Residual Varentropy of Lifetime Distributions

Maadani, S.[1], **Mohtashami Borzadaran, G.R.**[1], **and Rezaei Roknabadi, A.H.**[1]

[1] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract: This paper deal with the varentropy for the residual lifetime random variables. The influence of systems age on residual varentropy is investigated. It is shown that in some distributions such as uniform, exponential and generalized Pareto, residual varentropy is independent of systems age. These distributions have characterized using residual varentropy, and a new class of distributions is also introduced.

Keywords: Characterization, Generalized Pareto Family, Residual Varentropy, Varentropy.

1 Introduction

The Shannon entropy (1948) of a continuous random variable X, with density function f, is defined as follows

$$h(X) = -\int_S f(x)\log(f(x))dx, \tag{1}$$

where $h(X)$ is called differential entropy, and S is the support of X. It is obvious that the Shannon differential entropy of X is the expectation of information content $-\log(f(X))$.

In applied statistics, the moments of a random variable such as mean and variance have important roles in data analysis. Due to $-\log(f(X))$ is a random

[1]Maadani, S.: *madani.sa@mail.um.ac.ir*

variable, looking at its statistics including variance, higher moment, and so on can be valuable. The variance of information content $-\log(f(X))$ has been studied in some papers recently, and considerable results in finite-blocklength information theory have been achieved. This variance is called varentropy, and it is a important parameter to estimate the performance of optimal coding, determine the dispersion of sources and channel capacity in computer sciences. There are few papers about varentropy in statistical studies. Song (2001) investigated varentropy for comparing the measure of kurtosis in heavy tailed distributions. Liu (2007) presented some mathematical properties for varentropy. Zagrofos (2008) and Enomoto et al. (2013) proposed a goodness of fit test based on varentropy. See also Kontoyiannis and Verdu (2013), Fradelizi et al. (2016), and Erdal (2016).

Let X be a continuous random variable with density function f the varentropy of X is defined as follows

$$VE(X) = Var(-\log f(X)) = E[-\log f(X) - h(X)]^2, \tag{2}$$

where $VE(X)$ is the varentropy of the random variable X. The varentropy is the expectation of the squared deviation of the information content $-\log(f(X))$, from its mean. This is a measure that indicates how the information content is dispersed around the entropy. Song (2001) showed that the varentropy can be used to compare the tail and shape of different densities as an intrinsic measure of the shape of a distribution. Within density functions that have the fourth moment, μ_4, and variance σ^2, the varentropy provides similar information to the well known kurtosis measure, $\frac{\mu_4}{\sigma^4}$. If the standard measure of kurtosis can not be calculated, (several heavy-tailed distributions) such as Student's t with the degree of freedom less than four, Cauchy and Pareto distribution, the varentropy is a good measure instead of $\frac{\mu_4}{\sigma^4}$.

The varentropy can provide a partial order about the distribution tails. For example, if X has a Student's t distribution with degree of freedom $\nu = 1,2,3,4,5$, the varentropies are, 3.2899, 1.5978, 1.1595, 0.9661, and 0.8588, respectively. Therefore when ν is increased, the tails become lighter and the varentropy decreases, consequently.

Liu (2007) in his Ph.D. thesis introduced some mathematical characteristics of the varentropy. Liu called the varentropy by Information Volatility and showed that this measure can be characterized the uniform distribution, and used varentropy for separating the normal and gamma and a subfamily of the beta distribution.

If the lifetime of a system is considered as a random variable, the uncertainty measure of this system up to a specified time or afterword, has particular importance. These two important measures are referred to as past and residual entropies, respectively. That entropies have many applications, such as characterization of distributions, stochastic ordering, and so on.

Studying the varentropy for residual lifetime distributions is the aim of this paper. We will investigate the effect of systems age on it, and introduce some characterization using the residual varentropy. Also, a new class of the distributions by the residual varentropy is introduced.

2 Residual varentropy and characterization

Shannon entropy is used as a measure of uncertainty for a random variable in information theory. Nonetheless, if two random variables have the same entropy, there is a common question. Which of the entropies is the most suitable criterion for measuring uncertainty? For instance, the Shannon entropy is zero for standard uniform and also exponential distribution with parameter e. In reality, the question is, do both entropies calculate uncertainty equally accurately? If the concentration of information content is more around the entropy, then the entropy would be appropriate to measure the amount of uncertainty. This concentration can be calculated with the variance of $-\log f(X)$. It can be shown that the varentropy of the uniform distribution is zero and for exponential distribution is 1, so in the uniform distribution, entropy is more appropriate for measuring the uncertainty.

In lifetime studies, we usually have knowledge about the age of the system and we know, the system is still operating at the moment. If a system is known to have survived to age t, Clearly (2) is no longer useful for measuring the uncertainty about remaining lifetime of the system. Ebrahimi (1996) introduced a measure of uncertainty of residual lifetime distributions as follows

$$h(X,t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx, \tag{3}$$

where $h(X,t)$ is the residual entropy, $\bar{F}(\cdot)$ is survival function and (3) is expressed based on the Shannon entropy for random variable $\{X-t|X \geq t\}$.

For further study see also Ebrahimi and Kirmani (1996), Sankaran and Gupta (1999), Asadi and Ebrahimi (2000) and Abraham and Sankaran (2006). This entropy is the expectation of the random variable $-\log \frac{f(X)}{\bar{F}(t)}$ with respect to density function $g(x,t) = \frac{f(x)}{\bar{F}(t)}$, $x > t$. In this section, we introduce the residual varentropy for lifetime distribution. The residual varentropy is variance of $-\log \frac{f(X)}{\bar{F}(t)}$ and is noted by $VE(X,t)$.

Now the last question is raised again, if two residual lifetime random variables have the same uncertainty, which of them shows the uncertainty with more accurately? It is clear that the answer must be found by the residual varentropy. Therefore residual varentropy also indicate concentration of information content $-\log \frac{f(X)}{\bar{F}(t)}$ around the residual entropy, $h(X,t)$, and it answers the last question.

On the other hand, similar to Song's measure, the residual varentropy is able to compare the lifetime distribution in term of the heaviness tail, and gives us similar information to kurtosis measure for residual lifetime distributions.

Definition 2.1. Let X be a non-negative random variable with density function f, and $\{X-t|X \geq t\}$ be residual lifetime random variable, the residual varentropy is define as follows

$$VE(X-t|X \geq t) = VE(X,t) = Var\left(-\log \frac{f(X)}{\bar{F}(t)} \middle| X \geq t \right).$$

It is clear that $VE(X,0)$ is the varentropy of X.

Basically, the calculation of the variance of the random variable $-\log f(X)$

in not simple, and for the residual lifetime random variables is very difficult. Therefore we propose the using of the moment generating function (MGF) of the $-\log\frac{f(X)}{\bar{F}(t)}$ for calculating the $VE(X,t)$.

Proposition 2.2. *We define MGF of* $\log\frac{f(X)}{\bar{F}(t)}$ *as below*

$$L(X,t,\lambda) = E\left(e^{(\lambda-1)\log\frac{f(X)}{\bar{F}(t)}}\right) = \int_t^{\infty}\left(\frac{f(x)}{\bar{F}(t)}\right)^{\lambda}dx. \tag{4}$$

Then,

$$VE(X,t) = L''(X,t,1) - \left(L'(X,t,1)\right)^2, \tag{5}$$

where $L'(X,t,1)$ *and* $L''(X,t,1)$ *are first and second order derivatives of* $L(X,t,\lambda)$*, with respect to* λ*, respectively, at* $\lambda = 1$.

Using this proposition we calculated the residual varentropy in some lifetime distributions and compare varentropy with the residual varentropy. We see in uniform, exponential, Laplace and generalized Pareto distributions, the residual varentropy is independent of the systems age but in other distributions such as gamma, weibull, lognormal and so on the residual varentropy is dependent of t.

Example 2.3. Let X has gamma distribution with parametrers θ and λ and density function $f(x) = \frac{\lambda^{\theta}}{r(\theta)}x^{\theta-1}e^{-\lambda x}, \theta > 0, \lambda > 0, x > 0$, the residual varentropy using (5) is: $VE(X,t) = M[\lambda t - (\theta-1)(2\log(\lambda t) - 1)] - M^2 + 2M(\theta-1)$ $\Psi(\theta,\lambda t) + (\theta-1)^2\dot{\Psi}(\theta,\lambda t) + 2 - \theta$, and $M = \frac{\Gamma(\theta+1,\lambda t)}{\Gamma(\theta,\lambda t)} - \theta$,
where $\Gamma(a,b)$, $\Psi(a,b)$ and $\dot{\Psi}(a,b)$ are incomplete gamma, incomplete digamma and incomplete trigamma functions respectively.

This example implies that if $\theta = 2$ and $\lambda = 1$, $VE(X) = 0.63$ but if $t = 1$, $VE(X,1) = .76$. Therefore the residual varentropy is dependent of the systems age in this distribution.
If we calculate the derivative of the $VE(X,t)$, with respect to t, then:

$$VE'(X,t) = r(t)[VE(X,t) - (\log f(t) - E(\log f(X)|X \geq t)^2], \tag{6}$$

also it can be shown that

$$VE'(X,t) = r(t)\left[VE(X,t) - (\log r(t) + h(X,t))^2\right], \tag{7}$$

where $r(t) = \frac{f(t)}{\bar{F}(t)}$ is the hazard rate function and using (7) we have the following proposition

Proposition 2.4. *The residual varentropy is a constant function with respect to t if*

$$VE(X,t) = (\log r(t) + h(X,t))^2, \tag{8}$$

where $h(X,t)$ is the residual entropy of X, and $r(t)$ is the hazard rate function of it.

We shown that the residual varentropy is able to characterises some distributions. In the following Theorems we express this distributions.

Theorem 2.5. *X has a uniform distribution if and only if $VE(X,t) = 0$ for all $t > o$.*

Proof. Let $X \sim U(a,b)$ then $VE(X,t) = Var(\log \frac{1}{b-a} | X \geq t) = 0$. Also if $VE(X,t) = 0$, then we can show that $f(x) = \bar{F}(t)e^{-h(X,t)} = c$. □

Theorem 2.6. *X has exponential distribution if and only if $VE(X,t) = 1$.*

Proof. Let $X \sim Exp(\theta)$ then the random variable $\{X - t | X \geq t\}$ is identical in distribution with X. So $VE(X,t) = VE(X) = 1$. Vice versa if $VE(X,t) = 1$, then by using (7) and some mathematical computation it can be shown that $r(t) = c$. Therefore X has exponential distribution. □

One of the important distribution in reliability theory and survival analysis is the generalized Pareto distribution (GPD). This distribution was introduced by Pickands (1975). Its applications include use in the analysis of events, in the modeling of large insurance claims, as a failure-time distribution in reliability studies, and in any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives. if X has GPD, the distribution function of X as follows

$$F(x,k,\sigma) = 1 - (1 - \frac{kx}{\sigma})^{\frac{1}{k}}, \quad k \neq 0, \sigma > 0, \tag{9}$$

where k and σ are the shape and scale parameters, respectively. The support of X is $x > 0$ if $k \leq 0$, and $0 \leq x \leq \frac{\sigma}{k}$ if $k > 0$. In the special cases, if $k \to 0$, the GPD reduces to the exponential distribution with mean σ, and when $k = 1$, GPD has uniform distribution, and if $k < 0$, it has second kind of the Pareto distribution.

Theorem 2.7. *The continuous non-negative random variable X is GPD with distribution function* (9) *if and only if* $VE(X,t) = c \geq 0, c \neq 1$.

Proof. If the random variable X has a generalized Pareto distribution ,then the conditional distribution $X - t$ given $X \geq t$, is also generalized Pareto with the same value of k. It can be shown that, $VE(X,t) = VE(X) = (k-1)^2 = c, \;\; k \neq 0$. Vice versa, if $VE(X,t) = c$; therefore $VE'(X,t) = 0$ and by using (7) we have $h(X,t) = c - \log r(t)$. Asadi and Ebrahimi (2000) showed the last equation implies F is the generalized Pareto distribution. □

3 A class of distributions

Ebrahimi (1996) provided a class of lifetime distributions based on the measure of uncertainty of residual lifetime random variables as follows:

Definition 3.1. $\bar{F}$ has decreasing (increasing) uncertainty of residual life $DURL(IURL)$, if $h(X,t)$ is decreasing (increasing) in t.

He showed that if $\bar{F}$ is an increasing (decreasing) failure rate IFR (DFR), then it is also a $DURL$ $(IURL)$ and

$$r(t) \leq (\geq) \exp(1 - h(X,t)). \tag{10}$$

parallel to work of Ebrahimi(1996) we are going to introduce a class of lifetime distributions using residual varentropy. Various properties of this class will also provided.

Definition 3.2. $\bar{F}$ has increasing (decreasing) residual varentropy $IRVE(DRVE)$, if $VE(X,t)$ is an increasing (decreasing) in t, $t \geq 0$.

Remark 3.3. It is clear that if $\bar{F}$ is $IRVE(DRVE)$ we have:

$$VE(X,t) \geq (\leq) VE(X), \tag{11}$$

equality holds if (8) is established. We see that (11) is lower(upper) bound for residual varentropy in this situations.

Proposition 3.4. *For a non-negative random variable X, $\bar{F}$ has $IRVE(DRVE)$, if*

$$VE(X,t) \geq (\leq) (\log r(t) + h(X,t))^2. \tag{12}$$

Proof. using (7) , (12) easily obtained. □

Corollary 3.5. *Suppose that $\bar{F}$ is both $IRVE(DRVE)$ and $0 < f(0) < \infty$. then $VE(X) = (\log f(0) + h(X))^2$.*

Corollary 3.6. *If $\bar{F}$ has $IRVE(DRVE)$ in t, then*

$$VE(X) \geq (\leq) (\log f(0) + h(X))^2. \tag{13}$$

Therefore (13) is the lower (upper) bound for the VE in this distributions.

Corollary 3.7. *If $\bar{F}$ is IFR and $DRVE$ (DFR and $IRVE$), then*

$$VE(X,t) \leq (\geq) 1.$$

Proof. If X is IFR using (10) we have:

$$(\log r(t) + h(X,t))^2 \leq 1, \tag{14}$$

and if X is $DRVE$ (12) and (14) implies $VE(X,t) \leq 1$. Other inequality is similarly proved. □

Conclusion

In this paper, a measure has been proposed for evaluating uncertainty about the random variable of residual lifetime distribution with the name of residual varentropy. This measure is able to compare the kurtosis measure of residual lifetime distributions. It has been shown that the residual varentropy in some distributions is independent of the systems age, such as uniform, exponential, and generalized Pareto family. It also has been proven that the residual varentropy characterizes these distributions. Moreover, a new class of lifetime distributions has been introduced using residual varentropy. Future work in this direction may focus on characterizing by past varentropy.

References

[1] Abraham, B. and Sankaran, P. G. (2006). Renyi's entropy for residual lifetime distribution. *Statist. Papers*, **47**, 17–29.

[2] Asadi, M. and Ebrahimi, N. (2000). Residual entropy and its characterizations in terms of hazard function and mean residual life time function, *Statist. Probab. Lett.*, **49**, 263–269.

[3] Ebrahimi, N. (1996). How to measure uncertainty in the life time distributions. *Sankhyā Ser. A*, **58**, 48–57.

[4] Ebrahimi, N. and Kirmani, S. N. U. A. (1996). Some results on ordering of survival functions through uncertainty. *Statist. Probab. Lett.*, **29**, 167–176.

[5] Ebrahimi, N. and Pellerey, F. (1995). New partial ordering of survival functions based on the nation of uncertainty. *J. Appl. Probab.*, **32**, 202–211.

[6] Enomoto, R. Okamoto, N. and Seo, T. (2013). On the asymptotic normality of test statistics using Song's kurtosis. *J. Stat. Theory Pract.*, **7**, 102–119.

[7] Erdal, A. (2016). Varentropy decreases under the polar transform. *IEEE Trans. Inform. Theory*, **60**, 3390–3400.

[8] Fradelizi, M. Madiman, M. and Wang, L. (2016). Optimal concentration of information content for log-concave densities. High dimensional probability VII, *Progr. Probab.*, **71**, Springer, [Cham] 45–60.

[9] Kontoyiannis, I. and Verdu, S. (2013). Optimal lossless compression: Source varentropy and dispersion. *IEEE Int. Symposium on Information Theory*, 1739–1743.

[10] Liu, J. (2007). Information theoretic content and probability. PhD Thesis, *University of Florida.* ProQuest LLC, Ann Arbor, MI.

[11] Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist*, **3**, 119–131.

[12] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 623–656.

[13] Sankaran, P. G. and Gupta R. P. (1999). Characterization of the life time distributions using measure of uncertainty. *Calcutta Statistical Association Bulletin*, **49**, 159–166.

[14] Song, K. S. (2001). Renyi information, log likelihood and an intrinsic distribution measure, *J. Stat. Plan. Inference*, **93**, 51–69.

[15] Zografos, K. (2008). On Mardia's and Song's measures of kurtosis in elliptical distributions. *J. Multivariate Anal.*, **99**, 858–879.