JIRSS (2012)
Vol. 11, No. 2, pp 203-215

On the Discrete Cumulative Residual Entropy

S. Baratpour[1], Z. Bami[2]

[1]Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Iran.
[2]Department of Statistics, Mashhad Branch, Islamic Azad University, Iran.

Abstract. Recently, Rao et al. (2004) have proposed a new measure of uncertainty for a distribution function F, called cumulative residual entropy (CRE) and obtained some properties and applications of that. Asadi and Zohrevand (2007) have proposed that CRE has connected to some well-known reliability measures. In the present paper, we introduce cumulative residual entropy for discrete random variables and study some of its properties, and show how it is connected with some well-known measures such as mean residual life-time.

Keywords. Cumulative residual entropy, discrete distribution, discrete failure rate, mean residual life-time, seconde failure rate, survival function.

MSC: Primary: 62B10; Secondary: 94A17.

1 Introduction

Shannon (1948) introduced a measure of uncertainty in a discrete distribution based on the Boltzman entropy of classical statistical mechanics and called it entropy. Shannon entropy for discrete random variable X is defined by

$$H(X) = -\sum_x p(x) \log p(x),$$

S. Baratpour(✉)(baratpour@um.ac.ir), Z. Bami(zahrabami@gmail.com)
Received: December 12, 2011; Accepted: August 13, 2012

where $p(.)$ is probability mass function of X. With this, he opened up a new branch of mathematics with far reaching applications in many areas. To name a few: financial analysis, data compression (Salomon (1998)), statistics and information theory (Kullback (1959) and Cover and Thomas (1991)). In continuous cases, Shannon entropy is called relative entropy that is defined as follows

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) \mathrm{d}x,$$

where f is density function. However, the Shannon entropy has certain disadvantages. For example, it may take any value on the extended real line, it requires the knowledge of density function for non-discrete random variables, the discrete Shannon entropy dose not converge to its continuous analogous, and in order to estimate the Shannon entropy for a continuous density, one has to obtain the density estimation, which is not a trivial task. Rao et al. (2004) introduced an alternative of uncertainty called cumulative residual entropy (CRE). This measure is based on the cumulative distribution function F and is defined for non-negative random variables as follows

$$\xi(X) = -\int_{0}^{\infty} \bar{F}(x) \log \bar{F}(x) \mathrm{d}x,$$

where $\bar{F}(x) = 1 - F(x)$ is survival function. $\xi(X)$ measures the uncertainty contained in the survival function of X. The basic idea in their definition was to replace the density function by the survival function in Shannon's definition. CRE is more general than the Shannon entropy and possesses more general mathematical properties than the Shannon entropy. This measure is always non-negative and its definition is valid for both continuous and discrete cases. It can easily be computed from sample data and its estimation asymptotically converges to the true value.

The relation between CRE and Shanoon entropy is as follows.

• If Y be a non-negative random variable with pdf $g(y) = \frac{\bar{F}(y)}{E(X)}$, where $\bar{F}$ is the survival function of random variable X, then

$$H(Y) = \frac{1}{E(X)} \xi(X) + \log E(X). \tag{1}$$

Note that in this property, Y is called the equilibrium random variable assosiated to X (Rao et al. (2004)). Equality (1) clarify that CRE measures the uncertainty contained in survival function.

• CRE is equivalent to $E(MRL(X))$, where $MRL(X)$ is the mean residual life-time function (Asadi and Zohrevand (2007)).
CRE has applications in reliability engineering and computer vision, for more details see Rao (2005). Baratpour and Habibi Rad (2011) developed a consistent test for testing the hypothesis of exponentially against some alternatives. Asadi et al. (2007) showed how CRE connected with reliability measures such as the mean residual life-time. In this work we propose an alternative measure of uncertainty for random variable X with discrete distribution and call it discrete cumulative residual entropy (D-CRE) of X. The main objective of our study is to extend CRE to random variables with discrete distribution and connect it to some well-known reliability measure such as mean residual life-time.

Rest of the paper is organized as follows: Section 2 contains some basic definitions which are required in later sections. The reader may proceed to section 3, and refer to these results as needed. After that, in section 3, we define D-CRE and explain some properties of it. Then, we compute this measure in some examples. We show that D-CRE dominates the discrete Shannon entropy. Finally in section 4, we show that D-CRE is always greater than the expectation of the mean residual life-time of X which plays an important role in reliability and survival analysis. In this section, we also give a lower bound for the D-CRE of the statistical models which are in the class of new worse than used in expectation (NWUE) for discrete distributions.

2 Basic Relationships

In this section, we have some fundamental definitions that are needed in the next sections. Let X be a counting random variable with probability mass function $p(k) = P(X = k), \quad k = 0, 1, 2, ..b, \quad b \leq \infty$ and survival function

$$R(k) = P(X \geq k) = \sum_{x=k}^{b} P(X = x) = \sum_{x=k}^{b} p(x), \ k = 0, 1, 2, ..b, \quad b \leq \infty$$

Definition 2.1. The failure rate of X is defined as

$$h(k) = P(X = k \mid X \geq k) = \frac{P(X = k)}{P(X \geq k)} = \frac{p(k)}{R(k)}.$$

Therefore, we can write

$$h(k) = \frac{R(k) - R(k+1)}{R(k)} = 1 - \frac{R(k+1)}{R(k)}.$$

It may be noted that discrete setup ensures upper and lower bounds as 1 and 0, respectively, for the failure rate function. We refer the reader to Barlow and Marshal (1963) for some general results regarding these measures.
The most important difference between hazard rate function in continuous and discrete distribution is that $-\ln R(k) \neq \sum_{i=0}^{k} h(i)$. Thus, we have to determine another version of that. We refer the reader to Roy and Gupta (1999).

Definition 2.2. The second failure rate (SFR) of X is defined as

$$\begin{aligned} r(k) &= \log \frac{R(k-1)}{R(k)} \\ &= -\log P(X \geq k | X \geq k-1) \end{aligned} \tag{2}$$

In (2), $P(X \geq k|X \geq k-1)$ is the conditional probability that a device be alive at time k, given that it has not failed by $k-1$. By noting that $-\log(.)$ is a decreasing function, if $P(X \geq k|X \geq k-1)$ increases, then $r(.)$ will decrease.
It is easy to verify that the function $h(k)$ and $r(k)$ satisfy the relation of the form

$$r(k+1) = -\log(1-h(k)). \tag{3}$$

Hence, both $r(k)$ and $h(k)$ have the same monotonicity property.

The MRL plays an important role in reliability and survival analysis to model and analyze the data. Let in the following definition, X be a discrete random variable with survival function $R(k), k = k_1, k_2, ...(k_1 \leq k_2 \leq ...)$.

Definition 2.3. The MRL of X, which we denoted by $MRL(k)$ is defined as

$$MRL(k) = E(X-k \mid X \geq k) = [\sum_{k_j \geq k} k_j \frac{R(k_j) - R(k_{j+1})}{R(k)}] - k.$$

Assume $k = k_i$, then we have

$$MRL(k_i) = \sum_{j=i}^{\infty} \frac{(k_{j+1} - k_j) R(k_{j+1})}{R(k_i)}.$$

Let X be a discrete random variable with support $\{0, 1, ..., b\}, \quad b \leq \infty$, we have

$$MRL(k) = \sum_{i=k}^{b-1} \frac{R(i+1)}{R(k)}, \quad k = 0, 1, 2, ..., b, \quad b \leq \infty.$$

Classification of distribution with respect to ageing properties is a popular theme in reliability theory. A class of distributions which arises in the study of replacement and maintenance policies is the class of new better (worse) than used in expectation (NBUE (NWUE)) distributions.

Definition 2.4. A discrete random variable X with support $N = \{0, 1, 2, ..., b\}, b \leq \infty$ or its distribution is said to be D-NBUE (D-NWUE), if

$$R(k) \sum_{i=1}^{b} R(s) \geq (\leq) \sum_{i=k}^{b-1} R(i+1), \quad k = 0, 1, 2, ..., b-1. \tag{4}$$

We refer the readers to Esary and Marshal (1973) for some general results regarding this definition.

3 Discrete Cumulative Residual Entropy

In this section, we introduce an information measure similar to CRE but for non-negative discrete random variables.

Definition 3.1 Let X be a non-negative discrete random variable with survival function $R(x_i), \quad x_1 < x_2 < ... < x_b, \quad b \leq \infty$. The discrete cumulative residual entropy (D-CRE) of X is defined as

$$\begin{aligned} d\xi(X) &= -\sum_{i=1}^{b} P(X \geq x_i)(\log P(X \geq x_i))(x_i - x_{i-1}) \\ &= -\sum_{i=2}^{b} P(X \geq x_i)(\log P(X \geq x_i))(x_i - x_{i-1}). \end{aligned}$$

D-CRE measures the uncertainty contained in distribution function of discrete random variables. Thus, if $d\xi(X) \leq d\xi(Y)$, then we conclude that the uncertainty contained in distribution function of X is equal or less than the uncertainty contained in that of Y.

Some properties of D-CRE are as follows.

• D-CRE takes values in $[0, \infty]$. In particular, $d\xi(X) = 0$, if and only if X is a constant.

• If random variable X with support $N = \{0, 1, 2, ..., b\}$, $b \leq \infty$ be the discrete life-time of a device, then

$$d\xi(X) = -\sum_{i=1}^{b} P(X \geq i) \log P(X \geq i).$$

• If $Y = a + bX$, $a \geq 0$, $b > 0$, then $d\xi(Y) = b(d\xi(X))$.

• If X be uniformly distributed on $\{x_1, x_2, ..., x_b\}$, $b \leq \infty$, then

$$d\xi(X) = -\sum_{i=1}^{b} \frac{n-i}{n} \log \frac{n-i}{n}(x_i - x_{i-1}).$$

Example 3.1. (a) Consider a uniform distribution with the probability mass function

$$p(x) = \frac{1}{3}, \qquad x = 0, 1, 2,$$

then, its D-CRE is computed as follows

$$\begin{aligned} d\xi(X) &= -\sum_{i=1}^{2} \frac{1}{3}(3-i) \log(\frac{1}{3}(3-i)) \\ &= \log 3 - \frac{2}{3} \log 2 \\ &\simeq 0.63 \end{aligned}$$

(b) Let X have geometric distribution with density function:

$$P(x) = pq^x, \qquad x = 0, 1, ...$$

Then, D-CRE of the this distribution is

$$\begin{aligned} d\xi(X) &= -\sum_{i=1}^{\infty} q^i \log q^i \\ &= -\log q \sum_{i=1}^{\infty} iq^i \\ &= -\frac{q}{p^2} \log q \end{aligned}$$

We show below that D-CRE dominates the discrete Shannon entropy.

Theorem 3.1. *Let X be a discrete random variable with probability mass function and survival function $p(i)$ and $R(i)$, $i = 0, 1..., b$, $b \leq \infty$, respectively. Then,*

$$d\xi(X) \geq C \exp(\frac{H(X)}{1-p(0)}),$$

where $H(X)$ is discrete Shannon entropy of X and

$$\begin{aligned} C &= \exp\Big[\frac{1}{1-p(0)}\Big(p(0)\log p(0) + (1-p(0))\log(1-p(0)) \\ &+ \sum_{x=1}^{b} p(x)\log(R(x)|\log R(x)|)\Big)\Big]. \end{aligned}$$

Proof. Using the log-sum inequality, we have

$$\begin{aligned} \sum_{x=1}^{b} p(x)\log\frac{p(x)}{R(x)|\log R(x)|} &\geq \sum_{x=1}^{b} p(x)\log\frac{\sum_{x=1}^{b} p(x)}{\sum_{x=1}^{b} R(x)|\log R(x)|} \quad (5) \\ &= (1-p(0))\log\frac{(1-p(0))}{d\xi(X)}. \end{aligned}$$

If $d\xi(X)$ is infinite, then the proof is trivial. The left-hand side (LHS) in (5) equals to

$$-H(X) - p(0)\log p(0) - \sum_{x=1}^{b} p(x)\log[R(x)|\log R(x)|].$$

Thus,

$$\begin{aligned} \frac{H(X)}{1-p(0)} + \frac{1}{1-p(0)}\Big(p(0)\log p(0) + (1-p(0))\log(1-p(0)) \\ + \sum_{x=1}^{b} p(x)\log(R(x)|\log R(x)|)\Big) \leq \log d\xi(X). \quad (6) \end{aligned}$$

Exponentiating both sides of (6), the proof is completed. □

In the following theorem, based on the D-CRE, we find an upper bound for the expectation of the absolute value of the difference between two random variables.

Theorem 3.2. *If X and Y be two iid discrete random variables with support $\{0, 1, 2, ..., b\}$,*
$b \leq \infty$, then

$$E(|X - Y|) \leq 2d\xi(X).$$

Proof. By noting that X and Y are iid, we have

$$P(max(X, Y) \geq t)) = 2R(t) - R^2(t)$$

and

$$P(min(X, Y) \geq t) = R^2(t),$$

where $R(t) = P(X \geq t)$ is survival function of X. Thus,

$$\begin{aligned} E(|X - Y|) &= E(max(X, Y) - min(X, Y)) \\ &= \sum_{t=1}^{b} P(max(X, Y) \geq t) - \sum_{t=1}^{b} P(min(X, Y) \geq t) \\ &= 2\sum_{t=1}^{b}(R(t) - R^2(t)) \\ &\leq 2\sum_{t=1}^{b} P(X \geq t)|\log P(X \geq t)| \qquad (7) \\ &= 2d\xi(X), \end{aligned}$$

where (7) is obtained by $x(1 - x) \leq x|\log x|,\ 0 < x < 1$. □

4 Relations between D-CRE and some Measures in Reliability

In this section, we have some theorems and examples. In the following theorem, using stochastic ordering concept, an upper bound for the difference between two D-CRE is obtained.

Theorem 4.1. *Let X and Y be two discrete random variables with support $\{0, 1, 2, ..., b\}$,*
$b \leq \infty$ and finite expectations $E(X)$ and $E(Y)$, respectively. If $X \leq_{st} Y$, then

$$d\xi(X) - d\xi(Y) \leq E(X) \log \frac{E(Y)}{E(X)}.$$

Proof. By log-sum inequality, we have

$$\sum_{x=1}^{b} P(X \geq x) \log \frac{P(X \geq x)}{P(Y \geq x)} \geq \sum_{x=1}^{b} P(X \geq x) \log \frac{\sum_{x=1}^{b} P(X \geq x)}{\sum_{x=1}^{b} P(Y \geq x)}.$$

Thus,

$$d\xi(X) \leq -\sum_{x=1}^{b} P(X \geq x) \log P(Y \geq x) - E(X) \log \frac{E(X)}{E(Y)}$$

By definition of $X \leq_{st} Y$, we have $P(X \geq x) \leq P(Y \geq x)$, which concludes that

$$d\xi(X) \leq -\sum_{x=1}^{b} P(Y \geq x) \log P(Y \geq x) + E(X) \log \frac{E(Y)}{E(X)}.$$

Thus the proof is completed. □

The following corollary is concluded from Theorem 4.1.

Corollary 4.1. *Let X and Y be two discrete random variables with support $\{0, 1, 2, ..., b\}$,*
$b \leq \infty$ and finite expectations $E(X)$ and $E(Y)$, respectively. If $X \leq_{st} Y$ and
$E(X) = E(Y)$, then

$$d\xi(X) \leq d\xi(Y).$$

Theorem 4.2. *Let X be a discrete random variable with support $\{0, 1, 2, ..., b\}$, $b \leq \infty$ and SFR, r_j, then*

$$d\xi(X) = \sum_{i=1}^{b} \sum_{j=1}^{i} r(j) P(X \geq i).$$

Proof. To prove the result note that

$$\begin{aligned}
\sum_{i=1}^{b} \sum_{j=1}^{i} r(j) P(X \geq i) &= \sum_{i=1}^{b} \sum_{j=1}^{i} \log \frac{R(j-1)}{R(j)} P(X \geq i) \\
&= \sum_{i=1}^{b} \log \frac{R(0)}{R(i)} P(X \geq i) \\
&= -\sum_{i=1}^{b} P(X \geq i) \log P(X \geq i) \\
&= d\xi(X).
\end{aligned}$$

□

In the following theorem, we show in the case that the underlying distribution F is discrete, the D-CRE is greater than $E(MRL(X))$.

Theorem 4.3. *Let X be a discrete random variable with support $\{0,1,2,...,b\}$, $b \leq \infty$ and D-CRE, $d\xi(X)$, then*

$$d\xi(X) \geq E(MRL(X)), \tag{8}$$

and equality holds if and only if X be degenerate.

Proof. To prove the result, note that by Theorem 4.2,

$$\begin{aligned}
d\xi(X) &= \sum_{i=1}^{b}\sum_{j=1}^{i}\log\frac{R(j-1)}{R(j)}P(X\geq i)\\
&= \sum_{j=1}^{b}\sum_{i=j}^{b}\log\frac{R(j-1)}{R(j)}P(X\geq i)\\
&= \sum_{j=0}^{b-1}\log\frac{R(j)}{R(j+1)}\sum_{i=j+1}^{b}P(X\geq i)\\
&= \sum_{j=0}^{b-1}r(j+1)\sum_{i=j}^{b-1}P(X\geq i+1)\\
&= \sum_{j=0}^{b-1}[h(j)+\frac{\sum_{n=2}^{\infty}(h(j))^n}{n}]\sum_{i=j}^{b-1}P(X\geq i+1) \qquad (9)\\
&= \sum_{j=0}^{b-1}h(j)\sum_{i=j}^{b-1}P(X\geq i+1)+\sum_{j=0}^{b-1}[\sum_{n=2}^{\infty}\frac{(h(j))^n}{n}\sum_{i=j}^{b-1}P(X\geq i+1)]\\
&= E(MRL(X))+\epsilon\\
&\geq E(MRL(X)),
\end{aligned}$$

where $\epsilon = \sum_{j=0}^{b-1}[\sum_{n=2}^{\infty}\frac{(h(j))^n}{n}\sum_{i=j}^{b-1}P(X\geq i+1)]$ is nonnegative. In equality (9), we used equality (3) and power series of $-\ln(1-h(j))$. The proof of the if and only if part is trivial. Hence, the proof is complete. □

The representation in (8) is useful in the sense that in many statistical models one may has information about the MRL and its behaviour. Thus, we can find a lower bound for $d\xi(X)$. Let us look at the following example.

Example 4.1. Let X have geometric distribution with density function:

$$P(x) = pq^x, \qquad x = 0, 1, ...,$$

then its MRL is computed as follows:

$$\begin{aligned} MRL(k) &= \sum_{i=k}^{\infty} \frac{R(i+1)}{R(k)} \\ &= \sum_{i=k}^{\infty} \frac{q^{i+1}}{q^k} \\ &= \frac{q}{p}. \end{aligned}$$

Hence,

$$E(MRL(X)) = \frac{q}{p}.$$

In Example 3.1(b) we showed that,

$$d\xi(X) = -\frac{q}{p^2} \ln q.$$

Now by using equality $\ln(x) < x - 1$, for $0 < x < 1$, we get

$$-\frac{q}{p^2} \ln q > \frac{q}{p}.$$

Thus, $d\xi(X) > E(MRL(X))$

Remark 4.1. Let random variable X with support $N = \{0, 1, 2, ..., b\}, b \leq \infty$ be D-NWUE, then based on representation (4), we get

$$E(X) \leq E(MRL(X)) \leq d\xi(X).$$

This gives lower and upper bounds for $E(MRL(X))$ of the statistical models which are in the class of D-NWUE distributions. Also, when F is D-NWUE one can easily conclude that D-CRE of it is greater than mean of F. See the example below for more details.

Example 4.3. Let X be distributed as Waring with probability density function:

$$P(x) = \frac{(c-a)(a+x-1)!c!}{c(a-1)!(c+x)!}, \qquad x = 0, 1, ... \qquad c > a > 0.$$

It has decreasing D-FR (see Gupta et al. (1997)). Thus, it is NWUE and a lower bound for the D-CRE is given by

$$\frac{a}{c-a-1} \leq d\xi(X).$$

Acknowledgements

The authors would like to thank the referees for their careful reading and their useful comments which led to this considerably improved version.

References

Asadi, M. and Zohrevand, Y. (2007), On the dynamic cumulative residual entropy. Journal of Statistical Planning and Inference, **137**, 1931-1941.

Baratpour, S. and Habibi Rad, A. (2012), Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. Commun. Statist. Theo. Meth., to appear.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1963), Properties of probability distributions with monotonic hazard rate. J. Annals of Statistics, **34**(3), 348-350.

Cover, T. M. and Thomas, J. A. (1991), Elements of Information Theory, John Wiley.

Esary, J. D., Marshall, A. W., and Proschan, F. (1973), Shock Models and Wear Processes. J. The Annals of Probability, **1**(4), 627-649.

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1997), On the monotonic properties of discrete failure rates. Journal of Statistical planning and Inferences, **65**, 255-268.

Kullback, S. (1959), Information Theory and Statistics. Wiely: New York.

Rao, M., Chen, Y., Vemuri, B. C., and Wang, F. (2004), Cumulative residual entropy: a new measure of information. IEEE Trans. Inform. Theory, **50**(6), 1220-1228.

Rao, M. (2005), More on a new concept of entropy and information. J. Theoret. Probab., **18**, 967-981.

Roy, D. and Gupta, R. P. (1999), Characterizations and model selections through reliability measures in the discrete case. Statistics and Probability Letters, **43**, 197-206.

Salomon, D. (1998), Data Compression. Springer.

Shannon, C. E. (1948), A mathematical theory of communication. Bell System Tec. J., **27**, 379-432.