Statistics and Probability Letters 81 (2011) 1594–1598

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Characterizations of bivariate models using dynamic Kullback–Leibler discrimination measures

J. Navarro [a,*], S.M. Sunoj [b], M.N. Linu [b]

[a] Universidad de Murcia, Murcia, Spain

[b] Cochin University of Science and Technology, Cochin, Kerala, India

ARTICLE INFO

Article history:
Received 11 February 2011
Received in revised form 9 May 2011
Accepted 26 May 2011
Available online 14 June 2011

MSC:
62N05
62B10

Keywords:
Kullback–Leibler discrimination information
Weighted distributions
Conditionally specified models
Proportional hazard rate

ABSTRACT

In this paper, the residual Kullback–Leibler discrimination information measure is extended to conditionally specified models. The extension is used to characterize some bivariate distributions. These distributions are also characterized in terms of proportional hazard rate models and weighted distributions. Moreover, we also obtain some bounds for this dynamic discrimination function by using the likelihood ratio order and some preceding results.

1. Introduction

Let X and Y be two absolutely continuous random variables with the common support $S = (l, \infty)$ for $l \geq 0$. Denote by f, F and $\overline{F}$, the probability density function (PDF), the cumulative distribution function (CDF) and the survival (or reliability) function (SF) of X, respectively, and by g, G and $\overline{G}$, the corresponding functions of Y. As an information distance between X and Y, Kullback and Leibler (1951) proposed a directed divergence (also known as information divergence, information gain, relative entropy or discrimination measure) given by

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

This function is a measure of the similarity (closeness) between the two distributions and it plays an important role in information theory, reliability and other related fields. Further, note that if $f = g$ (a.e.), then $I_{X,Y} = 0$.

Length of time during a study period has been considered as a prime variable of interest in many fields such as reliability, survival analysis, economics, business, etc. In particular, consider an item under study, then the information about the residual (or past) lifetime is an important task in many applications. In such cases, the information measures are

* Corresponding address: Facultad de Matematicas, Universidad de Murcia, 30100 Murcia, Spain. Tel.: +34 868883509; fax: +34 868884182.
E-mail address: jorgenav@um.es (J. Navarro).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2011.05.016

functions of time, and thus they are dynamic. Based on this idea, Ebrahimi and Kirmani (1996) defined the Kullback–Leibler discrimination information measure of X and Y at time t by

$$I_{X,Y}(t)=\int_t^\infty \frac{f(x)}{\overline{F}(t)}\log\frac{f(x)\overline{G}(t)}{g(x)\overline{F}(t)}\mathrm{d}x.$$

Note that $I_{X,Y}(t)=I_{X_t,Y_t}$, where $X_t=(X-t|X>t)$ and $Y_t=(Y-t|Y>t)$ are the residual lifetimes associated to X and Y. A similar function is obtained in terms of the inactivity times $(t-X|X<t)$ and $(t-Y|Y<t)$ in Di Crescenzo and Longobardi (2004) (see also Maya and Sunoj, 2008). Interesting extensions to the multivariate case are obtained in Ebrahimi et al. (2007). For additional information on these measures, see Ebrahimi et al. (2010) and the references therein.

The concept of weighted distributions is usually considered in connection with modeling statistical data, where the usual practice of employing standard distributions is not found appropriate in some cases. In recent years, this concept has been applied in many areas of statistics, such as analysis of family size, human heredity, wildlife population study, renewal theory, biomedical, statistical ecology, reliability modeling, etc. Associated to a random variable X with PDF f and to a nonnegative real function w, we can define the weighted random variable X^w with density function

$$f^w(x)=\frac{w(x)f(x)}{E(w(X))},$$

where we assume $0<E(w(X))<\infty$. When $w(x)=x$, X^w is called the length (or size) biased random variable and it is denoted by X^*. For recent works on weighted distributions, we refer the reader to Bartoszewicz and Skolimowska (2006); Navarro et al. (2006); Blazej (2008); Maya and Sunoj (2008); Bartoszewicz (2009); Navarro and Sarabia (2010); Sunoj and Linu (in press).

The obtention of the joint distribution of (X,Y), when conditional distributions of $(X|Y=y)$ and $(Y|X=x)$ are known, has been an important problem dealt with by many researchers in the past. This approach of identifying a bivariate density using the conditionals is called the conditional specification of the joint distribution (see Arnold et al., 1999). These conditional models are often useful in many two component reliability systems, when the operational status of one component is known.

In the present paper, the Kullback–Leibler discrimination information measure $I_{X,Y}(t)$ proposed by Ebrahimi and Kirmani (1996) is extended to conditionally specified models. This extension is used to characterize some bivariate distributions. These distributions are also characterized in terms of proportional hazard rate models and weighted distributions. These results are given in Section 2. Moreover, in Section 3, we obtain bounds for this dynamic discrimination function by using the likelihood ratio order and some preceding results. The proof of the main result is given in Section 4.

2. Main results

Let (X_1,X_2) and (Y_1,Y_2) be two bivariate random vectors with joint PDF f and g, joint CDF F and G and joint SF $\overline{F}$ and $\overline{G}$, respectively. Let us assume that the common support is $S=(l,\infty)\times(l,\infty)$ for $l\geq 0$. Also let $f_i(s|t)$ and $g_i(s|t)$, $\overline{F}_i(s|t)$ and $\overline{G}_i(s|t)$ denote the PDF and the SF of $(X_i|X_{3-i}=t)$ and $(Y_i|Y_{3-i}=t)$, respectively, for $i=1,2$. Then we define the conditional Kullback–Leibler discrimination (CKLD) information functions as

$$I_{X_i,Y_i}(s|t)=\int_s^\infty \frac{f_i(x|t)}{\overline{F}_i(s|t)}\log\frac{f_i(x|t)\overline{G}_i(s|t)}{g_i(x|t)\overline{F}_i(s|t)}\mathrm{d}x$$

for $i=1,2$ and $s,t\geq l$. Note that

$$I_{X_i,Y_i}(s|t)=I_{(X_i|X_{3-i}=t),(Y_i|Y_{3-i}=t)}(s) \tag{1}$$

for $i=1,2$ and $s,t\geq l$. Hence, $I_{X_i,Y_i}(s|t)$ is the dynamic Kullback–Leibler discrimination information measure at time s defined by Ebrahimi and Kirmani (1996) but applied to the conditional random variables $(X_i|X_{3-i}=t)$ and $(Y_i|Y_{3-i}=t)$ for $i=1,2$. As in the univariate case, these functions measure the information distance between the residual lifetimes of the conditional distributions of the two random vectors. Of course, in the bivariate case, there are other interesting options (see, e.g., Ebrahimi et al., 2007).

In survival studies, the most widely used semi-parametric model is the proportional hazard rate (PHR) Cox model. Let X and Y be two random variables with the same support S and with hazard rate functions $h_X=f/\overline{F}$ and $h_Y=g/\overline{G}$, respectively. Then X and Y satisfy the PHR model when

$$h_Y(t)=\theta h_X(t),$$

for all $t\in S$. This relationship is also equivalent to

$$\overline{G}(t)=(\overline{F}(t))^\theta,$$

for all t (see Cox, 1959). Ebrahimi and Kirmani (1996) obtained the following result.

Theorem 1 (*Ebrahimi and Kirmani, 1996*)**.** *The function $I_{X,Y}(t)$ is constant if and only if X and Y satisfy the PHR model.*

In a similar way the random vectors (X_1, X_2) and (Y_1, Y_2) satisfy the conditional proportional hazard rate (CPHR) model (see Sankaran and Sreeja, 2007), when their respective conditional hazard rate functions satisfy

$$h_{(Y_i|Y_{3-i})}(s|t) = \theta_i(t) h_{(X_i|X_{3-i})}(s|t) \tag{2}$$

for $i = 1, 2$, where $\theta_i(t)$ is a nonnegative function of t. Then we can state the result as follows.

Theorem 2. *For $i = 1, 2$, the function $I_{X_i,Y_i}(s|t)$ only depends on t if and only if $(Y_i|Y_{3-i} = t)$ and $(X_i|X_{3-i} = t)$ satisfy the CPHR model (2).*

The proof is obtained from Theorem 1 and (1).

Next, let us consider the random vector (X_1^w, X_2^w) which has the bivariate weighted distribution associated to (X_1, X_2) and to two nonnegative real functions w_1 and w_2, that is, its joint PDF is

$$f^w(x_1, x_2) = \frac{w_1(x_1) w_2(x_2) f(x_1, x_2)}{E(w_1(X_1) w_2(X_2))}. \tag{3}$$

It is easy to see that the marginal random variable X_i^w has the (univariate) weighted distribution associated to X_i and X_i for $i = 1, 2$. In particular, the length biased bivariate random vector, denoted by (X_1^*, X_2^*), is obtained when $w_1(x) = w_2(x) = x$. There are other options in defining the bivariate weighted distribution which can be found in Navarro et al. (2006).

Now we can state the main result of the paper as follows.

Theorem 3. *Let (X_1^w, X_2^w) be a random vector which has the bivariate weighted distribution associated to (X_1, X_2) and to two nonnegative and differentiable functions w_1 and w_2. Let us assume that the support of (X_1, X_2) is $S = (l, \infty) \times (l, \infty)$ for $l \geq 0$. Then the following conditions are equivalent:*

(a) *(X_1^w, X_2^w) and (X_1, X_2) satisfy the CPHR model (2) for $i = 1, 2$.*
(b) *$I_{X_i, X_i^w}(s|t)$ only depends on t for $i = 1, 2$.*
(c) *The conditional reliability functions of (X_1, X_2) satisfy*

$$\log \overline{F}_i(s|t) = \frac{\log(w_i(s)/w_i(l))}{\theta_1(t) - 1}$$

for $i = 1, 2$.
(d) *(X_1, X_2) has the following joint PDF*

$$f(x_1, x_2) = c a_1 a_2 \frac{w_1'(x_1) w_2'(x_2)}{w_1^{a_1+1}(x_1) w_2^{a_2+1}(x_2)} \exp\left(-\phi a_1 a_2 \left(\log \frac{w_1(x_1)}{w_1(l)}\right)\left(\log \frac{w_2(x_2)}{w_2(l)}\right)\right)$$

for $x_1, x_2 \geq l$, where $c > 0$, $\phi \geq 0$ and $a_i > 1$ or $a_i < 0$ for $i = 1, 2$.

The proof is given in Section 4. Note that the conditions given in Theorem 3(c), imply that either $\log(w_i(s)/w_i(l))$ or $-\log(w_i(s)/w_i(l))$ should be cumulative hazard rate functions, that is, they should be nonnegative, increasing and they should go to ∞ when s goes to ∞. In the first case, w_i should be increasing in $[l, \infty)$ with $w_i(l) > 0$ and $w_i(\infty) = \infty$. In the second case, w_i should be decreasing, with $w_i(s) > 0$ for $s \in [l, \infty)$ and $w_i(\infty) = 0$. These conditions can also be written as

$$h_i(s|t) = \frac{w_i'(s)/w_i(s)}{1 - \theta_i(t)}$$

for $i = 1, 2$, that is, $(X_1|X_2 = t)$ and $(X_2|X_1 = t)$ satisfy the conditional proportional hazard rate model considered by Arnold and Strauss (1988). The reliability properties of this semi-parametric model can be seen in Navarro and Sarabia (in press). Actually, the model in (d) is just a truncated version of Arnold and Strauss model in the support $S = (l, \infty) \times (l, \infty)$ and when $l = 0$ both models coincide. Again we have two options, in the first one, $\lambda_i(s) = w_i'(s)/w_i(s)$ is a proper hazard rate function and, in the second one, $\lambda_i(s) = -w_i'(s)/w_i(s)$ is a proper hazard rate function. In the first option, we need $a_i > 1$ and in the second one $a_i < 0$, for $i = 1, 2$. The model in (d) contains several parametric models. In particular, when $l = 1$ and $w_1(x) = w_2(x) = x$ for $x > 1$, from Theorem 3, we can characterize the bivariate Pareto model with the following joint PDF

$$f(x_1, x_2) = \frac{c a_1 a_2}{x_1^{a_1+1} x_2^{a_2+1}} \exp(-\phi a_1 a_2 (\log x_1)(\log x_2))$$

for $x_1, x_2 \geq 1$, where $c > 0$, $a_1, a_2 > 1$ and $\phi \geq 0$.

3. Bounds

In this section, we obtain bounds for the CKLD functions by using the likelihood ratio (LR) order. The results are similar to that given in Di Crescenzo and Longobardi (2004). First, we need the definition of the LR order. If X and Y have PDF f and

g, respectively, then X is said to be less than Y in the likelihood ratio order (denoted by $X \le_{LR} Y$) if g/f is increasing in the union of their supports. Then we have the following results.

Theorem 4. *For $i = 1, 2$, if $(X_i|X_{3-i} = t) \le_{LR} (Y_i|Y_{3-i} = t)$, then*

$$I_{X_i,Y_i}(s|t) \le \log \frac{h_{X_i|X_{3-i}}(s|t)}{h_{Y_i|Y_{3-i}}(s|t)}.$$

Theorem 5. *For $i = 1, 2$, if w_i is increasing, then*

$$I_{X_i,X_i^w}(s|t) \le \log \frac{E(w_i(X_i)|X_i > s, X_{3-i} = t)}{w_i(s)}.$$

Theorem 6. *For $i = 1, 2$, if $(Y_i|Y_{3-i} = t) \le_{LR} (Z_i|Z_{3-i} = t)$, then*

$$I_{X_i,Y_i}(s|t) \ge I_{X_1,Z_1}(s|t) + \frac{h_{Z_i|Z_{3-i}}(s|t)}{h_{Y_1|Y_2}(s|t)}.$$

4. Proof of Theorem 3

The equivalence between (a) and (b) is a consequence of Theorem 2.

Let us prove that (a) implies (c). So, let us assume that (X_1^w, X_2^w) and (X_1, X_2) satisfy the CPHR model (2) for $i = 1, 2$. From the expression of the PDF of (X_1^w, X_2^w) given in (3), it is easy to prove that the PDF of $(X_i^w|X_{3-i} = t)$ is given by

$$f_i^w(s|t) = \frac{w_i(s)f_i(s|t)}{E(w_i(X_i)|X_{3-i} = t)}$$

for $i = 1, 2$, where $f_i(s|t)$ is the PDF of $(X_i|X_{3-i} = t)$. Then, for $i = 1, 2$, the hazard rate $h_i^w(s|t)$ of $(X_i^w|X_{3-i}^w = t)$ is given by

$$h_i^w(s|t) = \frac{w_i(s)f_i(s|t)}{\int_s^\infty w_i(x)f_i(x|t)\mathrm{d}x}. \tag{4}$$

Moreover, from (2), we have

$$h_i^w(s|t) = \theta_i(t)h_i(s|t)$$

and hence

$$\frac{w_i(s)f_i(s|t)}{\int_s^\infty w_i(x)f_i(x|t)\mathrm{d}x} = \theta_i(t)\frac{f_i(s|t)}{\overline{F}_i(s|t)}.$$

Therefore,

$$\frac{1}{\overline{F}_i(s|t)}\int_s^\infty w_i(x)f_i(x|t)\mathrm{d}x = \frac{1}{\theta_i(t)}w_i(s).$$

Then, differentiating both sides with respect to s, we obtain

$$-w_i(s)f_i(s|t) = \frac{1}{\theta_i(t)}(w_i'(s)\overline{F}_i(s|t) - w_i(s)f_i(s|t)),$$

that is,

$$h_i(s|t) = \frac{w_i'(s)}{(1-\theta_i(t))w_i(s)}.$$

Hence

$$\log \overline{F}_i(s|t) = -\int_l^s h_i(x|t)\mathrm{d}x = \frac{1}{(\theta_i(t)-1)}\log\frac{w_i(s)}{w_i(l)}$$

for $i = 1, 2$ and (c) holds.

Let us prove that (c) is equivalent to (d). The expressions given in (c) are equivalent to

$$h_i(s|t) = \frac{w_i'(s)/w_i(s)}{1-\theta_i(t)}$$

for $i = 1, 2$, that is, $(X_1|X_2 = t)$ and $(X_2|X_1 = t)$ satisfy the conditional proportional hazard rate model considered by Arnold and Strauss (1988) which is equivalent to (d).

Finally, let us prove that (d) implies (a). From the expression of the joint PDF given in (d), it is easy to prove that the conditional hazard rate functions are given by

$$h_i(s|t) = a_i\left(1-\phi a_{3-i}\log\frac{w_{3-i}(t)}{w_i(l)}\right)\frac{w_i'(s)}{w_i(s)}$$

for $i = 1, 2$. Moreover, the weighted version associated to w_1 and w_2 has the following joint PDF

$$f^w(x_1,x_2) = c\,\frac{w_1'(x_1)w_2'(x_2)}{w_1^{a_1}(x_1)w_2^{a_2}(x_2)}\exp\left(-\phi a_1 a_2\left(\log\frac{w_1(x_1)}{w_1(l)}\right)\left(\log\frac{w_2(x_2)}{w_2(l)}\right)\right)$$

which is also a model included in the type of the PDF given in (d) with parameters $a_1 - 1$ and $a_2 - 1$. Therefore, its hazard rate functions are

$$h_i^w(s|t) = (a_i-1)\left(1-\phi'(a_{3-i}-1)\log\frac{w_{3-i}(t)}{w_{3-i}(l)}\right)\frac{w_i'(s)}{w_i(s)}$$

for $i = 1, 2$. Hence (a) holds.

Acknowledgments

JN is partially supported by Ministerio de Ciencia y Tecnología and Fundación Séneca under grants MTM2009-08311 and 08627/PI/08. JN thanks the support of the Government of Kerala, India, under the ERUDITE scheme during his visit to Cochin University of Science and Technology.

References

Arnold, B.C., Castillo, E., Sarabia, J.M., 1999. Conditional Specification of Statistical Models. Springer Verlag, New York.
Arnold, B.C., Strauss, D., 1988. Bivariate distributions with conditionals in prescribed exponential families. J. R. Stat. Soc. Ser. B 53, 365–375.
Bartoszewicz, J., 2009. On a representation of weighted distributions. Statist. Probab. Lett. 79, 1690–1694.
Bartoszewicz, J., Skolimowska, M., 2006. Preservation of classes of life distributions and stochastic orders under weighting. Statist. Probab. Lett. 76, 587–596.
Blazej, P., 2008. Preservation of classes of life distributions under weighting with a general weight function. Statist. Probab. Lett. 78, 3056–3061.
Cox, D.R., 1959. The analysis of exponentially distributed lifetimes with two types of failure. J. R. Stat. Soc. Ser. B 21, 411–421.
Di Crescenzo, A., Longobardi, M., 2004. A measure of discrimination between past lifetime distributions. Statist. Probab. Lett. 67, 173–182.
Ebrahimi, N., Kirmani, S.N.U.A., 1996. A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions. Biometrika 83, 233–235.
Ebrahimi, N., Kirmani, S.N.U.A., Soofi, E.S., 2007. Multivariate dynamic information. J. Multivariate Anal. 98, 328–349.
Ebrahimi, N., Soofi, E.S., Soyer, R., 2010. Information measures in perspective. Int. Stat. Rev. 78, 383–412.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Statist. 22, 79–86.
Maya, S.S., Sunoj, S.M., 2008. Some generalized information measures in the context of weighted models. Statistica LXVIII (1), 71–84. anno.
Navarro, J., Ruiz, J.M., del Aguila, Y., 2006. Multivariate weighted distributions. Statistics 40, 51–54.
Navarro, J., Sarabia, J.M., 2010. Alternative definitions of bivariate equilibrium distributions. J. Statist. Plann. Inference 140, 2046–2056.
Navarro, J., Sarabia, J.M., 2011. Reliability properties of bivariate conditional proportional hazard rates models. J. Multivariate Anal. (in press). Published online first: doi:10.1016/j.jmva.2011.03.009.
Sankaran, P.G., Sreeja, V.N., 2007. Proportional hazard model for multivariate failure time data. Comm. Statist. Theory Methods 36, 1627–1641.
Sunoj, S.M., Linu, M.N., 2011. Dynamic cumulative residual Renyi's entropy. Statistics (in press). Published online first: doi:10.1080/02331888.2010.494730.