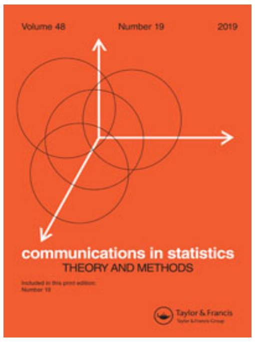

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Lin–Wong divergence and relations on type I censored data

A. Pakgohar, A. Habibirad & F. Yousefzadeh

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2019, VOL. 48, NO. 19, 4804–4819
https://doi.org/10.1080/03610926.2018.1494839

Lin–Wong divergence and relations on type I censored data

A. Pakgohar[a], A. Habibirad[a], and F. Yousefzadeh[b]

[a]Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran; [b]Department of Statistics, School of Mathematical Sciences and Statistics, University of Birjand, Birjand, Iran

ABSTRACT
Divergence measures are statistical tools designed to distinguish between the information provided by distribution functions of *f*(*x*) and *g*(*x*). The magnitude of divergence has been defined using a variety of methods such as Shannon entropy and other mathematical functions through a history of more than a century. In the present study, we have briefly explained the Lin–Wong divergence measure and compared it to other statistical information such as the Kullback-Leibler, Bhattacharyya and χ^2 divergence as well as Shannon entropy and Fisher information on Type I censored data. Besides, we obtain some inequalities for the Lin–Wong distance and the mentioned divergences on the Type I censored scheme. Finally, we identified a number of ordering properties for the Lin–Wong distance measure based on stochastic ordering, likelihood ratio ordering and hazard rate ordering techniques.

ARTICLE HISTORY
Received 31 July 2017
Accepted 13 June 2018

KEYWORDS
Bhattacharyya; Chi square; Distance measure; Fisher Information; Inequality; Kullback-Leibler; Lin–Wong; Stochastic Ordering

1. Introduction

Entropy is generally a measure of irregularity. This measure of variability for continuous variables is defined as

$$H_g(X) = -\int_{-\infty}^{\infty} g(x)\ \log g(x)\ dx. \tag{1}$$

Shannon (1948) applied this concept to the information theory for discrete variables, while an immediate extension leads to its continuous analog called the differential entropy. That is the differential entropy of a continuous random variable X with distribution density functions of $g(x)$ which was introduced by Cover and Thomas (1991). For convenience, we use $\log_e = \ln$ function when referring to all logarithm incidents in the rest of this article. Shannon entropy measure was extended into other fields and was implemented within significant applications. For instance, the Kullback-Leibler (KL) divergence measure which was defined by Kullback and Leibler (1951) is a well known measure of difference between probability distributions or information divergence, and it is based on the Shannon entropy. Furthermore, various types of divergence measures

CONTACT A. Habibirad ahabibi@um.ac.ir Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

are being used in addition to *KL* measure. One of them is the Lin–Wong (*LW*) divergence measure which was introduced by Lin and Wong (1990). However, some data distribution difference measures exist which were introduced before the *KL* divergence measure or the Shannon entropy. For example, the Bhattacharyya (*B*) and χ^2 distance measures. For more details, see Burgess (1928), Ferger (1931) and Deza and Deza (2009).

Here, we mainly focus on the *LW* divergence which is defined as

$$D_{\mathrm{LW}}(g||f) = \int_{-\infty}^{\infty} g(x) \ \ln \frac{2g(x)}{g(x)+f(x)} \ \mathrm{d}x. \tag{2}$$

However, the *LW* measure is in close relationship with the *KL*, and it is generally characterized and presented by within the relevant literature as a new directional distance measure which avoids the pitfalls in which observed with the *KL*. For instance the *KL* measure can be calculated infinity. For example, suppose that $\chi = (0, 1)$, be the unit interval and two densities $g(x) = 1$, and $f(x) = \frac{e^{-x}}{0.148}$, implies $D_{\mathrm{KL}}(g||f) = \infty$. In fact, many desirable features such as being finite and non negative, being bounded on one side (or two sides under certain conditions) are pointed out by the proposing authors. *LW* measure of divergence was further developed and extended by Lin (1991), Shioya and Da-Te (1995) as well Jain and Saraswat (2012). Furthermore, Abbasnejad et al. (2012) proposed an improvement of goodness of fit measure for exponential distributions based on this divergence as well, Khalili et al. (2017) studied some properties of *LW* on the past lifetime data.

In particular, there are studies focused on some specific data including stated calibration and detection limits. Environmental data, for example, often include values reported as these calibrations, such as a below detection limit along with the stated detection limit. The reader is referred to Kleinbaum and Klein (2006) and Helsel (2010). A sample contains censored observations if the information have been about some of the observations which they are below or above a specified value.

The density function for a Type I censored variable $\min(X, C)$, where C is the censoring point and is assumed to be a constant, is defined as

$$f_C(x) = \begin{cases} f(x) & \text{if} \quad x<C \\ \bar{F}(C) & \text{if} \quad x \geq C, \end{cases} \tag{3}$$

in which, $f(x), F(x), \bar{F}(x) = 1-F(x)$, represent density, distribution and survival functions, respectively.

The rest of this article is structured as follows. Section 2, expresses the *LW*, *KL*, *B* and χ^2 divergences on Type I censored data and review the properties of *LW* divergence measure on the mentioned censored data. Section 3, provides general introduction of the *LW* divergence measure and some other prominent measures of distance, along with a review of relations between them. In Section 4, we obtain some bounds for them from the well-known inequalities when applied on Type I censored data. Section 5, includes the *LW* divergence stochastic order properties and the bounds of it. Section 6, includes discussion and our conclusions.

2. *LW* divergence measure properties on type I censored data

Various types of distance measure, same as the *LW*, *KL*, *B* and χ^2 divergences are defined within the realms of information theory and statistics which are not in a same class but some relationships can be found between them. In following section we denote these divergences on Type I censored data. First at all, the LW divergence is expressed on Type I censored data as

$$D_{LW}^{C-I}(g||f) = \int_{-\infty}^{\infty} g_C(x)\ln\frac{2g_C(x)}{g_C(x)+f_C(x)}dx$$
$$= \int_{-\infty}^{C} g(x)\ln\frac{2g(x)}{g(x)+f(x)}dx + \bar{G}(C)\ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)}, \tag{4}$$

where $g(x)$ and $f(x)$ are probability density functions of $F(x)$ and $G(x)$, respectively, and $G(x)$ is absolutely continuous with respect to $F(x)$, besides, $\bar{G}(x)$ and $\bar{F}(x)$, represent survival functions respectively.

Obviously $\lim_{C\to\infty} D_{LW}^{C-I}(g||f) = D_{LW}(g||f)$.

Park and Shin (2014) regarded the KL divergence measure for Type I censored variable as

$$D_{KL}^{C-I}(g||f) = \int_{-\infty}^{C} g(x)\ln\frac{g(x)}{f(x)}dx + \bar{G}(C)\ln\frac{\bar{G}(C)}{\bar{F}(C)}. \tag{5}$$

Park (2016) expressions for the KL information on a mixer of Type I and II censoring scheme considering order statistics.

Finally, the Bhattacharyya (B) and χ^2 as a well-known distance measures can be defined on Type I censored data as

$$D_{B}^{C-I}(g||f) = \int_{-\infty}^{C} \sqrt{g(x)f(x)}\ dx + \sqrt{\bar{G}(C)\bar{F}(C)}. \tag{6}$$

$$D_{\chi^2}^{C-I}(g||f) = \int_{-\infty}^{C} \frac{g(x)^2}{f(x)}\ dx + \frac{\bar{G}(C)^2}{\bar{F}(C)} - 1. \tag{7}$$

Theorem 2.1, states the LW divergence measure on Type I censored data is increasing on C.

Theorem 2.1. Suppose that X is a continuous real-valued function on Type I censored variable. Then, the LW divergence measure is a monotonous increasing function of C.

Proof.

$$\frac{\partial D_{LW}^{C-I}(g||f)}{\partial C} = g(C)\left(\ln\frac{2g(C)}{g(C)+f(C)} - \ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)} - 1\right)$$
$$+\frac{(g(C)+f(C))\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)} = g(C)(Y(C) - \ln Y(C) - 1),$$

where $Y(C) = \frac{(g(C)+f(C))}{h_G(C)(\bar{G}(C)+\bar{F}(C))}$, and $h_G(C) = \frac{g(C)}{\bar{G}(C)}$. It would mentioned that

$$Y(C) - \ln Y(C) - 1 \geq 0 \qquad \forall \quad Y(C)>0. \tag{8}$$

In fact $Y(C) - \ln Y(C) - 1$, attains it's minimum at $Y(C) = 1$, with value zero.

By using the (8), we obtain that

$$\frac{\partial D_{LW}^{C-I}(g||f)}{\partial C} = g(C)(Y(C) - \ln Y(C) - 1) \geq 0. \tag{9}$$

Inequality (9) holds by (8) and $g(C) \geq 0$. Hence, the proof is complete. □

3. The *LW* divergence in relation with other information on type I censored data

A number of studies have addressed those relationships, for instance Kapur (1984), Lin (1991), Shioya and Da-te (1995), Dragomir (2003), Kumar and Taneja (2006). The divergence issue over this kind of data has been discussed by Joarder et al. (2011), Pakyari and Balakrishnan (2013), Park and Shin (2014), Pakyari and Resalati Nia (2017) amongst others. It is clear that the magnitude of H is less than that of the B. Dragomir (2003) showed the $D_{L-2} = B$ relationship, where $D_{L-2}(g||f) = \int \sqrt{g(x)f(x)}dx$.

Accordingly, we derive the relation between LW, KL and H divergence measures.

Trough Theorem 3.1, we obtain a boundary for the values of LW divergence using KL values obtained when applied on Type I censored data. Then we address the relation between LW divergence and Fisher information measures in the context of Type I censored data within the Theorem 3.2. Theorems 3.3 and 3.4 lead the relationship between LW divergence corresponding to B and χ^2 distance measures on Type I censored data.

Theorem 3.1. There is a relation between LW and KL divergence measures on Type I censored data which can be expressed as

$$D_{LW}^{C-I}(g||f) \leq \ln \frac{4}{3 - D_{KL}^{C-I}(g||f) + e^{-D_{KL}^{C-I}(g||f)}}. \tag{10}$$

Proof. Take the term $\ln \frac{2}{1+e^X}$, which has the property of concavity. From the Jensen's inequality we would have

$$\int_{-\infty}^{C} g(x) \ln \frac{2g(x)}{g(x)+f(x)} dx = \int_{-\infty}^{C} g(x) \ln \frac{2}{1+e^{\ln \frac{f(x)}{g(x)}}} dx \leq \ln \frac{2}{1+e^{\int_{-\infty}^{C} g(x)\ln \frac{f(x)}{g(x)} dx}}.$$

Similarly,

$$\bar{G}(C) \ln \frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)} \leq \ln \frac{2}{1+e^{\bar{G}(C)\ln \frac{\bar{F}(C)}{\bar{G}(C)}}}.$$

Therefore, we can write

$$\begin{aligned} D_{LW}^{C-I}(g||f) &\leq \ln \frac{2}{1+e^{\int_{-\infty}^{C} g(x)\ln \frac{f(x)}{g(x)} dx}} + \ln \frac{2}{1+e^{\bar{G}(C)\ln \frac{\bar{F}(C)}{\bar{G}(C)}}} \\ &= \ln \frac{4}{1+e^{\int_{-\infty}^{C} g(x)\ln \frac{f(x)}{g(x)} dx} + e^{\bar{G}(C)\ln \frac{\bar{F}(C)}{\bar{G}(C)}} + e^{-D_{KL}^{C-I}(g||f)}}. \end{aligned}$$

At the other hand, from the Mac Loren's expansion for $e^x \geq 1 + x$, we can get

$$e^{\int_{-\infty}^{C} g(x)\ln \frac{f(x)}{g(x)} dx} + e^{\bar{G}(C)\ln \frac{\bar{F}(C)}{\bar{G}(C)}} \geq 2 - D_{KL}^{C-I}(g||f).$$

So, the inequality (10) is established and the proof is complete. □

Remark 3.1. Similar to Theorem 3.1 there is a relation between LW and KL divergence measures on general data as follow

$$D_{LW}(g\|f) \leq \ln \frac{2}{1+e^{-D_{KL}(g\|f)}}.$$

Clearly, using the elementary geometric-arithmetic inequality, there is a relationship between LW distance value and KL divergence on Type I censored variable as follow

$$D_{LW}^{C-I}(g\|f) \leq \frac{1}{2} D_{KL}^{C-I}(g\|f). \tag{11}$$

The Fisher information is a way of measuring the amount of information that a random variable X, and defined as (see Lehmann and Casella 2006)

$$I(\theta) = \int \left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)^2 f(x;\theta)dx.$$

A relationship exists between the KL divergence and the Fisher information. Kullback (1959) described this relation as the following

$$D_{\mathrm{KL}}\big(f(x;\theta)\|f(x;\theta+\Delta\theta)\big) \approx \frac{(\Delta\theta)^2}{2} I(\theta).$$

Ferentinos and Papaioannou (1981), Gertsbakh (1995), Dabak and Johnson (2002), Park and Shin (2014) calculated the relationship between KL and Fisher information. Here we address a connection for LW divergence and the Fisher information measures on Type I censored variable.

The Fisher information about θ in X on Type I censored data is expressed as

$$\begin{aligned} I(\theta)^{C-I} &= \int_{-\infty}^{\infty} f_C(x;\theta)\left(\frac{\partial}{\partial\theta}\ln f_C(x;\theta)\right)^2 dx \\ &= \int_{-\infty}^{C} f(x;\theta)\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2 dx + \bar{F}(C;\theta)\left(\frac{\partial}{\partial\theta}\ln \bar{F}(C;\theta)\right)^2. \end{aligned} \tag{12}$$

It can easily be shown that

$$D_{\mathrm{KL}}^{C-I}\big(f(x;\theta)\|f(x;\theta+\Delta\theta)\big) \approx \frac{(\Delta\theta)^2}{2}\left(I(\theta)^{C-I}\right). \tag{13}$$

Here, we examine the relation between LW divergence and Fisher information measures in the context of Type I censored data.

Theorem 3.2. There is a relation between Fisher information and the LW divergence on Type I censored data as below

$$D_{LW}^{C-I}\big(f(x;\theta)\|f(x;\theta+\Delta\theta)\big) \approx \frac{(\Delta\theta)^2}{8} I(\theta)^{C-I}. \tag{14}$$

Proof. We would have

$$\int_{-\infty}^{C} f(x;\theta) \ln \frac{2f(x;\theta)}{f(x;\theta)+f(x;\theta+\Delta\theta)} \mathrm{d}x \approx \int_{-\infty}^{C} f(x;\theta) \ln \frac{f(x;\theta)}{f\left(x;\theta+\frac{\Delta\theta}{2}\right)} \mathrm{d}x.$$

Similarity, we can obtain

$$\bar{F}(C;\theta)\ln\frac{2\bar{F}(C;\theta)}{\bar{F}(C;\theta)+\bar{F}(C;\theta+\Delta\theta)}\mathrm{d}x\approx\bar{F}(C;\theta)\ln\frac{\bar{F}(C;\theta)}{\bar{F}\left(C;\theta+\frac{\Delta\theta}{2}\right)}\mathrm{d}x.$$

Hence, we have

$$D_{LW}^{C-I}(f(x;\theta)||f(x;\theta+\Delta\theta))\approx D_{KL}^{C-I}\left(f(x;\theta)||f\left(x;\theta+\frac{\Delta\theta}{2}\right)\right). \tag{15}$$

Because of by taking $\frac{\Delta\theta}{2}$ in (13) we would have

$$D_{LW}^{C-I}(f(x;\theta)||f(x;\theta+\Delta\theta))\approx D_{KL}^{C-I}\left(f(x;\theta)||f\left(x;\theta+\frac{\Delta\theta}{2}\right)\right)\approx\frac{(\Delta\theta)^2}{8}I(\theta)^{C-I}. \tag{16}$$

Comparing (13) and (14) states LW and KL divergences between $f_{\theta+\Delta\theta}$ and f_θ, is proportional to Fisher information at θ and $\theta+\Delta\theta$, where $\Delta\theta$ is a perturbation. Theorem 3.2 illustrates in effect that the divergence between two distributions with respect to LW and KL divergences is the same as that with respect to Fisher information on Type I censored case whenever the size difference between the two parameters $\theta+\Delta\theta$, and θ, is a perturbation.

Let $\phi:[0,)\to R$ be a convex function. Then Csiszár (1963) introduced the $\phi-$*divergence* functional as a generalized measure of information on the set of probability distribution. The $\phi-$*divergence* can be written as

$$D_\phi(g,f)=\int_{-\infty}^{\infty}f(x)\ \phi(t(x))\ dx, \tag{17}$$

where $t(x)=\frac{g(x)}{f(x)}$.

Some properties of $\phi-$*divergence* are as $\phi''(x)\geq 0$ and $\phi(1)=0$.

In following, we obtain a boundary for LW based on B distance considering $\phi-$*divergence* definition.

Theorem 3.3. There is a relationship between LW and B distance measures on Type I censored data as follow

$$D_{LW}^{C-I}(g||f)\geq D_B^{C-I}(g||f)-1. \tag{18}$$

Proof. Let $\phi_1(t(x))=t(x)\ln\frac{2t(x)}{t(x)+1}$ and $\phi_2(t(x))=1-\sqrt{t(x)}$. Because of $\phi_1(1)=\phi_2(1)=0$, besides $\phi''_1(t(x))$ and $\phi''_2(t(x))$, be greater than or equal to zero, we denote them as a ϕ divergence functional. Now take $\phi_3(t(x))=\phi_1(t(x))+\phi_2(t(x))$, that implies, $\phi_3(t(x))$ is functional Csiszár's measure. Therefore, by substituting $t(x)=\frac{g(x)}{f(x)}$, we can get

$$\int_{-\infty}^{C}g(x)\ln\frac{2g(x)}{g(x)+f(x)}\ dx\geq\int_{-\infty}^{C}\sqrt{g(x)f(x)}\ dx-F(C). \tag{19}$$

Similarly, by taking $t(C) = \frac{\bar{G}(C)}{\bar{F}(C)}$ in $\int_C^\infty f(x)\gamma(t(C))\ dx$, one can write

$$\bar{G}(C)\ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)} \geq \sqrt{\bar{G}(C)\bar{F}(C)} - \bar{F}(C). \tag{20}$$

Summing (19) and (20) gives (18). □

Theorem 3.4. Let X and Y be a continuous random variable on Type I censored data. Then there is a relationship between LW and χ^2 distance measures as follow

$$D_{LW}^{C-I}(g\|f) \leq \frac{D_{\chi^2}^{C-I}(g\|f)-1}{2}. \tag{21}$$

Proof. From Jensen inequality in view of ln function, besides, using (7), we have

$$D_{KL}^{C-I}(g\|f) \leq \ln\int_{-\infty}^{C} g(x)\frac{g(x)}{f(x)}\ dx + \ln\bar{G}(C)\frac{\bar{G}(C)}{\bar{F}(C)}$$

$$\leq \int_{-\infty}^{C}\frac{g^2(x)}{f(x)}\ dx + \frac{\bar{G}^2(C)}{\bar{F}(C)} - 2 = D_{\chi^2}^{C-I}(g\|f) - 1.$$

So, considering (11), the proof is complete. □

Remark 3.2. Similarity to Theorem 3.4 we have on general situation

$$D_{LW}(g\|f) \leq \frac{D_{\chi^2}(g\|f)}{2}.$$

4. Bounds for the *LW* divergence on type I censored data

In following section, we obtain some bounds for *LW* divergence measure on general and Type I censored scheme. The mentioned inequalities are given as Cassels and Diaz–Metcalf for integral version in which can be as a reverse Caushy-Schwarz inequality in various frameworks, e.g. integrals or inner product spaces as Hilbert space. For instance see Moslehian et al. (2011).

In 1950, The Cassels inequality can be expressed as (see Watson et al. 1997)

$$\frac{\int w(x)a^2(x)\ dx.\int w(x)b^2(x)\ dx}{\left(\int w(x)a(x)b(x)\ dx\right)^2} \leq \frac{(m+M)^2}{4mM}, \tag{22}$$

where $a(x)>0,\ b(x)>0,\ and\ \ w(x) \geq 0$, with $m = \min\left\{\frac{a(x)}{b(x)}\right\}$ and $M = \max\left\{\frac{a(x)}{b(x)}\right\}$.

Theorem 4.1. Let X and Y be continuous random variables on Type I censoring scheme with g(x) and f(x) probability mass functions respectively. Then

$$D_{LW}^{C-I}(g\|f) \leq \ln 2 - \frac{2m}{(m+M)}\sqrt{mM}\left(\sqrt{\bar{G}(C)} + \sqrt{G(C)}\right), \tag{23}$$

where $m = \min\left\{\sqrt{G(C)}\ln\frac{g(x)+f(x)}{g(x)}, \sqrt{\bar{G}(C)}\ln\frac{\bar{G}(C)+\bar{F}(C)}{\bar{G}(C)}\right\},\ \forall x \leq C,$ *and* $M = \max\left\{\sqrt{G(C)}\ln\frac{g(x)+f(x)}{g(x)}, \sqrt{\bar{G}(C)}\ln\frac{\bar{G}(C)+\bar{F}(C)}{\bar{G}(C)}\right\}.$

Proof. Considering $m = \min\{m_1, m_2\}$, $\forall x \le C$, and $M = \max\{M_1, M_2\}$, in which, $m_1 = \min\left\{\sqrt{G(C)}\ln\frac{g(x)+f(x)}{g(x)}\right\}$ $\forall x \le C, M_1 = \max\left\{\sqrt{G(C)}\ln\frac{g(x)+f(x)}{g(x)}\right\}$, $m_2 = M_2 = \sqrt{\bar{G}(C)}\ln\frac{\bar{G}(C)+\bar{F}(C)}{\bar{G}(C)}$ and taking $w(x) = g(x)$, $a(x) = \ln\frac{g(x)+f(x)}{g(x)}$, and $b(x) = \frac{1}{\sqrt{G(C)}}$, in (22) we have

$$\frac{G(C)\int_{-\infty}^{C} g(x)\left(\ln\frac{g(x)+f(x)}{g(x)}\right)^2 dx}{\left(\int_{-\infty}^{C} g(x)\ln\frac{g(x)+f(x)}{g(x)}\ dx\right)^2} \le \frac{(m+M)^2}{4mM}.$$

So

$$\begin{aligned}\left(\int_{-\infty}^{C} g(x)\ln\frac{g(x)+f(x)}{g(x)}\ dx\right)^2 &\ge \frac{4mM.G(C)}{(m+M)^2}\int_{-\infty}^{C} g(x)\left(\ln\frac{g(x)+f(x)}{g(x)}\right)^2 dx\\ &\ge \frac{4mM.G(C)}{(m+M)^2}\int_{-\infty}^{C} g(x)\left(\frac{m}{\sqrt{G(C)}}\right)^2 = \frac{4m^3M.G(C)}{(m+M)^2}.\end{aligned}$$

Therefore,

$$\int_{-\infty}^{C} g(x)\ln\frac{g(x)+f(x)}{g(x)}\ dx \ge \frac{2m}{(m+M)}\sqrt{m.M.G(C)}. \tag{24}$$

Similarly,

$$\bar{G}(C)\ln\frac{\bar{G}(C)+\bar{F}(C)}{\bar{G}(C)} \ge \frac{2m}{(m+M)}\sqrt{m.M.\bar{G}(C)}. \tag{25}$$

Hence, from (24) and (25) we have

$$\begin{aligned}&\int_{-\infty}^{C} g(x)\ln\frac{g(x)}{g(x)+f(x)}dx + \bar{G}(C)\ln\frac{\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)}\\ &\le -\frac{2m}{(m+M)}\sqrt{mM}\left(\sqrt{\bar{G}(C)}+\sqrt{G(C)}\right).\end{aligned}$$

Thus, (23) is hold. □

Remark 4.1. Clearly, for general scheme we can get

$$D_{LW}(g||f) \le \ln 2 - \frac{2m}{(m+M)}\sqrt{mM}, \tag{26}$$

where $0<m \le \ln\frac{g(x)+f(x)}{g(x)} \le M$.

Diaz and Metcalf (1963) illustrated the Diaz–Metcalf inequality for continuous version of real valued functions $t_1(x)$ and $t_2(x)$ (never zero) on the finite interval as follow

$$\int p(x)t_1^2(x)\ dx + ab\int p(x)t_2^2(x)\ dx \le (a+b)\int p(x)t_1(x)t_2(x)\ dx, \tag{27}$$

in which $a \le \frac{t_1(x)}{t_2(x)} \le b$, and $p(x)>0$ with $\int p(x)\ dx = 1$.

Theorem 4.2. There is the upper bound for LW divergence on Type I censored scheme as follow

$$D_{LW}^{C-I}(g||f) \le \frac{(a+b)-ab-1}{2}, \tag{28}$$

where a and b be minimum and maximum values of $\left\{\frac{g(x)}{f(x)}, \frac{\bar{G}(C)}{\bar{F}(C)}\right\}$, $x \le C$, *respectively.*

Proof. First, suppose that a and b be minimum and maximum the term $\left\{\frac{g(x)}{f(x)}, \frac{\bar{G}(C)}{\bar{F}(C)}\right\}$, respectively. Let $p(x) = g(x)$, $t_1(x) = \sqrt{\frac{g(x)}{f(x)}}$, $t_2(x) = \sqrt{\frac{f(x)}{g(x)}}$, so from (27) we have

$$\int_{-\infty}^{C} \frac{g^2(x)}{f(x)}\, dx + abF(C) \le (a+b)G(C). \tag{29}$$

Similarly, by taking $p(x) = g(x)$, $t_3(x) = \sqrt{\frac{\bar{G}(C)}{\bar{F}(C)}}$, $t_4(x) = \sqrt{\frac{\bar{F}(C)}{\bar{G}(C)}}$, in (27) we have

$$\frac{\overline{G^2}(C)}{\bar{F}(C)} + ab\bar{F}(C) \le (a+b)\bar{G}(C). \tag{30}$$

Therefore, from (7) one can write the χ^2 distance as we have

$$D_{\chi^2}^{C-I}(g||f) = \int_{-\infty}^{C} \frac{g^2(x)}{f(x)}\, dx + \frac{\overline{G^2}(C)}{\bar{F}(C)} - 1 \le (a+b) - ab - 1. \tag{31}$$

Thus, considering (21) and (31) implies (28). □

Remark 4.2. Consider to Remark 3.2 and Theorem 4.2 can be states on general data.

Example 4.1. Let X and Y be two exponentially distributed random variables with the mean values of 2θ and θ, respectively in which $\theta>0$ and suppose that C be the median point for G(x).

Clearly $G(C) = \frac{1}{2}$, $C = \frac{\ln\sqrt{2}}{\theta}$, and $F(C) = \frac{1}{\sqrt{2}}$ we have $\sqrt{\bar{G}(C)} \ln \frac{\bar{G}(C)+\bar{F}(C)}{\bar{G}(C)} \le \sqrt{\frac{1}{2}} \ln(1+\sqrt{2}) \simeq 0.623$, as well as, we can get

$$0.287 \simeq \sqrt{\frac{1}{2}} \ln\left(\frac{3}{2}\right) \le \sqrt{G(C)} \ln \frac{g(x)+f(x)}{g(x)} \le \sqrt{\frac{1}{2}} \ln\left(\frac{2+\sqrt{2}}{2}\right) \simeq 0.378. \quad \forall x \le C$$

Thus, considering Theorem 4.1, we would have $m = 0.287$ *and* $M = 0.623$. *Therefore,* $D_{LW}^{C-I}(g||f) \le 0.316$.

Furthermore, $\frac{\bar{G}(C)}{\bar{F}(C)} = \frac{\sqrt{2}}{2}$ *and* $\sqrt{2} \le \frac{g(x)}{f(x)} = 2e^{-\theta x} \le 2$. *Then, from Theorem 4.2 we would have* $a = \frac{\sqrt{2}}{2}$ *and* $b = 2$ *that implies* $D_{LW}^{C-I}(g||f) \le 0.146$.

Considering Theorems 4.1 and 4.1 we would have a sharper upper boundary for LW between $\{0, 0.146.\}$

5. Some boundaries for the *LW* divergence from conditional stochastic ordering on type I censored data

Stochastic (st) ordering concept is a prominent and useful approach in statistics and probability theory, which quantifies the concept of one random variable being "larger" than another. These are usually partial orders, so that one random variable X may be neither stochastically greater, equal or less than another random variable Y. Many different orders exist, which have various applications. This concept can be defined variously from the properties of probabilistic functions such as distribution, survival failure, probability density, etc.

Let X and Y be continuous random variables with distribution functions of $G(x)$ and $F(x)$, respectively. Also assume that they support (l_X, u_X), and $(l_{Y,}\ u_Y)$, where $-\infty \leq l_X \leq u_X \leq \infty$, and $-\infty \leq l_Y \leq u_Y \leq \infty$. Since X and Y are assumed to be strictly continuous, functions of probability density for those two variables are denoted as $g(x)$ and $f(x)$, respectively. The st ordering is the most common indicator which compares random variables to each other.

A lot of works has been done in the literature on ordering statistics. For a glimpse of this, see the books by David (1981), Arnold et al. (1992), as well, Shaked and Shanthikumar (1994). Furthermore, Takahasi (1988) showed some results on hazard rate ordering. According to the referenced notes we have the following definitions for the order of X and Y.

Definition 5.1. A random variable X is said to precede another random variable Y in sense of st order (denoted as $X \geq^{st} Y$), if and only if $\bar{G}(x) \geq \bar{F}(x)$. This is equivalent to $G(x) \leq F(x), \quad \forall x \epsilon \Re$.

Definition 5.2. X is said to be smaller than Y in terms of likelihood ratio order (denoted as $X \leq^{lr} Y$), if $\frac{g(x)}{f(x)}$, is an increasing function of x, $\forall x \geq 0$.

Definition 5.3. *X be smaller than Y in terms of hazard rate order (denoted as $X \leq^{hr} Y$) if $\frac{\bar{G}(x)}{\bar{F}(x)}$, is an increasing function of $x \in (-\infty, \max(u_X, u_Y))$.*

It should be noted that $X \geq^{hr} Y$ is equivalent to set of inequalities $P(X{-}t{>}x|X{>}t) \geq P(Y{-}t{>}x|Y{>}t)$, for all values of $x \geq 0$, and t.

What we already know about the relation between the KL and LW measures, is that the LW divergence value is defined on $[0, 1]$ and the magnitude of KL measure is always larger than twice the magnitude of this measure (Lin 1991).

The following Theorem addresses this difference under the st order condition.

Theorem 5.1. If $X \geq^{st} Y$, then the distance between the KL and LW divergence values would lie in the following interval on censored data

$$D_{KL}^{C-I}(g||f) + F(C)\ln\frac{G(C)}{F(C)} + 2\bar{G}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)} \leq D_{LW}^{C-I}(g||f) \leq D_{KL}^{C-I}(g||f) + G(C)\ln 2,$$

and if $X \leq^{st} Y$, then the interval would be

$$D_{KL}^{C-I}(g||f) + F(C)\ln\frac{F(C)}{G(C)} \leq D_{LW}^{C-I}(g||f) \leq D_{KL}^{C-I}(g||f) + G(C)\ln 2 + \bar{F}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)}.$$

Proof. In order to find the lower boundary, under the st condition we can write

$$D_{LW}^{C-I}(g\|f) \le \int_{-\infty}^{C} g(x)\ln\frac{g(x)}{f(x)}dx + \bar{G}(C)\ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)} + G(C)\ln 2$$
$$\overset{\bar{G}\ge\bar{F}}{\le} \int_{-\infty}^{C} g(x)\ln\frac{g(x)}{f(x)}dx + \bar{G}(C)\ln\frac{\bar{G}(C)}{\bar{F}(C)} + G(C)\ln 2$$
$$\le D_{KL}^{C-I}(g\|f) + G(C)\ln 2,$$

and for the upper boundary, we can get

$$D_{LW}^{C-I}(g\|f) \ge^{\bar{G}\ge\bar{F}} \int_{-\infty}^{C} g(x)\ln\frac{g(x)}{f(x)}\ dx + \int_{-\infty}^{C} g(x)\ln\frac{2f(x)}{g(x)+f(x)}dx + \bar{G}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)}$$
$$\ge D_{KL}^{C-I}(g\|f) + \int_{-\infty}^{C} g(x)\ln\frac{2f(x)}{g(x)+f(x)}dx + 2\bar{G}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)}.$$

Since, from the log-sum inequality, we can write

$$D_{LW}^{C-I}(g\|f) \ge D_{KL}^{C-I}(g\|f) + G(C)\ln\frac{2F(C)}{G(C)+F(C)} + 2\bar{G}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)}$$
$$\ge^{\bar{G}\ge\bar{F}} D_{KL}^{C-I}(g\|f) + G(C)\ln\frac{G(C)}{F(C)} + 2\bar{G}(C)\ln\frac{\bar{F}(C)}{\bar{G}(C)}.$$

The proof for condition $X \le^{st} Y$, is similar, hence we don't repeat. □

Meanwhile, in consideration of increasing and decreasing nature of above terms, boundaries can be found for $D_{LW}^{C-I}(g\|f)$.

Using the following theorem, we try to find an interval smaller than (0, 1) for the magnitude of LW divergence measure value considering the lr order condition.

Example 5.1. Suppose that X and Y be two exponentially distributed random variables with the means values of 2, and 1, respectively. It is clear that $X \ge^{st} Y$, from $\bar{G}(x) \ge \bar{F}(x)$. A graphical comparison of bounds mentioned in Theorem 5.1 for $C \in (0, 1]$ is shown in Figure 1.

Theorem 5.2. *If X and Y are lifetime variables, and* $X \le^{lr} Y$ *and* $C>0$, *then on censored data, we have* $D_{LW}^{C-I}(g\|f) \le G(C)\tau(C) + \bar{G}(C)\bar{\lambda}(C)$, *where* $\tau(C) = \ln\frac{2g(C)}{g(C)+f(C)}$ *and* $\bar{\lambda}(C) = \ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)}$.

For the case of $X \ge^{lr} Y$, and $C>0$, *the direction of inequality would be reversed.*

Proof. According to (4) for $X \le^{lr} Y$, and $\forall\ C>0$, we have

$$\int_0^C g(x)\ln\frac{2g(x)}{g(x)+f(x)}\ dx = -\int_0^C g(x)\ln\left(\frac{f(x)}{2g(x)}+\frac{1}{2}\right)dx$$
$$\overset{\frac{g(x)}{f(x)}\le\frac{g(C)}{f(C)}}{\le} \int_0^C g(x)\ln\frac{2g(C)}{g(C)+f(C)}dx.$$

So, $D_{LW}^{C-I}(g\|f) \le G(C)\tau(C) + \bar{G}(C)\bar{\lambda}(C)$.

The proof for the case of $X \geq^{lr} Y$, is similar. □

Example 5.2. Assume that X be an exponentially distributed random variable with the parameter 1, and, in the case of Cox Proportional Hazard model, take $\bar{F}(x) = \bar{G}^{\theta}(x)$ that yields $X \leq^{lr} Y$. By letting $\theta = 2$, Figure 2 shows a graphical comparison of bounds mentioned in Theorem 5.2 for $C \in (0, 1]$.

Theorem 5.3, aims to find a lower (upper) boundary for the LW measure value using the hr order condition.

Theorem 5.3. *If $x \geq 0$, and $y \geq 0$, for the case of $X \geq^{hr} Y$, we can have the boundary for $D_{LW}^{C-I}(g||f)$, on Type I censored data, as follow*

$$\left(\frac{F(C)\bar{G}(C) - G(C)\bar{F}(C)}{\bar{G}(C) + \bar{F}(C)} + G(C)\ln\frac{G(C)}{G(C) + F(C)}, F(C) + G(C)\ln\frac{\bar{G}(C) + \bar{F}(C)}{\bar{G}(C)}\right).$$

Proof. For $\forall x \leq C$, and by taking $Y(x) = \frac{(g(x)+f(x))}{h_G(x)(\bar{G}(x)+\bar{F}(x))}$, and $h_G(x) = \frac{g(x)}{\bar{G}(x)}$, from Theorem 2.1, and using the $X \geq^{hr} Y$, condition we would have $Y(x) \geq \frac{g(x)+f(x)}{g(x)}\frac{\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)}$, hence

$$\int_0^C g(x)(Y(x) - 1)\ dx \geq (G(C) + F(C))\frac{\bar{G}(C)}{\bar{G}(C) + \bar{F}(C)} - G(C) = \frac{F(C)\bar{G}(C) - G(C)\bar{F}(C)}{\bar{G}(C) + \bar{F}(C)},$$

furthermore,

$$\begin{aligned}
-\int_0^C g(x)\ln Y(x)\ dx &= \int_0^C g(x)\ln\frac{g(x)\left(\bar{G}(x) + \bar{F}(x)\right)}{(g(x) + f(x))\bar{G}(x)}\ dx \\
&\geq \int_0^C g(x)\ln\frac{g(x)}{g(x) + f(x)}\ dx \geq G(C)\ln\frac{G(C)}{G(C) + F(C)}.
\end{aligned}$$

Besides,

$$\int_0^C g(x)(Y(x) - 1)\ dx = \int_0^C g(x)\left(\frac{(g(x) + f(x))}{h_G(x)(\bar{G}(x) + \bar{F}(x)} - 1\right)\ dx \leq \int_0^C f(x)\ dx,$$

and

$$-\int_0^C g(x)\ln Y(x)\ dx \leq \int_0^C g(x)\ln\frac{\bar{G}(x) + \bar{F}(x)}{\bar{G}(x)}\ dx \overset{X \overset{hr}{\geq} Y}{\leq} G(C)\ln\frac{\bar{G}(C) + \bar{F}(C)}{\bar{G}(C)}.$$

So, the proof is complete. □

Example 5.3. Let X be an exponentially distributed random variable with the means value 1. Obviously $X \geq^{hr} Y$, following $\bar{F}(x) = \bar{G}^{\theta}(x)$ with mean value of 2. Figure 3, shows a graphical comparison of bounds mentioned in Theorem 5.3, for $C \in (0, 1]$.

Clearly, using the log-sum inequality we would have a lower boundary for the magnitude of LW divergence value as

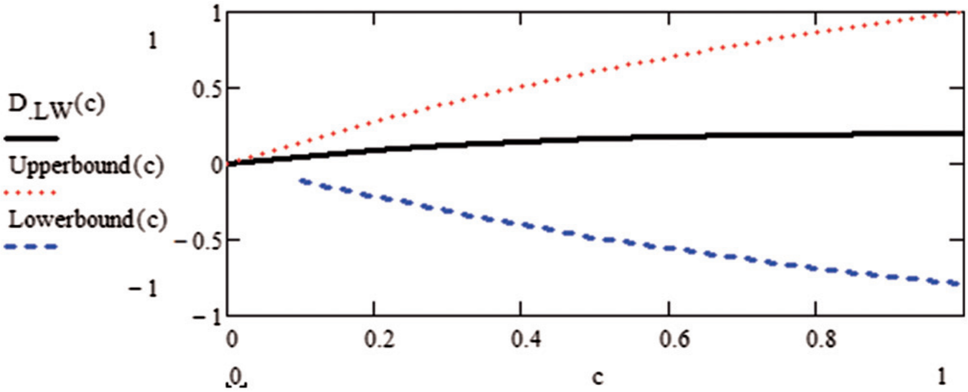

Figure 1. The lower and upper boundaries of Example 5.1 against $C \in (0, 1]$.

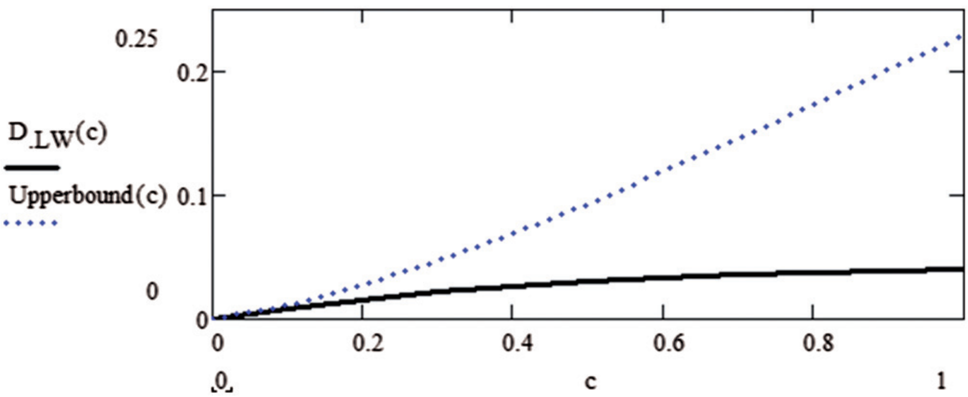

Figure 2. The lower and upper boundaries of Example 5.2 against $C \in (0, 1]$.

$$G(C)\ln\frac{2G(C)}{G(C)+F(C)}+\bar{G}(C)\ln\frac{2\bar{G}(C)}{\bar{G}(C)+\bar{F}(C)}.$$

Also, can be show that the value of LW divergence measure on Type I censored variable is equal to or less than that for the general LW divergence measure.

In Theorem 5.4, we denote the relation between LW divergence measure when applied to Type I censored data and the LW measure on general situation under the condition $X \overset{hr}{\geq} Y$.

Theorem 5.4. *If X and Y be two lifetime random variables and $X \geq^{hr} Y$, the boundary of $D_{LW}(g||f) - D_{LW}^{C-I}(g||f)$, is*

$$D_{LW}(g||f) - D_{LW}^{C-I}(g||f) \leq \int_C^{\infty} g(x)\ln\frac{h_G(x)}{h_F(x)}dx,$$

where, $h_F(x) = \frac{f(x)}{\bar{F}(x)}$, and $h_G(x) = \frac{g(x)}{\bar{G}(x)}$.

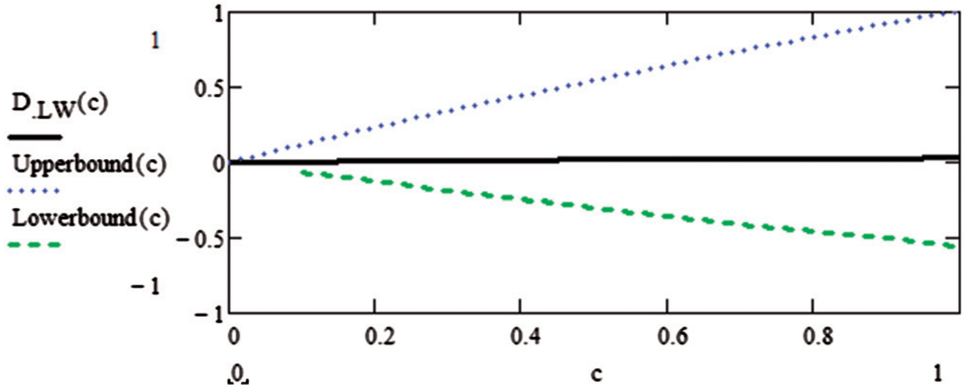

Figure 3. The lower and upper boundaries of Example 5.3 against $C \in (0, 1]$.

Proof. Clearly, $\int_C^\infty g(x) \ln \frac{2g(x)}{g(x)+f(x)}\ dx \le \int_C^\infty g(x) \ln \frac{2g(x)}{f(x)}\ dx$, furthermore, by taking the hr ordering properties, we have

$$\bar{G}(C) \ln \frac{2\bar{G}(C)}{\bar{G}(C) + \bar{F}(C)} = \int_C^\infty g(x) \ln \frac{2\bar{G}(C)}{\bar{G}(C) + \bar{F}(C)}\ dx \overset{X \overset{hr}{\geq} Y}{\geq} \int_C^\infty g(x) \ln \frac{2\bar{G}(x)}{\bar{G}(x) + \bar{F}(x)}\ dx.$$

Therefore, it would mentioned that

$$D_{LW}(g\|f) - D_{LW}^{C-I}(g\|f) \le \int_C^\infty g(x) \ln\left(\frac{h_G(x)}{h_F(x)}\left(\frac{\bar{G}(x)}{\bar{F}(x)} + 1\right)\right) dx = \int_C^\infty g(x) \ln \frac{h_G(x)}{h_F(x)} dx.$$

6. Conclusions

In this paper, we expressed the *LW* divergence measure on Type I censored data, which can be considered in computing the divergence on a sub interval between two probability functions. We found the *LW* divergence measure on Type I censored data is monotonous increasing of C. We have addressed the problem of finding relationships between various information and divergence measures such as the Fisher information, the *KL* and *B* and χ^2 distance measures. We have shown the *LW* divergence properties comparison to the censored Type I scheme and the generally situation. As well as we obtained some bounds for *LW* divergence corresponding the mentioned divergences and the well-known inequalities such as Cassels, Diaz–Metcalf and Kantorovich. Our findings indicate that the *LW* divergence value respect to *KL* and χ^2 is completely bounded, while the mentioned informations are semi-bounded. Next, this paper takes the st ordering concept such as st, lr and hr orders conditions into account to provide the relationship between the magnitude of the *LW* divergence value when applied to censored Type I data and other boundaries same as the value of the *LW* divergence when applied to general state data. This issue is a follow-up study of the subject *LW* and can be considered as a completion to the results regarding goodness of fit tests for normal and exponential in relation to Type I censored data which were previously done within another study of ours.

References

Abbasnejad, M., N. R. Arghami, and M. Tavakoli. 2012. A goodness of fit test for exponentiality based on Lin-Wong information. *Journal of the Iranian Statistical Society* 11 (2):191–202.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja. 1992. *A first course in order statistics*. New York: John Wiley.

Burgess, R. W. 1928. Introduction to the Mathematics of Statistics. *Bulletin American Mathematical Society* 34:372–3.

Cover, T. M., and J. A. Thomas. 1991. *Elements of information theory*. New York: John Wiley.

Csiszár, I. 1963. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoschen ketten. *Mathematical Institute of the Hungarian Academy of Sciences* 8:84–108.

Dabak, A. G., and D. H. Johnson. 2002. *Relations between KL distance and fisher information*. Dallas, TX: Texas Instruments DSP R-D Center.

David, H. A. 1981. *Order statistics*. 2nd ed. New York: John Wiley.

Deza, M. M., and E. Deza. 2009. *Encyclopedia of distances*. Berlin: Springer

Diaz, J. B., and F. T. Metcalf. 1963. Stronger forms of a class of inequalities of G. Pólya-G.Szegö, and LV Kantorovich. *Bulletin of the American Mathematical Society* 69 (3):415–8.

Dragomir, S. S. 2003. On the p-logarithmic and alpha-power divergence measures in information theory. *Panamerican Mathematical Journal* 13 (3):01–10.

Ferentinos, K., and T. Papaioannou. 1981. New parametric measures of information. *Information and Control* 51 (3):193–208.

Ferger, W. F. 1931. The nature and use of the harmonic mean. *Journal of the American Statistical Association* 26 (173):36–40.

Gertsbakh, I. 1995. On the fisher information in type I censored data. *Statistics and Probability Letters* 23 (4):297–306.

Helsel, D. 2010. Much ado about next to nothing: Incorporating nondetects in science. *The Annals of Occupational Hygiene* 54 (3):257–62.

Jain, K. C., and R. N. Saraswat. 2012. A new information inequality and its application in establishing relation among various f-divergence measures. *Journal of Applied Mathematics, Statistics and Informatics* 8 (1):17–32.

Joarder, A., H. Krishna, and D. Kundu. 2011. Inferences on Weibull parameters with conventional type I censoring. *Computational Statistics and Data Analysis* 55 (1):1–11.

Kapur, J. N. 1984. A comparative assessment of various measures of directed divergence. *Advances in Management Studies* 3 (1):1–16.

Khalili, M., A. Habibirad, and F. Yousefzadeh. 2017. Some properties of LinWong divergence on the past lifetime data. *Communications in Statistics-Theory and Methods* 47 (14):3464–76.

Kleinbaum, D. D., and M. Klein. 2006. *Survival analysis: A self-learning text*. Berlin: Springer Science and Business Media.

Kullback, S. 1959. *Statistics and information theory*. New York: Wiley.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22 (1):79–86.

Kumar, P., and I. J. Taneja. 2006. Generalized relative J-divergence measure and properties. *International Journal of Contemporary Mathematical Sciences* 1 (13):597–609.

Lehmann, E. L., and G. Casella. 2006. *Theory of point estimation*. Berlin: Springer Science and Business Media.

Lin, J. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 37 (1):145–51.

Lin, J., and S. K. M. Wong. 1990. A new directed divergence measure and its characterization. *International Journal of General System* 17 (1):73–81.

Moslehian, M. S., R. Nakamoto, and Y. Seo. 2011. A Diaz-Metcalf type inequality for positive linear maps and its applications. *The Electronic Journal of Linear Algebra* 22 (1):179–90.

Pakyari, R., and N. Balakrishnan. 2013. Testing exponentiality based on type I censored. *Journal of Statistical Computation and Simulation* 83 (12):2369–78.

Pakyari, R., and K. Resalati Nia. 2017. Testing goodness-of-fit for some lifetime distributions with conventional type-I censoring. *Communications in Statistics-Simulation and Computation* 46 (4):2998–3009.

Park, S. 2016. On the Kullback-Leibler information of hybrid censored data. *Communications in Statistics-Theory and Methods* 45 (15):4486–93.

Park, S., and M. Shin. 2014. Kullback-Leibler information of a censored variable and its applications. *Statistics* 48 (4):756–65.

Shaked, M., and J. G. Shanthikumar. 1994. *Stochastic orders and their applications*. New York: Academic Press.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal*. 27:379–423.

Shioya, H., and T. Da-Te. 1995. A generalization of Lin divergence and the derivation of a new information divergence. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 78 (7):34–40.

Takahasi, K. 1988. A note on hazard rates of order statistics. *Communications in Statistics-Theory and Methods* 17 (12):4133–6.

Watson, G. S., G. Alpargu, and G. P. Styan. 1997. Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor. *Linear Algebra and Its Applications* 264:13–53.