

On Dynamic Cumulative Entropy of Order Statistics

Richa Thapliyal, Vikas Kumar and H.C.Taneja

Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi-110042, India

Received: 14 Jul. 2012, Revised: 25 Oct. 2012, Accepted: 27 Oct. 2012

Published online: 1 Mar. 2013

Abstract: We have proposed measures of cumulative entropy and dynamic cumulative entropy based on order statistics and derived a characterization result that dynamic cumulative entropy of the i^{th} order statistics determines the distribution function uniquely. Some properties of the measures proposed have also been studied.

Keywords: Cumulative entropy, Order Statistics, Mean inactivity time, Reversed hazard rate, Survival function.

1 Introduction

In information theory, Shannon entropy [14] plays a vital role to measure the index of dispersion, volatility or uncertainty associated with a random variable X . Shannon entropy for a non-negative continuous random variable X is given by

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

where $f(x)$ is the pdf of random variable X .

Consider a system X that has survived upto time t . In order to calculate the uncertainty about the residual life of such a system, Shannon entropy (1) is not appropriate. Ebrahimi [7] introduced a new measure to ascertain the uncertainty about the residual life of a random variable $X_t = (X|X \geq t)$ given by

$$\begin{aligned} H(F;t) &= - \int_t^{\infty} \frac{f(x)}{\bar{F}(t)} \log \left(\frac{f(x)}{\bar{F}(t)} \right) dx \\ &= 1 - \frac{1}{\bar{F}(t)} \int_t^{\infty} f(x) \log \lambda_F(x) dx, \end{aligned} \tag{2}$$

where $\bar{F}(x) = 1 - F(x)$ is the survival function of X and $\lambda_F(x)$ is its failure rate or hazard rate defined by $f(x)/\bar{F}(x)$.

Clearly for $t = 0$, $H(F;0) = - \int_0^{\infty} f(x) \log f(x) dx$ represents the Shannon uncertainty contained in X . Di Cr- escenzo and Longobardi [6] studied the case when the system has been monitored on regular basis. Suppose that at any instant t , the system has been found to be dead. They introduced a measure known as past entropy to find the uncertainty about the time at which the system has failed. It is given by

$$\begin{aligned} \bar{H}(F;t) &= - \int_0^t \frac{f(x)}{F(t)} \log \left(\frac{f(x)}{F(t)} \right) dx \\ &= 1 - \frac{1}{F(t)} \int_0^t f(x) \log \tau_F(x) dx, \end{aligned} \tag{3}$$

where $\tau_F(x)$ is the reversed hazard rate of X given by $f(x)/F(x)$.

Rao et al. [13] specified the importance of survival function based measures over the density function based and gave some results based on these measures. Di Crescenzo and Longobardi [5] defined new measures of information known

* Corresponding author e-mail:

as cumulative and dynamic cumulative entropies based on the distribution function F of a random variable X , given respectively by

$$\mathcal{C}\xi(F) = - \int_0^{\infty} F(x) \log F(x) dx, \quad (4)$$

and

$$\mathcal{C}\xi(F;t) = - \int_0^t \frac{F(x)}{F(t)} \log \left(\frac{F(x)}{F(t)} \right) dx, \quad (5)$$

$t > 0, F(t) > 0$. In this paper, we study the measures of cumulative and dynamic cumulative entropies using order statistics. A $(n-k+1)$ -out-of- n system, which is an application of order statistics, is widely used in reliability theory and survival analysis.

The structure of $(n-k+1)$ -out-of- n system is such that it works if and only if atleast $(n-k+1)$ -out-of- n components are working. For $k=1$ and $k=n$ it reduces respectively to a series and a parallel system. Information theoretic aspects of order statistics have been studied widely. Wong and Chen [15] showed that the difference between average entropy of order statistics and the entropy of a data distribution is a constant. Park [12] showed some recurrence relations for entropy of order statistics. Ebrahimi et al. [8] studied some information properties of order statistics based on Shannon entropy and Kullback-Leibler [11] measure. Baratpour et al. [2,3] have derived characterization result for the i^{th} order statistics based on Shannon entropy and Renyi [13] entropy using Stone-Weistrass Theorem.

The organization of paper is as follows: In Section 2, we define measures of cumulative and dynamic cumulative entropies based on order statistics. Section 3 focuses on a characterization result. Some properties of these measures are discussed in Section 4. We conclude with the results obtained in Section 5.

2 Cumulative and dynamic cumulative entropy using order statistics

Suppose X_1, X_2, \dots, X_n are n independent and identically distributed observations from a distribution F , where $F(\cdot)$ is differentiable with a density $f(\cdot)$ which is positive in an interval and zero elsewhere. The order statistics of the sample is defined by the arrangement of X_1, X_2, \dots, X_n from the smallest to largest denoted as $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. Then the p.d.f. of the i^{th} order statistics $X_{i:n}$, for $i = 1, 2, \dots, n$, is given by

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} (F(x))^{i-1} (1-F(x))^{n-i} f(x), \quad (6)$$

where

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx, a > 0, b > 0, \quad (7)$$

is beta function with parameters a and b , refer to [1,4].

Analogous to (6) and (7), we define measures of cumulative and dynamic cumulative entropy of the i^{th} order statistics $X_{i:n}$ given respectively by

$$\mathcal{C}\xi(F_{i:n}) = - \int_0^{\infty} F_{i:n}(x) \log F_{i:n}(x) dx \quad (8)$$

and

$$\mathcal{C}\xi(F_{i:n};t) = - \int_0^t \frac{F_{i:n}(x)}{F_{i:n}(t)} \log \left(\frac{F_{i:n}(x)}{F_{i:n}(t)} \right) dx, t > 0, F_{i:n}(t) > 0. \quad (9)$$

Obviously, when $n = 1$, (10) and (11) reduces to (6) and (7).

Put $i = n$, in (10) and (11) we get

$$\mathcal{C}\xi(F_{n:n}) = - \int_0^{\infty} F_{n:n}(x) \log F_{n:n}(x) dx \quad (10)$$

and

$$\mathcal{C}\xi(F_{n:n};t) = - \int_0^t \frac{F_{n:n}(x)}{F_{n:n}(t)} \log \left(\frac{F_{n:n}(x)}{F_{n:n}(t)} \right) dx. \quad (11)$$

From (8), we have $f_{n:n}(x) = nF^{n-1}(x)f(x)$. Therefore, $F_{n:n}(x) = F^n(x)$. Using probability integral transformation $U = F(X)$, where U is uniformly distributed random variable in $[0, 1]$, in (12) and (13) we get

$$\mathcal{C}\xi(F_{n:n}) = -n \int_0^1 \frac{u^n \log u}{f(F^{-1}(u))} du. \tag{12}$$

and

$$\mathcal{C}\xi(F_{n:n};t) = \frac{-n}{F^n(t)} \int_0^{F(t)} \frac{u^n}{f(F^{-1}(u))} du + \frac{n \log F(t)}{F^n(t)} \int_0^{F(t)} \frac{u^n}{f(F^{-1}(u))} du. \tag{13}$$

In the example given below, we calculate the cumulative entropy of sample maxima for an exponentially distributed random variable.

Example 2.1 Let X be an exponentially distributed random variable with parameter θ , having pdf $f(x) = \theta e^{-\theta x}$, $\theta > 0$, $x \geq 0$. It is easy to see that $f(F^{-1}(u)) = \theta(1-u)$. Putting this value in (14), we obtain

$$\begin{aligned} \mathcal{C}\xi(F_{n:n}) &= -n \int_0^1 \frac{u^n \log u}{\theta(1-u)} du \\ &= \frac{n}{\theta} \left(\int_0^1 \frac{(1-u^n) \log u}{1-u} du - \int_0^1 \frac{\log u}{1-u} du \right). \end{aligned}$$

When we put $n = 1$ in the above integral, we get

$$\mathcal{C}\xi(F) = \frac{1}{\theta} \left(\frac{\pi^2}{6} - 1 \right).$$

3 A characterization result

We study the characterization result for the dynamic cumulative entropy of the i^{th} order statistics using the sufficient condition for the uniqueness of the solution of initial value differential equation

$$\frac{dy}{dx} = f(x,y), y(x_0) = y_0, \tag{14}$$

where f is a function of two variables whose domain is a region $D \subset R^2$, (x_0, y_0) is a point in D and y is the unknown function. By the solution of (15), we find a function which satisfies the following conditions: (i) ϕ is differentiable on I , (ii) the growth of ϕ lies in D , (iii) $\phi(x_0) = y_0$ and (iv) $\phi'(x) = f(x, \phi(x))$, for all $x \in I$. The following theorem together with other results will help in proving our characterization result.

Theorem 3.1 Let the function f be a continuous function defined in a domain $D \subset R^2$ and let f satisfy Lipschitz condition (with respect to y) in D , that is

$$|f(x, y_1) - f(x, y_2)| \leq k|y_1 - y_2|, k > 0, \tag{15}$$

for every point (x, y_1) and (x, y_2) in D . Then the function $y = \phi(x)$ satisfy the initial value problem $y' = f(x,y)$ and $\phi(x_0) = y_0, x \in I$, is unique.

We use the following Lemma, refer to [9], to present a sufficient condition which ensures that Lipschitz condition is satisfied in D .

Lemma 3.1 Suppose that the function f is continuous in a convex region $D \subset R^2$. Suppose that $\frac{\partial f}{\partial y}$ exists and it is continuous in D . Then, f satisfies Lipschitz condition in D .

Next, we give the main characterization result.

Theorem 3.2 Let X be a non-negative continuous random variable with distribution function $F(\cdot)$. Let $\mathcal{C}\xi(F_{i:n};t) < \infty, \forall t \geq 0$, denote the dynamic cumulative entropy of the i^{th} order statistics. Then, $\mathcal{C}\xi(F_{i:n};t)$ characterizes the distribution function.

Proof: Suppose there exists two functions F_1 and F_2 such that

$$\mathcal{C}\xi(F_{1i:n};t) = \mathcal{C}\xi(F_{2i:n};t),$$

for all $t \geq 0$ and for all $i \leq n$. We know that

$$\begin{aligned} \mathcal{C}\xi(F_{i:n};t) &= - \int_0^t \frac{F_{i:n}(x)}{F_{i:n}(t)} \log \left(\frac{F_{i:n}(x)}{F_{i:n}(t)} \right) dx \\ &= - \frac{1}{F_{i:n}(t)} \int_0^t F_{i:n}(x) \log F_{i:n}(x) dx + \mu_{F_{i:n}}(t) \log F_{i:n}(t), \end{aligned} \quad (16)$$

where

$$\mu_{F_{i:n}}(t) = \frac{1}{F_{i:n}(t)} \int_0^t F_{i:n}(x) dx$$

is the *mean inactivity time* of the i^{th} order statistics.

Taking derivative with respect to t on both sides of (18) and using the relationship between reversed hazard rate and mean inactivity time of the i^{th} order statistics given by

$$\tau_{F_{i:n}}(t) = \frac{1 - \mu'_{F_{i:n}}(t)}{\mu_{F_{i:n}}(t)}, \quad (17)$$

we have

$$\mu'_{F_{i:n}}(t) = \frac{\mu_{F_{i:n}}(t) - \mathcal{C}\xi(F_{i:n};t) - \frac{d}{dt} \mathcal{C}\xi(F_{i:n};t) \mu_{F_{i:n}}(t)}{\mu_{F_{i:n}}(t) - \mathcal{C}\xi(F_{i:n};t)}. \quad (18)$$

Suppose that

$$\mathcal{C}\xi(F_{1:i:n};t) = \mathcal{C}\xi(F_{2:i:n};t) = g(t) \text{ (say)}, \quad 1 \leq i \leq n, t \geq 0,$$

where F_1 and F_2 are two distribution functions. Then, $\forall t \geq 0$, from (20) we get

$$\mu'_{F_{1:i:n}}(t) = \psi(t, \mu_{F_{1:i:n}}(t)), \quad \mu'_{F_{2:i:n}}(t) = \psi(t, \mu_{F_{2:i:n}}(t))$$

where

$$\psi(t, y) = \frac{y - g(t) - g'(t)y}{y - g(t)}.$$

Using Theorem 3.1 and Lemma 3.1 we get $\mu_{F_{1:i:n}}(t) = \mu_{F_{2:i:n}}(t)$, for all $t \geq 0$ and for all $i \leq n$, which using (19) gives $\tau_{F_{1:i:n}}(t) = \tau_{F_{2:i:n}}(t)$, for all $t \geq 0$ and for all $i \leq n$. Hence the result follows.

Remark: For $n = 1$, Theorem 3.2 implies that the measure of dynamic cumulative entropy $\mathcal{C}\xi(F;t)$ characterizes the distribution function of X uniquely, which can be proved otherwise also on similar lines.

4 Some properties of cumulative entropy

In this section, we discuss some properties of cumulative and dynamic cumulative entropy of order statistics.

P.1 Let $\mathcal{C}\xi(F_{i:n}) < \infty$ denote the entropy of i^{th} order statistics, then

$$\mathcal{C}\xi(F_{i:n}) = E[(\mu_{F_{i:n}}(X))]. \quad (19)$$

Proof:

$$\begin{aligned} E[(\mu_{F_{i:n}}(X))] &= \int_0^\infty \left(\frac{1}{F_{i:n}(t)} \int_0^t F_{i:n}(x) dx \right) f_{i:n}(t) dt \\ &= \int_0^\infty \left(\int_0^t F_{i:n}(x) dx \right) \frac{f_{i:n}(t)}{F_{i:n}(t)} dt \\ &= \int_0^\infty \left(\int_x^\infty \frac{f_{i:n}(t)}{F_{i:n}(t)} dt \right) F_{i:n}(x) dx \\ &= - \int_0^\infty F_{i:n}(x) \log F_{i:n}(x) dx = \mathcal{C}\xi(F_{i:n}). \end{aligned}$$

In the following example we verify the above result for the first order statistics in case of exponential distribution.

Let X be an exponentially distributed random variable with p.d.f $f(x) = \theta e^{-\theta x}$, $\theta > 0$, $x \geq 0$. For $i = 1$, it can be easily seen that

$$\begin{aligned} E[(\mu_{F_{1:n}}(X))] &= \int_0^\infty \mu_{F_{1:n}}(x) f_{1:n}(x) dx \\ &= -\frac{1}{n\theta} - \frac{1}{n\theta} \int_0^1 \frac{\log u}{(1-u)} du \\ &= \frac{1}{n\theta} \left(\frac{\pi^2}{6} - 1 \right). \end{aligned}$$

Also we have

$$\begin{aligned} \mathcal{C}\xi(F_{1:n}) &= -\int_0^\infty F_{1:n}(x) \log F_{1:n}(x) dx \\ &= -\int_0^\infty (1 - e^{-n\theta x}) \log(1 - e^{-n\theta x}) dx \\ &= -\frac{1}{n\theta} - \frac{1}{n\theta} \int_0^1 \frac{\log u}{(1-u)} du \\ &= -\frac{1}{n\theta} + \frac{\pi^2}{6n\theta}, \end{aligned}$$

which is same as $E[(\mu_{F_{1:n}}(X))]$.

P.2 Let $\mathcal{C}\xi(F_{i:n}; t) < \infty$ denote cumulative entropy of the i^{th} order statistics. Then

$$\mathcal{C}\xi(F_{i:n}; t) = E[(\mu_{F_{i:n}}(X)) | X_{i:n} \leq t]. \tag{20}$$

Proof: The proof is similar to **P.1**.

P.3 The upper bound for the dynamic cumulative entropy of the i^{th} order statistics is given by

$$\mathcal{C}\xi(F_{i:n}; t) \leq \frac{\mathcal{C}\xi(F_{i:n})}{F_{i:n}(t)}.$$

Proof: We know that

$$\mathcal{C}\xi(F_{i:n}) = -\frac{1}{F_{i:n}(t)} \int_0^t F_{i:n}(x) \log F_{i:n}(x) dx + \frac{\log F_{i:n}(t)}{F_{i:n}(t)} \int_0^t F_{i:n}(x) dx. \tag{21}$$

For all $t \geq 0$, using $0 \leq \log F_{i:n}(t) \leq 1$ in the above equation we get

$$\begin{aligned} \mathcal{C}\xi(F_{i:n}; t) &\leq -\frac{1}{F_{i:n}(t)} \int_0^t F_{i:n}(x) \log F_{i:n}(x) dx \\ &\leq -\frac{1}{F_{i:n}(t)} \int_0^\infty F_{i:n}(x) \log F_{i:n}(x) dx. \end{aligned}$$

Hence

$$\mathcal{C}\xi(F_{i:n}; t) \leq \frac{\mathcal{C}\xi(F_{i:n})}{F_{i:n}(t)}$$

with equality as $t \rightarrow \infty$.

5 Conclusion and comments

The distribution function based information measures called cumulative entropy measures are in general more stable in comparison to density based measures. The dynamic measure of cumulative entropy (7) measures the inactivity time (or past lifetime) of distribution of a system found to be dead at time t . The entropy measures based on order statistics have been studied widely and are crucial for measuring uncertainty in statistical modeling. The dynamic cumulative entropy measures based on order statistics characterizes the underlying distribution uniquely and are also bounded above. The results studied can be of interest both from theoretical as well as practical point of views.

Acknowledgement

The first author is thankful to the Center for Scientific and Industrial Research, India, to provide financial assistance for this work.

References

- [1] Arnold , B. C.,Balakrishnan , N.,Nagaraja , H. N. *A first Course in Order Statistics*. New York: John Wiley and Sons,(1992).
- [2] Baratpour, S., Ahmadi, J., Arghami, N. R. Some characterizations based on entropy of order statistics and record values. *Communications in Statistics-Theory and Methods* 36, 47-57, (2007).
- [3] Baratpour, S., Ahmadi, J., Arghami, N. R. Characterizations based on Renyi entropy of order statistics and record values. *Journal of Statistical Planning and Inference*. 138, 2544-2551, (2008).
- [4] David, H.A., Nagaraja, H.N. *Order Statistics* (2003). New York : Wiley.
- [5] Di Crescenzo and Longobardi, M. On Cumulative Entropies. *Journal of Statistical Planning and Inference*. 139, 4072-4087, (2009).
- [6] Di Crescenzo, A. and Longobardi, M. Entropy-based measure of uncertainty in past lifetime distributions. *J. Appl. Prob.* 39, 434-440, (2002).
- [7] N. Ebrahimi. How to measure uncertainty in the residual lifetime distributions. *Sankhya A* 58, 48-57, (1996).
- [8] Ebrahimi, N., Soofi, E.S., Zahedi, H. Information properties of order statistics and spacings. *IEEE Trans. Inform. Theor.*(2004), 50: 177-183.
- [9] Gupta, R.C. and Gupta, R.D. Proportional reversed hazard rate model and its applications. *Journal of Statistical Planning and Inference* 137, 3525-3536, (2007).
- [10] Gupta, R.C. and Kirmani, S.N.U.A. Characterizations based on conditional mean function. *Journal of Statistical Planning and Inference* 138, 964-970, (2008).
- [11] Kullback, S. *Information Theory and Statistics*. Wiley, New York, (1959).
- [12] Park, S. The entropy of consecutive order statistics. *IEEE Trans. Inform. Theor.*, 41, 2003-2007, (1995).
- [13] Rao,M.,Chen,Y.,Vemuri,B.C.,Wang,F. Cumulative residual entropy:a new measure of information. *IEEE Trans. Inform. Theor.* 50, 1220-1228, (2004).
- [14] Renyi, A. On measure of entropy and information. *In Proceeding of the 4th Berkeley symposium on mathematical statistics and probability*. 1, 547-561, (1961).
- [15] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal* 279-423 and 623-656, (1948).
- [16] Wong , K.M., Chen.,S. The entropy of ordered sequences and order statistics. *IEEE Trans. Inform. Theor.* 36, 276-284, (1990).