ProbStat Forum, Volume 07, July 2014, Pages 45–54
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

On information theory and its applications

Sreenivasan Ravi[†]**, Ali Saeb**

Department of Studies in Statistics, University of Mysore, Manasagangotri, Mysore 570006, India

Abstract. In this note, we look at several definitions of entropy and some of their consequences in information theory. We also obtain the entropy and relative entropy for general error distribution which is used to model errors which are not normal.

1. Introduction

Entropy is a measure of uncertainty and is a central concept in the field of information theory. It appears that Nyquist (1924, 1928) and Hartley (1928) were the first to study entropy. Shannon (1948) studied the properties of information sources and of communication channels used to transmit the outputs of these information sources. Among the many articles and books wherein entropies have been discussed, we would like to mention Cover and Thomas (1991).

In this note, we look at several definitions of entropy and some of their consequences. In the next section, we mention several definitions of entropy. In Section 3, we look at entropy in Reliability. We discuss information theory and central limit theorem in Section 4. In Section 5, we obtain entropy for the general error distribution and also obtain relative entropy of this with normal and double exponential. Table 2 in Appendix gives a table of entropies for some transformations of random variables proved in Ravi and Saeb (2012) and tables 3 and 4 contain entropies of some standard discrete and continuous distributions some of which are from Johnson (2006), Lazo and Rathie (1978), Ebrahimi *et al.* (1999) and some from Ravi and Saeb (2012).

2. Definitions of Entropies

Shannon (1948) defined an entropy measure known as Shannon entropy. For a discrete random variable (rv) X with probability mass function $\{p_1, p_2, \ldots, p_n\}$, the Shannon entropy is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{1}$$

where the logarithm is to the base 2 and the entropy is expressed in bits of information. If the base of the logarithm is e, then the entropy is measured in nats. As a continuous analogue of (1) Cover and Thomas (1991) define the differential entropy of a continuous rv Y with probability density function (pdf) f as

$$H(Y) = H(f) = -\int f(y) \log f(y) dy,$$

Keywords. Shannon entropy, Relative entropy, Dynamic entropy, Renyi entropy, Reliability, Cumulative entropy, Central limit theorem, General error distribution.

Received: 17 December 2013; Accepted: 06 June 2014

[†]Corresponding author

Email address: `ravi@statistics.uni-mysore.ac.in, sreenivasanravi@yahoo.com` (Sreenivasan Ravi[†])

where, by convention, the integral is over those values for which the pdf f is positive.

For pdfs f and g, a related concept is the Kullback-Leibler distance or relative entropy, denoted as $D(f\|g)$ and defined by

$$\begin{aligned} D(f\|g) &= \int f(x)\log\left(\frac{f(x)}{g(x)}\right)dx, \\ &= \int f(x)\log f(x)dx - \int f(x)\log g(x)dx, \\ &= -H(X) + \Delta_Y(X). \end{aligned} \tag{2}$$

Whenever the support of f, supp $f \nsubseteq$ supp g then $D(f\|g)$ is considered to be equal to ∞.

Wyner and Ziv (1969) show that the entropy of a continuous rv X takes values in $\mathbb{R}$, and $E(X^2) < \infty$ implies that $H(X) < \infty$, but the converse may not hold. If $E|X|^k < \infty$, it can be shown that

$$H(X) \leq \frac{1}{k}\log\frac{2^k\Gamma^k(1/k)eE|X|^k}{k^{k-1}}, \quad k > 0. \tag{3}$$

The equality in (3) is attained by the maximum entropy distribution with density

$$f(x) = C(\theta)e^{-\theta|x|^k},$$

where θ is obtained by maximizing the function subject to the constraint $E|X|^k \leq \theta$ and $C(\theta)$ is the normalizing constant. For the case of $k = 2$, relation (3) gives $\frac{e^{2H(X)}}{2\pi e} \leq Var(X)$. The equality holds if and only if $f(\cdot)$ is normal distribution.

For a random vector $X \in \mathbb{R}^n$, entropy power denoted by $N(X)$, is defined as $N(X) = \dfrac{1}{2\pi e}\exp\left\{\dfrac{2}{n}H(X)\right\}$. In particular, $N(X) = |\Sigma|^{\frac{1}{n}}$ when X has multivariate normal distribution with variance-covariance matrix Σ. Dembo *et al.* (1991) derived inequalities for entropy power, for example, they showed that for independent $n-$dimensional random vectors X and Y, $2^{2H(X+Y)/n} \geq 2^{2H(X)/n} + 2^{2H(Y)/n}$, with equality if and only if X and Y are normally distributed with proportional covariance matrices. Madiman and Barron (2007) generalized the entropy power inequalities and examined monotonicity properties of information.

Since the pioneering work of Shannon (1948), the concept of entropy has been generalized by several researchers and is used in several disciplines and contexts. The entropy has also been used in various branches of statistics and related fields, and has become an integral part of probability and statistics. Renyi (1961) generalizes the Shannon entropy as the following which is called the Renyi entropy of order β :

$$H(X;\beta) = \frac{1}{1-\beta}\log\left(\int f^\beta(x)dx\right) = \frac{\beta}{1-\beta}\log(\|f\|_\beta), \tag{4}$$

for $0 < \beta < \infty$, $\beta \neq 1$, where $\|f\|_\beta = (\int f^\beta(x)dx)^{\frac{1}{\beta}}$. Renyi has pointed out that different sorts of problems may require different better measures of information. The Renyi entropy of order 2, $H(X;2)$, is called the collision entropy. It is to be noted here that, as $\beta \to 1$, the Renyi entropy tends to Shannon entropy, which can be seen as the negative expected log likelihood. Considering Renyi entropy as a function of β, $H(X;\beta)$ may be called the spectrum of Renyi information.

For pdfs f and g, the relative β entropy is defined as

$$D_\beta(f\|g) = \frac{1}{\beta-1}\log\left(\int f(x)\left(\frac{f(x)}{g(x)}\right)^\beta\right)dx.$$

By L'Hospital's rule, $\lim_{\beta\to 1} D_\beta(f\|g) = D(f\|g)$. Hayashi (2002) gives this definition and the limiting behaviour of D_β.

Khinchin (1957) generalized the Shannon entropy by choosing a convex function $\phi(\cdot)$ such that $\phi(1) = 0$ and defined the entropy as $H^\phi(X) = \int f(x)\phi(f(x))dx$. The Shannon entropy can be obtained from

Khinchin's entropy by choosing $\phi(x) = -\log x$. Burbea and Rao (1982) generalized Shannon entropy as $\int \phi(f(x))dx$ for some real concave function ϕ defined on $[0, \infty)$. Shannon entropy can be obtained from Burbea-Rao entropy by choosing $\phi(x) = -x \log x, \quad x > 0$. It is to be noted that Renyi entropy cannot be obtained as special case of either of the two generalized entropies discussed above. But it can be obtained as a particular case of the $(h, \phi)-$entropy defined by Salicru *et al.* (1993) and Menendez *et al.* (1997) as in the following theorem: Consider (Ω, P_θ) where $\theta \in \Theta$, an open subset of $\mathbb{R}^M$. Assume that there exists a pdf $f(x; \theta)$ for the probability P_θ with respect to a $\sigma-$finite measure μ. Then the $(h, \phi)-$entropy associated with $f(x; \theta)$ is given by $H_h^\phi(X) = h\left(\int_\Omega \phi(f(x; \theta))d\mu(x)\right)$, where either $\phi : [0, \infty) \to \mathbb{R}$ is concave and $h : \mathbb{R}^+ \to \mathbb{R}$ is increasing or $\phi : [0, \infty) \to \mathbb{R}$ is convex and $h : \mathbb{R}^+ \to \mathbb{R}$ is decreasing. Note also that if h is increasing and ϕ is convex or h is decreasing and ϕ is concave, $H_h^\phi(X)$ plays the role of a certainty function. In Table 1, we present some examples of certainty and $(h, \phi)-$ entropy measures.

Table 1: (h, ϕ) Entropy

Measure	$h(x)$	$\phi(x)$
Shannon (1948)	x	$-x \log x$
Renyi (1961)	$(1-\beta)^{-1} \log x$	x^β

Another measure of uncertainty proposed in the recent literature is the weighted entropy by Crescenzo and Longobardi (2006), defined as $H^w(X) = -\int_0^\infty x f(x) \log(f(x))dx$. As an alternative measure of uncertainty, Rao *et al.* (2004) proposed the cumulative residual entropy (CRE) of X defined by $\mathcal{E}(X) = -\int \overline{F}_X(t) \log(\overline{F}_X(t))dt$. Note that CRE is always non-negative and its definition is valid in the discrete and continuous cases. Other properties of this measure can be seen in Rao (2005). In analogy with CRE, Crescenzo and Longobardi (2009) define the cumulative entropy (CE) of a rv X with df F as $C\mathcal{E}(X) = -\int F(x) \log F(x)dx$. For a non-negative absolutely continuous rv X with survival function (sf) $\overline{F}$, Sunoj and Linu (2010) define the cumulative Renyi entropy of order β as

$$\mathcal{E}(X; \beta) = \frac{1}{1-\beta} \log\left(\int_0^\infty \overline{F}^\beta(x)dx\right), \quad \beta > 0, \quad \beta \neq 1. \tag{5}$$

When $\beta \to 1$, (5) reduces to $\mathcal{E}(X; 1) = -\int_0^\infty \overline{F}(x) \log(\overline{F}(x))dx$, which is cumulative entropy and hence possesses all the properties discussed in Rao *et al.* (2004).

3. Reliability based Entropy

Ebrahimi (1996) considered the Shannon entropy for the residual lifetime $X_t = [X - t | X > t]$ of a non negative rv X, called the residual entropy at time t and defined it as

$$H(X; t) = -\int_t^\infty \frac{f_X(x)}{\overline{F}_X(t)} \log\left(\frac{f_X(x)}{\overline{F}_X(t)}\right) dx.$$

Obviously $H(X; 0) = H(X)$. Abraham and Sankaran (2005) defined Renyi entropy of order β residual lifetime as

$$H(X; \beta; t) = \frac{1}{1-\beta} \log\left(\frac{\int_x^\infty f^\beta(t)dt}{\overline{F}^\beta(x)}\right), \quad \beta > 0, \quad \beta \neq 1.$$

Asadi and Zohrevand (2007) introduced the CRE for the residual lifetime distribution X_t. This function, called the dynamic cumulative residual entropy (DCRE), is defined by

$$\mathcal{E}(X; t) = -\int_t^\infty \frac{\overline{F}_X(x)}{\overline{F}_X(t)} \log\left(\frac{\overline{F}_X(x)}{\overline{F}_X(t)}\right) dx.$$

It is clear that $\mathcal{E}(X;0) = \mathcal{E}(X)$. The DCRE can be rewritten as

$$\mathcal{E}(X;t) = e_X(t)\log(\overline{F}_X(t)) - \frac{1}{\overline{F}_X(t)}\int_t^\infty \overline{F}_X(x)\log(\overline{F}_X(x))dx,$$

where $e_X(t) = E(X-t|X>t) = \frac{1}{\overline{F}(t)}\int_t^\infty \overline{F}(x)dx$, is the mean residual life (MRL) function. They showed that for any non-negative rv X, the CRE of X is the expectation of the MRL of X, that is, $\mathcal{E}(X) = E(e_X(X))$. They also proved that $\mathcal{E}(X;t) = E(e_X(X)|X>t)$.

For a non-negative rv X with an absolutely continuous sf $\overline{F}$, Sunoj and Linu (2010) define the dynamic cumulative Renyi residual lifetime entropy (DCRRE) of order β denoted by $\mathcal{E}(X;\beta;t)$ as

$$\mathcal{E}(X;\beta;t) = \frac{1}{1-\beta}\log\left(\int_t^\infty \frac{\overline{F}^\beta(x)}{\overline{F}^\beta(t)}dt\right), \quad \beta>0 \ \ \beta\neq 1,$$

which can be written as

$$(1-\beta)\mathcal{E}(X;\beta;t) = \log\left(\int_t^\infty \overline{F}^\beta(x)dx\right) - \beta\log\left(\overline{F}(t)\right).$$

Differentiating, we have

$$(1-\beta)\mathcal{E}'(X;\beta;t) = \beta r_X(t) - e^{-(1-\beta)\mathcal{E}(\beta;X;t)},$$

where $\mathcal{E}'(X;\beta;t)$ denotes the derivative of $\mathcal{E}(X;\beta;t)$ with respect to t and $r_X(t) = \frac{f_X(t)}{\overline{F}_X(t)}$ is the hazard rate of X. Obviously, when a system has completed t units of time, for different values of $\beta, \mathcal{E}(X;\beta;t)$ gives Renyis information for the remaining life of the system. Also, $\mathcal{E}(X;\beta;0) = \mathcal{E}(X;\beta)$.

4. Information Theory and Central Limit Theorem

For many random systems, the Shannon entropy plays a fundamental role in the analysis of how they evolve towards an equilibrium. It is interesting to study the convergence of the normalized sums of iid rvs to the Gaussian distribution: the central limit theorem. The idea of tracking the central limit theorem using Shannon entropy goes back to Linnik (1959) and Shimizu (1975), who used it to give a particular proof of the central limit theorem. Brown (1982), Barron (1986) and Takano (1987) were the first to prove a central limit theorem with convergence in the Shannon entropy sense. Artstein *et al.* (2004) and Johnson and Barron (2004) obtained the rate of convergence under some conditions on the density. According to Johnson and Vignat (2007), the Renyi entropy $H(X;\beta)$ is not maximum when X is normal, among all rvs with mean zero and variance 1. Therefore, the Renyi entropy is not monotonic increasing for the normalized sum. That the entropy convergence is increasing was proved by Artstein *et al.* (2004) and some simpler proofs can be found in Tulino and Verdu (2006) and Madiman and Barron (2007). Johnson (2006) is a good reference to the application of information theory to limit theorems, especially the central limit theorem. Cui and Ding (2010) showed the following for convergence of the Renyi entropy, where Y has normal distribution:

$$|H(S_n;\beta) - H(Y;\beta)| = \begin{cases} o\left(\frac{1}{n}\right), & if \ \ \beta\geq 2, \\ o\left(\frac{1}{\sqrt{n}}\right), & if \ \ \frac{3}{2}\leq\beta<2, \\ o\left(\frac{1}{n^{\beta-1}}\right), & if \ \ 1<\beta<\frac{3}{2}. \end{cases}$$

Barron (1986) showed the following: If $X_1, X_2, \ldots$ are iid with densities, zero mean and have finite variances $v_1, v_2, \ldots$ and $D(g_m\|\phi)$ is finite for some m then $\lim_{m\to\infty} D(g_m\|\phi) = 0$, where g_m is the density of $\sum_{i=1}^m X_i/\sqrt{mv}$. $D(f\|g)$ is not a metric; it is asymmetric and does not satisfy the triangle inequality. However, $D(f\|g)\geq 0$ with equality iff $f(x) = g(x)$ for almost all $x\in\mathbb{R}$. Furthermore (see Kullback (1967) for details),

$\|f-g\|_{L^1(dx)} \leq \sqrt{2D(f\|g)}$. The normal distribution maximises entropy amongst rvs of given variance, so the Central limit theorem corresponds to the entropy tending to its maximum. Fisher information is used to prove that the maximum is achieved. If W is a rv with differentiable density $g(u)$, we define the score function $\rho_W(u) = \frac{\partial}{\partial u}\log g(u) = \frac{g'(u)}{g(u)} I(g(u)>0)$ and the Fisher Information $J(W) = E\rho_W^2(W)$. If U is a rv with density f and variance 1, and Z_τ is $N(0,\tau)$, independent of U, then

$$D(f\|\phi) = \frac{1}{2}\int_0^\infty \left(J(U+Z_\tau) - \frac{1}{1+\tau}\right).$$

This is a rescaling of Lemma 1 of Barron (1986), which is an integral from of de Bruijn's identity. Also it is shown that if X is a rv with density f, and ϕ is standard normal density, then:

$$\begin{aligned}\sup_x |f(x)-\phi(x)| &\leq \left(1+\sqrt{\frac{6}{\pi}}\right)\sqrt{J(X)},\\ \int |f(x)-\phi(x)|dx &\leq 2d_H(f,\phi) \leq \sqrt{2J(X)},\end{aligned}$$

where $d_H(f,\phi)$ is the Hellinger distance $\left(\int |\sqrt{f(x)}-\sqrt{\phi(x)}|^2 dx\right)^{0.5}$.

5. Entropy and Relative entropy of General Error Distribution

Consider iid observations $X_1,\ldots,X_n$ with common distribution as the general error distribution, $n \geq 1$. The class of general error distributions (GEDs) of which normal distribution is a particular case, was introduced by Nelson (1991) for modelling the error in linear / regression models and also for modelling time series data with heavy tails when there is evidence to believe that the errors are not normal. The pdf of the standardized GED is given by

$$f(x) = \frac{ve^{-\frac{1}{2}|\frac{x}{\lambda}|^v}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}, \quad v>0,\ x\in\mathbb{R},$$

where $\lambda = \left(2^{-\frac{2}{v}}\frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}\right)^{\frac{1}{2}}$, $\Gamma(\cdot)$ denoting the gamma function. Nadarajah (2005) studies the properties of GED but the definition of GED used there is different from the one used here.

Lemma 5.1. *Let X be a rv with df GED. Then the Renyi entropy is*

$$H(X;\beta) = \frac{1}{1-\beta}\left[\log\beta^{-1/v} - (1-\beta)\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)\right].$$

And, the Shannon entropy is

$$H(X) = \frac{1}{v} - \log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right). \tag{6}$$

Proof. From (4), we have

$$\begin{aligned}H(X;\beta) &= \frac{1}{1-\beta}\log\left(\int_{-\infty}^\infty \left(\frac{ve^{-\frac{1}{2}|\frac{x}{\lambda}|^v}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)^\beta dx\right),\\ &= \frac{1}{1-\beta}\log\left(\beta^{-1/v}\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)^{\beta-1}\int_{-\infty}^\infty \frac{ve^{-\frac{1}{2}\left|\frac{x}{\beta^{-1/v}\lambda}\right|^v}}{\beta^{-1/v}\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}dx\right),\\ &= \frac{1}{1-\beta}\left[\log\beta^{-1/v} - (1-\beta)\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)\right].\end{aligned}$$

Taking limit as $\beta \to 1$, we obtain the Shannon Entropy as

$$\begin{aligned} H(X) &= \lim_{\beta\to 1} \frac{1}{1-\beta}\left[\log \beta^{-1/v} - (1-\beta)\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)\right], \\ &= \lim_{\beta\to 1} \frac{\log\beta}{v(\beta-1)} - \log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right), \\ &= \frac{1}{v} - \log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right). \end{aligned}$$

□

Theorem 5.2. *Let X be a rv with GED pdf f and Y be rv with the normal $(0,1)$ pdf ϕ. Then*

$$D(X\|Y;v) = -\frac{1}{v} + \log\left(\frac{v\sqrt{\Gamma(3/v)}}{2(\Gamma(1/v))^{3/2}}\right) + \log\sqrt{2\pi} + 1/2.$$

Proof. From (2), $D(X\|Y) = -H(X) + \Delta_Y(X)$, and we have

$$\begin{aligned} \Delta_Y(X) &= -\int_{-\infty}^{\infty} f(x)\log g(x)dx, \\ &= -\int_{-\infty}^{\infty} \frac{ve^{-\frac{1}{2}\left|\frac{x}{\lambda}\right|^v}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\log\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right)dx, \\ &= \log\sqrt{2\pi} + \int_{-\infty}^{\infty}\frac{x^2 ve^{-\frac{1}{2}\left|\frac{x}{\lambda}\right|^v}}{\lambda 2^{2+\frac{1}{v}}\Gamma(\frac{1}{v})}dx, \\ &= \log\sqrt{2\pi} + \frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\int_0^{\infty} x^2 e^{-\frac{x^v}{2\lambda^v}}dx. \end{aligned}$$

Taking $\frac{x^v}{2\lambda^v} = u$, we have $x = 2^{\frac{1}{v}}\lambda u^{1/v}$ and $dx = \frac{2^{\frac{1}{v}}\lambda}{v}u^{\frac{1}{v}-1}du$, so that

$$\begin{aligned} \Delta_Y(X) &= \log\sqrt{2\pi} + \frac{2^{\frac{2}{v}-1}\lambda^2}{\Gamma(\frac{1}{v})}\int_0^{\infty} u^{\frac{3}{v}-1}e^{-u}du, \\ &= \log\sqrt{2\pi} + \frac{2^{\frac{2}{v}-1}\lambda^2\Gamma(\frac{3}{v})}{\Gamma(\frac{1}{v})}, \\ &= \log\sqrt{2\pi} + \frac{1}{2}. \end{aligned} \tag{7}$$

From (6) and (7), we have

$$\begin{aligned} \mathcal{D}(X\|Y;v) &= -\frac{1}{v} + \log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right) + \log\sqrt{2\pi} + 1/2, \\ &= -\frac{1}{v} + \log\left(\frac{v\sqrt{\Gamma(3/v)}}{2(\Gamma(1/v))^{3/2}}\right) + \log\sqrt{2\pi} + 1/2. \end{aligned}$$

□

Remark 5.3. *Note that, if $v = 2$ then $D(X\|Y;2) = 0$, and GED is standard normal distribution.*

Theorem 5.4. *Let X be a rv with df double exponential with parameter θ with df f and Y be a rv with df GED and pdf g. Then*

$$D(X\|Y;v)=\log(\theta/2)-1-\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{\Gamma(1+v)}{2(\theta\lambda)^v}.$$

Proof. From (2), $D(X\|Y)=-H(X)+\Delta_Y(X)$, and we have

$$\begin{aligned}
\Delta_Y(X) &= -\int_{-\infty}^{\infty} f(x)\log g(x)dx,\\
&= -\int_0^{\infty}\frac{\theta}{2}e^{-\theta x}\log\left(\frac{ve^{-\frac{1}{2}(\frac{x}{\lambda})^v}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)dx-\int_{-\infty}^{0}\frac{\theta}{2}e^{\theta x}\log\left(\frac{ve^{-\frac{1}{2}(-\frac{x}{\lambda})^v}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)dx,\\
&= -\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{\theta}{4\lambda^v}\left(\int_0^{\infty}e^{-\theta x}x^v dx+\int_{-\infty}^{0}e^{\theta x}(-x)^v dx\right),\\
&= -\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{1}{2\lambda^v}\int_0^{\infty}\theta e^{-\theta x}x^v dx.
\end{aligned}$$

Taking $\theta x=u$, $dx=du/\theta$, and

$$\begin{aligned}
\Delta_Y(X) &= -\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{1}{2(\theta\lambda)^v}\int_0^{\infty}e^{-u}u^v du,\\
&= -\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{\Gamma(1+v)}{2(\theta\lambda)^v}. \qquad (8)
\end{aligned}$$

Using the entropy of double exponential distribution (see Table. 4) and (8), we have

$$\mathcal{D}(X\|Y;v) = \log(\theta/2)-1-\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)+\frac{\Gamma(1+v)}{2(\theta\lambda)^v}.$$

□

Remark 5.5. *Note that, if $v=1$ then $\lambda=1/2\sqrt{2}$ and $D(X\|Y;1)=0$, and GED is double Exponential with parameter $\sqrt{2}$.*

6. Appendix

Table 2: Transformation on Entropies

Random Variable	Transformation	Entropy
X is a rv	$Y=(X-b)/a$,	$H(Y)=-\log a+H(X)$
X is a positive rv	$Y=(X^c-b)/a$,	$H(Y)=-\log(a/c)+(c-1)E_X(\log X)+H(X)$
X is a negative rv	$Y=-((-X)^c-b)/a$,	$H(Y)=-\log(a/c)+(c-1)E_X(\log(-X))+H(X)$

Table 3: Entropies of some discrete distributions

Distribution	Probability Mass Function	Entropy
Poisson	$e^{-\lambda}\lambda^x/x!,\ \ x=0,1,\ldots,\ \ \lambda>0$	$\lambda-\lambda\log\lambda+E\log\Gamma(k+1)$
Bernoulli	$p^k q^{1-k},\ \ k=0,1,\ \ q=1-p$	$-q\log q-p\log p$
Binomial	$\binom{n}{x}p^k q^{n-k},\ \ x=0,1,\ldots,n,\ \ q=1-p$	$-np\log p-nq\log q-\log\Gamma(n+1)$ $+E(\log\Gamma(k+1)+\log\Gamma(n-k+1))$.
Geometric	$pq^k,\ \ k=0,1,\ldots$	$-(q\log q+p\log p)/p$
Uniform	$1/n,\ \ x=1,2,\ldots,n$	$\log n$

References

Abraham, B., Sankaran, P.G. (2005) Renyi's entropy for residual lifetime distribution, Statistical Papers 46, 17–30.

Artstein, S., Ball, K.M., Barthe, F., Naor, A. (2004) On the rate of convergence in the entropic central limit theorem, Probability Theory Related Fields 3, 381–390.

Artstein, S., Ball, K.M., Barthe, F., Naor, A. (2004) Solution of Shannon's problem on the monotonicity of entropy, Journal Amer Math. Soc 17, 975–982.

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

Barron, A.R. (1986) Entropy and the central limit theorem, The Annals of Probability 14(1), 336–342.

Brown, L.D. (1982) A proof of the central limit theorem motivated by the Cramer-Rao inequality, Statistics and probability: Essays in Honour of C.R. Rao, 141–148.

Burbea, J., Rao, C.R. (1982) On the convexity of some divergence measures based on entropy functions, IEEE Transactions on Information Theory 28, 489–495.

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

Crescenzo, A.Di, Longobardi, M. (2006) On weighted residual and past entropies, Science Math. Japan 64, 255–266.

Crescenzo, A. Di, Longobardi, M. (2009) On cumulative entropies, Journal of Statistical Planning and Inference 139, 4072–4087.

Cui, H., Ding, Y. (2010) The convergence of the Renyi entropy of the normalized sum of IID random variables, Statistics and Probability Letters 80, 1167–1173.

Dembo, A., Cover, M., Thomas, A. (1991) Information Theoretic Inequalities, IEEE Transactions on Information Theory.

Ebrahimi, N. (1996) How to measure uncertainty in the residual life time distribution, Sankhya Ser. A 58, 48–56.

Nader Ebrahimi, N., Maasoumi, E., Soofi, E. (1999) Ordering univariate distributions by entropy and variance, Journal of Econometrics 90, 317–336.

Hartley, R.T.V. (1928) Transmission of information, Bell System Technical Journal 7, 535–563.

Hayashi, M. (2002) Limiting behaviour of relative Renyi entropy in a nonregular location shift family, math. PR/0212077.

Johnson, O., Barron, A. (2004) Fisher information inequalities and the central limit theorem, Probab. Theory Relat. Fields 129, 391–409.

Johnson, O. (2006) Information Theory and the Central Limit Theorem, Imperical College Press.

Johnson, O., Vignat, Ch. (2007) Some results concerning maximum Renyi entropy distributions, Ann. Inst. H. Poincare Probab. Statist 43(1), 339–351.

Khinchin, A.J. (1957) Mathematical Foundation of Information Theory, Dover.

Kullback, S. (1967) A lower bound for discrimination information in terms of variation, IEEE Trans. Inform Theory 13, 126–127.

Lazo, A.C.G., Rathie, P.N. (1978) On the entropy of continuous distributions, IEEE Transactions on Information Theory 24, 120–122.

Linnik, Ju.V. (1959) An information theoretic proof of the central limit theorem with Lindeberg conditions, Theory Probability Application 4, 288–299.

Madiman, M., Barron, A.R. (2007) Generalized entropy power inequalities and monotonicity properties of information, IEEE Transaction Information Theory 53, 2317–2329.

Menendez, M.L., Morales, D., Pardo, L., Salicru, M. (1997) $(h,\phi)-$entropy differential metric, Applications of Mathematics 42, 81–98.

Nadarajah, S. (2005) A generalized normal distribution, Journal of Applied Statistics 32(7), 685–694.

Nelson, D.B. (1991) Conditional heteroscedasticity in asset returns: A new approach, Econometrica 59(2), 347–370.

Nyquist, H. (1924) Certain factors affecting telegraph speed, Bell System Technical Journal 3, 324–346.

Nyquist, H. (1928) Certain topics in telgraph transmission theory, IEEE Transactions 47, 617–644.

Rao, M., Chen, Y., Vemuri, B.C., Wang, F. (2004) Cumulative residual entropy: a new measure of information, IEEE Transactions on Information Theory 50, 1220–1228.

Rao, M. (2005) More on a new concept of entropy and information, Journal of Theoretical Probability 18, 967–981.

Ravi, S., Saeb, A. (2012) A note on entropies of l-max stable, p-max stable, generalized Pareto and generalized log Pareto distributions, ProbStat Forum, 05, July 2012, Pages 62–79.

Renyi, A. (1961) On Measures of Entropy and Information, Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, University of Colifornia Press, Vol. 1, 547–561.

Table 4: Entropies of some continuous distributions

Distribution	Density function	Shannon Entropy
General Normal	$\frac{v e^{-\frac{1}{2}\left\lvert\frac{x}{\lambda}\right\rvert^{v}}}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})},\ \ \lambda=\left(2^{-\frac{2}{v}}\frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}\right)^{\frac{1}{2}}$ $x\in\mathbb{R},\ v>0,\ \mu\in\mathbb{R}.$	$\frac{1}{v}-\log\left(\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma(\frac{1}{v})}\right)$
Gamma	$e^{-\theta x}\theta^{m}x^{m-1}/\Gamma(m),$ $m>0,\ \theta>0, x>0$	$m-\frac{\Gamma'(m)(m-1)\log e}{\Gamma(m)}+\frac{\log\Gamma(m)}{\theta}.$ $\Gamma'(m)/\Gamma(m)$ is digamma function
Cauchy	$c/(\pi(c^2+x^2)),$ $c>0,\ \ x>0$	$\log(4\pi c).$
Lévy	$\frac{c}{\sqrt{2\pi\lvert x\rvert^3}}\exp\left(-\frac{c^2}{2\lvert x\rvert}\right),$ $x>k,\ c>0, k>0$	$\log(16\pi e c^4)/2+3\gamma/2$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ \ x\in\mathbb{R},$	$\log(2\pi e)/2$
Exponential	$\lambda e^{-\lambda x},\ \ \lambda>0,\ \ x>0,$	$\lambda-\log\lambda$
Double Exponential	$\frac{\lambda}{2}e^{-\lambda x},\ x\geq 0,$ and $\frac{\lambda}{2}e^{\lambda x}\ \ x<0$	$1-\log(\lambda/2).$
Uniform $[0,c)$	$\frac{1}{c},\ \ c>0$	$\log c$
Fréchet	$\alpha x^{-\alpha-1}e^{-x^{-\alpha}},\ x>0,\ \ \alpha>0$	$-\log\alpha+\frac{\alpha+1}{\alpha}\gamma+1$
Weibull	$\alpha(-x)^{\alpha-1}e^{-(-x)^{\alpha}},\ x<0,\ \ \alpha>0$	$-\log\alpha+\frac{\alpha-1}{\alpha}\gamma+1$
Gumbel	$e^{-x}e^{-e^{-x}},\ x\in\mathbb{R}$	$1+\gamma$
Log Normal	$e^{-(\log x)^2/2}/(x\sqrt{2\pi}),\ x>0$	$\frac{1}{2}\log(2\pi e)$
Logistic	$e^{-x}/\sqrt{1+e^{-x}},\ x\in\mathbb{R}$	2
Log Fréchet	$\alpha x^{-1}(\log x)^{-\alpha-1}e^{-(\log x)^{-\alpha}},\ \ x>1$	$-\log\alpha+\frac{\alpha+1}{\alpha}+\Gamma(1-1/\alpha)+1,$ $\alpha>1$
Negative Log Fréchet	$\alpha(-x)^{-1}(-\log(-x))^{-\alpha-1}e^{-(-\log(-x))^{-\alpha}},$ $-1<x<0,$	$-\log\alpha+\frac{\alpha+1}{\alpha}-\Gamma(1-1/\alpha)+1,$ $\alpha>1$
Log Weibull	$\alpha x^{-1}(-\log x)^{\alpha-1}e^{-(-\log x)^{\alpha}},$ $0\leq x<1$	$-\log\alpha+\frac{\alpha-1}{\alpha}-\Gamma(1+1/\alpha)+1,$
Negative Log Weibull	$\alpha(-x)^{-1}(\log(-x))^{\alpha-1}e^{-(\log(-x))^{-\alpha}},$ $x<-1$	$-\log\alpha+\frac{\alpha-1}{\alpha}+\Gamma(1+1/\alpha)+1,$
Pareto	$1-x^{-\alpha},\ \ 1\leq x$	$-\log\alpha+(\alpha+1)/\alpha$
Negative Beta $(1,1)$	$1-(-x)^{\alpha},\ \ -1\leq x<0$	$-\log\alpha+(\alpha-1)/\alpha$
Log Pareto	$\alpha x^{-1}(\log x)^{-\alpha-1},\ \ x>e$	$-\log\alpha+(2\alpha^2-1)/(\alpha(\alpha-1)),$ $\alpha>1$
Negative Log Pareto	$-\alpha x^{-1}(-\log(-x))^{-\alpha-1},\ \ -e^{-1}\leq x$	$-\log\alpha-1/(\alpha(\alpha-1),$ $\alpha>1$
Log Negative Beta $(1,1)$	$\alpha x^{-1}(-\log x)^{\alpha-1},\ e^{-1}\leq x<1$	$-\log\alpha-1/(\alpha(\alpha+1))$
Negative Log Negative Beta $(1,1)$	$-\alpha x^{-1}(\log(-x))^{\alpha-1},\ -e\leq x<-1$	$-\log\alpha+(2\alpha^2-1)/(\alpha(\alpha+1))$

Salicru, M., Menendez, M.L., Morales, D., Pardo, L. (1993) Asymptotic distribution of $(h, \phi)-$entropies, Communications in Statistics (Theory and methods 22, 2015–2031.

Shannon, C.E. (1948) A mathematical theory of communications, Bell System Technical Journal 27, 379–423.

Shimizu, R. (1975) On Fisher's amount of information for location family, Statistical distributions in scientific work 3, 305–312.

Sunoj, S.M., Linu, M.N. (2010) Dynamic cumulative residual Renyi's entropy, Statistics 46(1), 1–16.

Takano, S. (1987) Convergence entropy central limit theorem, Yokohama Mathematical Journal 35, 143–148.

Tulino, A.M., Verdu, S. (2006) Monotonic decrease of the non Gaussianness of the sum of independent random variables: a simple proof, IEEE Transaction Information Theory 25, 4295–4297.

Wyner, A.D., Ziv, J. (1969) On communication of analog data from bounded source space, Bell System Technical Journal 48, 3139–3172.