J. Appl. Prob. **47,** 216–234 (2010)
Printed in England

GENERALIZED MEASURES OF DIVERGENCE IN SURVIVAL ANALYSIS AND RELIABILITY

FILIA VONTA,* *National Technical University of Athens*
ALEX KARAGRIGORIOU,** *University of Cyprus*

Abstract

Measures of divergence or discrepancy are used either to measure mutual information concerning two variables or to construct model selection criteria. In this paper we focus on divergence measures that are based on a class of measures known as Csiszár's divergence measures. In particular, we propose a measure of divergence between residual lives of two items that have both survived up to some time t as well as a measure of divergence between past lives, both based on Csiszár's class of measures. Furthermore, we derive properties of these measures and provide examples based on the Cox model and frailty or transformation model.

Keywords: Divergence measure; Csiszár's family of measures; Cox model; proportional reverse hazards model; frailty model

2000 Mathematics Subject Classification: Primary 62N05
Secondary 60E15; 90B25

1. Introduction

A measure of divergence is used as a way to evaluate the distance (divergence) between any two populations or functions. In the present work we concentrate on divergence measures that are based on a class of measures known as Csiszár's family of divergence measures or Csiszár's φ-divergence (see Csiszár (1963) and Ali and Silvey (1966)).

An issue of fundamental importance in statistics is the investigation of information measures. These measures are classified in different categories and measure the quantity of information contained in the data with respect to a parameter θ, the divergence between two populations or functions, the information we get after the execution of an experiment, and other important information according to the application they are used for. Traditionally, the measures of information are classified into four main categories, namely divergence type, entropy type, Fisher type, and Bayesian type.

Measures of divergence between two probability distributions have a very long history, beginning with the pioneering work of Pearson (1900), Mahalanobis (1936), Lévy (1925) and Kolmogorov (1933). Among the most popular measures of divergence are the Kullback–Leibler measure of divergence (see Kullback and Leibler (1951)) and the Csiszár's φ-divergence family of measures (see Csiszár (1963) and Ali and Silvey (1966)). A unified analysis has been provided in Cressie and Read (1984), who introduced the power divergence family of statistics that depends on a parameter λ and is used for goodness-of-fit tests for multinomial distributions.

Received 31 March 2009; revision received 21 October 2009.
* Postal address: Department of Mathematics, National Technical University of Athens, 9 Iroon Polytechneiou, Zografou Campus, 15780 Athens, Greece. Email address: vonta@math.ntua.gr
** Postal address: Department of Mathematics and Statistics, University of Cyprus, University Campus, CY-1678 Nicosia, Cyprus.

The Cressie and Read family includes among others the well-known Pearson's X^2 divergence measure and, for multinomial models, the log-likelihood ratio statistic. Recently, the BHHJ divergence measure was proposed in Basu *et al.* (1998) and generalized to the BHHJ family of measures in Mattheou *et al.* (2009). The BHHJ family depends on an index a which controls the trade-off between robustness and efficiency when the measure is used as an estimating criterion for robust parameter estimation.

Ebrahimi and Kirmani (1996b) introduced a measure of discrepancy between the lifetimes X and Y of two items at time t. In survival analysis or in reliability we might know the current age t of a biomedical or technical system. We need to take this information into consideration when we compare two systems or populations. Ebrahimi and Kirmani (1996b) achieved this by replacing the distribution functions of the random variables X and Y in the Kullback–Leibler divergence of X and Y (see Kullback and Leibler (1951)) by the distributions of their residual lifetimes. Di Crescenzo and Longobardi (2004) defined a dual measure of divergence which constitutes a distance between past-life distributions.

In this paper we focus on the distance between lifetimes and propose generalized measures of divergence between residual lives of two items that have both survived up to some time t as well as between past lives, based on Csiszár's family of measures. In Section 2 we define the proposed measures of divergence for Csiszár's class of functions, which are referred to as φ-distances between lifetimes. In Section 3 we examine properties of these measures of divergence, and in Section 4 we find various discrimination measures in cases like the proportional hazards model, the proportional reverse hazards model, and the frailty or transformation model. For the latter case, we provide the divergence between the distribution functions associated with the Cox and frailty models as well as the φ-distance between the respective residual and past lifetimes.

2. The proposed generalized divergence measures

Let X and Y be absolutely continuous, nonnegative random variables that describe the lifetimes of two items. Let $f(x)$, $F(x)$, and $\overline{F}(x)$ be the density function, the cumulative distribution function, and the survival function of X, respectively. Also, let $g(x)$, $G(x)$, and $\overline{G}(x)$ be the density function, the cumulative distribution function, and the survival function of Y, respectively. Let $h_X(x) = f(x)/\overline{F}(x)$ and $h_Y(x) = g(x)/\overline{G}(x)$ be the hazard rate functions of X and Y, and let $\tau_X(x) = f(x)/F(x)$ and $\tau_Y(x) = g(x)/G(x)$ be the reversed hazard rate functions of X and Y. Without loss of generality, we assume throughout the paper that the support of f and g is $(0, +\infty)$.

The Kullback–Leibler distance between F and G (see Kullback and Leibler (1951)) is defined by

$$I_{X,Y} = \int_0^\infty f(x) \log\left(\frac{f(x)}{g(x)}\right) \mathrm{d}x,$$

where log denotes the natural logarithm. A generalization of this distance is defined as

$$I^{\varphi}_{X,Y} = \int_0^\infty g(x)\varphi\left(\frac{f(x)}{g(x)}\right) \mathrm{d}x,$$

which is known as Csiszár's family of measures of divergence.

When the function φ is defined as $\varphi(u) = u \log u$ or $\varphi(u) = u \log u + 1 - u$, then the above measure reduces to the Kullback–Leibler measure. If $\varphi(u) = \frac{1}{2}(1-u)^2$, Csiszár's measure yields the Pearson's chi-square divergence (also known as Kagan's divergence; see

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

Kagan (1963)). If

$$\varphi(u) := \varphi_1(u) = \frac{u^{a+1} - u - a(u-1)}{a(a+1)},$$

we obtain the Cressie and Read power divergence (see Cressie and Read (1984)), $a \neq 0, -1$. If $\varphi(u) = (1 - \sqrt{u})^2$, we obtain the Matusita's divergence (see Matusita (1967)), and if $\varphi(u) = -\log(u) + u - 1$ or $\varphi(u) = -\log(u)$, we obtain the Kullback–Leibler divergence between G and F, also known as the minimum discrimination information (see Pardo (2006, p. 4)).

Other functions that we consider are

$$\varphi(u) := \varphi_2(u) = u^{1+a} - \left(1 + \frac{1}{a}\right)u^a + \frac{1}{a}, \qquad a \neq 0,$$

and

$$\varphi(u) := \varphi_3(u) = 1 - \left(1 + \frac{1}{a}\right)u + \frac{u^{1+a}}{a}, \qquad a \neq 0.$$

The latter is related to a recently proposed measure of divergence (the BHHJ power divergence proposed in Basu *et al.* (1998)), while both $\varphi_2(\cdot)$ and $\varphi_3(\cdot)$ are special cases of the BHHJ family of divergence measures proposed in Mattheou *et al.* (2009):

$$I_X^\alpha(g, f) = \mathrm{E}_g\left(g^\alpha(X)\varphi\left(\frac{f(X)}{g(X)}\right)\right) = \int g^{1+\alpha}(z)\varphi\left(\frac{f(z)}{g(z)}\right)\mathrm{d}\mu, \qquad \alpha \geq 0,$$

where μ represents the Lebesgue measure. Appropriately chosen functions $\varphi(\cdot)$ give rise to special measures mentioned above, while, for $\alpha = 0$, the BHHJ family reduces to the Csiszár family.

Ebrahimi and Kirmani (1996b) introduced the following measure of discrepancy between X and Y at time t:

$$I_{X,Y}(t) = \int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}\right)\mathrm{d}x, \qquad t > 0. \tag{1}$$

A dual measure which constitutes a distance between past lifetimes was defined in Di Crescenzo and Longobardi (2004) as

$$\overline{I}_{X,Y}(t) = \int_0^t \frac{f(x)}{F(t)} \log\left(\frac{f(x)/F(t)}{g(x)/G(t)}\right)\mathrm{d}x, \qquad t > 0. \tag{2}$$

We now propose two new measures of discrepancy which are based on the Csiszár's φ-divergence family, namely, the *φ-distance between residual lifetimes*,

$$I_{X,Y}^\varphi(t) = \int_t^\infty \frac{g(x)}{\overline{G}(t)} \varphi\left(\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}\right)\mathrm{d}x, \qquad t > 0, \tag{3}$$

and the *φ-distance between past lifetimes*,

$$\overline{I}_{X,Y}^\varphi(t) = \int_0^t \frac{g(x)}{G(t)} \varphi\left(\frac{f(x)/F(t)}{g(x)/G(t)}\right)\mathrm{d}x, \qquad t > 0, \tag{4}$$

where the function φ belongs to a class of functions Φ with the following properties:

(i) $\varphi(x)$ is continuous, differentiable, and convex for $x \geq 0$;

(ii) $\varphi(1) = 0$;

(iii) $\varphi'(1) = 0$.

From the above assumptions we deduce that $\varphi(x) \geq 0$ for all $x > 0$, $\varphi'(x) > 0$ for $x > 1$, and $\varphi'(x) < 0$ for $x < 1$.

Let us now denote by $f_t(x) = f(x)/F(t)$, $0 < x < t$, the probability density function (PDF) of $X_t = [X \mid X \leq t]$ (or $[t - X \mid X \leq t]$) and by $g_t(x) = g(x)/G(t)$, $0 < x < t$, the PDF of $Y_t = [Y \mid Y \leq t]$ (or $[t - Y \mid Y \leq t]$). Similarly, let $f^t(x) = f(x)/\overline{F}(t)$, $x > t$, denote the PDF of $X^t = [X \mid X > t]$ (or $[X - t \mid X > t]$) and let $g^t(x) = g(x)/\overline{G}(t)$, $x > t$, denote the PDF of $Y^t = [Y \mid Y > t]$ (or $[Y - t \mid Y > t]$). Finally, we define $F_t := \mathrm{P}(X_t \leq x) = F(x)/F(t)$ and $G_t := \mathrm{P}(Y_t \leq x) = G(x)/G(t)$ for $0 < x < t$ to be the cumulative distribution functions (CDFs) of X_t and Y_t. Similarly, we define F^t and G^t to be the CDFs of the residual lifetimes X^t and Y^t.

Remarks. Let $\Phi^\star \subset \Phi$ be the class of functions $\varphi(\cdot)$ satisfying properties (i) and (ii) above. It is not difficult to see that, for every function $\varphi \in \Phi^\star$ which is differentiable at 1, we can always define the function $\psi(u) = \varphi(u) - \varphi'(1)(u-1)$ such that $\psi(\cdot) \in \Phi$, $\psi'(1) = 0$, $I^{\psi}_{X,Y}(t) = I^{\varphi}_{X,Y}(t)$, and $\overline{I}^{\psi}_{X,Y}(t) = \overline{I}^{\varphi}_{X,Y}(t)$. Observe that, for the function $\Phi^\star \ni \varphi(u) = u\log(u)$, the discrepancy measure (3) is reduced to the measure (1) and, equivalently, the discrepancy measure (4) is reduced to the measure (2). In this case, the function $\psi \in \Phi$ for which $I^{\psi} = I^{\varphi}$ is given by $\psi(u) = u \log u + 1 - u$. As a result, from now on and without loss of generality, we focus solely on functions that belong to the Csiszár family, that is, the class of functions Φ that satisfy properties (i)–(iii). Finally, observe that, for the function $\varphi(u) = -\log(u) + u - 1$, the discrepancy measure (3) is equal to the measure $I_{Y,X}(t)$, and the discrepancy measure (4) becomes $\overline{I}_{Y,X}(t)$.

3. Properties of the proposed measures

In this section we present a number of properties of the proposed generalized measures. First we examine the nonegativity property of the proposed measures as well as their connection to the standard Csiszár family of measures. Owing to the assumptions made about the function φ and Jensen's inequality we have

$$\mathrm{E}_{g^t}\left(\varphi\left(\frac{f^t}{g^t}\right)\right) \geq \varphi\left(\mathrm{E}_{g^t}\left(\frac{f^t}{g^t}\right)\right) = 0.$$

Therefore, $I^{\varphi}_{X,Y} \geq 0$ and, similarly, $\overline{I}^{\varphi}_{X,Y} \geq 0$, with equality holding in both cases if and only if $f^t(x) = g^t(x)$ almost everywhere (a.e.) for $x > t$ and $f_t(x) = g_t(x)$ a.e. for $0 < x < t$, respectively.

It is also very easy to see that

$$\lim_{t\to\infty} \overline{I}^{\varphi}_{X,Y}(t) = I^{\varphi}_{X,Y} = \lim_{t\to 0} I^{\varphi}_{X,Y}(t).$$

Let us now define some stochastic orders useful in our case (see Shaked and Shanthikumar (1994, Chapter 1) for definitions and basic results).

Definition. Let X and Y be nonnegative, absolutely continuous random variables.

(a) If $\overline{F}(x) \leq \overline{G}(x)$ for all real x or, equivalently, $F(x) \geq G(x)$ for all real x, then $X \leq_{\mathrm{st}} Y$, the usual stochastic ordering.

(b) If $f(t)/g(t)$ is increasing in t for $t > 0$ then $X \geq_{\mathrm{lr}} Y$, the likelihood ratio ordering.

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

(c) If $f(t)/g(t)$ is increasing in t for $t > 0$ then $\tau_X(t) \geq \tau_Y(t)$ for $t > 0$, the reverse hazard ordering, denoted by $X \geq_{\text{rh}} Y$, which is true if and only if $X_t \geq_{\text{st}} Y_t$ for all $t > 0$.

(d) If $f(t)/g(t)$ is increasing in t for $t > 0$ then $h_X(t) \leq h_Y(t)$ for $t > 0$, the hazard rate ordering, denoted by $X \geq_{\text{hr}} Y$, which is true if and only if $X^t \geq_{\text{st}} Y^t$ for all $t > 0$.

It is true that

$$X \geq_{\text{lr}} Y \Rightarrow X \geq_{\text{hr}} Y \Rightarrow X \geq_{\text{st}} Y,$$
$$X \geq_{\text{lr}} Y \Rightarrow X \geq_{\text{rh}} Y \Rightarrow X \geq_{\text{st}} Y.$$

The next theorem provides bounds for the proposed measures.

Theorem 1. *For the measures $I^{\varphi}_{X,Y}(t)$ and $\overline{I}^{\varphi}_{X,Y}(t)$, we have*

$$\text{(i)} \quad I^{\varphi}_{X,Y}(t) < \varphi(0) + \lim_{r\to\infty} \frac{\varphi(r)}{r} \tag{5}$$

and

$$\text{(ii)} \quad \overline{I}^{\varphi}_{X,Y}(t) < \varphi(0) + \lim_{r\to\infty} \frac{\varphi(r)}{r}. \tag{6}$$

Proof. (i) For every convex function φ, it is known that

$$\varphi(t^*) \leq \varphi(0) + t^* \lim_{r\to\infty} \frac{\varphi(r)}{r} \tag{7}$$

for $t^* \geq 0$, with the strict inequality holding for strictly positive t^*. Therefore, the measure $I^{\varphi}_{X,Y}(t)$, which is equal to

$$\int_t^{\infty} \frac{g(x)}{\overline{G}(t)} \varphi\left(\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}\right) \mathrm{d}x, \qquad t > 0,$$

by (7), is strictly less than

$$\int_t^{\infty} \frac{g(x)}{\overline{G}(t)} \left(\varphi(0) + \frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)} \lim_{r\to\infty} \frac{\varphi(r)}{r}\right) \mathrm{d}x,$$

and the result is immediate. Part (ii) is proved similarly.

Remarks. (a) It should be pointed out that bounds (5) and (6) are in full accordance with the bound obtained for the measures of divergence between two distribution functions (see Proposition 1.1 of Pardo (2006)).

(b) Let t_0 be the point where $f(t_0)/\overline{F}(t) = g(t_0)/\overline{G}(t)$. If we assume that $t_0 < t$ then a lower bound can also be obtained for the measure $I^{\varphi}_{X,Y}(t)$, that is,

$$I^{\varphi}_{X,Y}(t) \geq \varphi\left(\frac{h_X(t)}{h_Y(t)}\right) \geq 0.$$

If, on the other hand, we assume that $t_0 > t$ then a lower bound can be obtained for the measure $\overline{I}^{\varphi}_{X,Y}(t)$, that is,

$$\overline{I}^{\varphi}_{X,Y}(t) \geq \varphi\left(\frac{\tau_X(t)}{\tau_Y(t)}\right) \geq 0.$$

Recall that, for $f(x)/g(x)$ increasing in x, $h_X(x)/h_Y(x) \leq 1$ and $\tau_X(x)/\tau_Y(x) \geq 1$ for all $x > 0$.

The effect of bijective transformations on the generalized divergence measures is discussed in the following two theorems.

Theorem 2. *Let X and $X^\star$ have support $(0,\infty)$, and also let ψ be a bijective transformation from $(0,\infty)$ to $(0,\infty)$. If ψ is strictly increasing then*

$$\text{(i)}\quad \overline{I}^{\varphi}_{\psi(X),\psi(X^\star)}(t)=\overline{I}^{\varphi}_{X,X^\star}(\psi^{-1}(t)),\qquad t>0,$$

and

$$\text{(ii)}\quad I^{\varphi}_{\psi(X),\psi(X^\star)}(t)=I^{\varphi}_{X,X^\star}(\psi^{-1}(t)),\qquad t>0.$$

Proof. For (i), we have

$$\begin{aligned}\overline{I}^{\varphi}_{\psi(X),\psi(X^\star)}(t) &= \int_0^t \frac{g_{\psi(X^\star)}(y)}{G_{\psi(X^\star)}(t)}\varphi\bigg(\frac{f_{\psi(X)}(y)/F_{\psi(X)}(t)}{g_{\psi(X^\star)}(y)/G_{\psi(X^\star)}(t)}\bigg)\,\mathrm{d}y\\ &= \int_0^t \frac{g(\psi^{-1}(y))|\mathrm{d}\psi^{-1}(y)/\mathrm{d}y|}{G(\psi^{-1}(t))}\varphi\bigg(\frac{f(\psi^{-1}(y))|\mathrm{d}\psi^{-1}(y)/\mathrm{d}y|/F(\psi^{-1}(t))}{g(\psi^{-1}(y))|\mathrm{d}\psi^{-1}(y)/\mathrm{d}y|/G(\psi^{-1}(t))}\bigg)\,\mathrm{d}y,\end{aligned}$$

which, by a change of variable $x=\psi^{-1}(y)$, becomes

$$\int_0^{\psi^{-1}(t)} \frac{g(x)}{G(\psi^{-1}(t))}\varphi\bigg(\frac{f(x)/F(\psi^{-1}(t))}{g(x)/G(\psi^{-1}(t))}\bigg)\,\mathrm{d}x.$$

This completes the proof. Part (ii) is proved similarly.

Theorem 3. *Let X and $X^\star$ have support $(0,\infty)$ and distribution functions F and G, respectively, and also let ψ be a bijective transformation from $(0,\infty)$ to $(0,\infty)$.*

(i) *If ψ is strictly decreasing then $\overline{I}^{\varphi}_{\psi(X),\psi(X^\star)}(t)$ is given as*

$$\int_0^{\psi^{-1}(t)} \frac{g(x)}{\vartheta_{X^\star}(\psi^{-1}(t))G(\psi^{-1}(t))}\varphi\bigg(\frac{f(x)/(\vartheta_X(\psi^{-1}(t))F(\psi^{-1}(t)))}{g(x)/(\vartheta_{X^\star}(\psi^{-1}(t))G(\psi^{-1}(t)))}\bigg)\,\mathrm{d}x$$

for $t>0$.

(ii) *If ψ is strictly decreasing then $I^{\varphi}_{\psi(X),\psi(X^\star)}(t)$ is given as*

$$\int_{\psi^{-1}(t)}^{\infty} \frac{g(x)}{\overline{G}(\psi^{-1}(t))/\vartheta_{X^\star}(\psi^{-1}(t))}\varphi\bigg(\frac{f(x)/(\overline{F}(\psi^{-1}(t))/\vartheta_X(\psi^{-1}(t)))}{g(x)/(\overline{G}(\psi^{-1}(t))/\vartheta_{X^\star}(\psi^{-1}(t)))}\bigg)\,\mathrm{d}x$$

for $t>0$, where $\vartheta_X(t)=\overline{F}_X(t)/F_X(t)$ is known in reliability theory as the 'odds function' (see Kirmani and Gupta (2001)).

Proof. Since ψ is decreasing, we have

$$G_{\psi(X^\star)}(t)=\frac{\overline{G}(\psi^{-1}(t))}{G(\psi^{-1}(t))}G(\psi^{-1}(t))=\vartheta_{X^\star}(\psi^{-1}(t))G(\psi^{-1}(t)).$$

Similarly,

$$F_{\psi(X)}(t)=\vartheta_X(\psi^{-1}(t))F(\psi^{-1}(t)).$$

Then the distance $\overline{I}^{\varphi}_{\psi(X),\psi(X^\star)}(t)$ equals

$$\int_0^{\psi^{-1}(t)} \frac{g(x)}{\vartheta_{X^\star}(\psi^{-1}(t))G(\psi^{-1}(t))}\varphi\left(\frac{f(x)/(\vartheta_X(\psi^{-1}(t))F(\psi^{-1}(t)))}{g(x)/(\vartheta_{X^\star}(\psi^{-1}(t))G(\psi^{-1}(t)))}\right)\mathrm{d}x.$$

The second part is proved similarly.

Theorem 4. *Let three random variables X_1, X_2, and Y have PDFs f_1, f_2, and g, respectively. If $X_2^t \geq_{\mathrm{st}} X_1^t$ and $f_1(x)/g(x)$ is increasing in x, then*

$$I^{\varphi}_{X_1,Y}(t) \leq I^{\varphi}_{X_2,Y}(t), \qquad t > 0.$$

If $X_1^t \geq_{\mathrm{st}} X_2^t$ and $f_2(x)/g(x)$ is increasing in x, the inequality is reversed.

Proof. Consider the difference under question,

$$I^{\varphi}_{X_1,Y}(t) - I^{\varphi}_{X_2,Y}(t) = \mathrm{E}_{g^t}\left(\varphi\left(\frac{f_1^t}{g^t}\right) - \varphi\left(\frac{f_2^t}{g^t}\right)\right). \tag{8}$$

From the mean value theorem, for

$$\min\left\{\frac{f_1^t(x)}{g^t(x)}, \frac{f_2^t(x)}{g^t(x)}\right\} \leq \frac{f_\star^t}{g_\star^t} \leq \max\left\{\frac{f_1^t(x)}{g^t(x)}, \frac{f_2^t(x)}{g^t(x)}\right\},$$

the expectation in (8) is equal to

$$\int_t^\infty \varphi'\left(\frac{f_\star^t}{g_\star^t}\right)\left(\frac{f_1^t}{g^t} - \frac{f_2^t}{g^t}\right)\mathrm{d}G^t = \int_t^\infty \varphi'\left(\frac{f_\star^t}{g_\star^t}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x.$$

Without loss of generality, assume that there is one change of sign in the difference $f_1^t(x) - f_2^t(x)$ from positive to negative at $t_0 > t$. The opposite case or more general cases can be treated similarly. The above integral could be written as

$$\int_t^{t_0} \varphi'\left(\frac{f_\star^t}{g_\star^t}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x + \int_{t_0}^\infty \varphi'\left(\frac{f_\star^t}{g_\star^t}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x. \tag{9}$$

For $x \in [t, t_0)$, $f_1^t(x) - f_2^t(x) > 0$, and since φ' is an increasing function, the first integral in (9) is less than or equal to

$$\int_t^{t_0} \varphi'\left(\max\left\{\frac{f_1^t(x)}{g^t(x)}, \frac{f_2^t(x)}{g^t(x)}\right\}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x = \int_t^{t_0} \varphi'\left(\frac{f_1^t(x)}{g^t(x)}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x. \tag{10}$$

For $x \in (t_0, \infty)$, $f_2^t(x) - f_1^t(x) > 0$, and since $-\varphi'(f_\star^t/g_\star^t) \leq -\varphi'(\min\{f_1^t(x)/g^t(x), f_2^t(x)/g^t(x)\}) = -\varphi'(f_1^t(x)/g^t(x))$, the second integral in (9) is less than or equal to

$$\int_{t_0}^\infty \varphi'\left(\frac{f_1^t(x)}{g^t(x)}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x.$$

Finally, (9) is less than or equal to

$$\int_t^\infty \varphi'\left(\frac{f_1^t(x)}{g^t(x)}\right)(f_1^t(x) - f_2^t(x))\,\mathrm{d}x. \tag{11}$$

By an integration by parts, (11) becomes

$$\int_t^\infty \left(\frac{F_2(x)}{\overline{F}_2(t)} - \frac{F_1(x)}{\overline{F}_1(t)}\right) \mathrm{d}\left(\varphi'\left(\frac{f_1^t(x)}{g^t(x)}\right)\right), \tag{12}$$

and since X_2^t is stochastically larger than X_1^t, φ' is increasing, and $f_1(x)/g(x)$ is increasing in x, the integral in (12) is nonpositive. This completes the proof.

From the above we deduce that, under the assumptions of Theorem 4, at any point t the reference distribution G^t is closer to F_1^t than to F_2^t.

Theorem 5. *Let three random variables X_1, X_2, and Y have PDFs f_1, f_2, and g, respectively. If $X_{2,t} \geq_{\mathrm{st}} X_{1,t}$ and $f_1(x)/g(x)$ is increasing in x, then*

$$\overline{I}^{\varphi}_{X_1,Y}(t) \leq \overline{I}^{\varphi}_{X_2,Y}(t), \qquad t > 0.$$

If $X_{1,t} \geq_{\mathrm{st}} X_{2,t}$ and $f_2(x)/g(x)$ is increasing in x, the inequality is reversed.

Proof. An argument similar to that used in the proof of Theorem 4 gives the results.

From the above we deduce that, under the assumptions of Theorem 5, at any point t the reference distribution G_t is closer to $F_{1,t}$ than to $F_{2,t}$.

Theorem 6. *Let three random variables X, Y_1, and Y_2 have PDFs f, g_1, and g_2, respectively. If $Y_1^t \geq_{\mathrm{st}} Y_2^t$ and $f(x)/g_1(x)$ is increasing in x, then*

$$I^{\varphi}_{X,Y_1}(t) < I^{\varphi}_{X,Y_2}(t) + \lim_{r\to\infty} \frac{\varphi(r)}{r}\left(1 - \frac{h_X(t)}{h_{Y_1}(t)}\right), \qquad t > 0. \tag{13}$$

Proof. Observe that the difference $I^{\varphi}_{X,Y_1}(t) - I^{\varphi}_{X,Y_2}(t)$ can be written as

$$\mathrm{E}_{g_1^t}\left(\varphi\left(\frac{f^t}{g_1^t}\right)\right) - \mathrm{E}_{g_2^t}\left(\varphi\left(\frac{f^t}{g_1^t}\right)\right) + \mathrm{E}_{g_2^t}\left(\varphi\left(\frac{f^t}{g_1^t}\right) - \varphi\left(\frac{f^t}{g_2^t}\right)\right). \tag{14}$$

The difference in the first two terms above, by (7) with the strict inequality, becomes strictly less than

$$\int_t^\infty \left(\varphi(0) + \frac{f^t}{g_1^t}\lim_{r\to\infty}\frac{\varphi(r)}{r}\right)\mathrm{d}(G_1^t - G_2^t) = \lim_{r\to\infty}\frac{\varphi(r)}{r}\left(\int_t^\infty f^t(x)\,\mathrm{d}x - \int_t^\infty f^t(x)\frac{g_2^t(x)}{g_1^t(x)}\,\mathrm{d}x\right).$$

Since, by our assumption that $f^t(x)/g_1^t(x)$ is increasing in x, we have $f^t(x)/g_1^t(x) \geq f^t(t)/g_1^t(t)$ for all $x \in (t, \infty)$, so that the above relation is less than or equal to

$$\lim_{r\to\infty}\frac{\varphi(r)}{r}\left(1 - \int_t^\infty \frac{h_X(t)}{h_{Y_1}(t)} g_2^t(x)\,\mathrm{d}x\right) = \lim_{r\to\infty}\frac{\varphi(r)}{r}\left(1 - \frac{h_X(t)}{h_{Y_1}(t)}\right) \geq 0.$$

From the mean value theorem, for

$$\min\left\{\frac{f^t(x)}{g_1^t(x)}, \frac{f^t(x)}{g_2^t(x)}\right\} \leq \frac{f_\star^t}{g_\star^t} \leq \max\left\{\frac{f^t(x)}{g_1^t(x)}, \frac{f^t(x)}{g_2^t(x)}\right\},$$

the last expectation in (14) is equal to

$$\mathrm{E}_{g_2^t}\left(\varphi'\left(\frac{f_\star^t}{g_\star^t}\right)\left(\frac{f^t}{g_1^t} - \frac{f^t}{g_2^t}\right)\right) = \int_t^\infty (g_2^t(x) - g_1^t(x))\varphi'\left(\frac{f_\star^t}{g_\star^t}\right)\frac{f^t(x)}{g_1^t(x)}\,\mathrm{d}x.$$

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

An argument similar to that used in the proof of Theorem 4 shows that the above integral is less than or equal to

$$\int_t^\infty (g_2^t(x) - g_1^t(x))\varphi'\bigg(\frac{f^t(x)}{g_1^t(x)}\bigg)\frac{f^t(x)}{g_1^t(x)}\,\mathrm{d}x.$$

By an integration by parts, the above integral becomes

$$\int_t^\infty \bigg(\frac{G_1(x)}{\overline{G}_1(t)} - \frac{G_2(x)}{\overline{G}_2(t)}\bigg)\,\mathrm{d}\bigg(\varphi'\bigg(\frac{f^t(x)}{g_1^t(x)}\bigg)\frac{f^t(x)}{g_1^t(x)}\bigg) \le 0,$$

since Y_1^t is stochastically larger than Y_2^t, $f(x)/g_1(x)$ is increasing in x, and φ' is increasing. This completes the proof.

Theorem 7. *Let three random variables X, Y_1, and Y_2 have PDFs f, g_1, and g_2, respectively. If $Y_{1,t} \ge_{\mathrm{lr}} Y_{2,t}$ and $f(x)/g_1(x)$ is increasing in x, then*

$$\overline{I}^{\varphi}_{X,Y_1}(t) < \overline{I}^{\varphi}_{X,Y_2}(t) + \lim_{r\to\infty}\frac{\varphi(r)}{r}\bigg(1 - \frac{\tau_{Y_2}(t)}{\tau_{Y_1}(t)}\bigg), \qquad t > 0. \tag{15}$$

Proof. Observe that the difference $\overline{I}^{\varphi}_{X,Y_1}(t) - \overline{I}^{\varphi}_{X,Y_2}(t)$ can be written as

$$\mathrm{E}_{g_{1,t}}\bigg(\varphi\bigg(\frac{f_t}{g_{1,t}}\bigg)\bigg) - \mathrm{E}_{g_{2,t}}\bigg(\varphi\bigg(\frac{f_t}{g_{1,t}}\bigg)\bigg) + \mathrm{E}_{g_{2,t}}\bigg(\varphi\bigg(\frac{f_t}{g_{1,t}}\bigg) - \varphi\bigg(\frac{f_t}{g_{2,t}}\bigg)\bigg). \tag{16}$$

As in the proof of Theorem 6, by using (7) with the strict inequality, the difference in the first two terms in (16) is strictly less than

$$\lim_{r\to\infty}\frac{\varphi(r)}{r}\bigg(\int_0^t f_t(x)\,\mathrm{d}x - \int_0^t f_t(x)\frac{g_{2,t}(x)}{g_{1,t}(x)}\,\mathrm{d}x\bigg).$$

Since by our assumption that $g_{1,t}(x)/g_{2,t}(x)$ is increasing in x, we have $g_{2,t}(x)/g_{1,t}(x) \ge g_{2,t}(t)/g_{1,t}(t)$ for all $x \in (0, t)$, so that the above relation is less than or equal to

$$\lim_{r\to\infty}\frac{\varphi(r)}{r}\bigg(1 - \int_0^t \frac{\tau_{Y_2}(t)}{\tau_{Y_1}(t)} f_t(x)\,\mathrm{d}x\bigg) = \lim_{r\to\infty}\frac{\varphi(r)}{r}\bigg(1 - \frac{\tau_{Y_2}(t)}{\tau_{Y_1}(t)}\bigg) \ge 0.$$

Using an argument similar to that used for the corresponding term in the proof of Theorem 6, the last expectation in (16) is nonpositive. This completes the proof.

Remark. Theorems 6 and 7 examine the relation between the distributions F^t, G_1^t, and G_2^t, and the distributions F_t, $G_{1,t}$, and $G_{2,t}$. It is interesting to point out here that whenever $\lim_{r\to\infty}\varphi(r)/r \le 0$, (13) and (15) simplify to

$$I^{\varphi}_{X,Y_1}(t) < I^{\varphi}_{X,Y_2}(t) \quad \text{and} \quad \overline{I}^{\varphi}_{X,Y_1}(t) < \overline{I}^{\varphi}_{X,Y_2}(t),$$

respectively. The above results imply that the reference distribution F^t is closer to G_1^t than to G_2^t and that the reference distribution F_t is closer to $G_{1,t}$ than to $G_{2,t}$. Such cases are encountered when, among others, φ is taken to be equal to $\varphi_2(\cdot)$ for $a < 0$, or $\varphi_3(\cdot)$ for $a < -1$, or $\varphi(u) = -\log(u)$, which corresponds to the minimum discrimination information measure, or $\varphi(u) = (u-1)^2/(u+1)^2$, which is the function that corresponds to the measure of Balakrishnan and Sanghvi (1968).

4. Divergence measures in survival and reliability models

In this section we provide the formulae for the proposed discrimination measures between two random variables for the case of the proportional hazards and reverse hazards models. Then we evaluate the divergence between the Cox proportional hazards model and the frailty or transformation model.

4.1. Proportional hazards and proportional reverse hazards models

We will concentrate first on the case of proportional hazards. Let X and Y be random variables with distribution functions F and G, respectively, for which it holds that

$$\overline{G}(x) = (\overline{F}(x))^{\theta} \quad \text{for all } x > 0 \text{ and } \theta > 0. \tag{17}$$

See Cox (1972) for details and Ebrahimi and Kirmani (1996a) for a connection with divergence measures.

Theorem 8. (i) *The discrimination measure $I^{\varphi}_{X,Y}(t)$ between the random variables X and Y which satisfy the proportional hazards assumption (17) is independent of t and is given by*

$$I^{\varphi}_{X,Y}(t) = \int_0^1 \varphi\left(\frac{1}{\theta y^{\theta-1}}\right) \mathrm{d}y^{\theta}. \tag{18}$$

(ii) *If $I^{\varphi}_{X,Y}(t)$ is independent of t and given by (18), then there exists a constant $\theta > 0$ such that (17) holds.*

Proof. We have

$$\begin{aligned} I^{\varphi}_{X,Y}(t) &= \int_t^{\infty} \frac{g(x)}{\overline{G}(t)} \varphi\left(\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}\right) \mathrm{d}x \\ &= \int_t^{\infty} \frac{\theta(\overline{F}(x))^{\theta-1} f(x)}{(\overline{F}(t))^{\theta}} \varphi\left(\frac{f(x)/\overline{F}(t)}{\theta(\overline{F}(x))^{\theta-1} f(x)/(\overline{F}(t))^{\theta}}\right) \mathrm{d}x \\ &= \int_t^{\infty} \theta\left(\frac{\overline{F}(x)}{\overline{F}(t)}\right)^{\theta-1} \frac{f(x)}{\overline{F}(t)} \varphi\left(\frac{1}{\theta}\left(\frac{\overline{F}(t)}{\overline{F}(x)}\right)^{\theta-1}\right) \mathrm{d}x, \end{aligned}$$

which, by a change of variable $y = \overline{F}(x)/\overline{F}(t)$, is equal to

$$\int_0^1 \varphi\left(\frac{1}{\theta y^{\theta-1}}\right) \mathrm{d}y^{\theta},$$

which is independent of t.

The reverse is shown by following the reverse steps of the proof.

Now we will focus on the proportional reverse hazards model, which is defined as

$$G(x) = (F(x))^{\theta} \quad \text{for all } x > 0 \text{ and } \theta > 0. \tag{19}$$

See Gupta *et al*. (1998), Di Crescenzo (2000), and Di Crescenzo and Longobardi (2004) for this model and its connection to divergence measures.

Theorem 9. (i) *The discrimination measure $\overline{I}^{\varphi}_{X,Y}(t)$ between the random variables X and Y which satisfy the proportional reverse hazards assumption (19) is independent of t and is* *given by*

$$\overline{I}^{\varphi}_{X,Y}(t) = \int_0^1 \varphi\left(\frac{1}{\theta y^{\theta-1}}\right) \mathrm{d}y^{\theta}. \tag{20}$$

(ii) *If* $\overline{I}^{\varphi}_{X,Y}(t)$ *is independent of* t *and given by (20), then there exists a constant* $\theta > 0$ *such that (19) holds.*

Proof. We have

$$\begin{aligned}\overline{I}^{\varphi}_{X,Y}(t) &= \int_0^t \frac{g(x)}{G(t)} \varphi\left(\frac{f(x)/F(t)}{g(x)/G(t)}\right) \mathrm{d}x \\ &= \int_0^t \frac{\theta(F(x))^{\theta-1} f(x)}{(F(t))^{\theta}} \varphi\left(\frac{f(x)/F(t)}{\theta(F(x))^{\theta-1} f(x)/(F(t))^{\theta}}\right) \mathrm{d}x \\ &= \int_0^t \theta\left(\frac{F(x)}{F(t)}\right)^{\theta-1} \frac{f(x)}{F(t)} \varphi\left(\frac{1}{\theta}\left(\frac{F(t)}{F(x)}\right)^{\theta-1}\right) \mathrm{d}x,\end{aligned}$$

which, by a change of variable $y = F(x)/F(t)$, is equal to

$$\int_0^1 \varphi\left(\frac{1}{\theta y^{\theta-1}}\right) \mathrm{d}y^{\theta},$$

which is independent of t.

The reverse is shown by following the reverse steps of the proof.

4.1.1. *Examples and applications.* If we consider the function $\varphi(u) = -\log(u) + u - 1$, the discrimination measure $I^{\varphi}_{X,Y}(t)$ for model (17) and the discrimination measure $\overline{I}^{\varphi}_{X,Y}(t)$ for model (19) simplify to

$$\log(\theta) - \frac{\theta - 1}{\theta},$$

which is known as the minimum discrimination information (MDI) between X and Y. Obviously, when $\theta = 1$, the distance between X and Y becomes 0.

If we consider the function $\varphi(u) = u\log(u) - u + 1$, the discrimination measure $I^{\varphi}_{X,Y}(t)$ for model (17) and the discrimination measure $\overline{I}^{\varphi}_{X,Y}(t)$ for model (19) simplify to

$$-\log(\theta) + \theta - 1,$$

which is the Kullback–Leibler (KL) divergence. Obviously, when $\theta = 1$, the distance between X and Y becomes 0. (The graph of the KL distance between X and Y as a function of θ is given in Figure 3, below.)

If we consider the Cressie and Read (CR) function $\varphi_1(\cdot)$, the discrimination measure $I^{\varphi_1}_{X,Y}(t)$ for model (17) and the discrimination measure $\overline{I}^{\varphi_1}_{X,Y}(t)$ for model (19) become

$$\frac{1}{a(1+a)}\left(\frac{1}{\theta^{a}(a(1-\theta)+1)} - 1\right). \tag{21}$$

The graph of the CR distance between X and Y for various values of the index a, including the value $a = \frac{2}{3}$, which is considered to be the best choice of a for the CR function (for details, see Cressie and Read (1984)), is given in Figure 1. As a tends to 0, the measure tends to $-\log(\theta) + \theta - 1$, which is the KL divergence.

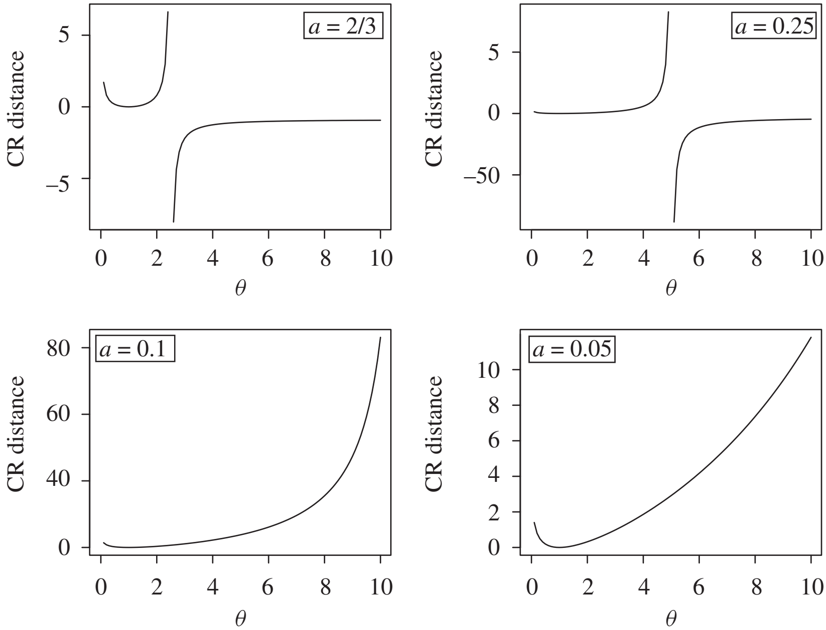

FIGURE 1: The CR discrimination measure as a function of the index a and the parameter $\theta \in (0, 10)$. The limit for $a \to 0$ (not shown) is the KL measure.

If we consider the BHHJ function

$$\varphi_2(u) = u^{1+a} - \left(1 + \frac{1}{a}\right)u^a + \frac{1}{a}, \qquad a > 0,$$

the discrimination measures $I^{\varphi_2}_{X,Y}(t)$ and $\overline{I}^{\varphi_2}_{X,Y}(t)$ for models (17) and (19) take the form

$$\frac{1}{\theta^a(1 - a(\theta - 1))} - \frac{a+1}{a}\frac{1}{\theta^{a-1}(\theta - a(\theta - 1))} + \frac{1}{a}.$$

The graph of the BHHJ distance between X and Y as a function of θ, based on the function φ_2 for various values of the index a is given in Figure 2. Note that, when a tends to 0, the measure tends to the MDI measure between X and Y, that is, $\log(\theta) - (\theta - 1)/\theta$.

If we consider the function

$$\varphi_3(u) = 1 - \left(1 + \frac{1}{a}\right)u + \frac{u^{1+a}}{a}, \qquad a > 0,$$

the discrimination measure $I^{\varphi_3}_{X,Y}(t)$ for model (17) and the discrimination measure $\overline{I}^{\varphi_3}_{X,Y}(t)$ for model (19) become

$$1 - \frac{a+1}{a} + \frac{1}{a}\frac{1}{\theta^a(1 + a(1 - \theta))}. \tag{22}$$

As a is tending to 0, the measure tends to $-\log(\theta) + \theta - 1$, which is the KL information. Observe that function φ_3 is the same as the CR function φ_2, apart from a factor $1/(1+a)$. Note that the same relationship holds for the measures as well, that is, the measure that corresponds to Cressie and Read given in (21) is the measure given in (22) multiplied by a factor $1/(1+a)$. As a goes to 0, they both tend to the KL divergence. The graph for the φ_3 divergence is given in Figure 3.

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

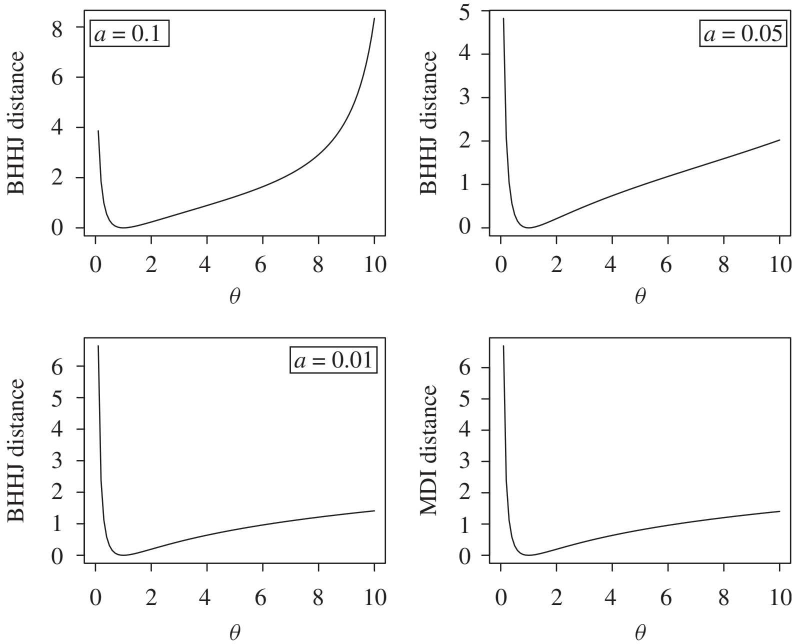

FIGURE 2: The BHHJ discrimination measure as a function of the index a and the parameter $\theta \in (0, 10)$. The limit for $a \to 0$ is the MDI measure.

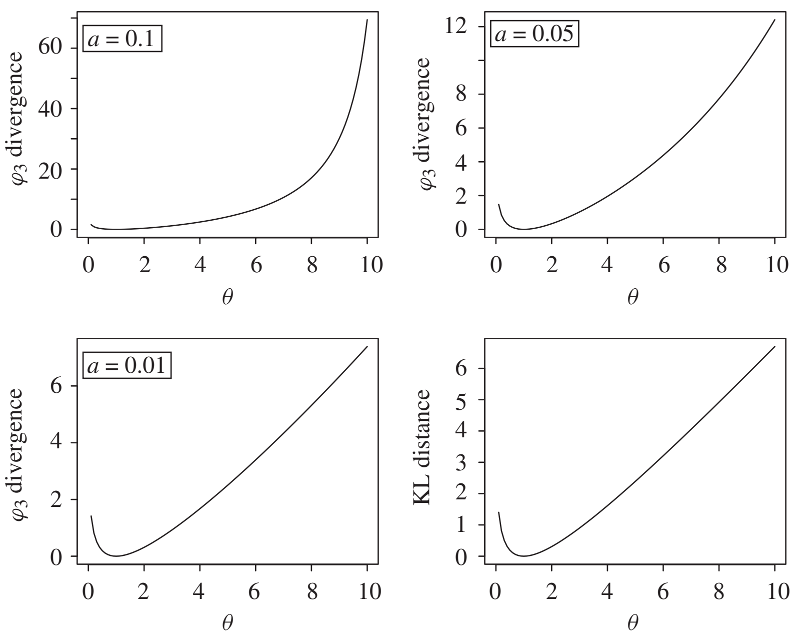

FIGURE 3: The φ_3 discrimination measure as a function of the index a and the parameter $\theta \in (0, 10)$. The limit for $a \to 0$ is the KL measure.

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

4.2. Frailty or transformation model versus Cox proportional hazards model

Let X and Y be random variables with distribution functions F_1 and F_2, probability density functions f_1 and f_2, and survival functions S_1 and S_2, respectively. Let H be the baseline cumulative hazard function, and let h be the baseline intensity hazard function. Let X follow a Cox model (see Cox (1972)) under which

$$S_1(x) = \mathrm{e}^{-\theta H(x)}, \qquad \theta > 0,$$

and let Y follow a frailty model (see Vonta (1996)) under which

$$S_2(x) = \mathrm{e}^{-G(\theta H(x))}, \qquad \theta > 0, \tag{23}$$

where the function G is assumed to be concave and increasing with $G(0) = 0$ and $G(\infty) = \infty$.

In this section we provide the usual divergence between the distributions of X and Y using the KL divergence (Theorem 10, below) and Csiszár's divergence (Theorem 11, below), as well as the φ-distance between the respective past (Theorem 12, below) and residual lifetimes (Theorem 13, below) of X and Y.

Theorem 10. *The discrimination measure $I_{X,Y}$ between the random variables X and Y which follow the Cox proportional hazards model and the frailty or transformation model (23), respectively, is given by*

$$\int_0^\infty \mathrm{e}^{-y}(G(y) - \log(G'(y)))\,\mathrm{d}y - 1.$$

Proof. For the KL discrimination measure, we have

$$\begin{aligned} I_{X,Y} &= \int_0^\infty f_1(x)\log\biggl(\frac{f_1(x)}{f_2(x)}\biggr)\,\mathrm{d}x \\ &= \int_0^\infty \mathrm{e}^{-\theta H(x)}\theta h(x)\log\biggl(\frac{\mathrm{e}^{-\theta H(x)}\theta h(x)}{\mathrm{e}^{-G(\theta H(x))}G'(\theta H(x))\theta h(x)}\biggr)\,\mathrm{d}x \\ &= \int_0^\infty \mathrm{e}^{-\theta H(x)}(G(\theta H(x)) - \theta H(x) - \log(G'(\theta H(x))))\,\mathrm{d}(\theta H(x)), \end{aligned}$$

which, by the change of variable $y = \theta H(x)$, becomes

$$\int_0^\infty \mathrm{e}^{-y}(G(y) - y - \log(G'(y)))\,\mathrm{d}y = \int_0^\infty \mathrm{e}^{-y}(G(y) - \log(G'(y)))\,\mathrm{d}y - 1.$$

Theorem 11. *The discrimination measure $I^{\varphi}_{X,Y}$ between the random variables X and Y which follow the Cox proportional hazards model and the frailty or transformation model (23), respectively, is given by*

$$\int_0^\infty \mathrm{e}^{-G(y)}G'(y)\varphi\biggl(\frac{\mathrm{e}^{-y}}{\mathrm{e}^{-G(y)}G'(y)}\biggr)\,\mathrm{d}y.$$

Proof. It is easy to see that the $I^{\varphi}_{X,Y}$ measure is equal to

$$\int_0^\infty \mathrm{e}^{-G(\theta H(x))}G'(\theta H(x))\theta h(x)\varphi\biggl(\frac{\mathrm{e}^{-\theta H(x)}\theta h(x)}{\mathrm{e}^{-G(\theta H(x))}G'(\theta H(x))\theta h(x)}\biggr)\,\mathrm{d}x,$$

which, by the change of variable $y = \theta H(x)$, takes the desired form.

4.2.1. *The measure* $I^{\varphi}_{X,Y}(t)$.

Theorem 12. *The discrimination measure* $I^{\varphi}_{X,Y}(t)$ *between the random variables* X *and* Y *which follow the Cox proportional hazards model and the frailty or transformation model (23), respectively, is given by*

$$\int_{\theta H(t)}^{\infty} \frac{\mathrm{e}^{-G(y)} G'(y)}{\mathrm{e}^{-G(\theta H(t))}} \varphi\biggl(\frac{\mathrm{e}^{-y}/\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(y)} G'(y)/\mathrm{e}^{-G(\theta H(t))}}\biggr) \,\mathrm{d}y \tag{24}$$

for $t > 0$.

Proof. The measure $I^{\varphi}_{X,Y}(t)$ for $t > 0$ is equal to

$$\int_{t}^{\infty} \frac{\mathrm{e}^{-G(\theta H(x))} G'(\theta H(x))}{\mathrm{e}^{-G(\theta H(t))}} \varphi\biggl(\frac{\mathrm{e}^{-\theta H(x)}/\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(\theta H(x))} G'(\theta H(x))/\mathrm{e}^{-G(\theta H(t))}}\biggr) \,\mathrm{d}(\theta H(x)),$$

which, by a change of variable, yields (24).

Examples. For the function $\varphi(u) = -\log(u) + u - 1$, the measure defined in (24) becomes

$$\frac{1}{\mathrm{e}^{-G(\theta H(t))}} \int_{\theta H(t)}^{\infty} \mathrm{e}^{-G(y)} G'(y)(y - G(y) + \log(G'(y))) \,\mathrm{d}y + G(\theta H(t)) - \theta H(t),$$

since $\int_{\theta H(t)}^{\infty} \mathrm{e}^{-G(y)} G'(y) \,\mathrm{d}y = -\int_{\theta H(t)}^{\infty} \mathrm{d}\mathrm{e}^{-G(y)} = \mathrm{e}^{-G(\theta H(t))}$, and it is further simplified to

$$\begin{aligned} &\frac{1}{\mathrm{e}^{-G(\theta H(t))}} \int_{\theta H(t)}^{\infty} \mathrm{e}^{-G(y)} G'(y)(y + \log(G'(y))) \,\mathrm{d}y - \theta H(t) - 1 \\ &\quad = \frac{1}{\mathrm{e}^{-G(\theta H(t))}} \int_{\theta H(t)}^{\infty} (\mathrm{e}^{-G(y)} + \mathrm{e}^{-G(y)} G'(y) \log(G'(y))) \,\mathrm{d}y - 1, \end{aligned}$$

by integration by parts.

If we consider the function $\varphi(u) = u\log(u) - u + 1$, the measure defined in (24) becomes

$$\frac{1}{\mathrm{e}^{-\theta H(t)}} \int_{\theta H(t)}^{\infty} \mathrm{e}^{-y}(-y + G(y) - \log(G'(y))) \,\mathrm{d}y - G(\theta H(t)) + \theta H(t)), \tag{25}$$

which, by integration by parts, becomes

$$\frac{1}{\mathrm{e}^{-\theta H(t)}} \int_{\theta H(t)}^{\infty} \mathrm{e}^{-y}(G'(y) - \log(G'(y))) \,\mathrm{d}y - 1. \tag{26}$$

If we consider the function φ_1 of Cressie and Read then the discrimination measure (24) becomes

$$\frac{1}{a(1+a)}\biggl(\frac{\mathrm{e}^{-aG(\theta H(t))}}{\mathrm{e}^{-(a+1)\theta H(t)}} \int_{\theta H(t)}^{\infty} \frac{\mathrm{e}^{-(a+1)y}}{\mathrm{e}^{-aG(y)}(G'(y))^{a}} \,\mathrm{d}y - 1\biggr). \tag{27}$$

When a tends to 0, the above measure tends, as expected, to the measure that corresponds to the function $\varphi(u) = u\log(u) - u + 1$ and is given in (25). This measure is the KL measure at time t between two random variables X and Y that follow the Cox model and the frailty or transformation model.

If we consider the function φ_2, the discrimination measure (24) becomes

$$\begin{aligned}
&\int_{\theta H(t)}^{\infty} \frac{\mathrm{e}^{-G(y)}G'(y)}{\mathrm{e}^{-G(\theta H(t))}} \bigg\{ \bigg(\frac{\mathrm{e}^{-y}/\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(y)}G'(y)/\mathrm{e}^{-G(\theta H(t))}} \bigg)^{a} \\
&\qquad\qquad \times \bigg(\frac{\mathrm{e}^{-y}/\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(y)}G'(y)/\mathrm{e}^{-G(\theta H(t))}} - \frac{a+1}{a} \bigg) + \frac{1}{a} \bigg\} \,\mathrm{d}y \\
&= \frac{\mathrm{e}^{-aG(\theta H(t))}}{\mathrm{e}^{-a\theta H(t)}} \bigg\{ \int_{\theta H(t)}^{\infty} \frac{1}{\mathrm{e}^{-G(\theta H(t))}} \frac{\mathrm{e}^{-(a+1)y}}{\mathrm{e}^{-aG(y)}(G'(y))^{a}} \,\mathrm{d}y \\
&\qquad - \frac{a+1}{a} \int_{\theta H(t)}^{\infty} \frac{1}{\mathrm{e}^{-G(\theta H(t))}} \frac{\mathrm{e}^{-ay}}{\mathrm{e}^{-(a-1)G(y)}(G'(y))^{a-1}} \,\mathrm{d}y \bigg\} + \frac{1}{a}.
\end{aligned}$$

As a tends to 0, the above discrimination measure tends to the measure that corresponds to the function $\varphi(u) = -\log(u) + u - 1$, given in (31), below, which is the MDI at time t between two random variables X and Y that follow the Cox model and the frailty or transformation model.

Recall that the function φ_3 is equal to the CR function multiplied by a factor of $1 + a$ and, therefore, it is easy to prove that the discrimination measure (24), in the case of φ_3, is the measure (27) multiplied by a factor of $1 + a$. The two measures (related to φ_1 and φ_3) become equal and equal to the KL divergence given in (25) as a tends to 0.

4.2.2. *The measure* $\overline{I}^{\varphi}_{X,Y}(t)$.

Theorem 13. *The discrimination measure* $\overline{I}^{\varphi}_{X,Y}(t)$ *between the random variables* X *and* Y *which follow the Cox proportional hazards model and the frailty or transformation model (23), respectively, is given by*

$$\int_{0}^{\theta H(t)} \frac{\mathrm{e}^{-G(y)}G'(y)}{1-\mathrm{e}^{-G(\theta H(t))}} \varphi\bigg(\frac{\mathrm{e}^{-y}/1-\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(y)}G'(y)/1-\mathrm{e}^{-G(\theta H(t))}} \bigg) \,\mathrm{d}y \tag{28}$$

for $t > 0$.

Proof. The measure $\overline{I}^{\varphi}_{X,Y}(t)$ for $t > 0$ is equal to

$$\int_{0}^{t} \frac{\mathrm{e}^{-G(\theta H(x))}G'(\theta H(x))}{1-\mathrm{e}^{-G(\theta H(t))}} \varphi\bigg(\frac{\mathrm{e}^{-\theta H(x)}/1-\mathrm{e}^{-\theta H(t)}}{\mathrm{e}^{-G(\theta H(x))}G'(\theta H(x))/1-\mathrm{e}^{-G(\theta H(t))}} \bigg) \,\mathrm{d}(\theta H(x)),$$

which, by a change of variable, yields (28).

Examples. For the function $\varphi(u) = -\log(u) + u - 1$, the measure defined in (28) becomes

$$\frac{1}{1-\mathrm{e}^{-G(\theta H(t))}} \int_{0}^{\theta H(t)} \mathrm{e}^{-G(y)}G'(y)(y - G(y) + \log(G'(y)))\,\mathrm{d}y - \log\bigg(\frac{1-\mathrm{e}^{-G(\theta H(t))}}{1-\mathrm{e}^{-\theta H(t)}} \bigg). \tag{29}$$

If we consider the function $\varphi(u) = u\log(u) - u + 1$, the measure defined in (28) becomes

$$\frac{1}{1-\mathrm{e}^{-\theta H(t)}} \int_{0}^{\theta H(t)} \mathrm{e}^{-y}(-y + G(y) - \log(G'(y)))\,\mathrm{d}y + \log\bigg(\frac{1-\mathrm{e}^{-G(\theta H(t))}}{1-\mathrm{e}^{-\theta H(t)}} \bigg). \tag{30}$$

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

If we consider the function φ_1 of Cressie and Read then the discrimination measure (28) becomes

$$\frac{1}{a(1+a)}\biggl(\frac{(1-\mathrm{e}^{-G(\theta H(t))})^a}{(1-\mathrm{e}^{-\theta H(t)})^{a+1}}\int_0^{\theta H(t)}\frac{\mathrm{e}^{-(a+1)y}}{\mathrm{e}^{-aG(y)}(G'(y))^a}\,\mathrm{d}y-1\biggr). \tag{31}$$

When a tends to 0, the above measure tends, as expected, to the measure that corresponds to the function $\varphi(u)=u\log(u)-u+1$ and is given in (30). This measure is the KL measure between the past lifetimes of two random variables X and Y that follow the Cox model and the frailty or transformation model, respectively.

For the φ_2 function, it is easy to see that the discrimination measure (28) takes the form

$$\begin{aligned}\frac{(1-\mathrm{e}^{-G(\theta H(t))})^a}{(1-\mathrm{e}^{-\theta H(t)})^a}\biggl\{&\int_0^{\theta H(t)}\frac{1}{1-\mathrm{e}^{-G(\theta H(t))}}\,\frac{\mathrm{e}^{-(a+1)y}}{\mathrm{e}^{-aG(y)}(G'(y))^a}\,\mathrm{d}y\\ &-\frac{a+1}{a}\int_0^{\theta H(t)}\frac{1}{1-\mathrm{e}^{-G(\theta H(t))}}\,\frac{\mathrm{e}^{-ay}}{\mathrm{e}^{-(a-1)G(y)}(G'(y))^{a-1}}\,\mathrm{d}y\biggr\}+\frac{1}{a}.\end{aligned}$$

As a tends to 0, the above discrimination measure tends to the measure that corresponds to the function $\varphi(u)=-\log(u)+u-1$, given in (29), which is the MDI between the past lifetimes of two random variables X and Y that follow the Cox model and the frailty or transformation model.

Recall that function φ_3 is equal to the CR function multiplied by a factor of $1+a$ and, therefore, the discrimination measure (28) becomes, for φ_3, equal to the one given in (31) multiplied by a factor of $1+a$. The two measures (relative to φ_1 and φ_3) become equal and equal to the KL divergence given in (30) as a tends to 0.

4.2.3. *Examples of the function G.* If $G(x)=x$, which is the function that corresponds to the Cox model, the discrimination measures defined in this section become, as expected, equal to 0 (recall that $\varphi(1)=0$).

If we consider the function $G(x,c)=(1/c)\log(1+cx)$ for $c>0$, which corresponds to a gamma-distributed frailty with mean 1 and variance c, then we have the measure

$$I_{X,Y}=\frac{1+c}{c}\int_0^\infty \mathrm{e}^{-x}\log(1+cx)\,\mathrm{d}x-1=(1+c)\int_0^\infty\frac{\mathrm{e}^{-x}}{1+cx}\,\mathrm{d}x-1,$$

by integration by parts.

If we consider the function $G(x,b)=2\sqrt{b(x+b)}-2b$ for $b>0$, which corresponds to an inverse Gaussian-distributed frailty with mean 1 and variance $1/2b$, then we have the measure

$$\begin{aligned}I_{X,Y}&=\int_0^\infty \mathrm{e}^{-x}\biggl(2\sqrt{b(b+x)}-2b-\log\biggl(\frac{\sqrt{b}}{\sqrt{b+x}}\biggr)\biggr)\,\mathrm{d}x-1\\ &=-2b-\frac{1}{2}\log b+2\sqrt{b}\int_0^\infty \mathrm{e}^{-x}(b+x)^{1/2}\,\mathrm{d}x+\frac{1}{2}\int_0^\infty \mathrm{e}^{-x}\log(b+x)\,\mathrm{d}x-1\\ &=-2b-\frac{1}{2}\log b+2\sqrt{b}\mathrm{e}^b\int_b^\infty \mathrm{e}^{-y}y^{1/2}\,\mathrm{d}y+\frac{1}{2}\mathrm{e}^b\int_b^\infty \mathrm{e}^{-y}\log y\,\mathrm{d}y-1.\end{aligned}$$

Similarly, we can evaluate the other discrimination measures defined in this paper, namely, $I^{\varphi}_{X,Y}$, $I^{\varphi}_{X,Y}(t)$, and $\overline{I}^{\varphi}_{X,Y}(t)$. For example, for the functions $\varphi(u)=u\log(u)+1-u$ and

https://doi.org/10.1239/jap/1269610827 Published online by Cambridge University Press

$G(x, c) = (1/c)\log(1 + cx)$, we have, from (26),

$$I^{\varphi}_{X,Y}(t) = \frac{1}{\mathrm{e}^{-\theta H(t)}}\bigg(\int_{\theta H(t)}^{\infty} \mathrm{e}^{-y}\Big((1+cy)^{-1} - \log(1+cy)^{-1}\Big)\,\mathrm{d}y\bigg) - 1$$
$$= \log(1+\theta H(t)) + \frac{c+1}{\mathrm{e}^{-\theta H(t)}}\int_{\theta H(t)}^{\infty} \frac{\mathrm{e}^{-y}}{1+cy}\,\mathrm{d}y - 1,$$

by integration by parts.

5. Discussion

In this paper we defined measures of divergence between past lifetimes and between residual lifetimes of two items that are known to have survived up to a time t. The measures are based on a class of functions known as Csiszár's class of functions that satisfy some regularity conditions. We examined properties of these measures and also calculated the proposed divergence measures between two distributions in a number of examples, such as the divergence measure between two random variables that satisfy the proportional hazards assumption as well as the proportional reverse hazards assumption.

It is worth pointing out that both measures of divergence, that is, between residual or past lifetimes, are independent of t in the cases of proportional hazards or proportional reverse hazards. More specifically, the measure $I^{\varphi}_{X,Y}(t)$ for the proportional hazards case and $\overline{I}^{\varphi}_{X,Y}(t)$ for the proportional reverse hazards case are equal and, therefore, the information that we obtain for two items that have survived up to a point t is the same if we look in the past or in the future as long as an appropriate proportionality holds for their distributions. The two measures are also independent of how long the two items have survived.

In Figure 1 we provided the divergence measure based on the Cressie and Read (CR) function for various values of the index a, including the value $a = \frac{2}{3}$, which is claimed to be the optimum for tests in multinomial populations (see Cressie and Read (1984)). If we let a approach 0 then the CR divergence resembles very closely that of Figure 3 due to the fact that, as $a \to 0$, the CR measure reduces to the KL measure. In Figure 2 we presented the BHHJ divergence measure for the function $\varphi_2(u)$ for various values of the index a. It is interesting to point out that this divergence is almost identical with the graph of the minimum discrimination information divergence. This is due to the fact that the information based on φ_2 as a tends to 0 converges to the minimum discrimination information. The divergence based on the function φ_3 was presented in Figure 3. The graph also included the Kullback–Leibler distance between the residual lives of X and Y that follow the model (17) or between past lives of X and Y that follow the model (19), given that both items have survived up to time t. It was pointed out that the KL divergence is the limit as $a \to 0$ of both the φ_3 function and the CR divergence function.

We have also provided the distance between two random variables X and Y that follow a Cox model and a frailty or transformation model, respectively. The frailty or transformation model is an extension of the Cox model which arises when we introduce a random effect into the Cox model in order to explain possible heterogeneity present in the population. This distance depends on a function G which is determined by the distribution of the frailty or random effect and satisfies some regularity conditions. We further furnished the φ-divergence between past lifetimes and residual lifetimes of X and Y when they have survived up to some time t.

Note that the proposed measures of divergence could be used for goodness-of-fit tests. We could utilize these divergence measures to compare two lifetime distributions or to decide which model is closest to the true model between competing candidate models. This part is left for future research.

Acknowledgement

The authors would like to thank an anonymous referee for comments and suggestions that greatly improved the quality of the paper.

References

ALI, S. M. AND SILVEY, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. R. Statist. Soc. B* **28,** 131–142.

BALAKRISHNAN, V. AND SANGHVI, L. D. (1968). Distance between populations on the basis of attribute. *Biometrics* **24,** 859–865.

BASU, A., HARRIS, I. R., HJORT, N. L. AND JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85,** 549–559.

COX, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc. B* **34,** 187–202.

CRESSIE, N. AND READ, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. R. Statist. Soc. B* **46,** 440–464.

CSISZÁR, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat on Markhoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8,** 84–108.

DI CRESCENZO, A. (2000). Some results on the proportional reversed hazards model. *Statist. Prob. Lett.* **50,** 313–321.

DI CRESCENZO, A. AND LONGOBARDI, M. (2004). A measure of discrimination between past lifetime distributions. *Statist. Prob. Lett.*, **67,** 173–182.

EBRAHIMI, N. AND KIRMANI, S. N. U. A. (1996a). A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions. *Biometrika* **83,** 233–235.

EBRAHIMI, N. AND KIRMANI, S. N. U. A. (1996b). A measure of discrimination between two residual life-time distributions and its applications. *Ann. Inst. Statist. Math.* **48,** 257–265.

GUPTA, R. C., GUPTA, P. L. AND GUPTA, R. D. (1998). Modeling failure time data by Lehmann alternatives. *Commun. Statist. Theory Meth.* **27,** 887–904.

KAGAN, A. M. (1963). On the theory of Fisher's amount of information. *Soviet Math. Dokl.* **4,** 991–993.

KIRMANI, S. N. U. A. AND GUPTA, R. C. (2001). On the proportional odds model in survival analysis. *Ann. Inst. Statist. Math.* **53,** 203–216.

KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Instituto Italiano degli Attuari* **4,** 83–91.

KULLBACK, S. AND LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22,** 79–86.

LÉVY, P. (1925). *Calcul des Probabilites.* Gauthiers-Villars, Paris.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Acad. Sci. India* **2,** 49–55.

MATTHEOU, K., LEE, S. AND KARAGRIGORIOU, A. (2009). A model selection criterion based on the BHHJ measure of divergence. *J. Statist. Planning Infer.* **139,** 228–235.

MATUSITA, K. (1967). On the notion of affinity of several distributions and some of its applications. *Ann. Inst. Statist. Math.* **19,** 181–192.

PARDO, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC, Boca Raton, FL.

PEARSON, K. (1900). On the criterion that a given system of deviation from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.* **50,** 157–172.

SHAKED, M. AND SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston, MA.

VONTA, F. (1996). Efficient estimation in a non-proportional hazards model in survival analysis. *Scand. J. Statist.* **23,** 49–61.