Aspects of Residual Information Measures for Weighted Distributions

G. R. Mohtashami Borzadaran- Samira Goodarzi
Department of Statistics
The University of Birjand
Birjand-IRAN
Email : gmb1334@yahoo.com

Abstract

The concepts of weighted distributions have been introduced by Rao (1965). A weighted function will be denoted by $w(x)$ and $g(x,\theta) = \frac{w(x)f(x,\theta)}{E(w(X))}$ where $E_\theta(w(X)) = \int_D w(x)dF(x)$, and $f(.,\theta)$ is the distribution of random variable X and g is the pdf of the weighted distribution.

Characterization results for the residual information measures are given here in view of the weighted distributions. We also derive relationship among residual information measures and reliability measures such as hazard rate.The residual divergence between two positive random variables are studied and finding link results relevant to information theory and reliability theory. Some examples that lead us to results related to information measures are derived for order statistics, record value, proportional hazard, proportional reversed hazard, Lorenz curve and hazard rate as special cases of weighted families.

Ebrahimi and Kirmani(1996) defined the uncertainty of residual lifetime distributions, then Asadi et. al. (2005) obtained some results related to minimum dynamic discrimination information and maximum dynamic entropy models.We obtain results concerning their relations with life distributions and information measures and give some examples for weighted families. Some inequalities, relations and partial ordering for weighted reliability measures are also presented.

Keywords: Information Measures, Residual Entropy, Weighted Family, Residual Kullback-Leibler Information, Residual Hellinger Distances, Residual Information Measures, Cumulative Residual Entropy, Characterization.

1 Introduction

After the creation of C. E. Shannon (1948), a number of research papers, and monographs discussing and extending Shannon's original work have appeared. Among them Dragomir (2003), Kagan, Linnik & Rao (1973), and Kullback (1959) are using and extending results due to information measures. Recently proposed a dynamic measure based on differential geometry applicable to residual life time. This measure has been used for the classification and ordering of survival function. Ebrahimi & Kirmani (1996), Nanda et al (2006) and Asadi et al (2005) gave an overview of some aspects of residual Renyi divergence and residual Kullback-Leibler information and residual entropy. Further implications and properties of the dynamic measures such as above and the uncertainty ordering, proportional hazard model through a measure of discrimination between two residual life distributions on the basis of the measures that are mentioned are obtained by at least one of the above references.

In this paper, characterization results for residual entropy, residual information measures are obtained. Link between these type of information measures, reliability measures and weighted families are derived. Examples for some special cases is another direction of this work.

2 Preliminaries

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and f be a measurable function from Ω to $[0, \infty)$, such that $\int_\Omega f d\mu = 1$. The Shannon entropy (or simply the entropy) of f relative to μ, is defined by

$$H(f, \mu) = -\int_\Omega f \ln f d\mu, \text{ (with } f \ln f = 0 \text{ if } f = 0), \tag{1}$$

and assumed to be defined for which $f \ln f$ is integrable. If X is an r.v. with pdf f, then we refer to H as the entropy of X and denotes also it by the notation H_X. In the case μ is a version of counting measure, (1) leads us to a specialized version that introduced by Shannon (1948) as $H_X = -\sum_{i=1}^n p_i \ln p_i$ where $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. One of the important issues in many applications of probability theory is finding an appropriate measure of distance between two probability distributions. A number of divergence measure for this purpose have been studied in a lot of references related to various type of information measures such as Dragomir (2003). These measures have applied in a variety of fields. Consider F and G be two distributions which are absolutely continuous w.r.t. measure μ and $\frac{dF}{d\mu} = f$ and $\frac{dG}{d\mu} = g$. We have the following definitions:

Kullback Leibler Information :

$$D_{KL}(F, G) = \int_\chi \ln \frac{f(x)}{g(x)} f(x) d\mu. \tag{2}$$

χ^2- **Divergence :**

$$D_{\chi^2}(F, G) = \int_\chi \frac{[f(x) - g(x)]^2}{f(x)} d\mu. \tag{3}$$

Bhattacharyya Distance and Hellinger Distance:

$$D_{Bh}(F, G) = \int_\chi \sqrt{g(x) f(x)} d\mu, D_H(F, G) = 2[1 - D_{Bh}(F, G)]. \tag{4}$$

$\alpha-$**Divergence :**

$$D_\alpha(F, G) = \frac{1}{1 - \alpha^2} \int_\chi \{1 - \frac{g^{\frac{1+\alpha}{2}}(x)}{f^{\frac{1+\alpha}{2}}(x)}\}^2 f(x) d\mu, \tag{5}$$

The following results are related to the above measures :

- It is easy to see that $D_H(F, G) \leq 2$. Via Taylor expansion and approximation, we can get, $D_{KL}(F, G) \approx \frac{1}{2} D_{\chi^2}(F, G)$, $D_J(F, G) \approx \frac{1}{2}[D_{\chi^2}(F, G) + D_{\chi^2}(G, F)]$, $D_{\chi^2}(F, G) \approx 4D_H(F, G)$ and $D_{\chi^2}(F, G) \geq D_H(F, G)$.
- Sometimes we are interested in the distances that is introduced in (2) to (5), between the distributions $F = F_{\theta_1}$ and $G = F_{\theta_2}$. Between the corresponding samples distributions which we denote $F_{\theta_1}^n$ and $F_{\theta_2}^n$, the distances are meaningful for arbitrary distributions and have no relation to the nature of spaces.
- The chi-squared divergence $D_{\chi^2}(F, G) = 2D_\alpha(F, G)$ on taking $\alpha = -3$ in (5). Also, the Hellinger distance $D_H(F, G) = \frac{1}{2} D_\alpha(F, G)$ on taking $\alpha = 0$ in (5). The Hellinger distance and Bhattacharyya distance are symmetric and has all properties of metric.

The relative information generating function of f given the reference measure g is defined as,

$$R(F, G, \gamma) = \int_\chi (\frac{f}{g}(x))^{\gamma - 1} f(x) dx, \tag{6}$$

where $\gamma \geq 1$ and the integral is convergent on noting that $R(F, G, 1) = 1$. In particular, $R'(F, G, 1)$ is just Kullback Leibler information, and $-R'(F, 1, 1)$ and $R(F, 1, 2)$ are Shannon entropy and second order entropy respectively.

The power divergence measure (PWD) which gathers most of the interesting specification is indexed by

$$PWD(F, G, \lambda) = \frac{1}{\lambda(\lambda + 1)} \int_{\chi} \{[\frac{f(x)}{g(x)}]^{\lambda} - 1\} f(x) d\mu, \tag{7}$$

The power divergence family implies different well-known divergence measures for different values of λ. PWD for $\lambda = -2, -1, -.5, 0, 1$, implies Neyman Chi-square, Kullback Leibler, squared Hellinger distance, Likelihood disparity and Pearson Chi-square divergence respectively. Note that $PWD(F, G, \lambda) = \frac{1}{\lambda(\lambda+1)}[R(F, G, \lambda + 1) - 1]$.

3 Information Measures in view of the Weighted Families

On considering weighted function $w(x, \beta)$ which is a non-negative function with parameter β represent a family of distributions with pdf, $g(x, \beta, \theta) = \frac{w(x,\beta)f(x,\theta)}{E[w(X,\beta)]}$, which is called a weighted version of distribution. $w(x, \beta) = x$ is called sized-biased distribution. Order statistics, record value, residual lifetime of a stationary renewal process, selection samples, hazard rate, reversed hazard rate, proportional hazard model, reversed proportional hazard model, Lorenz curve and probability weighted moments are some special cases of weighted families. Among them, we will concentrate on order statistics, record values and some special cases of probability weighted moments in some information measures.

We can consider $f_1(x) = \frac{w_1(x)f(x)}{Ew_1(X)}$ and $g_1(x) = \frac{w_2(x)f(x)}{Ew_2(X)}$ as the weighted distributions of f (see Rao 1965). Note that all the expectations are w.r.t pdf f. Then, for the above measures the following results are important and noticeable :

For the order statistics distributions, we consider $w_1(x) = \frac{1}{\beta(i,n-i+1)}[F(x)]^{i-1}[1-F(x)]^{n-i}$ and $w_2(x) = \frac{1}{\beta(j,n-j+1)}[F(x)]^{j-1}[1 - F(x)]^{n-j}$, hence

- D_{KL} is increasing function of n, for $i = j + 1$, it is increasing for $j > \frac{n}{2}$ and decreasing for $j < \frac{n}{2}$ when $w_1(x) = 1$. D_{χ^2} for $w_1(x) = 1$ and ($j = 1$ and $j = n$) is increasing function of n.

- D_{Bh} is increasing function of n, In the case that $i = j + 1$, $j = 1$ and $j = n - 1$ lead us to an decreasing and increasing function of n respectively. D_H is decreasing function of n, In the case that $i = j + 1$, $j = 1$ and $j = n - 1$ lead us to an increasing and decreasing function of n respectively.

- For the order statistics distributions that mentioned, we have the relative information generating function when $i = 1$, is increasing function of $\gamma \geq 1$, decreasing function of n.

- It is clear that $i = j$ is equivalent to $w_1(x) = w_2(x)$ for $\forall x$, hence,$R(F_1, G_1, w, w, f, \gamma) = \frac{n-i}{\Gamma(n+1)}$ which is not depending on γ, decreasing function of n and i.

- For $w_1(x) = \frac{1}{\beta(i,n-i+1)}[F(x)]^{i-1}[1 - F(x)]^{n-i}$ and $w_2(x) = 1$, ($i = 1$ or $i = n$) lead us to $R(F_1, G_1, w_1, 1, f, \gamma)$, is increasing function of n and γ.

A class of moments, called probability weighted moments (PWM) as $M_{l,j,k} = E[X^l(F(X)^j(1 - F(X))^k]$. We can consider $\frac{w_1(x)}{Ew_1(X)} = \frac{X^l(F(X)^j(1-F(X))^k}{M_{l,j,k}}$ and $\frac{w_2(x)}{Ew_2(X)} = \frac{X^{l'}(F(X)^{j'}(1-F(X))^{k'}}{M_{l',j',k'}}$. In here we will concentrate on the case that $l = l' = 0$.

- For $j = j' = 0, l' = l + 1$ and $j = j' = 1, l' = l + 1$, D_{KL} increasing function w.r.t. ($l < -2$ or $l > -1$) and ($l < -3$ or $l > -1$) respectively. The same result can be find for $l = l' = 0, j' = j + 1$ and $l = l' = 1, j' = j + 1$. Also, ($j' = l' = 0, j = l + 1$) and ($l = n - j, l' = n - j'$) are other special cases that is related to order statistics. ($j = j' = 0, l' = l + 1$), implies that D_{χ^2} decreasing function w.r.t. l.

- For $j = j' = 0, l' = l + 1$ and $j = j' = 1, l' = l + 1$, D_{Bh} increasing function w.r.t. l. D_H decreasing function w.r.t. l.
- $j = j'$ implies that $D_\alpha(F_1, G_1)$ has a simple form that you can find the behaviour of it w.r.t. j, l and l'.
- The proportional hazard model is expressed by the following relations between the survival function of the random lifetimes, $\overline{F}_1(x) = (\overline{F}(x))^\alpha, \alpha > 0, \overline{F}_2(x) = (\overline{F}(x))^\beta, \beta > 0$. Hence, relative information generating function is increasing function of α and decreasing function of β when $\alpha < \beta$. The proportional reversed hazard model is expressed by the similar relations between the survival function of the random lifetimes, $F_1(x) = (F(x))^\alpha, \alpha > 0, F_2(x) = (F(x))^\beta, \beta > 0$ hence, relative information generating function for $\alpha > \beta$ is increasing function of α, decreasing function of β, and for $\beta = 1$, is increasing function of α when $\alpha > 1$.

Let $\{X_i, i \geq 1\}$ be a sequence of independent identically distributed random variables having cdf F and pdf f. An observation X_j will be called an upper record value with pdf $g(u) = \frac{[s(u)]^{n-1}}{\Gamma(n)} f(u)$, where $s(u) = -\ln \overline{F}(u)$, $\overline{F}(u) = 1 - F(u)$, and $U = X_{T_n}$ such that $T_n = min\{j, j > T_{n-1}, X_j = X_{T_{n-1}}\}$. Let $f_1(x) = \frac{[s(x)]^{i-1}}{\Gamma(i)} f(x)$ and $f_2(x) = \frac{[s(x)]^{j-1}}{\Gamma(j)} f(x)$ be the pdf of the two upper record distributions, then:

- D_{KL} for $i = j + 1$, is equal to $-\ln j + \psi(j + 1)$. Also, $D_{KL} = -\ln \Gamma(i) + (i - 1)\psi(i)$ when $w_2(x) = 1$. D_{χ^2} for $w_1(x) = 1$ is equal $\frac{\Gamma(2j-1)}{[\Gamma(j)]^2} - 1$.
- D_{Bh} is increasing function of j, when $i = j + 1$ and increasing function of n. $i = j + 1$, $j = 1$ and $j = n - 1$ lead us to an decreasing and increasing function of n respectively. For $w_1(x) = 1$ we can find again simpler statements than the general case. D_H is decreasing function of j, when $i = j + 1$.
- For lower record value we have pdf the same as upper record value with $s(u) = -\ln F(u)$ and for above measures, we have the same as achieved via upper record value.
- $i - j = 1$, implies that $R(F_1, G_1, w_1, 1, f, \gamma)$ is decreasing function w.r.t. γ and j.
- For PWD we have the results via relation between relative information generating function and PWD.

4 Residual Entropy in view of the Weighted Families

If a unit is known to have survived up to an age t, Ebrahimi (1996) defined residual entropy of the nonnegative continuous random variable X as

$$H(F, t) = -\int_t^\infty \ln[\frac{f(x)}{\overline{F}(t)}] \frac{f(x)}{\overline{F}(t)} dx, \tag{8}$$

where $\overline{F}(t)$ is the survival function of X. If we put $t = 0$, then we get $H(X, 0)$ is the Shannon entropy. Nanda et al (2006) defined

$$H_1^\beta(F, t) = \frac{1}{\beta - 1}[1 - \int_t^\infty [\frac{f(x)}{\overline{F}(t)}]^\beta]dx, \tag{9}$$

and

$$H_2^\beta(F, t) = \frac{1}{1 - \beta} \ln[\int_t^\infty (\frac{f(x)}{\overline{F}(t)})^\beta dx], \tag{10}$$

where $H_1^\beta(F, t)$ and $H_2^\beta(F, t)$ are first kind residual entropy of order β and second kind residual entropy of order β of the random variable X respectively. It can be noted that as $\beta \longrightarrow 1$, then

(9) and (10) reduce to resodual entropy that defined in (8). $H_1^\beta(F,t)$ and $H_2^\beta(F,t)$ can always be made non-negative by choosing appropriate β. In the following results of this note, we will consider $g(x)=\frac{w(x)f(x)}{Ew(X)}$ as a weighted version of f and $A(t)=\frac{\int_t^\infty w(x)f(x)dx}{\overline{F}(t)}$.

- Let X and Y be two nonnegative random variables having densities f and g and distribution functions F and G and survival functions $\overline{F}$ and $\overline{G}$ respectively as defined in previous, then, X is said to have less uncertainty than Y if $H(F,t)\leq H(G,t)$ for all $t\geq 0$. We write $X\leq^{LU} Y$.
  X is said to be less than Yin (first kind residual entropy of order β (written $X\leq^{\beta(1)} Y$) if $H_1^\beta(F,t)\leq H_1^\beta(G,t)$ for all $t>0$. X is said to be less than Yin (first kind residual entropy of order β (written $X\leq^{\beta(2)} Y$) if $H_2^\beta(F,t)\leq H_2^\beta(G,t)$ for all $t>0$.
  Let $w(x)\leq A(t)$ for $\forall x>t$, then $H_i^\beta(F,t)\leq H_i^\beta(G,t)$ for $i=1,2$ that g is a weighted version of distribution F.
- X is said to be larger than Y in likelihood ratio ordering ($X\geq^{LR} Y$) if $\frac{f(x)}{g(x)}$ is a non-decreasing function of $x\geq 0$. $w(x)$ non-increasing in x implies that $X\geq^{LR} Y$.
- Let $X\leq^{LR} Y$ and $\lambda_F(x)$ or $\lambda_G(x)$ be non-decreasing in x. Then it follows that $X\leq^{LU} Y$. So, let $w(x)$ non-decreasing and $\lambda_F(x)$ or $\lambda_G(x)$ be non-decreasing in x, then, $X\leq^{LU} Y$. Also, Let $\frac{A(x)}{w(x)}$, $-\lambda_F(x)$ and $A(x)$ be non-decreasing function of x, $X\leq^{LU} Y$.
- If $\lambda_F(x)\geq E(w(X))[\ln\lambda_F(x)+\ln\frac{w(x)}{A(t)}$ for $\forall t\geq 0, \forall x>t$ and $\lambda_F(x)$ be non-decreasing in x, then, $X\leq^{LU} Y$.
- A natural question whether residual entropy like mean residual life and hazard rate characterizes survival function or distribution function. Ebrahimi (1996) proved that $H(X,t)<\infty, t\geq 0$, uniquely determine the distribution function. $H_i^\beta(X,t)$ is increasing in t, then, $H_i^\beta(X,t)$ uniquely determines $\overline{F}(t)$, for $i=1,2$. (Nanda et al 2006).
- A non-negative random variable is said to have decreasing (increasing) uncertainty in residual life (DURL(IURL)) if $H(X,t)$ is decreasing (increasing). A non-negative random variable is said to have DURL(IURL) of first kind of order β (DURLF(β)) if $H_1^\beta(X,t)$ is decreasing in $t\geq 0$. A non-negative random variable is said to have DURL(IURL) of second kind of order β (DURLS(β)) if $H_2^\beta(X,t)$ is decreasing in $t\geq 0$. In the above definition if we replace the word "decreasing" by "increasing", then we call them IURLF(β) and IURLS(β) respectively.
- The first system is very strongly better than the second system if $X\leq^{LU} Y$ and $X\geq^{LR} Y$. So, let $w(x)$ and $\frac{w(x)}{A(x)}$ be non-decreasing functions of x,then, $H(F,t)-H(G,t)$ is increasing in t.
- X is said to be stochastically than Y ($X\leq^{ST} Y$) if $\overline{F}(x)\leq\overline{G}(x)$ for all $x\geq 0$. Hence, for weighted case, if $A(x)\geq E(w(X))$ for $\forall x$, then, $X\leq^{ST} Y$.

5 Residual Information Measures for the Weighted Families

In view of Ebrahimi (1996), we defined the above measures for the case that after the unit has survived for time t. So, Assume that the set $[t,\infty)$ be the suitable support of distributions and F and G be two distributions which are absolutely continuous w.r.t. measure μ and $\frac{dF}{d\mu}=f$ and $\frac{dG}{d\mu}=g$. We have the following definitions:

Residual Kullback Leibler Information :

$$D_{KL}(F,G,t)=\int_t^\infty \ln\frac{\frac{f(x)}{\overline{F}(t)}}{\frac{g(x)}{\overline{G}(t)}}\frac{f(x)}{\overline{F}(t)}d\mu, \tag{11}$$

Residual χ^2- Divergence :

$$D_{\chi^2}(F,G,t)=\int_t^\infty \frac{[\frac{f(x)}{\overline{F}(t)}-\frac{g(x)}{\overline{G}(t)}]^2}{\frac{f(x)}{\overline{F}(t)}}d\mu, \quad (12)$$

Residual Bhattacharyya Distance and Residual Hellinger Distance:

$$D_{Bh}(F,G,t)=\int_t^\infty \sqrt{\frac{g(x)}{\overline{G}(t)}\frac{f(x)}{\overline{F}(t)}}d\mu, D_H(F,G,t)=2[1-D_{Bh}(F,G,t)]. \quad (13)$$

Residual $\alpha-$Divergence :

$$D_\alpha(F,G,t)=\frac{1}{1-\alpha^2}\int_t^\infty\{1-\frac{[\frac{g(x)}{\overline{G}(t)}]^{\frac{1+\alpha}{2}}}{[\frac{f(x)}{\overline{F}(t)}]^{\frac{1+\alpha}{2}}}\}^2\frac{f(x)}{\overline{F}(t)}d\mu, \quad (14)$$

The following results are related to the above measures :

- It is easy to see that $D_H(F,G,t) =\leq 2$. Via Taylor expansion and approximation, we can get, $D_{KL}(F,G,t)\approx\frac{1}{2}D_{\chi^2}(F,G,t)$, $D_J(F,G,t)\approx\frac{1}{2}[D_{\chi^2}(F,G,t)+D_{\chi^2}(G,F,t)]$, $D_{\chi^2}(F,G,t)\approx 4D_H(F,G,t)$.
- Sometimes we are interested in the residual distances that is introduced in (11) to (14), between the distributions $F=F_{\theta_1}$ and $G=F_{\theta_2}$. Between the corresponding samples distributions which we denote $F^n_{\theta_1}$ and $F^n_{\theta_2}$, the distances are meaningful for arbitrary distributions and have no relation to the nature of spaces. Hence, results can be applicable similar the case that $t=0$ for any t, but not easier than the case $t=0$.
- The residual chi-squared divergence $D_{\chi^2}(F,G,t)=2D_\alpha(F,G,t)$ on taking $\alpha=-3$ in (5). Also, the residual Hellinger distance $D_H(F,G,t)=\frac{1}{2}D_\alpha(F,G,t)$ on taking $\alpha=0$ in (5).

The relative information generating function of f given the reference measure g is defined as,

$$R(F,G,\gamma,t)=\int_t^\infty(\frac{\frac{f(x)}{\overline{F}(t)}}{\frac{g(x)}{\overline{G}(t)}})^{\gamma-1}\frac{f(x)}{\overline{F}(t)}dx, \quad (15)$$

where $\gamma\geq 1$ and the integral is convergent on noting that $R(F,G,1,t)=1$. In particular, $R'(F,G,1,t)$ which is just residual Kullback Leibler information and $R'(F,G,1,0)+R'(G,F,1,0)$ is residual $J-$divergence between F and G. $-R'(F,1,1,t)$ and $R(F,1,2,t)$ are residual Shannon (1948) entropy and residual second order entropy respectively.

The residual power divergence measure (PWD) is indexed by

$$PWD(F,G,t)=\frac{1}{\lambda(\lambda+1)}\int_t^\infty\{[\frac{\frac{f(x)}{\overline{F}(t)}}{\frac{g(x)}{\overline{G}(t)}}]^\lambda-1\}f(x)d\mu, \quad (16)$$

The power divergence family implies different well-known divergence measures for different values of λ. PWD for $\lambda=-2,-1,-.5,0,1$, implies residual Neyman Chi-square, residual Kullback Leibler, residual squared Hellinger distance, residual Likelihood disparity and residual Pearson Chi-square divergence respectively. Note that $PWD(F,G,t)=\frac{1}{\lambda(\lambda+1)}[R(F,G,\lambda+1,t)-1]$.

- Let ϕ be an invertible increasing function, then, $D_\alpha(F_1,G_1,\phi^{-1}(t))=D_\alpha(\phi(F_1),\phi(G_1),t)$. Because,

$$\begin{aligned} D_\alpha(\phi(F),\phi(G),t) &= \frac{1}{1-\alpha^2}\int_{\phi^{-1}(t)}^\infty\{1-\frac{[\frac{g(y)}{\overline{G}(\phi^{-1}(t))}]^{\frac{1+\alpha}{2}}}{[\frac{f(y)}{\overline{F}(\phi^{-1}(t))}]^{\frac{1+\alpha}{2}}}\}^2\frac{f(y)}{\overline{F}(\phi^{-1}(t))}dy \\ &= D_\alpha(F,G,\phi^{-1}(t)). \quad (17)\end{aligned}$$

It is clear that for residual Kullback Leibler information (Ebrahimi et al 1996), residual χ^2 divergence, residual Bhattachryya distance, residual Hellinger distance as special cases of (17) the result hold.

- $D_\alpha(F_1, G_1, t)$ is independent of t if and only if F_1 and G_1 have proportional hazard rate. On noting that the only if is easy but for if case, assume that $D_\alpha(F_1, G_1, t) = b$ that b is constant. It is clear that for residual Kullback Leibler information (Ebrahimi et al 1996b), residual χ^2 divergence, residual Bhttacharyya distance, residual Hellinger distance as special cases of the above result. Note that results achieved via the technique that applied in the Asadi et al (2005).

- Also, the residual relative information generating function and residual power divergence measure, is independent of t if and only if F_1 and G_1 have proportional hazard rate.

- Suppose $\frac{A(x)}{w(x)}$ be increasing in x and both F_1 and G_1 are new better than used (F is said to be new better than used if $\overline{F}(x+y) \leq \overline{F}(x)\overline{F}(y), \forall x, y$, and F is said to be worse than used if $\overline{F}(x+y) \geq \overline{F}(x)\overline{F}(y), \forall x, y,)$. Then $D_{KL}(F_1, G_1, t) \geq D_{KL}(F_1, G_1, 0$. When both F_1 and G_1 are worse than used, then $D_{KL}(F_1, G_1, t) \leq D_{KL}(F_1, G_1, 0$.

We can consider $f_1(x) = \frac{w_1(x)f(x)}{Ew_1(X)}$ and $g_1(x) = \frac{w_2(x)f(x)}{Ew_2(X)}$ as the weighted distributions of f. Then, the above measures are expressed as :

Residual Kullback Leibler Information :

$$D_{KL}(F_1, G_1, t) = \frac{1}{A_1(t)\overline{F}(t)} \int_t^\infty \ln[\frac{w_1(x)}{w_2(x)}] w_1(x) f(x) dx + \ln[\frac{A_1(t)}{A_2(t)}], \tag{18}$$

where $A_i(t)\overline{F}(t) = \int_t^\infty w_i(x) f(x) dx$ for $i = 1, 2$.

Residual χ^2- Divergence :

$$D_{\chi^2}(F_1, G_1, t) = \frac{1}{A_2(t)\overline{F}(t)} \int_t^\infty [\frac{[w_2(x)]^2}{w_1(x)} \frac{A_1(t)}{A_2(t)}] f(x) dx. \tag{19}$$

Residual Bhattacharyya Distance and Residual Hellinger Distance:

$$D_{Bh}(F_1, G_1, t) = \frac{1}{\overline{F}(t)} \int_t^\infty \sqrt{\frac{w_1(x)w_2(x)}{A_1(t)A_2(t)}} f(x) dx, D_H(F_1, G_1, t) = 2 - 2D_{Bh}(F_1, G_1, t). \tag{20}$$

Residual $\alpha-$Divergence :

$$D_\alpha(F_1, G_1, t) = \frac{1}{(1-\alpha^2)A_1(t)\overline{F}(t)} \int_t^\infty \{1 - [\frac{w_2(x)}{w_1(x)} \frac{A_1(t)}{A_2(t)}]^{(\frac{1+\alpha}{2})}\}^2 w_1(x) f(x) dx. \tag{21}$$

- For the case that $w_1(x) = 1$, $A_1(t) = 1$, statements in (18) to (21) change to simple statements that their calculation is easier than the previous statements.

- For the above residual information measures, for weights like, order statistics, record value, proportional hazard rate, reversed proportional hazard rate, hazard rate, selection samples,.. we can find the values of these residual measures and some properties of them in special cases. Some of them lead us to calculating the integrals via incomplete gamma and incomplete beta functions.

6 Conclusions

In many reliability and survival analysis problem the current age of the item under study must be taken into account by information measures of the lifetime distribution. In this paper, we concentrate on information measures in view of weighted distribution and obtained some statements and characterization results due to them. Also, measures such as residual Kullback-Leibler information, residual Hellinger distance, residual χ^2 distance, residual Bhattacharyya distance and residual Renyi $\alpha-$ information measures are defined and discussed some properties of them specially related to weighted cases. Among the weighted cases, for these idea we find some examples. Some properties of these examples due to residual measures discussed at the end of this note.

7 Further Works

Further works related to this research are as follows:

- Applying the weighted version of exponential and natural exponential family in view of the residual measures that is defined in this paper.
- More details and results of the links with entropy and Fisher information in terms of hazard rate, relationship between residual entropy function in view of Ebrahimi & Kirmani (1996) and Asadi et al (2005) results for some weighted families.
- Finding the same results in a multivariate set-up in view of weighted families that may be nice.
- We can apply similar results due to the above discrimination measures between two past lives.
- The asymptotic behaviour of Kullback-Leibler information, the Hellinger distance and the Chi-square are identical when the ratio of the density function is near one. These three distances are used extensively in parametric families of distributions to quantify the distance between measures from the same family indexed by different parameters. Borovkov showed how these distances are related to the Fisher information in the limit as the difference in the parameters go to zero. Is there a similar opinion related to residual Kullback-Leibler information, residual Hellinger distance and the residual Chi-square measures ?

References

[1] Asadi, M., Ebrahimi, N. & Soofi, E. (2005). Dynamic generalized information measures. *Stat. Prob Letters,* **71,** 85-98.

[2] Dragomir, S. S. (2003). On the $p-$ logarithmic and $\alpha-$ power divergence measures in information theory. *arXiv:math.PR/0304240 v1.*

[3] Ebrahimi, N & Kirmani S. N. U. A. (1996). A measure of discrimination between two residual life-time distributions and its applications. *Ann. Inst. Statist. Math.*, **Vol. 48, No. 2**, 257-265.

[4] Kagan, A. M., Linnik, Yu. V. & Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

[5] Kullback, S. (1959). *Information and Statistics.* Wiley, New York.

[6] Nanda, Asok. K. & Paul, P. (2006). Some results on generalized residual entropy. *Information Sciences*, **176,** 24-47.

[7] Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment, in *Classical and Contagious Discrete Distributions,* G. Patil Ed., Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

[8] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.