Probability in the Engineering and Informational Sciences, 7, 1993, 413–420. Printed in the U.S.A.

A NOTE ON THE RESIDUAL ENTROPY FUNCTION

PIETRO MULIERE
Dipartimento di Economia Politica e Metodi Quantitativi
Università di Pavia, via San Felice 5
Pavia, Italy, 27100

GIOVANNI PARMIGIANI
Institute of Statistics and Decision Sciences
Duke University
Durham, North Carolina 27708-0251

NICHOLAS G. POLSON
Graduate School of Business
University of Chicago, 1101 East 58th Street
Chicago, Illinois 60637

Interest in the informational content of truncation motivates the study of the residual entropy function, that is, the entropy of a right truncated random variable as a function of the truncation point. In this note we show that, under mild regularity conditions, the residual entropy function characterizes the probability distribution. We also derive relationships among residual entropy, monotonicity of the failure rate, and stochastic dominance. Information theoretic measures of distances between distributions are also revisited from a similar perspective. In particular, we study the residual divergence between two positive random variables and investigate some of its monotonicity properties. The results are relevant to information theory, reliability theory, search problems, and experimental design.

1. INTRODUCTION AND BASIC DEFINITIONS

This paper focuses on the informational content of truncation. Consider a positive random variable X with cumulative distribution function $F(x)$ and den-

 0269-9648/93 \$5.00 + .00

sity $f(x)$. The survival function is then $\bar{F}(x) = 1 - F(x)$, the failure rate is $r(x) = f(x)/\bar{F}(x)$, and the entropy functional is

$$H(X) \equiv -\int_0^\infty \log f(x)\, dF(x) \tag{1}$$

(see Cover and Thomas [2] and Verdugo Lazo and Rathie [6]).

In many practical problems information about a random variable X takes the form $X > t$. We are interested in studying the entropy of the random variable in question as a function of t when the new information is available. Conditional on $X > t$, the survival function is $\bar{F}_t(x) = \bar{F}(x)/\bar{F}(t)$, $x > t$; consequently, the entropy of X conditional on the new information is

$$h(t) \equiv H(X|X > t) = -\int_t^\infty \log\left[\frac{f(x)}{\bar{F}(t)}\right] dF_t(x). \tag{2}$$

This quantity, regarded as a function of t, is the main object of study of this paper. We refer to it as the residual entropy function.

There is an important relationship between the residual entropy function and the failure rate, given by

$$h(t) = 1 - \int_t^\infty \log r(x)\, dF_t(x). \tag{3}$$

For $t = 0$ this was shown by Teitler, Rajagopal, and Ngai [5]. The extension to arbitrary t follows by a simple integration-by-parts argument.

Let us illustrate h with two simple examples.

Example 1: Take X to be uniform in the unit interval. Then $f_t(x) = 1/(1 - t)$, and, from the definition,

$$h(t) = \log(1 - t). \tag{4}$$

As intuitive, the residual entropy decreases upon conditioning on $X > t$.

Example 2: Take X to be a Pareto random variable; that is, let

$$\bar{F}(x) = \left(\frac{b}{b + x}\right)^a, \tag{5}$$

with a and b positive. This distribution arises in applications as the predictive distribution of the waiting time of a Poisson process when the rate at which event occurs is an unknown λ. It is additionally assumed that λ has a gamma distribution, with scale parameter b and shape parameter a. For a Pareto, $r(x) = a/(b + x)$. Therefore,

$$h(t) = 1 - \int_t^\infty \log\left(\frac{a}{b + x}\right) \frac{a(b + t)^a}{(b + x)^{a+1}}\, dx = 1 - \frac{1}{a} - \log\left(\frac{a}{b + t}\right). \tag{6}$$

Interestingly, in this case the residual entropy function is a linear function of the logarithm of the failure rate. The residual entropy is increasing because conditioning on $X > t$ introduces new information that increases the probability mass assigned to lower values of the unknown rate λ.

It is natural to compare the residual entropy function with the residual expected lifetime $m(t) = E\{X|X > t\}$, obtained by applying right truncation to the mean functional. The residual expected lifetime is related to the failure rate by

$$m(t) = \int_t^\infty \frac{1}{r(x)}\, dF_t(x). \tag{7}$$

A simple inequality relating $h(t)$ and $m(t)$ is

$$h(t) = 1 + \int_t^\infty \log\left(\frac{1}{r(x)}\right) dF_t(x) \le 1 + \log m(t) \tag{8}$$

by Jensen's inequality.

Right truncation is a special case of conditioning on the outcome of $E = I_{X>t}$. In analogy to Eq. (2), one can then define the following:

$$g(t) \equiv H(X|X \le t) = -\int_0^t \log\left[\frac{f(x)}{F(t)}\right] \frac{f(x)}{F(t)}\, dx. \tag{9}$$

There is functional relation between h and g. It can be derived as follows. There are two ways of decomposing the entropy of the bivariate random variable (X,E). First, by conditioning on E we have

$$H(X,E) = H(E) + F(t)g(t) + \bar{F}(t)h(t), \tag{10}$$

where $H(E) = -F(t)\log F(t) - \bar{F}(t)\log \bar{F}(t)$. Second, by conditioning on X and noting that $H(E|X) = 0$, we have $H(X,E) = H(X)$. Now, in terms of the residual entropy function, $H(X) = h(0)$. Therefore, $h(0) = H(E) + F(t)g(t) + \bar{F}(t)h(t)$, or

$$h(0) - h(t) = H(E) + F(t)[g(t) - h(t)]. \tag{11}$$

Example 3: In search problems it is often necessary to find a partition point t for which both outcomes of E are equally informative. This is obtained by solving $g(t) = h(t)$. With an increasing $h(t)$ no solution can exist, as $H(E) > 0$. With a decreasing $h(t)$ the solution is $t = h^{-1}[h(0) - H(E)]$. In the uniform case this gives the well-known solution $t = \frac{1}{2}$.

The remainder of the paper is organized as follows. In Section 3 we show that under smoothness conditions the residual entropy function characterizes uniquely the distribution of the random variable. We then study the relationship between monotonicity of the failure rate and monotonicity of the residual entropy function. In Section 4 we consider the stochastic ordering of positive random variables induced by residual entropy. Relationships with existing orderings are investigated. In Section 5 we extend consideration to the Kullback-Leibler divergence between two independent random variables and consider the divergence conditional on right truncation of both variables at the same point. We study the relationship between monotonicity of the residual divergence and monotonicity of the ratio of the failure rates of the two random variables.

2. CHARACTERIZATION AND MONOTONICITY PROPERTIES

A basic question of interest is whether or not the residual entropy function characterizes the probability distribution of a random variable. Kotlarski [4] showed that $E\{\phi(X)|X > t\}$ characterizes the cumulative distribution whenever ϕ is differentiable and monotone. From Eq. (3), Kotlarski's result implies that the residual entropy characterizes the distribution whenever the failure rate is monotone.

Here we present a more general result, assuming only that the failure rate is differentiable. We begin by writing the derivatives of the residual entropy function in a convenient recursive form. For ease of reference we summarized the first and second derivatives in the following lemma. The proof, which is omitted, involves Eq. (3) and straightforward differentiation and rearranging.

LEMMA 1:

$$h'(t) = r(t)[h(t) - 1 + \log r(t)] \quad (12)$$

$$h''(t) = r'(t)[h(t) + \log r(t)] + h'(t)r(t) \quad (13)$$

THEOREM 1: *Let X and Y be two positive* r.v.'s *with* c.d.f.'s *F and G, nonzero failure rates r_F and r_G, and residual entropy functions h_F and h_G, respectively. Then*

$$h_F(t) = h_G(t)\ \forall t > 0 \quad \Leftrightarrow \quad F(t) = G(t)\ \forall t > 0. \quad (14)$$

PROOF:

($\Leftarrow$) For all $t > 0$, $F(t) = G(t)$, implies $r_F(t) = r_G(t)$. From Eq. (3), $h_F(t) = h_G(t)$.

($\Rightarrow$) As X and Y have densities, and r_F and r_G are differentiable, from Eq. (3), h_F and h_G are twice differentiable. Therefore, $h_F(t) = h_G(t)$ implies $h'_F(t) = h'_G(t)$ and $h''_F(t) = h''_G(t)$ for all $t > 0$. Consider a fixed t_0. Let $h = h_F(t_0) = h_G(t_0)$, $u = r_F(t_0)$, and $v = r_G(t_0)$. Using Eq. (12), $h'_F(t) = h'_G(t)$ becomes

$$u[h - 1 + \log u] - v[h - 1 + \log v] = 0. \quad (15)$$

As t_0 is arbitrary, it only remains to show that, for fixed h, the only proper solution to Eq. (15) is $u = v$. Differentiating both sides of Eq. (15) with respect to u, with h fixed,

$$[h + \log u] - [h + \log v]\,\frac{dv}{du} = 0. \tag{16}$$

Now, let $u' = dr_F/dt$ and $v' = dr_G/dt$. Then $dr_G/dr_F = v'/u'$. From Eq. (16),

$$[h + \log u] = [h + \log v]\,\frac{v'}{u'}. \tag{17}$$

On the other hand, from $h_F''(t) = h_G''(t)$ and Eq. (13),

$$v'[h + \log v] + vh' = u'[h + \log u] + uh'. \tag{18}$$

If $h' \neq 0$, by substituting in Eq. (17) we have $u = v$. If $h' = 0$, because the failure rates are nonzero,

$$h - 1 + \log u = h - 1 + \log v, \tag{19}$$

so that again $u = v$. In turn, this is sufficient for $F = G$ (see, e.g., Barlow and Proschan [1]. ■

In Examples 1 and 2 a monotone failure rate resulted in a monotone residual entropy function. In the remainder of this section, we explore the general implications between monotonicity of failure rate and residual entropy function. In particular, we show that monotonicity of the failure rate implies monotonicity of the residual entropy function. Further conditions are necessary for the converse to be true. Let h be twice differentiable, and define

$$h^*(t) = 1 - \int_t^\infty \log\left[\frac{r(x)}{r(t)}\right] dF_t(x). \tag{20}$$

THEOREM 2:

(a) *If $r(t)$ is increasing (decreasing) $h(t)$ is decreasing (increasing).*

(b) *If $h(t)$ is increasing and concave, $r(t)$ is decreasing. If $h(t)$ is decreasing and convex, then $r(t)$ is increasing (decreasing) if and only if $h^*(t) > 0$.*

PROOF:

(a) If $r(t)$ is increasing (decreasing), then, from Eq. (3), $\log r(t) < (>)\ 1 - h(t)$. Thus, the conclusion follows directly from Eq. (12).

(b) Using Eq. (13) and rearranging, we obtain

$$h'' = r'(t)h^*(x) + r(t)h'(t). \tag{21}$$

As $h^*(t) = 1 + h'(t)/r(t)$, $h'(t) > 0$ implies $h^*(t) > 0$, so that $h''(t) < 0$ implies $r'(t) < 0$. Now let $h'(t) < 0$ and $h''(t) > 0$. Then it must be $r'(t)h^*(x) > 0$, whence the result. ■

4. STOCHASTIC ORDERING BY RESIDUAL ENTROPY

The residual entropy function can be used to define an ordering among distributions. In particular, if X and Y have distributions F and G, $X \geq_h Y$ whenever $h_X(t) \geq h_Y(t)$.

Many notions of dominance between random variables have been studied in the literature (see Gaede [3] for a survey). It is natural to compare entropy dominance to failure rate dominance ($X \geq_{FR} Y$ whenever $r_X(t) \geq r_Y(t)$) and mean residual life dominance ($X \geq_M Y$ whenever $m_X(t) \leq m_Y(t)$). Exponential random variables are ranked the same way by all three orderings.

As is well known, failure rate dominance implies mean residual life dominance. In this section we discuss an example showing that the entropy ordering is neither implied by nor implies the failure rate ordering. The same holds for all orderings implied by the failure rate ordering, including the residual life ordering.

Example 4: For independence of residual entropy and failure rate ordering, consider a random variable with piecewise constant failure rate of the form $r(t) = \lambda_1$, $0 < t \leq a$, and $r(t) = \lambda_2$, $t > a$. Then the residual entropy is

$$h(t) = \begin{cases} 1 - \log \lambda_1 - e^{-\lambda_1(a-t)} \log \dfrac{\lambda_2}{\lambda_1}, & 0 < t \leq a \\ 1 - \log \lambda_2, & t > a. \end{cases} \tag{22}$$

Take two random variables with failure rate of the preceding form. In particular, let $r_F(t) = 1$, $0 < t \leq 1$, and $r_F(t) = e^4$, $t > 1$, and let $r_G(t) = e$, $0 < t \leq 1$, and $r_G(t) = e^4$, $t > 1$. Clearly $X \geq_{FR} Y$. However, for $t \leq 1$, $h_F(t) = 1 - 4e^{-(1-t)}$ and $h_G(t) = (1-4)e^{-(1-t)e}$, so that $h_F(t) < h_G(t)$ and $X \not\geq_h Y$. Moreover, we have $X \leq_h Y$, without having $X \leq_{FR} Y$.

The ordering reversal displayed by this example shows that the entropy ordering is not implied by nor implies the failure rate ordering.

5. RESIDUAL DIVERGENCE

Let X and Y be two positive random variables with c.d.f.'s F and G, respectively. Assume X and Y are independent and mutually absolutely continuous. We define residual Kullback-Leibler divergence between F and G at t the Kullback-Leibler divergence between F and G conditional on $X > t$ and $Y > t$. Formally,

$$d(t) = \int_t^\infty \log\left(\frac{g(x)\bar{F}(x)}{f(x)\bar{G}(x)}\right) dG_t(x). \tag{23}$$

It is useful to reexpress the residual divergence as a function of the failure rates of the distributions involved.

LEMMA 2: *If* $\lim_{x\to\infty} \bar{G}(x)\log \bar{F}(x) = 0$, *then*

$$d(t) = \int_t^\infty \left[\frac{r_F(x)}{r_G(x)} - \log\left(\frac{r_F(x)}{r_G(x)}\right)\right] dG_t(x) + \log\left(\frac{\bar{F}(t)}{\bar{G}(t)}\right) - 1. \tag{24}$$

PROOF: Write

$$d(t) = \int_t^\infty \log\left(\frac{r_G(x)}{r_F(x)}\right) dG_t(x) + \int_t^\infty \log\left(\frac{\bar{G}(x)}{\bar{F}(x)}\right) dG_t(x) + \log\left(\frac{\bar{F}(t)}{\bar{G}(t)}\right). \tag{25}$$

Integrating by parts,

$$\begin{aligned}
-\int_t^\infty \log\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dG_t(x) &= \int_t^\infty \log\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) d\bar{G}_t(x) \\
&= \frac{\bar{G}(x)}{\bar{G}(t)} \log\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)\bigg|_t^\infty \\
&\quad - \int_t^\infty \frac{\bar{F}(x)g(x) - f(x)\bar{G}(x)}{\bar{G}(t)\bar{F}(x)}\, dx \\
&= \int_t^\infty \frac{r_F(x)}{r_G(x)}\, dG_t(x) - 1.
\end{aligned}$$

Rearranging, we obtain Eq. (24). ∎

THEOREM 3: *Let*

$$\lim_{x\to\infty} \left[\frac{r_F(x)}{r_G(x)} - \log\left(\frac{r_F(x)}{r_G(x)}\right)\right] \bar{G}(x) = 0. \tag{26}$$

(a) *If* $r_F(t) \le r_G(t)$, $t > 0$, *then* $r_F(t)/r_G(t)$ *increasing (decreasing) implies* $d(t)$ *decreasing (increasing).*

(b) *If* $r_F(t) \ge r_G(t)$, $t > 0$, *then* $r_F(t)/r_G(t)$ *increasing (decreasing) implies* $d(t)$ *increasing (decreasing).*

PROOF: Differentiating Eq. (23) and rearranging,

$$\begin{aligned}
d'(t) = r_G(t)\bigg[&\int_t^\infty \left[\frac{r_F(x)}{r_G(x)} - \log\left(\frac{r_F(x)}{r_G(x)}\right)\right] dG_t(x) \\
&- \left[\frac{r_F(t)}{r_G(t)} - \log\left(\frac{r_F(t)}{r_G(t)}\right)\right] + 1 - \frac{r_F(t)}{r_G(t)}\bigg] \\
= r_G(t)\bigg\{&\int_t^\infty \left(\frac{r_F(x)}{r_G(x)} - \frac{r_F(t)}{r_G(t)}\right) dG_t(x) + 1 - \frac{r_F(t)}{r_G(t)} \\
&+ \int_t^\infty \left[\log\left(\frac{r_F(t)}{r_G(t)}\right) - \log\left(\frac{r_F(x)}{r_G(x)}\right)\right] dG_t(x)\bigg\} \\
\equiv r_G(t)\,[I_1 + I_2 + I_3]&,
\end{aligned}$$

where $I_2 = 1 - r_F(t)/r_G(t)$.

(a) When $r_F(t) < r_G(t)$, if $r_F(t)/r_G(t)$ is increasing (decreasing) $I_1 + I_2 <$ ($>$) 0 and $I_3 <$ ($>$) 0 so that d is decreasing (increasing).

(b) When $r_F(t) > r_G(t)$, if $r_F(t)/r_G(t)$ is increasing (decreasing) $I_1 + I_2 >$ ($<$) 0 and $I_3 >$ ($<$) 0 so that d is increasing (decreasing). ■

Corollary 1: *Let X be IFR (DFR) and Y be exponential. Then, if $r_F(t) > 1$, d is increasing (decreasing). If $r_F(t) > 1$, d is decreasing (increasing).*

Acknowledgments

We thank Eshan Soofi and the referee for useful suggestions.

References

1. Barlow, R.E. & Proschan, F. (1975). *Statistical theory of reliability and life testing. Probability models.* New York: Holt Reinhardt and Winston.
2. Cover, T.M. & Thomas, J.A. (1991). *Elements of information theory.* New York: Wiley.
3. Gaede, K.-W. (1991). Stochastic orderings in reliability. In K. Mosler & M. Scarsini (eds.), *Stochastic orders and decisions under risk.* Hayward, CA: Institute of Mathematical Statistics, pp. 123–140.
4. Kotlarski, I. (1972). Characterization of probability distributions by conditional properties. *Sankhya, A* 34: 461–466.
5. Teitler, R., Rajagopal, A.K., & Ngai, U. (1986). Maximum entropy and reliability distributions. *IEEE Transactions on Reliability* R-35: 391–395.
6. Verdugo Lazo, A.C.G. & Rathie, P.N. (1978). On the entropy of continuous probability distributions. *IEEE Transactions on Information Theory* IT24-1: 120–122.