{CC} - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_{CC}^{-1}(\boldsymbol{\lambda}))$,
(d) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_{IPW} - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_{IPW}^{-1}(\boldsymbol{\lambda}))$,
(e) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_{CCC} - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_{CCC}^{-1}(\boldsymbol{\lambda}))$,
(f) $\sqrt{n}(\hat{\boldsymbol{\lambda}}_{CIPW} - \boldsymbol{\lambda}) \xrightarrow{D} N_3(\mathbf{0}, J_{CIPW}^{-1}(\boldsymbol{\lambda}))$,

where $J_j(\boldsymbol{\lambda}) = n^{-1} I_j(\boldsymbol{\lambda})$ *with* $j = F, C, CC, IPW, CCC$ *and* $CIPW$.

Proof of Theorem 3.1 is given in Appendix 1. This theorem determines the asymptotic properties of the Bivariate Rayleigh distribution parameters. For example, we can use this theorem to construct some confidence intervals for the parameters.

Now, we can obtain properties of the stress-strength parameter based on Theorem 3.1 by the following theorem:

Theorem 3.2: *For large n,*

(a) $\sqrt{n}(\hat{R}_F - R) \xrightarrow{D} N(0, \sigma_F^2)$,
(b) $\sqrt{n}(\hat{R}_C - R) \xrightarrow{D} N(0, \sigma_C^2)$,
(c) $\sqrt{n}(\hat{R}_{CC} - R) \xrightarrow{D} N(0, \sigma_{CC}^2)$,
(d) $\sqrt{n}(\hat{R}_{IPW} - R) \xrightarrow{D} N(0, \sigma_{IPW}^2)$,
(e) $\sqrt{n}(\hat{R}_{CCC} - R) \xrightarrow{D} N(0, \sigma_{CCC}^2)$,
(f) $\sqrt{n}(\hat{R}_{CIPW} - R) \xrightarrow{D} N(0, \sigma_{CIPW}^2)$, *where,* $\sigma_j^2 = B^{\mathrm{T}} J_j(\boldsymbol{\lambda}) B$ *with* $j = F, C, CC, IPW, CCC$, $CIPW$ *and* $B = (\partial R/\partial\lambda_1,\ \partial R/\partial\lambda_2,\ \partial R/\partial\lambda_3)$.

Proof of Theorem 3.2 is given in Appendix 2. This theorem determines the asymptotic properties of the stress-strength component of Bivariate Rayleigh distribution. By using this theorem, we can analyze the R criterion.

4. Simulation study

Before applying our proposed methods to the soccer data set, it is necessary to see the performance of introduced methods. In this section, we consider Monte Carlo simulation studies in two cases to investigate the performance of the models under different types of incomplete datasets. First, we study a model with an equal percentage of missing and censored data for different methods. In the second study, we consider a model with different percentage of incomplete data. Therefore, in some cases, the rates of missing data and censored data will not be equal. Also, in any Monte Carlo simulation study, we will repeat the algorithm 5000 times.

4.1. Study 1

Suppose the bivariate components (X, Y) follows from the $BVR(1,1,1)$. We consider the fixed right censoring time in 1.2. Also, we suppose that the following two missing mechanisms produce missing data in the random variable Y:

$$\Pr(\delta_i = 1) = 0.89, \quad i = 1, 2, \ldots, n, \tag{21}$$

$$\Pr(\delta_i = 1 \mid X_i = x_i) = \frac{1}{1 + e^{-1-1.5x_i-1.5x_i^2}}, \quad i = 1, 2, \ldots, n, \tag{22}$$

where the first mechanism is MCAR and the second mechanism is MAR. Therefore, some data of random variable Y will be missing or right censored and some individuals of random variable X will be right censored. In this study, we have calculated Absolute Bias (Bias), Mean Square Error (MSE), Confidence Interval Width (CIW) and Coverage Probability (CP) of the stress-strength reliability parameter. The results of Study 1 are given in Table 2 for 5000 Monte Carlo repetitions. Moreover, Missing Mechanism (MM) and the Rate of Incomplete Data (RID) in random variables X and Y are given in Table 2. To see the sample size effect on inferences, we have used two different sample sizes in simulations. Results of the study with sample size 50 are given in the upper row and the results of the study with sample size 100 are given in lower row (inside of brackets).

In this study, we can say that all the methods have good performances. This result follows from the low rate of incomplete data. However, the IPW method has a little better performance than other methods when the missing mechanism is MAR. Also, by increasing the sample size to 100, the error of all methods reduced. On the other hand, by comparing the results of censored based methods with missing based methods, we can see a bigger effect of missing data in comparing with that of censored data to estimate parameters. This means that the accuracy of estimates in censored based methods is better than that of the estimates in missing based methods.

Table 2. Study 1 by using two different sample sizes, 50 and 100.

Methods	MM	RID in X	RID in Y	Bias	MSE	CIW	CP
C	–	0.056	0.055	0.0405	0.0028	0.2372	0.9758
				(0.0276)	(0.0012)	(0.1684)	(0.9808)
CC	MCAR	0.000	0.109	0.0531	0.0044	0.2496	0.9298
				(0.0365)	(0.0021)	(0.1780)	(0.9452)
IPW	MCAR	0.000	0.109	0.0524	0.0043	0.2353	0.9150
				(0.0359)	(0.0020)	(0.1678)	(0.9332)
CC	MAR	0.000	0.094	0.0532	0.0045	0.2486	0.9360
				(0.0378)	(0.0022)	(0.1771)	(0.9410)
IPW	MAR	0.000	0.094	0.0509	0.0041	0.2357	0.9242
				(0.0351)	(0.0020)	(0.1680)	(0.9392)
CCC	MCAR	0.056	0.159	0.0434	0.0032	0.2512	0.9758
				(0.0291)	(0.0013)	(0.1785)	(0.9802)
CIPW	MCAR	0.056	0.159	0.0440	0.0032	0.2361	0.9614
				(0.0297)	(0.0014)	(0.1679)	(0.9720)
CCC	MAR	0.056	0.145	0.0429	0.0031	0.2502	0.9746
				(0.0295)	(0.0014)	(0.1777)	(0.9788)
CIPW	MAR	0.056	0.145	0.0419	0.0030	0.2362	0.9676

Notes: Where the results of the study by using sample size 100 are given in brackets. This study contains a different percentage of incomplete data (MM: Missing Mechanism; RID: Rate of Incomplete Data; CIW: Confidence Interval Width; CP: Coverage Probability).

4.2. Study 2

In this study, we assume the bivariate components (X, Y) follow from the $BVR(0.25, 0.5, 1)$. We assume that the random right censoring times for method C follow from the following patterns:

$$X_{c_i} \sim \text{Uniform}(1.2, 1.5), \quad i = 1, 2, \dots, n, \tag{23}$$

$$Y_{c_i} \sim \text{Uniform}(1, 1.5), \quad i = 1, 2, \dots, n. \tag{24}$$

when we estimate parameters by CC and IPW methods, the following missing mechanisms are assumed to produce missing data in random variable Y:

$$\Pr(\delta_i = 1) = 0.79, \quad i = 1, 2, \dots, n, \tag{25}$$

$$\Pr(\delta_i = 1 \mid X_i = x_i) = \frac{1}{1 + e^{-0.75 - 0.75x_i}}, \quad i = 1, 2, \dots, n. \tag{26}$$

Moreover, we consider the following random right censoring time and missing mechanism for data, when we estimate parameters by CCC and $CIPW$ methods:

$$X_{c_i} \sim \text{Uniform}(1.2, 1.5), \quad i = 1, 2, \dots, n, \tag{27}$$

$$Y_{c_i} \sim \text{Uniform}(1.2, 1.8), \quad i = 1, 2, \dots, n, \tag{28}$$

$$\Pr(\delta_i = 1) = 0.94, \quad i = 1, 2, \dots, n, \tag{29}$$

$$\Pr(\delta_i = 1 \mid X_i = x_i) = \frac{1}{1 + e^{-2x_i - 1.4}}, \quad i = 1, 2, \dots, n. \tag{30}$$

By the above patterns, we will have the same percentage of incomplete data in any cases. In addition, in this study, the rate of incomplete data for some methods have been increased from 10 percent to 21 percent in comparison to that of Study 1. The results of Study 2 are given in Table 3 for the 5000 Monte Carlo repetitions.

From Table 3, one can conclude that the IPW based methods have a good performance in comparison to those of CC based methods. These good performances of the IPW based methods are obvious when the missing mechanism is MAR. Also, the performance of the IPW based methods is not too different with respect to the change of mechanisms. By comparison of performances of CC method with CCC method and IPW method with $CIPW$ method, one can conclude that missing data have a bigger effect than that of censored data on estimates. Also, by increasing the sample sizes, the accuracy of the estimates increased.

5. Applications

This paper is organized based on the Bivariate Rayleigh distribution, but one can use similar methods to analyze data with other bivariate distributions. Our proposed methods can be applied in complete and incomplete data cases. We used these methods to infer about fully observed data or a dataset with right-censored observations and missing data. Also, we used our methods to infer soccer data. However, we can use this method in similar sports, in economic data to investigate investments in heterogeneous assets, in clinical researches to study the mortality risk of lung cancer for men versus women, in industrial studies to calculate the lifetime of different types of steels and etc. Indeed, we can use these methods in bivariate parametric inferences.

Table 3. Study 2 by using two different sample sizes, 50 and 100.

Methods	MM	RID in X	RID in Y	Bias	MSE	CIW	CP
C	–	0.106	0.106	0.0498	0.0046	0.2253	0.9154
				(0.0350)	(0.0021)	(0.1600)	(0.9210)
CC	MCAR	0.000	0.210	0.0669	0.0087	0.2644	0.8826
				(0.0440)	(0.0041)	(0.1889)	(0.9242)
IPW	MCAR	0.000	0.210	0.0666	0.0087	0.2344	0.8532
				(0.0436)	(0.0041)	(0.1675)	(0.8952)
CC	MAR	0.000	0.211	0.0691	0.0095	0.2668	0.8858
				(0.0468)	(0.0048)	(0.1905)	(0.9206)
IPW	MAR	0.000	0.211	0.0650	0.0083	0.2351	0.8666
				(0.0430)	(0.0040)	(0.1678)	(0.9030)
CCC	MCAR	0.106	0.102	0.0531	0.0062	0.2301	0.9166
				(0.0338)	(0.0025)	(0.1632)	(0.9444)
CIPW	MCAR	0.106	0.102	0.0578	0.1014	0.2265	0.9114
				(0.0333)	(0.0025)	(0.1590)	(0.9440)
CCC	MAR	0.106	0.102	0.0537	0.0066	0.2325	0.9216
				(0.0334)	(0.0025)	(0.1649)	(0.9530)
CIPW	MAR	0.106	0.102	0.0577	0.0064	0.2245	0.9138
				(0.0333)	(0.0024)	(0.1591)	(0.9480)

Notes: where results of the study by using sample size 100 are given in brackets. This study contains approximately 21 percentage of incomplete data (MM: Missing Mechanism; RID: Rate of Incomplete Data; CIW: Confidence Interval Width; CP: Coverage Probability).

6. Real data study

Table 4 represents football dataset for the La Liga league in the seasons 2013–2014, 2014–2015 and 2015–2016[1]. The time of the first goal of assumed player (Cristiano Ronaldo) in each match is shown with Y (stress component) and the time of the second goal of the match which is scored by any player is shown with X (strength component). In

Table 4. Real data: Time of the first goal of the player of interest (Y) and the second goal of the match by any player (X) in Real Madrid's matches in La Liga for three seasons.

2013–2014				2014–2015				2015–2016			
X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
26	90^+	25	10	30	88	26	83^+	90^+	90^+	9	9
90^+	90^+	74	56	10	·	2	·	39	90^+	71	90^+
45	45	73	75^+	27	27	37	·	17	7	12	12
39	64	25	·	35	28	19	90^+	90^+	90^+	60	90^+
19	33	27	·	21	27	74	90^+	67	90^+	10	3
51	51	72	·	39	39	70	70	90^+	90^+	33	33
90^+	90^+	28	82	40	4	65	51	83	90^+	90^+	90^+
61	90	49	11	37	12	90^+	90^+	30	30	38	34
90	90	90^+	23	36	36	39	90^+	23	8	50	50
79	90^+	20	55	32	1	32	32	14	14	87	90^+
27	32	19	14	41	82	31	31	22	90^+	64	64
31	3	55	15	44	44	72	69	39	90^+	56	85
26	12	66	·	84	90^+	32	20	82	82	18	19
61	3	53	·	66	35	68	90	16	38	40	90
36	·	85	9^+	40	80	17	90^+	90^+	90^+	69	90^+
39	62^+	52	6	53	15	48	90^+	10	30	14	·
34	40	59	90	29	90^+	38	35	49	42	90^+	·
82	82	63	·	68	64	25	90^+	45	90^+	42	28
90^+	90^+	86	·			74	58	22	90^+	25	8
						23	12				

Also, censored data are considered using '+' sign and missed data are considered using '·' sign.

this case, (X, Y) is a bivariate continuous random vector for which any cases may happen, this means that $X < Y$, $X > Y$ and $X = Y$ are possible. Also, the matches that the special player was absent are considered as missing data. On the other hand, the matches that an experiment has been failed or the matches that the player of interest has been substituted with no score are considered as censored data.

Some of the values are right censoring and some of the information of random variable Y are missing. Therefore, we use the CCC and CIPW methods to estimate stress-strength criterion. Maximum likelihood estimates of stress-strength criterion and its 95% confidence interval based on the *CCC* method are 0.2586 and $(0.1757, 0.3416)$, respectively. Also, MLE and confidence interval based on the *CIPW* for this criterion are 0.2457 and $(0.1686, 0.3228)$, respectively.

By the results of *CIPW* method, we can say that our assumed player receives a goal by the 0.2457 chance in any matches before than any other player. In any soccer matches, at least 22 players make role. This means that our assumed player has a big role in the results of any matches as he scores the first goal of any matches by the 0.2457 chance. Also, one can use this criterion to check the performance of the other soccer players.

Figure 1 shows the marginal distribution of fitted model over the fully observed data. Where the ML estimation of $(\lambda_1, \lambda_2, \lambda_3)$, based on *CCC* and *CIPW* methods are $(0.00022, 0.00010, 0.00007)$ and $(0.00021, 0.00010, 0.00008)$, respectively.

Figure 2 shows the marginal distribution of variables versus empirical distributions. Our assumed components are dependent and empirical distribution is drawn by independence assumption. These caused some difference between the marginal distributions of bivariate Rayleigh distribution and the empirical distributions. This difference is obvious in the

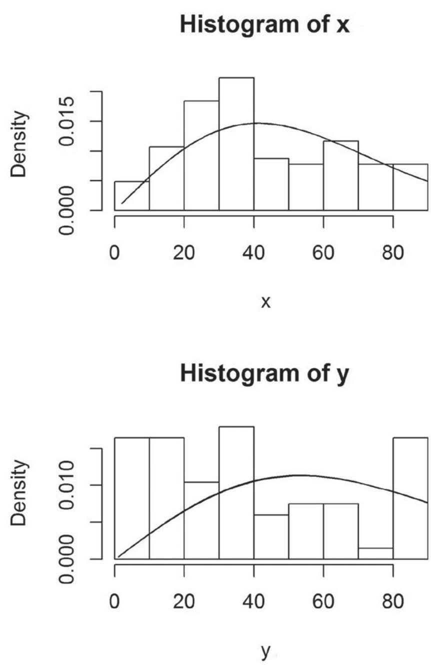

Figure 1. Marginal fitted densities of bivariate Rayleigh distribution by the maximum likelihood method for the football dataset.

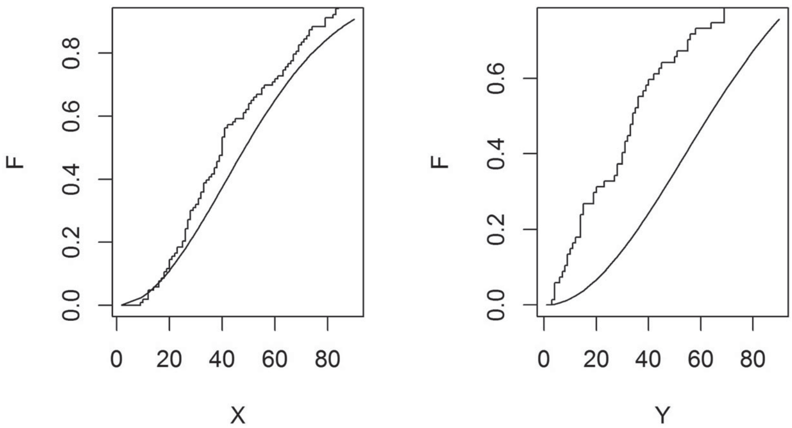

Figure 2. Marginal distributions versus the empirical distributions.

marginal distribution of component Y , but it is not more different for X component. However, these shapes give us an insight about the good performance of the bivariate Rayleigh distribution which is fitted to the football dataset under the censored and missing data problems.

7. Discussion and conclusion

Based on the statistics and soccer professionals experiences our assumed player is one of the best forward players in history. As mentioned in the last section, he may score a goal in any match with 0.2457 probability before the other players. Therefore, to study the performance of the other forward players, one can estimate stress-strength probability of that player and compare it with the Cristiano Ronaldo's stress-strength probability. Obviously, if this stress-strength probability is closed to 0.2457 or bigger than that, one can say the player is a top forward player. Also, if this stress-strength probability is less than 0.2457, one can say that the player is not as a good forward player.

To obtain Fisher information matrix based on the different methods, we need to obtain expectations such as $E(\delta C_x C_y h(X,Y))$ where $h(\cdot)$ is a real-valued function of X and Y . Moreover, it is assumed that C_x and C_y are independent of each other and their distributions are not functions of X and Y. By this assumption, we can conclude that:

$$E(\delta C_x C_y h(X,Y)) = E(C_x)E(C_y)E(\delta h(X,Y)). \tag{31}$$

Also, under the MCAR assumption, we can conclude that:

$$E(\delta h(X,Y)) = E(\delta)E(h(X,Y)). \tag{32}$$

Moreover, under the MAR assumption, we can conclude that:

$$E\left(\frac{\delta}{\pi(X,Y)}h(X,Y)\right) = E\left(E\left(\frac{\delta}{\pi(X,Y)}h(X,Y)\right)\Big| X=x, Y=y\right) = E(h(X,Y)), \tag{33}$$

where in the above equation, first equality follows from the conditional expectation property and the second equality follows from the MAR mechanism property. In the IPW method, we have used a nonparametric method to estimate unknown function $\pi(\cdot)$. In this method, choosing kernel function is not very critical but choosing the bandwidth parameter is very important. How to select the bandwidth parameter is the open problem of these statistical studies. We have chosen to use the Cross Validation method to estimate unknown parameter h.

In this paper, unlike Tong [15] and Constantine *et al.* [2], we assumed that variables could be dependent. Meintanis [11] and Pak *et al.* [12] have used full observations for their studies and they have ignored most of the incomplete data that were effective in the analyzes. But unlike them, we have used all of the information and we have not ignored any data. It can be said that in spite of other researchers which they ignored the effects of missing data in studies, we have used all of the information in the analysis.

In this paper, we have considered the ability of scoring goals in the matches by players. This ability may depend on some factors such as matches location, team quality, match period and etc. Distinguishing effective factors on the ability of scoring goals needs some other statistical methods such as linear models and factor analysis. Our aim was not to identify these factors. our aim was to estimate the stress-strength probability of a player (ability to score the first goal before than any player) under any conditions.

Note

1. For data source see also following site: 'http://www.espnfc.com/spanish-primera-division/15/statistics/performance'.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] A. Al-khedhairi, A. Sarhan, and L. Tadj, *Estimation of the generalized Rayleigh distribution parameters*, Int. J. Reliab. Appl. 7 (2007), pp. 1–12.
[2] K. Constantine, S. Tse, and M. Karson, *The estimation of $P(Y < X)$ in gamma case*, Comm. Stat. – Simul. Comput. 15 (1986), pp. 365–388.
[3] H.A. Dina, *Bivariate generalized Rayleigh distribution*, J. Appl. Sci. Res. 9 (2013), pp. 5403–5411.
[4] T.S. Ferguson, *A Course in Large Sample Theory*, Chapman and Hall, New York, 1996.
[5] T.G. Horvitz and T.J. Thompson, *A generalization of sampling without replacement from a finite universe*, J. Amer. Statist. Assoc. 47 (1952), pp. 663–685.
[6] P.K. Jana, *Estimation for $P(Y < X)$ in the bivariate exponential case due to Marshal and Olkin*, J. Indian Statist. Assoc. 32 (1994), pp. 35–37.
[7] E.S. Jeevanand, *Bayes estimate of the reliability under stress-strength model for the Marshal and Olkin bivariate exponential distribution*, IAPQR Trans. 23 (1998), pp. 133–136.
[8] C.S. Kakade, D.T. Shirke, and D. Kundu, *Inference for $P(Y < X)$ in exponentiated Gumbel distribution*, J. Stat. Appl. 3 (2008), pp. 121–133.
[9] D. Kundu and R.D. Gupta, *Estimation of $P(Y < X)$ for generalized exponential distributions*, Metrika 61 (2005), pp. 291–308.
[10] R.J.A. Little and D.B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed., Wiley-Interscience, New York, 2002.
[11] S.G. Meintanis, *Test of fit for Marshal-Olkin distributions with applications*, J. Stat. Plan. Inference 137 (2007), pp. 3954–3963.

[12] A. Pak, N.B. Khoolenjani, and A.A. Jafari, *Inference on $P(Y < X)$ in bivariate Rayleigh distribution*, Comm. Statist. Theory Methods 43 (2014), pp. 4881–4892.
[13] D.B. Rubin, *Inference and missing data*, Biometrika 63 (1976), pp. 581–592.
[14] J. Shao, *Mathematical Statistics*, 2nd ed., Texts in Statistics, Springer, New York, 2003.
[15] H. Tong, *On the estimation of $P(Y < X)$ for exponential distributions*, IEEE Trans. Reliab. 26 (1977), pp. 54–56.
[16] S. Wang and C.Y. Wang, *A note on kernel assisted estimators in missing covariate regression*, Statist. Probab. Lett. 55 (2001), pp. 439–449.
[17] L.P. Zhao and S. Lipsitz, *Design and analysis of two stage-studies*, Stat. Med. 11 (1992), pp. 769–782.

Appendices

Appendix 1. Proof of Theorem 3.1

Proof of Theorem 3.1 follows from the asymptotic properties of maximum likelihood estimates and Multivariate Central Limit Theorem under some regularity conditions given in Wang and Wang [16]. Also, in parts (d) and (f), using Slutsky's theorem, we have the results based on the following equations, respectively:

$$\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_i} \boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \boldsymbol{\psi}_F(\boldsymbol{\lambda} \mid x_i, y_i) + o_P(\mathbf{1}), \tag{A1}$$

$$\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_i} \boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \boldsymbol{\psi}_C(\boldsymbol{\lambda} \mid x_i, y_i) + o_P(\mathbf{1}), \tag{A2}$$

where $o_P(\mathbf{1})$ is the probability notation of order and its dimension is equal to dimension of estimating equations. In our case its dimension is equal to 3.

Appendix 2. Proof of Theorem 3.2

The results of Theorem 3.2 are concluded from Theorem 3.1 and the invariance property of asymptotically consistent normal estimates under continuous transformations. To see more similar details about Theorems 3.1 and 3.2, see also Shao [14] and Ferguson [4].